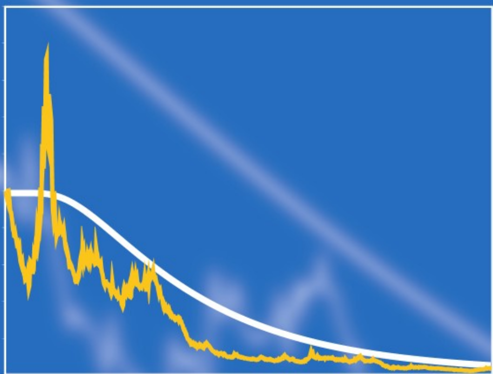
Springer Finance

Eckhard Platen
David Heath

# A Benchmark Approach to Quantitative Finance



Springer

Springer Finance

# *Springer Finance*

*Springer Finance* is a programme of books addressing students, academics and practitioners working on increasingly technical approaches to the analysis of financial markets. It aims to cover a variety of topics, not only mathematical finance but foreign exchanges, term structure, risk management, portfolio theory, equity derivatives, and financial economics.

*Ammann M.*, Credit Risk Valuation: Methods, Models, and Application (2001)
*Back K.*, A Course in Derivative Securities: Introduction to Theory and Computation (2005)
*Barucci E.*, Financial Markets Theory. Equilibrium, Efficiency and Information (2003)
*Bielecki T.R. and Rutkowski M.*, Credit Risk: Modeling, Valuation and Hedging (2002)
*Bingham N.H. and Kiesel R.*, Risk-Neutral Valuation: Pricing and Hedging of Financial Derivatives (1998, 2nd ed. 2004)
*Brigo D. and Mercurio F.*, Interest Rate Models: Theory and Practice (2001, 2nd ed. 2006)
*Buff R.*, Uncertain Volatility Models – Theory and Application (2002)
*Carmona R.A. and Tehranchi M.R.*, Interest Rate Models: An Infinite Dimensional Stochastic Analysis Perspective (2006)
*Dana R.-A. and Jeanblanc M.*, Financial Markets in Continuous Time (2003)
*Deboeck G. and Kohonen T. (Editors)*, Visual Explorations in Finance with Self-Organizing Maps (1998)
*Delbaen F. and Schachermayer W.*, The Mathematics of Arbitrage (2005)
*Elliott R.J. and Kopp P.E.*, Mathematics of Financial Markets (1999, 2nd ed. 2005)
*Fengler M.R.*, Semiparametric Modeling of Implied Volatility (2005)
*Filipović D.*, Term-Structure Models (2009)
*Fusai G. and Roncoroni A.*, Implementing Models in Quantitative Finance (2008)
*Geman H., Madan D., Pliska S.R. and Vorst T. (Editors)*, Mathematical Finance – Bachelier Congress 2000 (2001)
*Gundlach M. and Lehrbass F. (Editors)*, CreditRisk$^+$ in the Banking Industry (2004)
*Jeanblanc M., Yor M., Chesney M.*, Mathematical Methods for Financial Markets (2009)
*Jondeau E.*, Financial Modeling Under Non-Gaussian Distributions (2007)
*Kabanov Y.A. and Safarian M.*, Markets with Transaction Costs (2010)
*Kellerhals B.P.*, Asset Pricing (2004)
*Külpmann M.*, Irrational Exuberance Reconsidered (2004)
*Kwok Y.-K.*, Mathematical Models of Financial Derivatives (1998, 2nd ed. 2008)
*Malliavin P. and Thalmaier A.*, Stochastic Calculus of Variations in Mathematical Finance (2005)
*Meucci A.*, Risk and Asset Allocation (2005, corr. 2nd printing 2007, Softcover 2009)
*Pelsser A.*, Efficient Methods for Valuing Interest Rate Derivatives (2000)
*Platen E. and Heath D.*, A Benchmark Approach to Quantitative Finance (2006, corr. printing 2010)
*Prigent J.-L.*, Weak Convergence of Financial Markets (2003)
*Schmid B.*, Credit Risk Pricing Models (2004)
*Shreve S.E.*, Stochastic Calculus for Finance I (2004)
*Shreve S.E.*, Stochastic Calculus for Finance II (2004)
*Yor M.*, Exponential Functionals of Brownian Motion and Related Processes (2001)
*Zagst R.*, Interest-Rate Management (2002)
*Zhu Y.-L., Wu X., Chern I.-L.*, Derivative Securities and Difference Methods (2004)
*Ziegler A.*, Incomplete Information and Heterogeneous Beliefs in Continuous-time Finance (2003)
*Ziegler A.*, A Game Theory Analysis of Options (2004)

Eckhard Platen · David Heath

# A Benchmark Approach to Quantitative Finance

Springer

Eckhard Platen
School of Finance and Economics
and School of Mathematical Sciences
University of Technology, Sydney
PO Box 123
Broadway, NSW 2007
Australia
eckhard.platen@uts.edu.au

David Heath
Centre for Mathematics and its Applications
Mathematical Sciences Institute
Australian National University
Canberra, ACT 0200
Australia
david.heath@maths.anu.edu.au

ISBN 978-3-540-26212-1 e-ISBN 978-3-540-47856-0
DOI 10.1007/978-3-540-47856-0
Springer Heidelberg Dordrecht London New York

Library of Congress Control Number: 2006932290

Mathematics Subject Classification (2000): 90A12, 60G30, 62P20
JEL Classification: G10, G13



*Cover design*: WMXDesign GmbH

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

# Authors' Comments on the Corrected Second Printing

The original printing of the book appeared in 2006. Its very positive reception has led to its being sold out in less than three years. Springer's decision to reprint the book gave us the opportunity to correct minor mathematical and typographical errors in the original printing of the monograph. We would appreciate receiving any suggestions for further improvements and thank all those readers who have pointed out misprints and errors to us.

August 2009

*Eckhard Platen*
*David Heath*

# Preface

In recent years products based on financial derivatives have become an indispensable tool for risk managers and investors. Insurance products have become part of almost every personal and business portfolio. The management of mutual and pension funds has gained in importance for most individuals. Banks, insurance companies and other corporations are increasingly using financial and insurance instruments for the active management of risk. An increasing range of securities allows risks to be hedged in a way that can be closely tailored to the specific needs of particular investors and companies. The ability to handle efficiently and exploit successfully the opportunities arising from modern quantitative methods is now a key factor that differentiates market participants in both the finance and insurance fields. For these reasons it is important that financial institutions, insurance companies and corporations develop expertise in the area of *quantitative finance*, where many of the associated quantitative methods and technologies emerge.

This book aims to provide an *introduction to quantitative finance.* More precisely, it presents an introduction to the mathematical framework typically used in financial modeling, derivative pricing, portfolio selection and risk management. It offers a unified approach to risk and performance management by using the *benchmark approach*, which is different to the prevailing paradigm and will be described in a systematic and rigorous manner.

This approach uses the *growth optimal portfolio* as numeraire and the real world probability measure as pricing measure. The existence of an equivalent risk neutral probability measure is not required, which is one of the aspects distinguishing the approach in this book from other more conventional texts in the area. It is our experience that many practitioners find the use of the *real world* probability measure attractive for *pricing* because it is natural and pricing can still be carried out even under circumstances when a risk neutral probability measure cannot exist.

We have attempted to write a multi-purpose book that provides information and methods for a wide range of professionals, researchers and graduate students. It is designed for three groups of readers. In the first instance it should provide useful information to financial analysts and practitioners in the investment, banking and insurance industries. Other professionals at financial software companies, hedge funds, consultants, regulatory authorities and government agencies may significantly benefit from using this book. Secondly, the book aims to introduce those with a reasonable basic mathematical background to the area of quantitative finance. Engineers, computer scientists, numerical analysts, physicists, theoretical chemists, biologists, astrophysicists, statisticians, econometricians, actuaries and other readers should be able to gain access to the field through the book. Thirdly, researchers in financial mathematics will find the later parts of the book interesting and possibly challenging. In particular, the monograph aims to stimulate further developments of the benchmark approach.

The material presented is a self-contained introduction that could be part of a coursework masters or PhD program in quantitative finance. The areas of probability and statistics, stochastic calculus, optimization and numerical methods relevant to finance are all introduced. The book has been designed in a modular way with cross references so that it can also be used as a handbook allowing relevant definitions, formulas and results to be easily looked up.

The monograph is divided into fifteen chapters. The first two chapters summarize fundamental results from probability and statistics which are essential for quantitative finance. Some statistical analysis on the log-return distribution of indices is included at the end of Chap. 2.

The Chaps. 3 and 4 introduce stochastic processes. The stochastic calculus needed for financial modeling using stochastic differential equations is presented in Chaps. 5 to 7. Stochastic differential equations with jumps are introduced from a finance perspective. Some of the material goes beyond what can be found in standard textbooks.

In Chap. 8 basic financial derivatives are introduced from a hedging perspective. European call and put options are priced via the corresponding Black-Scholes partial differential equation. The sensitivities of these option prices to movements in parameter values are studied. Hedge simulations are performed, which illustrate derivative pricing and hedging.

Chapter 9 presents various alternative pricing methodologies. First, the concept of *real world pricing* is introduced. Several other pricing methods are shown to be special cases of real world pricing. These include actuarial pricing, risk neutral pricing and pricing under change of numeraire. The existence of an equivalent risk neutral probability measure is *not* required under the benchmark approach. The chapter concludes by introducing the Girsanov theorem, the Bayes rule and the Feynman-Kac formula.

Chapter 10 develops a unified modeling framework for continuous financial markets under the benchmark approach. It presents a range of new concepts and ideas that do not fit under the presently prevailing approaches. A *diversification theorem* is derived, which shows under some regularity condition that diversified portfolios approximate the growth optimal portfolio. This allows us to interpret a diversified market index as a proxy for the growth optimal portfolio.

Chapter 11 derives results on portfolio optimization via the maximization of Sharpe ratios. The capital asset pricing model (CAPM), the Markowitz efficient frontier, two fund separation and results on expected utility maximization, utility indifference pricing, derivative pricing and hedging are also presented in this chapter.

The modeling of stochastic volatility of stock market indices under the benchmark approach is discussed in Chap. 12. This analysis includes the pricing of index derivatives under models that do not admit an equivalent risk neutral probability measure. More general volatility models than those permitted under the standard risk neutral approach are covered.

In Chap. 13 it is shown that the discounted growth optimal portfolio follows the dynamics of a time transformed squared Bessel process of dimension four. Making the drift of the discounted growth optimal portfolio a function of time, yields the *minimal market model*. Derivative prices which follow under this parsimonious model appear to be rather realistic. Long term derivatives can be realistically priced. These prices deviate significantly from those obtained under risk neutral pricing because the hypothetical risk neutral measure has after several years a total mass that is significantly less than one. Extensions of the minimal market model with random scaling are considered.

In Chap. 14 models are analyzed that permit jumps to model event risk. Most of the results of previous chapters are generalized to jump diffusion markets. Two market models illustrate differences in derivative pricing under the standard risk neutral and the benchmark approach.

Finally, in Chap. 15 a brief introduction is given from a unifying perspective to basic numerical methods for quantitative finance. This introduction covers scenario simulation, Monte Carlo simulation, tree based methods and finite difference methods. A binomial tree method is developed for the benchmark approach and finite difference methods are explained as numerical methods for systems of coupled ordinary differential equations.

Selected *exercises* at the end of each chapter should enable the reader to further develop skills and test the understanding of the subject. *Solutions* to these exercises are included at the end of the book. The material can be taught at different levels. The first sections in most chapters provide a less technical presentation of the subject. At the end of some sections or chapters (*)-subsections or (*)-sections have been included. These are more technical in nature and are usually not necessary for a first reading.

The formulas are numbered according to the chapter and section where they appear. Assumptions, theorems, lemmas, definitions and corollaries are numbered sequentially in each section. The most common notations are listed at the beginning of the book and an *index of keywords* is given at its end. Some readers may find the *author index* at the end of the book useful.

Substantial work is involved in studying the material presented. This should not be underestimated by the reader. Actively solving exercises is strongly recommended. The reward for this demanding work will be a sound understanding of essential methods in quantitative finance with an emphasis on the benchmark approach.

The authors would like to thank several colleagues and PhD students for many valuable suggestions on the manuscript, including Nicola Bruti-Liberati, Hans Bühlmann, Carl Chiarella, Boris Choy, Morten Christensen, Marc Craddock, Ernst Eberlein, Robert Elliott, Kevin Fergusson, Chris Heyde, John van der Hoek, Hardy Hulley, Monique Jeanblanc, Leah Kelly, Truc Le, Shane Miller, Alex Novikov, Alun Pope, Wolfgang Runggaldier and Marc Yor. The authors would like to express their deep gratitude to Katrin Platen, who organized all technical work on the book, in particular, many figures. She carefully and patiently type set the countless versions of the extensive manuscript. Finally, we like to thank Catriona Byrne from Springer Verlag for her excellent work and for encouraging us to write this book.

It is greatly appreciated if readers could forward any errors, misprints or suggested improvements to: eckhard.platen@uts.edu.au

The interested reader is likely to find updated information about the benchmark approach, as well as, teaching material related to the book on the webpage of the first author under "Benchmark_Approach":

http://www.business.uts.edu.au/
finance/staff/Eckhard/Benchmark_Approach.html

Sydney, March 2006

*Eckhard Platen*
*David Heath*

# Acknowledgements

Continuing support is acknowledged by the Quantitative Finance Research Centre, the Department of Mathematical Sciences and the School of Finance & Economics at the University of Technology Sydney and the Centre for Mathematical Analysis and the Centre for Financial Mathematics at the Australian National University. Research on the book was supported by the ARC grants DP0559879 and DP0343913. Research on the benchmark approach benefited greatly from extended visits by the first author to the ETH Zurich, Humboldt University, Isaac Newton Institute, Kyoto University, Hitotsubashi University and Columbia University. Generous support by these institutions is acknowledged.

The financial data for most of the figures were provided by Datastream Thomson Financial. Figures 13.1.1–13.3.2 are based on reconstructed long term data from Global Financial Data. For providing the implied volatility surfaces in Figs. 12.1.3 and 12.1.4, the authors thank Marek Musiela from BNP Paribas. The implied volatility surface in Fig. 12.3.5 is based on publicly available data from CBOT. The forward rates in Fig. 10.4.1 were published by the US Federal Reserve Bank. With the permission of Routledge, a part of Taylor & Francis, Tables 2.6.1 and 2.6.2, together with Figs. 2.6.1–2.6.3, are taken from the paper Fergusson & Platen (2006) that appears in *Applied Mathematical Finance*. Taylor & Francis permitted also the inclusion of Figs. 12.2.1–12.2.4 and 12.3.1–12.3.7 from the papers Heath & Platen (2002a) and Heath & Platen (2006) in *Quantitative Finance*. Figure 10.6.5 was kindly provided by Frank Howard. With the permission of Springer Verlag, Figs. 13.4.1–13.4.5 are taken from Heath & Platen (2005b), published in *Asia-Pacific Financial Markets*. It is acknowledged that MATHEMATICA is a trademark of Wolfram Research and MSCI is a trademark of Morgan Stanley Capital International.

# Contents

# Basic Notation

| | |
|---|---|
| $\mu_X$ | mean of $X$; 21, 22 |
| $\sigma_X^2$, $\mathrm{Var}(X)$ | variance of $X$; 23, 24 |
| $\beta_X$ | skewness of $X$; 25 |
| $\kappa_X$ | kurtosis of $X$; 26 |
| $\underline{\kappa}_X$ | excess kurtosis; 28 |
| $\mathrm{Cov}(X,Y)$ | covariance of $X$ and $Y$; 39 |
| $\inf\{\cdot\}$ | greatest lower bound; 94, 129 |
| $\sup\{\cdot\}$ | smallest upper bound; 61, 79, 94, 128, 129 |
| $\max(a,b) = a \vee b$ | maximum of $a$ and $b$; 170 |
| $\min(a,b) = a \wedge b$ | minimum of $a$ and $b$; 170 |
| $\boldsymbol{x}^\top$ | transpose of a vector or matrix $\boldsymbol{x}$; 40 |
| $\boldsymbol{x} = (x^1, x^2, \ldots, x^d)^\top$ | column vector $\boldsymbol{x} \in \Re^d$ with $i$th component $x^i$; 44 |
| $\lvert\boldsymbol{x}\rvert$ | absolute value of $\boldsymbol{x}$ or Euclidean norm; 20, 22, 49 |
| $\boldsymbol{A} = [a^{i,j}]_{i,j=1}^{k,d}$ | $(k \times d)$-matrix $\boldsymbol{A}$ with $ij$th component $a^{i,j}$; 40 |
| $\det(\boldsymbol{A})$ | determinant of a matrix $\boldsymbol{A}$; 40 |
| $\boldsymbol{A}^{-1}$ | inverse of a matrix $\boldsymbol{A}$; 41, 46 |
| $(\boldsymbol{x}, \boldsymbol{y})$ | inner product of vectors $\boldsymbol{x}$ and $\boldsymbol{y}$; 49 |
| $\mathcal{N} = \{1, 2, \ldots\}$ | set of natural numbers; 5 |
| $\infty$ | infinity; 2 |

| | |
|---|---|
| $(a, b)$ | open interval $a < x < b$ in $\Re$; 8 |
| $[a, b]$ | closed interval $a \le x \le b$ in $\Re$; 12 |
| $\Re = (-\infty, \infty)$ | set of real numbers; 8 |
| $\Re^+ = [0, \infty)$ | set of nonnegative real numbers; 39 |
| $\Re^d$ | $d$-dimensional Euclidean space; 38 |
| $\Omega$ | sample space; 4 |
| $\emptyset$ | empty set; 4 |
| $A \cup B$ | the union of sets $A$ and $B$; 4 |
| $A \cap B$ | the intersection of sets $A$ and $B$; 4 |
| $A \backslash B$ | the set $A$ without the elements of $B$; 124, 258, 359 |
| $\mathcal{E} = \Re \backslash \{0\}$ | $\Re$ without origin; 124, 564 |
| $[X, Y]_t$ | covariation of processes $X$ and $Y$ at time $t$; 178 |
| $[X]_t$ | quadratic variation of process $X$ at time $t$; 172 |
| $n! = 1 \cdot 2 \cdot \ldots \cdot n$ | factorial of $n$; 10, 62 |
| $[a]$ | largest integer not exceeding $a \in \Re$; 522 |
| i.i.d. | independent identically distributed; 55 |
| a.s. | almost surely; 6, 56 |
| $f'$ | first derivative of $f: \Re \to \Re$; 14 |
| $f''$ | second derivative of $f: \Re \to \Re$; 219, 421 |
| $f: Q_1 \to Q_2$ | function $f$ from $Q_1$ into $Q_2$; 8 |
| $\frac{\partial u}{\partial x^i}$ | $i$th partial derivative of $u: \Re^d \to \Re$; 39 |
| $\left(\frac{\partial}{\partial x^i}\right)^k u$ | $k$th order partial derivative of $u$ with respect to $x^i$; 39 |
| $\exists$ | there exists; 128 |
| $F_X(\cdot)$ | distribution function of $X$; 8 |
| $f_X(\cdot)$ | density function of $X$; 11 |
| $\phi_X(\cdot)$ | characteristic function of $X$; 35 |
| $\mathbf{1}_A$ | indicator function for event $A$ to be true; 9 |

| | |
|---|---|
| $N(\cdot)$ | Gaussian distribution function; 14 |
| $\Gamma(\cdot)$ | gamma function; 15 |
| $\Gamma(\cdot;\cdot)$ | incomplete gamma function; 15 |
| $(\bmod c)$ | modulo $c$; 548 |
| $\mathcal{A}$ | collection of events, sigma-algebra; 5 |
| $\underline{\mathcal{A}}$ | filtration; 162 |
| $E(X)$ | expectation of $X$; 21, 22 |
| $E(X \mid \mathcal{A})$ | conditional expectation of $X$ under $\mathcal{A}$; 32, 33 |
| $P(A)$ | probability of $A$; 4 |
| $P(A \mid B)$ | probability of $A$ conditioned on $B$; 6 |
| $\in$ | element of; 1 |
| $\notin$ | not element of; 4 |
| $\neq$ | not equal; 5 |
| $\approx$ | approximately equal; 72, 169 |
| $a \ll b$ | $a$ is significantly smaller than $b$; 426, 515 |
| $\lim_{N\to\infty}$ | limit as $N$ tends to infinity; 2 |
| $\liminf_{N\to\infty}$ | lower limit as $N$ tends to infinity; 93, 94 |
| $\limsup_{N\to\infty}$ | upper limit as $N$ tends to infinity; 93, 94 |
| $\imath$ | square root of $-1$, imaginary unit; 35, 149 |
| $\delta(\cdot)$ | Dirac delta function at zero; 143, 146 |
| $\boldsymbol{I}$ | unit matrix; 44 |
| $\operatorname{sgn}(x)$ | sign of $x \in \Re$; 42 |
| $\mathcal{L}_T^2$ | space of square integrable, progressively measurable functions on $[0,T] \times \Omega$; 191 |
| $\mathcal{B}(U)$ | smallest sigma-algebra on $U$; 124 |
| $\ln(a)$ | natural logarithm of $a$; 1 |
| MM | Merton model; 252 |
| MMM | minimal market model; 251 |

| | |
|---|---|
| EWI | equi-value weighted index; 397 |
| MSCI | Morgan Stanley capital weighted world stock accumulation index; 332 |
| ODE | ordinary differential equation; 151, 239 |
| SDE | stochastic differential equation; 207, 235, 237 |
| PDE | partial differential equation; 143 |
| PIDE | partial integro differential equation; 358 |
| WSI | world stock index; 399 |
| $I_\nu(\cdot)$ | modified Bessel function of the first kind with index $\nu$; 16 |
| $K_\lambda(\cdot)$ | modified Bessel function of the third kind with index $\lambda$; 17, 18 |
| $\mathcal{V}$ | set of nonnegative portfolios; 373 |
| $\mathcal{V}^+$ | set of strictly positive portfolios; 369 |
| $\bar{\mathcal{V}}^+_{S_0}$ | set of strictly positive, discounted fair portfolios with initial value $S_0$; 419 |

# 1

# Preliminaries from Probability Theory

This chapter reviews some important results from probability theory and fixes notation. First we introduce discrete and continuous random variables and their distributions. Then we discuss functionals of random variables such as moments. Furthermore, we introduce certain classes of distributions and also multivariate distributions together with copulas.

## 1.1 Discrete Random Variables and Distributions

In financial markets one can observe the prices of assets such as stocks, commodities, currencies, futures, bonds etc. It is a challenge to model these random quantities in a satisfactory manner.

### Log-Returns

Let us assume that we observe an asset price at times $t_i = i\Delta$ for $i \in \{0, 1, \ldots\}$ with time step size $\Delta > 0$. The time $\Delta$ between two successive observations is typically the length of one day. If $X_{t_i}$ denotes the asset price at time $t_i$, then the *log-return* $R_{t_i}$ at this time is defined as

$$R_{t_i} = \ln(X_{t_{i+1}}) - \ln(X_{t_i}) = \ln\left(\frac{X_{t_{i+1}}}{X_{t_i}}\right) \tag{1.1.1}$$

for $i \in \{0, 1, \ldots\}$.

We define the daily log-return of an asset price as the daily increment of the natural logarithm of this price because, as we shall see later on, this reflects well the growth nature of economies and financial markets. Typically log-returns exhibit considerable variability.

We focus in this book on the modeling of log-returns while we introduce the basic concepts of probability, statistics, stochastic processes, stochastic calculus and stochastic differential equations. It will turn out that stochastic

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

differential equations provide an ideal mathematical framework for the modeling of financial quantities. In this context log-returns will also allow us to apply the powerful tools of stochastic calculus. This is not so conveniently achieved when using, so-called, *returns* that are of the form

$$\tilde{R}_{t_i} = \frac{X_{t_{i+1}} - X_{t_i}}{X_{t_i}}$$

and closely approximate log-returns when these are small. As we shall see, log-returns are more tractable in continuous time.

### Relative Frequencies and Probabilities

Let us interpret an asset's log-return $R_{t_i}$ as the *outcome* of an experiment based on observations of the data. Suppose, for simplicity, that we classify the log-returns as strictly negative, zero or positive. We denote these *elementary outcomes* or *states* by $\omega_1$, $\omega_2$, $\omega_3$, indicating that we observe a negative, zero or strictly positive log-return, respectively. We call the set of outcomes or states $\Omega = \{\omega_1, \omega_2, \omega_3\}$ the *sample space* for our experiment.

If we repeat our experiment $N$ times, that is, we observe for a stock daily log-returns on $N$ different days, and count the number $N(\omega_i)$ of times, that the outcome $\omega_i$ occurs, we can form the *relative frequency*

$$f_i(N) = \frac{N(\omega_i)}{N}.$$

For smaller $N$ this number usually varies considerably. As $N$ becomes larger, our experience would indicate that the relative frequency should approach a limit $p_i$, written as

$$\lim_{N\to\infty} f_i(N) = p_i,$$

which we call the *probability* of outcome $\omega_i$.

To illustrate the above example let us look at the daily IBM share price in US dollars over the period from 1977 until 1997, which is shown in Fig. 1.1.1. The corresponding log-returns are plotted in Fig. 1.1.2. In Fig. 1.1.3 we then display the relative frequencies $f_1(t_i)$, $f_2(t_i)$ and $f_3(t_i)$, $i \in \{0, 1, \ldots\}$, of negative, zero and strictly positive log-returns, respectively, during the time period. Note that after some wild fluctuations for small time $t$, at the beginning of the period, the relative frequency for negative log-returns stabilizes around a value close to $p_1 = 0.465$. Similarly, we obtain at the end of the period a value $p_3 = 0.463$ for the relative frequency of strictly positive log-returns. The value $p_2 = 0.072$ is then obtained for the rather small probability of zero log-returns. Clearly, we have $0 \leq p_i \leq 1$ for each $i \in \{1, 2, 3\}$ and $\sum_{i=1}^{3} p_i = 1$, that is, the probabilities $p_1$, $p_2$ and $p_3$ add up to one.


**Fig. 1.1.1.** IBM share price from 1977 until 1997

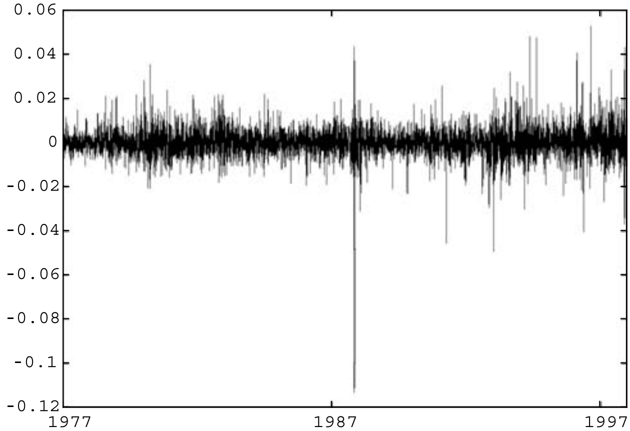



**Fig. 1.1.2.** Log-returns of IBM stock

## Probability Space

To analyze a model one is often interested in combinations of outcomes. We call such a combination an *event* if we can identify it either by its occurrence or its non-occurrence. Obviously, if a subset $A$ of the set of outcomes $\Omega$ is an event, then its *complement* $A^c = \{\omega_i \in \Omega : \omega_i \notin A\}$, which denotes the set of all $\omega_i$ from the sample space $\Omega$ that do not belong to the set $A$, must also be an event. In the case of the above example we might consider the event $A = \{\omega_1, \omega_2\}$ that corresponds to the occurrence of either a negative or zero log-return. The complement of this event is then $A^c = \{\omega_i \in \Omega : \omega_i \notin \{\omega_1, \omega_2\}\} = \{\omega_3\}$. This is the event $\{\omega_3\}$ of a strictly positive log-return.

In particular, the whole sample space $\Omega$ is an event, which is called the *sure event* since one of its outcomes must always occur. The complement of $\Omega$

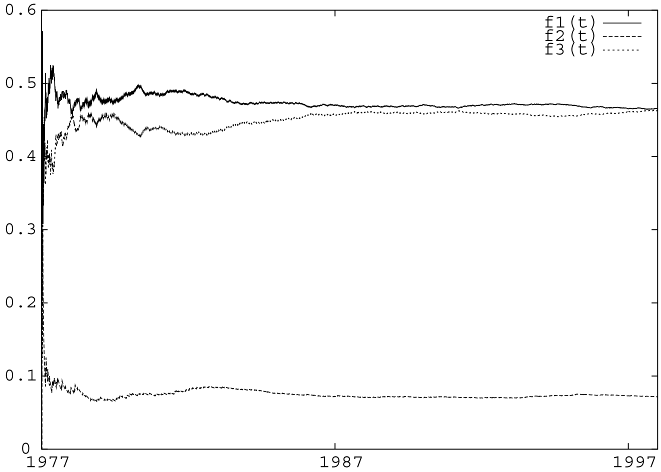



**Fig. 1.1.3.** Relative frequency over time

is the *empty set* $\emptyset$, which is also defined as an event but never occurs. If $A$ and $B$ are events, then the event $A \cup B$ occurs if either $A$ or $B$ occurs, whereas the event $A \cap B$ occurs if both $A$ and $B$ occur. With $A = \{\omega_1, \omega_2\}$, as in our example, and the event $B = \{\omega_2\}$ indicating a zero log-return we note that $A \cup B = \{\omega_1, \omega_2\} \cup \{\omega_2\} = \{\omega_1, \omega_2\}$ stands for an event consisting of either negative or zero log-returns and $A \cap B = \{\omega_1, \omega_2\} \cap \{\omega_2\} = \{\omega_2\}$ is the event which indicates only a zero log-return.

In the above discussion we have only mentioned experiments with a finite number of outcomes. However, the introduction of probabilities based on an infinite set of outcomes and the use of relative frequencies to define probabilities can lead to conceptual subtleties and other mathematical problems. To resolve these difficulties, Kolmogorov developed in the late 1920s an axiomatic approach to probability theory. In this approach the probabilities represent numbers assigned to corresponding events. In what follows we shall employ this axiomatic framework.

Let us denote by $P(A)$ the *probability* of the occurrence of an event $A$ that is taken from the *collection of events* $\mathcal{A}$ that corresponds to the sample space $\Omega$. Then from corresponding properties of relative frequencies we would expect these probabilities to satisfy the following relationships

$$0 \leq P(A) \leq 1, \tag{1.1.2}$$

$$P(A^c) = 1 - P(A), \tag{1.1.3}$$

$$P(\emptyset) = 0, \qquad P(\Omega) = 1, \tag{1.1.4}$$

and

$$P(A \cup B) = P(A) + P(B) \tag{1.1.5}$$

if $A$ and $B$ are exclusive, that is $A \cap B = \emptyset$ for events $A$ and $B$ taken from $\mathcal{A}$.

The above relationships allow us, for a given finite sample space $\Omega = \{\omega_1, \omega_2, \ldots, \omega_n\}$, consistently to allocate probabilities to each event. One can deduce that

$$\bigcup_{i=1}^{n} A_i \quad \text{and} \quad \bigcap_{i=1}^{n} A_i$$

are events if $A_1, A_2, \ldots, A_n$ are events, and that

$$P\left(\bigcup_{i=1}^{n} A_i\right) = \sum_{i=1}^{n} P(A_i)$$

if $A_1, A_2, \ldots, A_n$ are *mutually exclusive*, that is if $A_i \cap A_j = \emptyset$ for all $i, j \in \{1, 2, \ldots, n\}$ with $i \neq j$.

For the above example suppose we assign the probabilities $p_i = P(\{\omega_i\})$ for each outcome $\omega_i$, $i \in \{1, 2, 3\}$, as obtained from frequency records. Then the event $A = \{\omega_1, \omega_2\}$ of non-strictly positive outcomes has, according to (1.1.5), the probability

$$P(A) = P(\{\omega_1, \omega_2\}) = P(\{\omega_1\} \cup \{\omega_2\}) = P(\{\omega_1\}) + P(\{\omega_2\}) = p_1 + p_2.$$

The essential probabilistic information that characterizes an experiment can be succinctly summarized in the corresponding triplet $(\Omega, \mathcal{A}, P)$ consisting of the sample space $\Omega$, the collection of events $\mathcal{A}$ and the probability measure $P$, where these have to satisfy certain relationships. In the above analysis we have considered finite collections of events. To cover the case of infinite collections we must specify these properties to avoid contradictions. We assume that the collection of events $\mathcal{A}$ is a *sigma-algebra*, which means that

$$\Omega \in \mathcal{A}, \tag{1.1.6}$$

$$\text{if } A \in \mathcal{A} \text{ then } A^c \in \mathcal{A}, \tag{1.1.7}$$

$$\text{if } A \in \mathcal{A} \text{ and } B \in \mathcal{A} \text{ then } A \cup B \in \mathcal{A} \tag{1.1.8}$$

$$\text{and if } A_i \in \mathcal{A} \text{ for any } i \in \mathcal{N} = \{1, 2, \ldots\} \text{ then } \left(\bigcup_{i=1}^{\infty} A_i\right) \in \mathcal{A}. \tag{1.1.9}$$

In the case of infinite collections, equation (1.1.5) is replaced by what is called *countably additive probabilities*. This means,

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i) \tag{1.1.10}$$

for any sequence $(A_i)_{i \in \mathcal{N}}$ of mutually exclusive events.

It can be shown by DeMorgan's law that a sigma-algebra is closed under finite and countable intersections of events. In addition, if a set function defined on a sigma-algebra satisfies (1.1.2) and (1.1.10) with $P(\Omega) = 1$, then

(1.1.3)–(1.1.5) also hold and hence this set function would be a probability measure.

A triplet $(\Omega, \mathcal{A}, P)$ is then called a *probability space* and the relations (1.1.2)–(1.1.5) can be shown to form a consistent set of rules for modeling probabilities in this space. This kind of structure will be used for all of our modeling work described in this book. Since the models that we can construct will always remain abstract objects, they can reflect reality only to a limited extent. It will be our aim to introduce more and more flexible mathematical structures that provide the potential to model successfully complex stochastic phenomena in finance. However, the reader should never believe that there is anything like a perfect model. Even if some model were to become very successful, the market would regularly demand further modifications and extensions to the model.

The relations (1.1.2)–(1.1.5) allow us to prove in a straightforward manner that if $A, B \in \mathcal{A}$ and $A \subseteq B$ then

$$P(A) \leq P(B). \tag{1.1.11}$$

Furthermore, if $A, B \in \mathcal{A}$ then

$$P(A \cap B^c) = P(A) - P(A \cap B). \tag{1.1.12}$$

There may be some events $A$ with $P(A) = 0$. These are then called *null events*. On the other hand, there may be some event $B$ for which $P(B) = 1$. In this case we say $B$ has occurred *almost surely* (a.s.) or *with probability one*.

## Probabilities

The probability $P(A)$ of an event $A$ can be interpreted as a measure of the likelihood that $A$ occurs. If we have some additional information, such as that another event has occurred, then our estimate of this likelihood may change. For instance, if we know in the above example that the event $A = \{\omega_1, \omega_2\}$ of having no strictly positive log-return has occurred, then conditioned on this information, the conditional probabilities of observing negative or zero log-returns will add up to one. We denote by $P(\{\omega_1\} \,\big|\, A)$ the *conditional probability* that a negative log-return, the outcome $\omega_1$, will be observed, given that the event $A = \{\omega_1, \omega_2\}$ has occurred. Note that this conditional probability can be expressed by the ratio

$$P(\{\omega_1\} \,\big|\, A) = \frac{P(\{\omega_1\} \cap A)}{P(A)} = \frac{P(\{\omega_1\})}{P(\{\omega_1, \omega_2\})},$$

where $P(A) > 0$. This relation is readily suggested from the ratio of relative frequencies

$$\frac{f_1(N)}{f_1(N) + f_2(N)} = \frac{\frac{N(\omega_1)}{N}}{\frac{N(\omega_1)}{N} + \frac{N(\omega_2)}{N}} = \frac{N(\omega_1)}{N(\omega_1) + N(\omega_2)},$$

where $N(\omega_1)$ and $N(\omega_2)$ denote the number of outcomes $\omega_1$ and $\omega_2$, respectively, that have occurred out of $N$ repetitions of the experiment.

In general, the *conditional probability* $P(A \,|\, B)$ for the event $A$ given that the event $B$ has occurred is defined by the formula

$$P(A \,|\, B) = \frac{P(A \cap B)}{P(B)} \tag{1.1.13}$$

provided $P(B) > 0$. This formula is also called the *Bayes formula.* As a consequence of (1.1.13) one obtains

$$P(A)\, P(B \,|\, A) = P(B)\, P(A \,|\, B), \tag{1.1.14}$$

which is sometimes called *Bayes' Theorem.*

Conditional probabilities have similar properties to ordinary probabilities, for instance, they sum to one, when conditioned on the same $B$.

The likelihood for the occurrence of an event could be unaffected by whether or not another event $B$ has occurred. In such a case the conditional probability $P(A \,|\, B)$ should equal $P(A)$, which implies together with (1.1.13) that

$$P(A \cap B) = P(A)\, P(B). \tag{1.1.15}$$

We say that the events $A$ and $B$ are *independent* if and only if (1.1.15) holds. By assuming $P(B) > 0$ and rearranging formula (1.1.15) we see that events $A$ and $B$ are independent if

$$P(A) = \frac{P(A \cap B)}{P(B)}. \tag{1.1.16}$$

For instance, if we extend slightly our example and consider the log-returns from two different days to be independent, then the event characterizing the log-return from the first day does not affect the event that describes the log-return for the second day. In this example the second log-return is assumed to be not influenced by the outcome of the first log-return and vice versa.

More generally we say that $m$ events $A_1, A_2, \ldots, A_m$ are *independent* if

$$P(A_{i_1} \cap A_{i_2} \cap \ldots \cap A_{i_k}) = P(A_{i_1})\, P(A_{i_2}) \, \ldots \, P(A_{i_k}) \tag{1.1.17}$$

for all $k \in \mathcal{N}$ and non-empty subsets $\{i_1, i_2, \ldots, i_k\}$ of the set of indices $\{1, 2, \ldots, m\}$.

One can show that if $A_1, A_2, B \in \mathcal{A}$ and $P(B) > 0$, then

$$P\left(A_1 \cap A_2^c \,\middle|\, B\right) = P\left(A_1 \,\middle|\, B\right) - P\left(A_1 \cap A_2 \,\middle|\, B\right). \tag{1.1.18}$$

A sequence of events $(A_i)_{i \in \mathcal{N}}$ with $A_i \in \mathcal{A}$ for all $i \in \mathcal{N}$ is called a *partition* of $\Omega$ if

$$\bigcup_{i=1}^{\infty} A_i = \Omega, \tag{1.1.19}$$

and $A_\ell \cap A_m = \emptyset$ for all $\ell \neq m$. This allows us to formulate the following statement on the *total probability.* If $(A_i)_{i \in \mathcal{N}}$ is a partition of $\Omega$ with $P(A_i) > 0$ for all $i \in \mathcal{N}$, then for any event $B \in \mathcal{A}$ one obtains the representation

$$P(B) = \sum_{i=1}^{\infty} P\left(B \,\middle|\, A_i\right)\, P(A_i). \tag{1.1.20}$$

This formula can be very helpful for calculating the probabilities of certain events.

### Random Variables and Distributions

We are often interested in assigning some numerical quantity to the outcomes of a probabilistic experiment. For instance, in our stock log-return example, the quantity $X(\omega)$ might take the value 1 for a strictly positive log-return, 0 for a zero log-return and $-1$ for a negative log-return.

These assigned quantities correspond to the values taken by a function $X : \Omega \to \Re$, where $\Re = (-\infty, \infty)$ is the *set of real numbers.* In our example we have

$$X(\omega) = \begin{cases} 1 & \text{for} \quad \omega = \omega_1 \\ 0 & \text{for} \quad \omega = \omega_2 \\ -1 & \text{for} \quad \omega = \omega_3. \end{cases} \tag{1.1.21}$$

More generally, given a probability space $(\Omega, \mathcal{A}, P)$ we say, that a function $X : \Omega \to \Re$ is an $\mathcal{A}$-*measurable function* or a *random variable* if the set $\{\omega \in \Omega :\, a < X(\omega) \leq b\}$ is an event for each $a, b \in \Re$ with $a < b$. This means that this set is an element of $\mathcal{A}$. Using this definition it can be shown that if $X$ is a random variable, then it holds for any Borel subset of the real line

$$X^{-1}(B) = \{\omega \in \Omega :\, X(\omega) \in B\} \in \mathcal{A},$$

see Shiryaev (1984). We say that two random variables $X$ and $Y$ are *independent* if the corresponding events $\{\omega \in \Omega :\, X(\omega) \leq a\}$ and $\{\omega \in \Omega :\, Y(\omega) \leq b\}$ are independent for all $a, b \in \Re$.

Now it is appropriate to introduce for a random variable $X$ its *distribution function* $F_X : \Re \to [0, 1]$ that is defined for each real valued $x \in \Re$ by the relation

$$\begin{aligned} F_X(x) &= P(\{\omega \in \Omega :\, X(\omega) \leq x\}) \\ &= P(X \leq x). \end{aligned} \tag{1.1.22}$$

Here we have used in the last term an abbreviated notation for the probability of an event, which will also be used in other parts of the book. In Fig. 1.1.4 we show the three probabilities, $p_1 = 0.465$, $p_2 = 0.072$ and $p_3 = 0.463$ for the stock log-return example with possible outcomes $-1, 0, 1$, respectively, that is

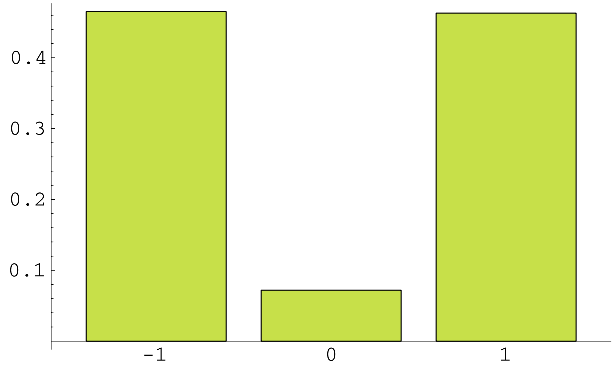



**Fig. 1.1.4.** Probabilities for the stock log-return example

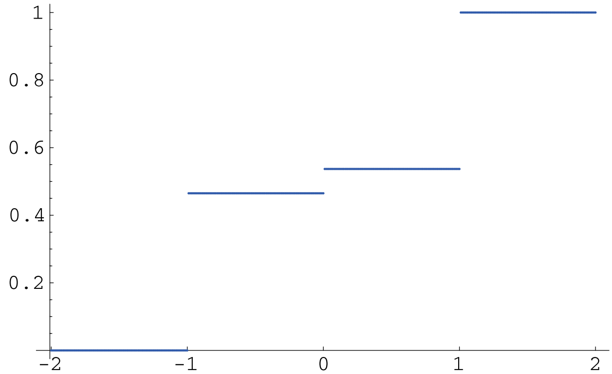



**Fig. 1.1.5.** Distribution for the stock log-return example

$$P(X = x) = \begin{cases} p_1 & \text{for} \quad x = -1 \\ p_2 & \text{for} \quad x = 0 \\ p_3 & \text{for} \quad x = 1. \end{cases} \tag{1.1.23}$$

The distribution function is then according to (1.1.22) given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < -1 \\ p_1 & \text{for} \quad -1 \leq x < 0 \\ p_1 + p_2 & \text{for} \quad 0 \leq x < 1 \\ 1 & \text{for} \quad 1 \leq x \end{cases} \tag{1.1.24}$$

for $x \in \Re$, which we plot in Fig. 1.1.5.

## Two-Point Distribution

A simple random variable is the *indicator function* $\mathbf{1}_A : \Omega \to \{0, 1\}$ of an event $A \in \mathcal{A}$, where

$$\mathbf{1}_A(\omega) = \begin{cases} 1 & \text{for} \quad \omega \in A \\ 0 & \text{for} \quad \omega \notin A. \end{cases} \tag{1.1.25}$$

Here the corresponding distribution function is of the form

$$F_{\mathbf{1}_A}(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - P(A) & \text{for} \quad 0 \leq x < 1 \\ 1 & \text{for} \quad 1 \leq x, \end{cases} \tag{1.1.26}$$

where $P(A)$ denotes the probability of the event $A$. This is an example of a *two-point random variable* which takes two distinct real values $x_1$ and $x_2$ with probabilities $p_1$ and $p_2 = 1 - p_1$, respectively, where $x_1 < x_2$.

It can be shown that for any random variable $X$ the limit of the value of the distribution function $F_X(x)$ for $x$ tending to minus infinity, $x \to -\infty$, is zero. That is

$$\lim_{x \to -\infty} F_X(x) = 0. \tag{1.1.27}$$

Similarly, it can be verified that

$$\lim_{x \to \infty} F_X(x) = 1 \tag{1.1.28}$$

and $F_X(x)$ is non-decreasing in $x \in \Re$.

The above examples indicate that a distribution function does not have to be continuous. However, one can show that it is always *right-continuous*, that is

$$\lim_{h \to 0+} F_X(x + h) = F_X(x) \tag{1.1.29}$$

for all $x \in \Re$.

### Poisson Distribution

An important discrete random variable is the *Poisson random variable* $X$ characterized by its *mean* $\lambda > 0$. It can be used to model, for instance, the number of trades per day that occur for a given stock or the number of bankruptcies that occur during a year. A Poisson random variable $X$ takes values $0, 1, \ldots$ without any upper bound. The corresponding probabilities $p_n = P(X = n)$ are the *Poisson probabilities* that are given by

$$p_n = \frac{\lambda^n}{n!} \exp\{-\lambda\} \tag{1.1.30}$$

for $n \in \{0, 1, \ldots\}$, where $\lambda > 0$, $n! = 1 \cdot 2 \cdot \ldots \cdot n$ for $n \in \mathcal{N}$ and $0! = 1$. These probabilities are displayed in Fig. 1.1.6 for the *intensity parameter* $\lambda = 2$. We write $X \sim P(\lambda)$ to indicate that $X$ has a Poisson distribution with intensity $\lambda$.

Let $\Omega = \mathcal{N} = \{1, 2, \ldots\}$ denote the set of natural numbers. A *discrete real valued random variable* $X$ is a measurable function from $\Omega$ into a finite or

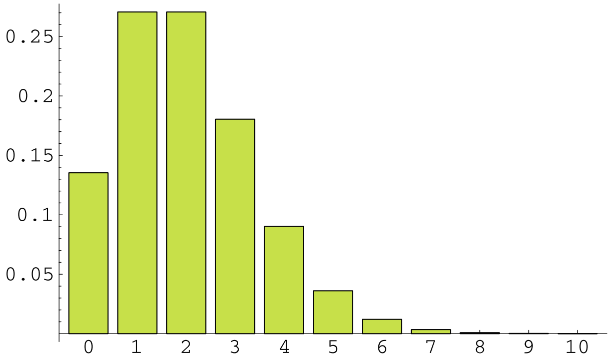



**Fig. 1.1.6.** Poisson probabilities for $\lambda = 2$

possibly infinite set of distinct real values $x_1 < x_2 < \cdots < x_n < \cdots$ with probabilities $p_n = P(X = x_n)$ for $n \in \mathcal{N}$. Its distribution function $F_X$ has the representation

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < x_1 \\ \sum_{i=1}^{n} p_i & \text{for} \quad x_n \leq x < x_{n+1}, \end{cases} \tag{1.1.31}$$

for $n \in \mathcal{N}$. $F_X$ is a right-continuous step-function with steps of height $p_n$ at $x = x_n$. For this random variable the set $\{x_1, x_2, \ldots\}$ could be used as the sample space $\Omega$, with all of its subsets being events.

## 1.2 Continuous Random Variables and Distributions

The modeling of events in a financial context often requires random variables that take any value in $\Re = (-\infty, \infty)$ or subintervals of $\Re$. We call a random variable $X$ a *continuous random variable* if the probability $P(X = x)$ is zero for all $x \in \Re$. If $X$ is a continuous random variable, then the corresponding distribution function $F_X$ will also be continuous.

In cases where the distribution function $F_X$ is differentiable, there exists a nonnegative function $f_X$, called the *density function*, such that

$$f_X(x) = \frac{dF_X(x)}{dx} \tag{1.2.1}$$

for all $x \in \Re$. However, $F_X$ could be differentiable *Lebesgue almost everywhere*, that is except possibly on a set of Lebesgue measure zero. It can be shown that if $F_X$ is absolutely continuous, then it can be expressed as integral of the form

$$F_X(x) = \int_{-\infty}^{x} f_X(s)\, ds \tag{1.2.2}$$

for all $x \in \Re$, where $f_X$ is the corresponding density function.

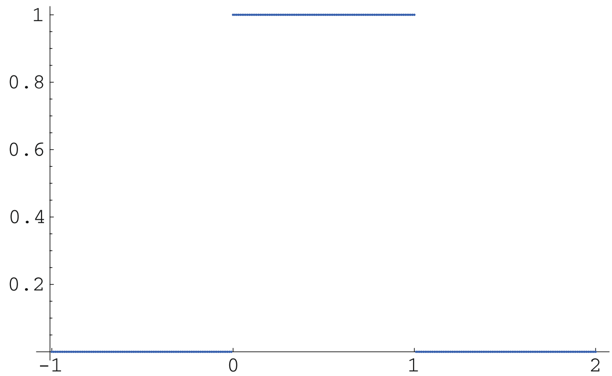



**Fig. 1.2.1.** The uniform density with $a = 0$ and $b = 1$

We shall now describe some commonly occurring examples of continuous random variables.

### Uniform Distribution

Consider a random variable $X$ which takes values only in a finite interval $[a, b)$, such that the probability of its being in a given subinterval is proportional to the length of the subinterval. Then the distribution function is given by

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < a \\ \dfrac{x-a}{b-a} & \text{for} \quad a \leq x < b \\ 1 & \text{for} \quad b \leq x, \end{cases}$$

which is differentiable everywhere except at $x = a$ and $x = b$. The corresponding density function is then of the form

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x \notin [a, b) \\ \dfrac{1}{b-a} & \text{for} \quad x \in [a, b). \end{cases} \tag{1.2.3}$$

We say that the random variable $X$ is in this case *uniformly distributed* on $[a, b)$ and use the abbreviation $X \sim U(a, b)$ to denote this fact. For example, log-returns of a stock could be modeled by a $U(-a, a)$ distributed random variable with a parameter $a > 0$ that describes the largest possible absolute log-return. The density for a $U(0, 1)$ distributed random variable is shown in Fig. 1.2.1.

### Exponential Distribution

The waiting time between two events when there is no memory kept on the time when the first event occurred, for instance, bankruptcies, catastrophes

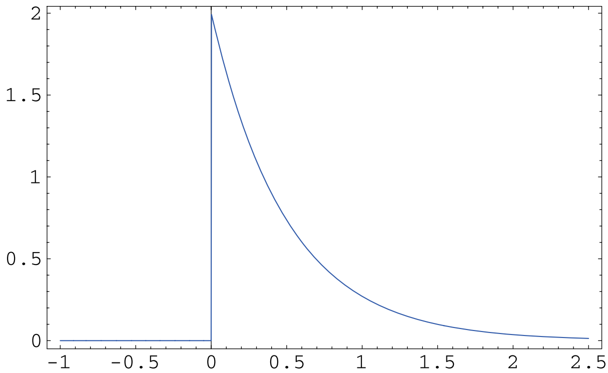



**Fig. 1.2.2.** The exponential density for intensity $\lambda = 2$

or changes in credit ratings, can be often modeled by a random variable $X$ with an *exponential distribution* given by the distribution function

$$F_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ 1 - \exp\{-\lambda\, x\} & \text{for} \quad x \geq 0 \end{cases} \tag{1.2.4}$$

for some intensity parameter $\lambda > 0$. $F_X$ is differentiable everywhere except when $x = 0$ and has as corresponding density function

$$f_X(x) = \begin{cases} 0 & \text{for} \quad x < 0 \\ \lambda \exp\{-\lambda\, x\} & \text{for} \quad x \geq 0. \end{cases} \tag{1.2.5}$$

We write $X \sim Exp(\lambda)$ to indicate that $X$ is an exponentially distributed random variable. A larger intensity parameter $\lambda$ means that it is more likely that the waiting time between two events is shorter. In Fig. 1.2.2 we plot the density of the exponential distribution for the intensity $\lambda = 2$.

## Gaussian Distribution

The *Gaussian density function* given by

$$f_X(x) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} \tag{1.2.6}$$

for $x \in \Re$ has a bell-shaped graph which is symmetric about $x = \mu$. In Fig. 1.2.3 we show the density of an $N(0,1)$ distributed random variable which is also called a *standard Gaussian random variable*. The corresponding standard Gaussian distribution function $F_X(x)$ is everywhere differentiable and has a sigmoidal-shaped graph, see Fig. 1.2.4. A random variable $X$ with the density function (1.2.6) is called a *Gaussian random variable* and we summarize this fact by writing $X \sim N(\mu, \sigma^2)$.

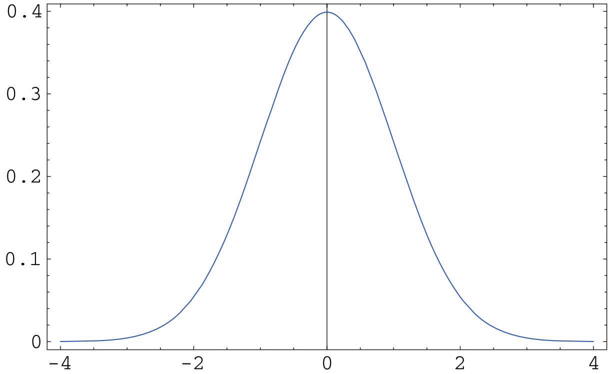



**Fig. 1.2.3.** The standard Gaussian density

Gaussian random variables occur so commonly in many applications, including financial ones, that they are often said to be *normally distributed*. The log-returns of stocks have been widely modeled as normally distributed random variables resulting in the well-known *lognormal asset price model* which we shall discuss later in detail. For this standard market model the increments of the logarithm of the stock price, the log-returns, are assumed to be normally distributed.

Unfortunately, the Gaussian distribution has no explicit analytic representation. Since it is often used in finance, for instance, in option pricing and Value at Risk calculations, it is useful to have an accurate approximation for the standard Gaussian distribution function $N : \Re \to (0, 1)$. This function can be approximated, for instance, by the expression

$$\begin{aligned} N(x) = \int_{-\infty}^{x} N'(z)\, dz = 1 - 0.5\,(1 &+ 0.0498673470\, x + 0.0211410061\, x^2 \\ &+ 0.0032776263\, x^3 + 0.0000380036\, x^4 \\ &+ 0.0000488906\, x^5 + 0.0000053830\, x^6)^{-16} \\ &+ \varepsilon(x), \end{aligned} \tag{1.2.7}$$

for $x \geq 0$, where we have an error term $\varepsilon(x)$ with $|\varepsilon(x)| < 0.00000015$, as established in Abramowitz & Stegun (1972). To obtain values for $N(x)$ for $x < 0$ we can use the relation $N(x) = 1 - N(-x)$. Here $N'(\cdot)$ denotes the standard Gaussian density function

$$N'(x) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\, x^2 \right\} \tag{1.2.8}$$

for $x \in \Re$. In Fig. 1.2.4 we graph the standard Gaussian distribution function. For statistical and other studies it is helpful to know that, for $X \sim N(\mu, \sigma^2)$,

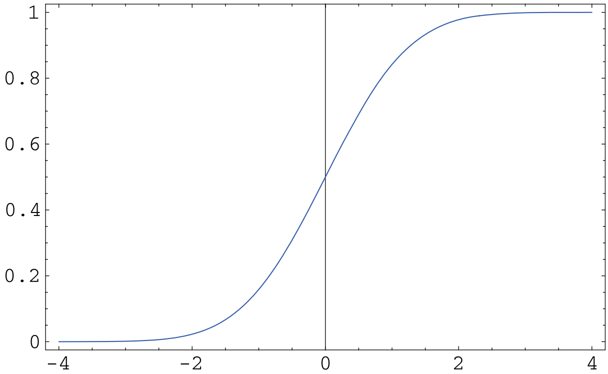



**Fig. 1.2.4.** The standard Gaussian distribution

we have, so-called, *k-sigma rules*, where $|X - \mu| < k\,\sigma$ approximately with probability 0.95 for $k = 2$, 0.9973 for $k = 3$ and 0.99994 for $k = 4$.

## Gamma Distribution

A *gamma distributed* random variable $X$ takes only positive real values and has a density function

$$f_X(x) = \frac{\alpha^p}{\Gamma(p)} \exp\{-\alpha\, x\}\, x^{p-1} \tag{1.2.9}$$

for $0 < x < \infty$ and parameters $\alpha > 0$ and $p > 0$. Here $\Gamma$ denotes the *gamma function* given by

$$\Gamma(p) = \int_0^\infty t^{p-1}\, e^{-t}\, dt \tag{1.2.10}$$

for $p > 0$. We use the abbreviation $X \sim G(p, \alpha)$ to indicate that a random variable $X$ is gamma distributed with the density function (1.2.9) for given parameters $\alpha$ and $p$. We plot in Fig. 1.2.5 the density of the gamma distribution for $\alpha = 0.5$ and $p = 2$.

In the special case $\alpha = 0.5$ the gamma distribution is equivalent to the *chi-square distribution* with $n = 2p$ degrees of freedom. For $n \in \mathcal{N}$ this distribution is obtained as that of a random variable $X$, that is the sum of the squares of $n = 2p$ independent standard Gaussian random variables. We abbreviate this by writing $X \sim \chi^2(n)$. Thus, Fig. 1.2.5 also shows a chi-square density with $n = 4$ degrees of freedom.

Let $X$ denote a chi-square distributed random variable with $n$ degrees of freedom. Its distribution function has the form

$$F_X(x) = \chi^2(x; n) = \int_0^x \frac{\exp\left\{-\frac{u}{2}\right\} \left(\frac{u}{2}\right)^{\frac{n}{2}-1}}{2\,\Gamma\left(\frac{n}{2}\right)}\, du = 1 - \frac{\Gamma\left(\frac{x}{2}; \frac{n}{2}\right)}{\Gamma\left(\frac{n}{2}\right)} \tag{1.2.11}$$

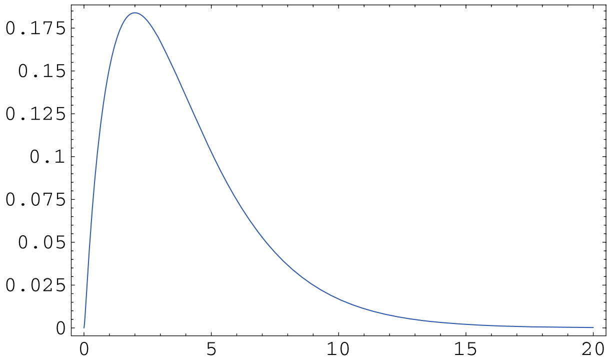



**Fig. 1.2.5.** The gamma density for $\alpha = 0.5$ and $p = 2$

for $x \geq 0$, where

$$\Gamma(u; a) = \int_u^\infty t^{a-1} \exp\{-t\} \, dt \tag{1.2.12}$$

is the *incomplete gamma function* for $u \geq 0$, $a > -1$, see Abramowitz & Stegun (1972) and Johnson, Kotz & Balakrishnan (1995).

## Non-Central Chi-Square Distribution

For a *non-central chi-square distributed* random variable $X \sim \chi^2(n, \ell)$ with $n \geq 0$ degrees of freedom and non-centrality parameter $\ell > 0$ its distribution function has the form

$$F_X(x) = \chi^2(x; n, \ell) = \sum_{k=0}^\infty \frac{\exp\left\{-\frac{\ell}{2}\right\} \left(\frac{\ell}{2}\right)^k}{k\,!} \left(1 - \frac{\Gamma\left(\frac{x}{2}; \frac{n+2k}{2}\right)}{\Gamma\left(\frac{n+2k}{2}\right)}\right) \tag{1.2.13}$$

for $x \geq 0$. In some sense, the non-central chi-square distribution is a weighted sum of central chi-square distributions with Poisson probabilities as weights. The corresponding density function is given as

$$f_X(x) = \frac{1}{2} \left(\frac{x}{\ell}\right)^{\frac{n}{4} - \frac{1}{2}} \exp\left\{-\frac{\ell + x}{2}\right\} I_{\frac{n}{2}-1}\left(\sqrt{\ell\, x}\right), \tag{1.2.14}$$

for $x > 0$. Here $I_\nu(\cdot)$ is the *modified Bessel function of the first kind* with index $\nu$, which is of the form

$$I_\nu(z) = \left(\frac{z}{2}\right)^\nu \sum_{j=0}^\infty \frac{\left(\frac{z^2}{4}\right)^j}{j\,!\,\Gamma(j + \nu + 1)}. \tag{1.2.15}$$

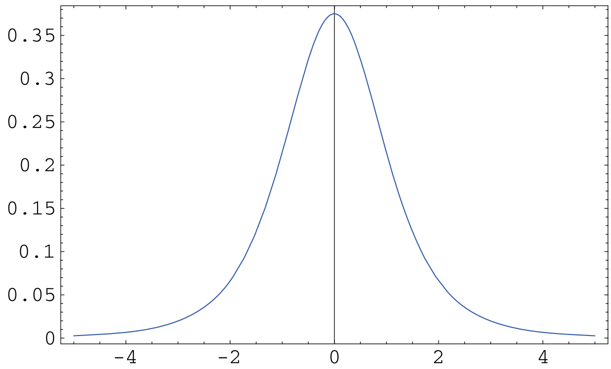



**Fig. 1.2.6.** Student $t$ density for $n = 4$ degrees of freedom

## Central Student $t$ Distribution

Let $Y \sim N(0,1)$ be a standard Gaussian distributed random variable and $Z \sim \chi^2(n)$ be an independent chi-square distributed random variable with $n > 0$ degrees of freedom. Then the random variable

$$X = \frac{Y}{\sqrt{\frac{Z}{n}}} \tag{1.2.16}$$

turns out to be a *central Student t*, or in short a *Student t*, distributed with $n$ degrees of freedom. Its density function is given by

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi\, n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}, \tag{1.2.17}$$

for $x \in \Re$. We write $X \sim t(n)$ if the random variable $X$ has a Student $t$ distribution with $n$ degrees of freedom. In Fig. 1.2.6 we plot the density of the Student $t$ distribution for $n = 4$ degrees of freedom. As will be shown later, this distribution seems to model log-returns of indices extremely well.

It is interesting to express the Student $t$ distribution function $F_{t(n)}(x)$ in terms of rational and trigonometric functions for small integers $n$, see Shaw (2005). For $n = 1$ one obtains in this way the standard *Cauchy distribution*

$$F_{t(1)}(x) = \frac{1}{2} + \frac{1}{\pi} \tan^{-1}(x), \tag{1.2.18}$$

where $\tan^{-1}(\cdot)$ expresses the inverse function of $\tan(\cdot)$. Further Student $t$ distribution functions are given by

$$F_{t(2)}(x) = \frac{1}{2} + \frac{x}{2\sqrt{x^2+2}}, \tag{1.2.19}$$

$$F_{t(3)}(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{\sqrt{3}}\right) + \frac{\sqrt{3}\,x}{\pi\,(x^2+3)}, \tag{1.2.20}$$

$$F_{t(4)}(x) = \frac{1}{2} + \frac{x\,(x^2+6)}{2\,(x^2+4)^{\frac{3}{2}}}, \tag{1.2.21}$$

$$F_{t(5)}(x) = \frac{1}{2} + \frac{1}{\pi}\tan^{-1}\left(\frac{x}{\sqrt{5}}\right) + \frac{\sqrt{5}\,x\,(3x^2+25)}{3\,\pi\,(x^2+5)^2}, \tag{1.2.22}$$

$$F_{t(6)}(x) = \frac{1}{2} + \frac{x\,(2x^4+30x^2+135)}{4\,(x^2+6)^{\frac{5}{2}}}. \tag{1.2.23}$$

**Symmetric Generalized Hyperbolic Distribution (*)**

Various authors have proposed asset price models with log-returns that relate to the rich class of *symmetric generalized hyperbolic* (SGH) distributions. This class of distributions was extensively examined by Barndorff-Nielsen (1977), see Hurst & Platen (1997) for a study on log-returns. We shall use this class later on to identify the distribution that fits best observed log-returns.

The SGH density function for a random variable $X$ has the form

$$f_X(x) = \frac{1}{\delta\,K_\lambda(\alpha\,\delta)}\sqrt{\frac{\alpha\,\delta}{2\,\pi}}\left(1+\frac{(x-\mu)^2}{\delta^2}\right)^{\frac{1}{2}\left(\lambda-\frac{1}{2}\right)} K_{\lambda-\frac{1}{2}}\left(\alpha\,\delta\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right) \tag{1.2.24}$$

for $x \in \Re$, where $\lambda \in \Re$ and $\alpha, \delta \geq 0$. We set $\alpha \neq 0$ if $\lambda \geq 0$ and $\delta \neq 0$ if $\lambda \leq 0$. Here $K_\lambda(\cdot)$ is the *modified Bessel function of the third kind* with index $\lambda$, see Abramowitz & Stegun (1972). It can be defined by the integral representation

$$K_\lambda(z) = \frac{1}{2}\int_0^\infty u^{\lambda-1}\,\exp\left\{-\frac{1}{2}\,z\left(u+\frac{1}{u}\right)\right\}du \tag{1.2.25}$$

for $z \in (0,\infty)$. For $\lambda = \eta + \frac{1}{2}$, where $\eta$ is a nonnegative integer, one has the explicit expression

$$K_{\eta+\frac{1}{2}}(z) = \sqrt{\frac{\pi}{2\,z}}\,\exp\{-z\}\sum_{\ell=0}^{\eta}\frac{(\eta+\ell)!}{(\eta-\ell)!\,\ell!}\,(2\,z)^{-\ell}. \tag{1.2.26}$$

The SGH density is a four parameter density. The parameter $\mu$ is a *location parameter*. The two *shape parameters* for its tails are $\lambda$ and $\bar{\alpha} = \alpha\,\delta$, defined so that they are invariant under scale transformations. The other parameters contribute to the scaling of the density. We define the parameter $c$ as the *unique scale parameter* such that

$$c^2 = \begin{cases} \frac{2\,\lambda}{\alpha^2} & \text{if } \delta = 0 \text{ for } \lambda > 0,\ \bar{\alpha} = 0, \\ \frac{\delta^2\,K_{\lambda+1}(\bar{\alpha})}{\bar{\alpha}\,K_\lambda(\bar{\alpha})} & \text{otherwise.} \end{cases} \tag{1.2.27}$$

It can be shown that as $\lambda \to \pm\infty$ and/or $\bar{\alpha} \to \infty$ the SGH density asymptotically approaches the Gaussian density.

To illustrate certain typical SGH densities we shall describe four special cases of the SGH density in the sequel. These coincide with log-return densities of important asset price models suggested in the literature.

### Student *t* Density (*)

Praetz (1972) and Blattberg & Gonedes (1974) proposed for log-returns a Student $t$ density with degrees of freedom $n > 0$. This is also the log-return density that arises from observations over long periods of time generated by the *minimal market model* (MMM), which will be derived in Chap. 13, see also Platen (2001). This density is obtained from the above SGH density for the shape parameters $\lambda = -\frac{1}{2}n < 0$ and $\bar{\alpha} = 0$, where $\alpha = 0$ and $\delta = \varepsilon\sqrt{n}$. Using these parameter values the Student $t$ density function for $X$ has then the form

$$f_X(x) = \frac{\Gamma(\frac{n+1}{2})}{\varepsilon\sqrt{\pi\, n}\,\Gamma(\frac{n}{2})} \left(1 + \frac{(x-\mu)^2}{\varepsilon^2\, n}\right)^{-\frac{n+1}{2}} \tag{1.2.28}$$

for $x \in \Re$, where $\Gamma(\cdot)$ is again the gamma function, see (1.2.10). Equation (1.2.28) expresses a generalization of the probability density (1.2.17) of a central Student $t$ distributed random variable with $n$ degrees of freedom. The Student $t$ density is a three parameter density. The degree of freedom $n = -2\lambda$ is the shape parameter, with smaller $n$ implying larger tail heaviness for the density. This means that there is a larger probability of extreme values. Furthermore, when the degrees of freedom increase, that is $n \to \infty$, then the Student $t$ density asymptotically approaches the Gaussian density. We plot in Fig. 1.2.7 the central Student $t$ density in logarithmic scale in dependence on the degrees of freedom $n$.

### Normal-Inverse Gaussian Density (*)

Barndorff-Nielsen (1995) proposed log-returns to follow a normal-inverse Gaussian mixture distribution. The corresponding density arises from the SGH density when the shape parameter $\lambda = -\frac{1}{2}$ is chosen. For this parameter value it follows by (1.2.24) that the probability density function of $X$ is then

$$f_X(x) = \frac{\sqrt{\bar{\alpha}}\,\exp\{\bar{\alpha}\}}{c\,\pi} \left(1 + \frac{(x-\mu)^2}{\bar{\alpha}\,c^2}\right)^{-\frac{1}{2}} K_1\left(\bar{\alpha}\sqrt{1 + \frac{(x-\mu)^2}{\bar{\alpha}\,c^2}}\right) \tag{1.2.29}$$

for $x \in \Re$, where $c^2 = \frac{\delta^2}{\bar{\alpha}}$. The normal-inverse Gaussian density is a three parameter density. The parameter $\bar{\alpha}$ is the shape parameter for the tails with smaller $\bar{\alpha}$ implying larger tail heaviness. Furthermore, when $\bar{\alpha} \to \infty$ the normal-inverse Gaussian density asymptotically approaches the Gaussian density. Figure 1.2.8 shows the normal-inverse Gaussian density in logarithmic scale in dependence on the shape parameter $\bar{\alpha}$.

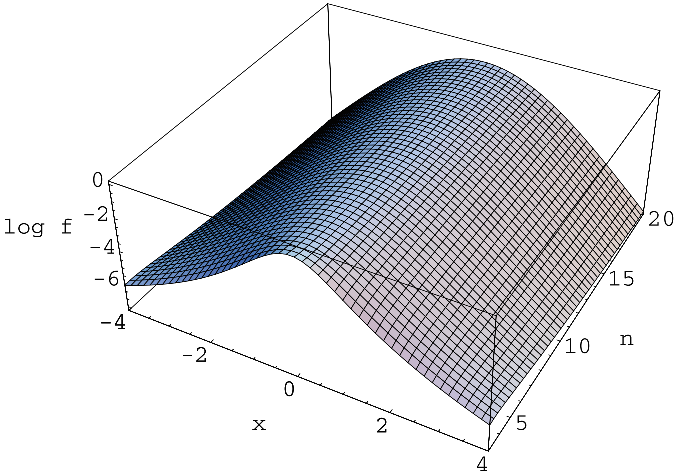



**Fig. 1.2.7.** Student $t$ density under log scale

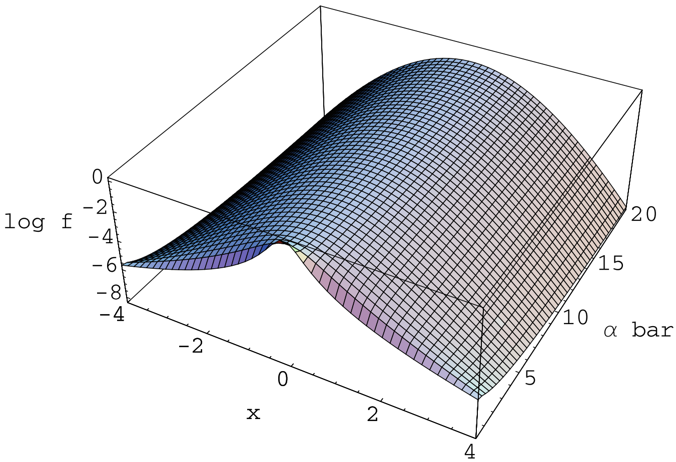



**Fig. 1.2.8.** Normal-inverse Gaussian density under log scale

## Hyperbolic Density (*)

Eberlein & Keller (1995) and Küchler, Neumann, Sørensen & Streller (1999) proposed models, where log-returns appear to be hyperbolicly distributed. This occurs for the choice of the shape parameter $\lambda = 1$ in the SGH density. Using this parameter value the probability density function of $X$ is

$$f_X(x) = \frac{1}{2\,\delta\,K_1(\bar{\alpha})}\,\exp\left\{-\bar{\alpha}\,\sqrt{1+\frac{(x-\mu)^2}{\delta^2}}\right\} \tag{1.2.30}$$

for $x \in \Re$, where

$$\delta^2 = \frac{c^2\,\bar{\alpha}\,K_1(\bar{\alpha})}{K_2(\bar{\alpha})}.$$

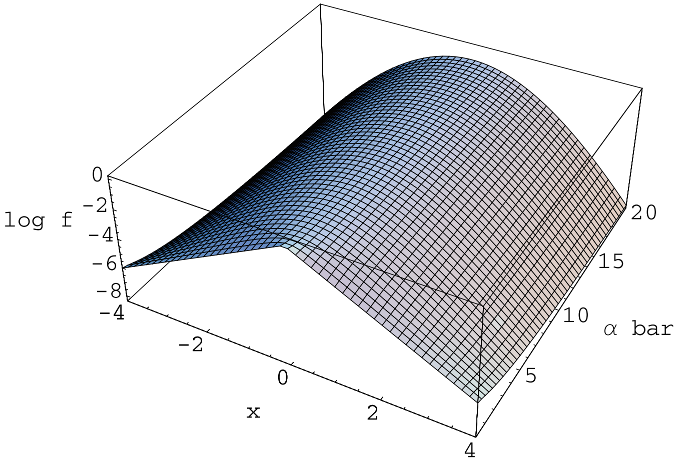



**Fig. 1.2.9.** Hyperbolic density under log scale

The hyperbolic density is a three parameter density. The parameter $\bar{\alpha}$ is the shape parameter with smaller $\bar{\alpha}$ implying larger tail heaviness. Furthermore, when $\bar{\alpha} \to \infty$ the hyperbolic density asymptotically approaches the Gaussian density. In Fig. 1.2.9 we graph the hyperbolic density in a logarithmic scale.

### Variance Gamma Density (*)

Madan & Seneta (1990) proposed that log-returns are distributed with a normal-variance gamma mixture distribution. This case is obtained when the shape parameters are such that $\lambda > 0$ and $\bar{\alpha} = 0$, that is, $\delta = 0$ and $\alpha = \frac{\sqrt{2\,\lambda}}{c}$. With these parameter values the probability density function of $X$ is

$$f_X(x) = \frac{\sqrt{\lambda}}{c\sqrt{\pi}\,\Gamma(\lambda)\,2^{\lambda-1}} \left(\sqrt{2\,\lambda}\,\frac{|x-\mu|}{c}\right)^{\lambda-\frac{1}{2}} K_{\lambda-\frac{1}{2}}\left(\sqrt{2\,\lambda}\,\frac{|x-\mu|}{c}\right) \quad (1.2.31)$$

for $x \in \Re$. The variance gamma density is a three parameter density. The parameter $\lambda$ is the shape parameter with smaller $\lambda$ implying larger tail heaviness. Furthermore, when $\lambda \to \infty$ the variance gamma density asymptotically approaches the Gaussian density. Figure 1.2.10 plots the logarithm of the variance gamma density.

The densities of the Student $t$, normal inverse Gaussian, hyperbolic and variance gamma distribution look very similar when plotted directly. However, their tail densities highlight significant differences. One can see, for instance, that for large $\bar{\alpha}$ and/or large $|\lambda|$ the densities are all close to the Gaussian density. Therefore, we have plotted the corresponding densities in logarithmic scale. In general, it is a challenging problem to identify for log-returns the type of distributions that fits best observed data, as will be discussed later on.

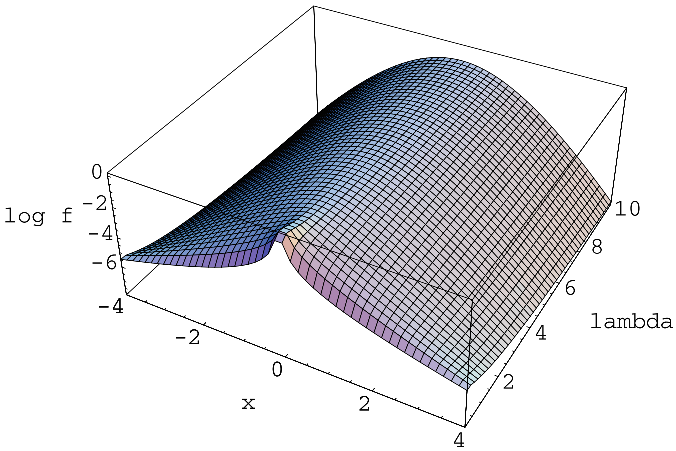



**Fig. 1.2.10.** Variance gamma density under log scale

## 1.3 Moments of Random Variables

Figure 1.1.2 clearly shows that stock log-returns can vary considerably. Therefore, it is important to provide measures for the variability of random variables. Moments, which we shall introduce in the following, provide the most common variability measures.

### Mean

The first of these moments is the arithmetic average that is weighted by the likelihood of occurrence. It is usually called the *mean*, *expectation* or simply *first moment* of the given random variable $X$ and is denoted by $E(X)$. For a *discrete random variable* $X$ the *mean* is defined as

$$\mu_X = E(X) = \sum_{i=0}^{\infty} x_i \, p_i, \tag{1.3.1}$$

where the summation is over all indices of the possible values taken by the random variable. This definition of the mean is readily suggested by the relative frequency interpretation of the probabilities that we discussed in Sect. 1.1.

For example, in the case of a two-point distributed random variable $X$, which takes the value $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$, we have the mean

$$\mu_X = x_1 \, p_1 + x_2 \, (1 - p_1) = x_2 + (x_1 - x_2) \, p_1. \tag{1.3.2}$$

Another example is obtained by computing the mean for the Poisson distribution with the probabilities (1.1.30). Here we have for $X \sim P(\lambda)$ the mean

$$\mu_X = \sum_{i=0}^{\infty} i \, \frac{\lambda^i}{i!} \, \exp\{-\lambda\} = \lambda. \tag{1.3.3}$$

When a *continuous random variable* has a probability density $f_X$, then the corresponding expression for its *mean* is

$$\mu_X = E(X) = \int_{-\infty}^{\infty} x \, f_X(x) \, dx. \tag{1.3.4}$$

One may say that the product $f_X(x)\,dx$ approximates the probability that $X$ takes its value in the interval $(x, x + dx)$. Note the similarity between (1.3.4) and (1.3.1).

Since $X$ is a random variable defined on $\Omega$, then (1.3.1) and (1.3.4) can both be equivalently expressed as an integral with respect to the measure $P$. That is, we can write

$$E(X) = \int_{\Omega} X(\omega) \, dP(\omega). \tag{1.3.5}$$

Of course, the above definitions for the mean assume that the summation over the possibly infinite series (1.3.1) and the integral (1.3.4) actually exist, that is, they are finite and well defined for each subset of $\Omega$. This is not always the case, as can be seen from Exercise 1.12 at the end of this chapter. To ensure that the corresponding means are well defined and exist, a necessary and sufficient condition is that $X$ is *integrable*, that is,

$$E(|X|) = \int_{\Omega} |X(\omega)| \, dP(\omega) < \infty. \tag{1.3.6}$$

If $E(|X|) = \infty$, then we say $X$ is not integrable and $E(X)$ does not exist. However, there is no problem in formally defining the mean, even if $E(X) < \infty$ or $E(|X|) = \infty$.

Furthermore, for $p \geq 1$ we say that $X$ is *p-integrable*, if

$$E(|X|^p) = \int_{\Omega} |X(\omega)|^p \, dP(\omega) < \infty. \tag{1.3.7}$$

In particular, if (1.3.7) holds for the case $p = 2$ we call the random variable $X$ *square integrable.*

Let us now compute the means of certain continuous random variables introduced in Sect. 1.2:

The mean of a $U(a, b)$ uniformly distributed random variable $X$ is according to (1.3.4) and (1.2.3) of the form

$$\mu_X = \int_{-\infty}^{\infty} x \, f_X(x) \, dx = \int_a^b \frac{x}{b-a} \, dx = \frac{(a+b)}{2}. \tag{1.3.8}$$

For a random variable $X \sim Exp(\lambda)$ with the exponential distribution (1.2.4) one obtains

$$\mu_X = \int_0^\infty x\,\lambda\,\exp\{-\lambda\,x\}\,dx = \frac{1}{\lambda}. \tag{1.3.9}$$

For a Gaussian distributed random variable $X \sim N(\mu, \sigma^2)$ with density (1.2.6) its mean is given by

$$\mu_X = \int_{-\infty}^\infty \frac{x}{\sqrt{2\,\pi}\,\sigma}\,\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \mu. \tag{1.3.10}$$

One can show that a gamma distributed random variable $X \sim G(p, \alpha)$ with density (1.2.9) has mean

$$\mu_X = \int_0^\infty x\,\frac{\alpha^p}{\Gamma(p)}\,\exp\{-\alpha\,x\}\,x^{p-1}\,dx = \frac{p}{\alpha}. \tag{1.3.11}$$

Finally, we mention that a central Student $t$ distributed random variable $X \sim t(n)$ with $n > 1$ degrees of freedom has mean zero, that is

$$\mu_X = \int_{-\infty}^\infty x\,\frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\,\sqrt{\pi\,n}}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 0. \tag{1.3.12}$$

We could add further examples but it should now be clear how to obtain the mean of a random variable with a given density.

## Variance

A measure for the spread around the mean $\mu_X$ of the values taken by a random variable $X$ is given by its *variance* $\sigma_X^2$, denoted also by $\mathrm{Var}(X)$, which is defined as

$$\sigma_X^2 = \mathrm{Var}(X) = E\left((X - E(X))^2\right) = E((X-\mu_X)^2), \tag{1.3.13}$$

provided that the expression (1.3.13) is finite. Consequently, the variance, also called the *second central moment*, is always nonnegative. The square root of the variance, $\sigma_X = \sqrt{\sigma_X^2}$, is called the *standard deviation* of $X$. Note that if $\mathrm{Var}(X) = 0$, then

$$P(X = E(X)) = 1. \tag{1.3.14}$$

For a two-point distributed random variable $X$, taking values $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$, its variance is given by

$$\sigma_X^2 = p_1\,(1-p_1)\,(x_2 - x_1)^2. \tag{1.3.15}$$

For a Poisson distributed random variable $X \sim P(\lambda)$ with intensity $\lambda$ we obtain from (1.1.30) and (1.3.3) the variance

$$\sigma_X^2 = \sum_{i=0}^{\infty} (i - \lambda)^2 \, \frac{\lambda^i}{i\,!} \, \exp\{-\lambda\} = \lambda, \tag{1.3.16}$$

which equals its mean as given by (1.3.3).

It is easy to check that a $U(a, b)$ uniformly distributed random variable $X$ with density (1.2.3) and mean (1.3.8) has variance

$$\sigma_X^2 = \int_a^b \left(x - \frac{a+b}{2}\right)^2 \frac{1}{b-a} \, dx = \frac{(b-a)^2}{12}. \tag{1.3.17}$$

The variance of an exponentially distributed random variable $X \sim Exp(\lambda)$ is, according to (1.2.4) and (1.3.9), given by

$$\sigma_X^2 = \int_0^\infty \left(x - \frac{1}{\lambda}\right)^2 \lambda \, \exp\{-\lambda \, x\} \, dx = \lambda^{-2}. \tag{1.3.18}$$

An $N(\mu, \sigma^2)$ distributed Gaussian random variable $X$ with density (1.2.6) can be shown to have a variance that equals $\sigma^2$, that is

$$\sigma_X^2 = \int_{-\infty}^\infty (x - \mu)^2 \, \frac{1}{\sqrt{2\pi}\,\sigma} \, \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = \sigma^2. \tag{1.3.19}$$

The variance of a gamma distributed random variable $X \sim G(p, \alpha)$ with density (1.2.9) is of the form

$$\sigma_X^2 = \int_0^\infty \left(x - \frac{p}{\alpha}\right)^2 \frac{\alpha^p}{\Gamma(p)} \, \exp\{-\alpha \, x\} \, x^{p-1} \, dx = \frac{p}{\alpha^2}. \tag{1.3.20}$$

Finally, for a central Student $t$ distributed random variable $X \sim t(n)$ we obtain from (1.2.17) and (1.3.12) the variance

$$\sigma_X^2 = \int_{-\infty}^\infty x^2 \, \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\,\sqrt{\pi\,n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = \frac{n}{n-2}, \tag{1.3.21}$$

as long as we have degrees of freedom $n > 2$. A central Student $t$ distribution with $n \leq 2$ degrees of freedom has no finite variance.

### Skewness

Some random variables have probability densities with non-symmetric shapes. One way to measure their asymmetry is to compute the skewness $\beta_X$ of the corresponding density. The *skewness* of a random variable $X$ is measured using the centralized and normalized third moment, that is

$$\beta_X = E\left(\left(\frac{X - \mu_X}{\sigma_X}\right)^3\right). \tag{1.3.22}$$

For a random variable $X$ we say that its density is called *positively skewed* if $\beta_X > 0$, *negatively skewed* if $\beta_X < 0$ and *symmetric* if $\beta_X = 0$.

For a two-point distributed random variable, taking values $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$, we obtain, using (1.3.2) and (1.3.15), the expression

$$\beta_X = \sqrt{p_1\,(1-p_1)}\,(2p_1 - 1). \tag{1.3.23}$$

Consequently, there is no skewness for $p_1 = 0.5$ in the two-point distribution.

For a Poisson distributed random variable $X$ with probabilities given in (1.1.30) its skewness, using (1.3.3) and (1.3.16), has the form

$$\beta_X = \sum_{i=0}^{\infty} \left(\frac{i-\lambda}{\sqrt{\lambda}}\right)^3 \frac{\lambda^i}{i\,!}\,\exp\{-\lambda\} = \frac{1}{\sqrt{\lambda}}, \tag{1.3.24}$$

which means that the corresponding Poisson distribution is positively skewed.

Furthermore, we note from (1.2.3), (1.3.8) and (1.3.17) that a $U(a,b)$ uniformly distributed random variable $X$ has zero skewness since

$$\beta_X = \int_a^b \left(\frac{x - \frac{a+b}{2}}{\frac{(b-a)}{\sqrt{12}}}\right)^3 \frac{1}{b-a}\,dx = 0. \tag{1.3.25}$$

This confirms the view that the shape of the uniform density in Fig. 1.2.1 is symmetric around its mean. On the other hand, an exponentially distributed random variable $X$ with density (1.2.5) can be shown to have fixed skewness with value

$$\beta_X = \int_0^\infty \left(\frac{x - \frac{1}{\lambda}}{\frac{1}{\lambda}}\right)^3 \lambda\,\exp\{-\lambda\,x\}\,dx = 2, \tag{1.3.26}$$

see also Fig. 1.2.2.

One can show for an $N(\mu, \sigma^2)$ distributed Gaussian random variable $X$, using (1.3.10) and (1.3.19), that its density (1.2.6) is symmetric and thus has no skewness. That is, we have

$$\beta_X = \int_{-\infty}^\infty \left(\frac{x-\mu}{\sigma}\right)^3 \frac{1}{\sqrt{2\pi}\,\sigma}\,\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = 0. \tag{1.3.27}$$

The Gaussian distribution is obviously not a reasonable distribution if one has to model a strongly skewed random variable.

By (1.3.11) and (1.3.20) a gamma distributed random variable $X \sim G(p, \alpha)$ has positive skewness

$$\beta_X = \int_0^\infty \left(\frac{x - \frac{p}{\alpha}}{\frac{\sqrt{p}}{\alpha}}\right)^3 \frac{\alpha^p}{\Gamma(p)}\,\exp\{-\alpha\,x\}\,x^{p-1}\,dx = \frac{2}{\sqrt{p}} \tag{1.3.28}$$

for $p > 0$. This is also indicated by inspection of its density, as displayed in Fig. 1.2.5.

Finally, we mention that the density of a central Student $t$ distributed random variable $X \sim t(n)$ with $n > 3$ degrees of freedom is symmetrically skewed, that is,

$$\beta_X = \int_{-\infty}^{\infty} \left( \frac{x}{\sqrt{\frac{n}{n-2}}} \right)^3 \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi\, n}} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 0. \tag{1.3.29}$$

This fact is also apparent from the shape of the density shown in Fig. 1.2.6.

## Kurtosis

Extreme values of returns are very important in a range of financial applications. A large negative log-return of a stock index, which may arise in a major market correction, can considerably change the overall short term performance of a portfolio. The likelihood of such extreme values can be reflected by the *kurtosis* $\kappa_X$, which is the centralized and normalized fourth moment, that is

$$\kappa_X = E\left( \left( \frac{X - \mu_X}{\sigma_X} \right)^4 \right). \tag{1.3.30}$$

For a two-point distributed random variable $X$ taking values $x_1$ with probability $p_1$ and $x_2$ with probability $p_2 = 1 - p_1$ we obtain, using (1.3.2) and (1.3.15),

$$\kappa_X = \frac{\left(\frac{1}{3} - p_1 + p_1^2\right)}{3 p_1 (1 - p_1)}. \tag{1.3.31}$$

A Poisson distributed random variable $X$ with intensity $\lambda$ yields according to (1.3.3) and (1.3.16) a kurtosis of the form

$$\kappa_X = \sum_{i=0}^{\infty} \left( \frac{i - \lambda}{2} \right)^4 \frac{\lambda^i}{i\,!} \exp\{-\lambda\} = 3 + \frac{1}{\lambda}. \tag{1.3.32}$$

The kurtosis of a $U(a, b)$ uniformly distributed random variable $X$ by (1.3.8) and (1.3.17) is given by the constant

$$\kappa_X = \int_b^a \left( \frac{x - \frac{a+b}{2}}{\frac{b-a}{\sqrt{12}}} \right)^4 dx = 1.8. \tag{1.3.33}$$

For an exponentially distributed random variable $X$ it can be shown, using (1.3.9) and (1.3.18), that it has a constant kurtosis with

$$\kappa_X = \int_0^{\infty} \left( \frac{x - \frac{1}{\lambda}}{\frac{1}{\lambda}} \right)^4 \lambda \exp\{-\lambda\, x\}\, dx = 9. \tag{1.3.34}$$

An $N(\mu, \sigma^2)$ distributed Gaussian random variable $X$ has by (1.3.10) and (1.3.19) the constant kurtosis

$$\kappa_X = \int_{-\infty}^{\infty} \left(\frac{x-\mu}{\sigma}\right)^4 \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx = 3. \tag{1.3.35}$$

If the kurtosis $\kappa_X$ of a random variable $X$ is greater than 3, then this random variable, its density and also its distribution are called *leptokurtic*.

The kurtosis $\kappa_X$ of a gamma distributed random variable $X \sim G(\alpha, \beta)$ is by (1.3.11) and (1.3.20) of the value

$$\kappa_X = \int_0^{\infty} \left(\frac{x-\frac{p}{\alpha}}{\frac{\sqrt{p}}{\alpha}}\right)^4 \frac{\alpha^p}{\Gamma(p)} \exp\{-\alpha\, x\}\, x^{p-1}\, dx = \frac{3(p+2)}{p}, \tag{1.3.36}$$

which is larger for smaller $p > 0$. Thus a gamma distributed random variable is leptokurtic.

Finally, by (1.3.12) and (1.3.21) we have for a Student $t$ distributed random variable $X \sim t(n)$ the kurtosis

$$\kappa_X = \int_0^{\infty} \left(\frac{x}{\sqrt{\frac{n}{n-2}}}\right)^4 \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\,\sqrt{\pi\, n}} \left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}} dx = 3\left(\frac{n-2}{n-4}\right). \tag{1.3.37}$$

This is finite only for $n > 4$ degrees of freedom. This type of random variable is also leptokurtic. The Student $t$ density approaches asymptotically a Gaussian density as $n \to \infty$. This is also reflected in its limiting kurtosis of three as $n \to \infty$.

In Fig. 1.3.1 we plot the kurtosis

$$\kappa_X = \frac{3\, K_\lambda(\bar{\alpha})\, K_{\lambda+2}(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})^2} \tag{1.3.38}$$

for $(\bar{\alpha}, \lambda) \in [0, \infty) \times \Re$ of a symmetric generalized hyperbolic distributed random variable, with density given in (1.2.24), in dependence on the two shape parameters $\lambda$ and $\bar{\alpha}$. Note that the kurtosis is not finite for a Student $t$ distribution with degrees of freedom not greater than four. The hyperbolic distribution yields only a kurtosis of six, which limits its applicability as a log-return distribution because a much higher kurtosis is typically observed for log-returns.

It is an empirical stylized fact, which we shall document later on, that the probability densities of log-returns of stock indices, stock prices and exchange rates have much thicker tails than that of a Gaussian density, which means they are leptokurtic. In some cases the kurtosis of a fitted model may not even be finite. For convenience Table 1.3.1 summarizes the moments for several distributions discussed previously.

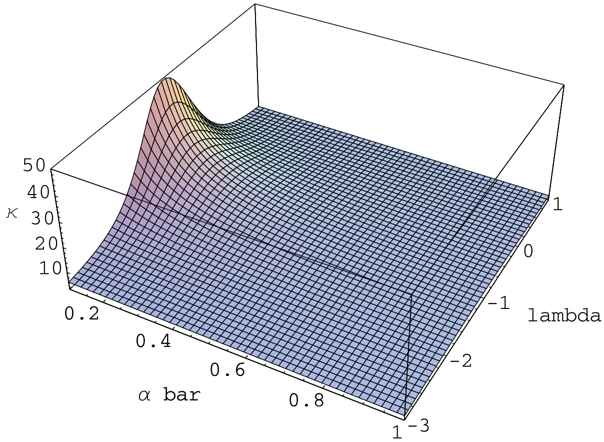



**Fig. 1.3.1.** Kurtosis of SGH random variable in dependence on shape parameters $\lambda$ and $\bar{\alpha}$

**Table 1.3.1.** Moments of some distributions

| $X$ distributed as | | $\mu_X$ | $\sigma_X^2$ | $\beta_X$ | $\kappa_X$ |
|---|---|---|---|---|---|
| Poisson | $P(\lambda)$ | $\lambda$ | $\lambda$ | $\lambda^{\frac{1}{2}}$ | $3+\lambda^{-1}$ |
| Uniform | $U(a,b)$ | $\frac{a+b}{2}$ | $\frac{(b-a)^2}{12}$ | 0 | 1.8 |
| Exponential | $Exp(\lambda)$ | $\lambda^{-1}$ | $\lambda^{-2}$ | 2 | 9 |
| Normal | $N(\mu,\sigma^2)$ | $\mu$ | $\sigma^2$ | 0 | 3 |
| Gamma | $G(p,\alpha)$ | $\frac{p}{\alpha}$ | $\frac{p}{\alpha^2}$ | $2\,p^{-\frac{1}{2}}$ | $3\left(\frac{p+2}{p}\right)$ |
| Chi-square | $\chi^2(n)$ | $n$ | $2\,n$ | $\frac{2\sqrt{2}}{n}$ | $3\,\frac{(n+4)}{n}$ |
| Central Student $t$ | $t(n)$ | 0 | $\frac{n}{n-2}$ | 0 | $3\left(\frac{n-2}{n-4}\right)$ |

Finally, let us mention that sometimes the notion of *excess kurtosis* $\underline{\kappa}_X$ of a random variable $X$ is used. This is simply the difference between the kurtosis $\kappa_X$ and the value 3 for the Gaussian kurtosis, that is

$$\underline{\kappa}_X = \kappa_X - 3. \tag{1.3.39}$$

## Higher Order Moments

In general, a new random variable is obtained when we transform or combine random variables by functions or arithmetic operations. For a general transformation of a random variable, however, we need to observe some restrictions on the transforming function $g$. These restrictions follow from measurability constraints to ensure that the resulting variable is still a random variable as defined in Sect. 1.1. More precisely, the function $g$ should be Borel measurable. This is the case when $g$ is, for instance, continuous or piecewise continuous. For more details on these issues the reader is referred to Shiryaev (1984).

When $Y = g(X)$ is a random variable, its expected value, or mean, is

$$E(g(X)) = \sum_{i \in \mathcal{N}} g(x_i)\, p_i \tag{1.3.40}$$

when $X$ is discrete, or

$$E(g(X)) = \int_{-\infty}^{\infty} g(x)\, f_X(x)\, dx \tag{1.3.41}$$

when $X$ is continuous with a density function $f_X$. It should be noted that these expectations may be undefined or infinite.

Typical functions of a random variable $X$ are the polynomials $g(x) = x^p$ or $g(x) = (x - \mu_X)^p$ for integers $p \geq 1$. The resulting expected value of $Y = g(X)$ is then called the $p$th *moment*

$$\alpha_p = E(X^p) \tag{1.3.42}$$

or the $p$th *central moment*

$$m_p = E((X - \mu_X)^p), \tag{1.3.43}$$

respectively. For instance, the *variance*

$$\sigma_X^2 = m_2 = \mathrm{Var}(X) = E\left((X - \mu_X)^2\right) \tag{1.3.44}$$

is the *second central moment* of $X$. We have the following important relationships between moments and central moments:

$$m_1 = 0, \quad m_2 = \alpha_2 - \alpha_1^2, \quad m_3 = \alpha_3 - 3\,\alpha_1\,\alpha_2 + 2\,\alpha_1^3$$
$$m_4 = \alpha_4 - 4\,\alpha_1\,\alpha_3 + 6\,\alpha_1^2\,\alpha_2 - 3\,\alpha_1^4. \tag{1.3.45}$$

If we use the transform function

$$g(x) = \left(\frac{x - \mu_X}{\sigma_X}\right)^p, \tag{1.3.46}$$

then we obtain the $p$th *normalized central moment*, $p \geq 1$. As previously mentioned, the *skewness* $\beta_X$ is the third normalized central moment and the *kurtosis* $\kappa_X$ is the fourth normalized central moment. Obviously, the first normalized central moment is zero and the second normalized central moment equals one.

Moments provide important information about the given random variable. Note that the higher order moments need not always provide additional information. For example, the Gaussian distribution is completely characterized by its first two moments, its mean $\mu$ and variance $\sigma^2$.

For an $N(\mu, \sigma^2)$ Gaussian distributed random variable $X$ one can show that its $p$th normalized central moment has the form

$$E\left(\left(\frac{X - \mu}{\sigma}\right)^p\right) = \begin{cases} 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2j-1) & \text{for} \quad p = 2j \\ 0 & \text{for} \quad p = 2j - 1, \end{cases} \tag{1.3.47}$$

where $j \in \mathcal{N}$.

The Poisson distribution is already fully characterized by its mean $\lambda$. For a gamma distributed random variable $X \sim \Gamma(p;\alpha)$ the $r$th moment has the form

$$E(X^r) = \frac{\Gamma(p+r)}{\alpha^r\,\Gamma(p)} \tag{1.3.48}$$

for $\alpha > 0,\ p > 0$ and $r > -p$. With (1.3.46) and (1.3.20) we then obtain in this case the $r$th *normalized moment*

$$E\left(\left(\frac{X}{\sigma_X}\right)^r\right) = p^{-\frac{r}{2}}\,\frac{\Gamma(p+r)}{\Gamma(p)} \tag{1.3.49}$$

for $r > -p$, which does not depend on $\alpha$.

## Properties of Moments

General properties of moments can be used to gain an understanding of, and insight into, many of the problems that arise in quantitative finance. Using basic properties of integrals, or equivalently those of infinite series in the discrete case, the first moment, see (1.3.5), inherits the *additivity property*. That is

$$E(a\,X_1 + b\,X_2) = a\,E(X_1) + b\,E(X_2) \tag{1.3.50}$$

for any two random variables $X_1$, $X_2$ and any two real numbers $a$, $b$, provided the expectations are finite.

When $P(X_1 \leq X_2) = 1$, then we have for the first moment the *monotonicity property*

$$E(X_1) \leq E(X_2). \tag{1.3.51}$$

Moreover, *Jensen's inequality*

$$g(E(X)) \leq E(g(X)) \tag{1.3.52}$$

holds for any convex function $g : \Re \to \Re$, which is a function satisfying the relation

$$g(\lambda\,x + (1-\lambda)\,y) \leq \lambda\,g(x) + (1-\lambda)\,g(y)$$

for all $x,\ y \in \Re$ and $\lambda \in [0,1]$. In particular, for $g(x) = |x|,\ g(x) = x^2$ and $g(x) = \exp\{x\}$ this implies

$$|E(X)| \leq E\left(|X|\right) \tag{1.3.53}$$

$$|E(X)| \leq \sqrt{E(X^2)}. \tag{1.3.54}$$

and

$$\exp\{E(X)\} \leq E(\exp\{X\}). \tag{1.3.55}$$

If $E(|X|^s)$ is finite for some $s > 0$, then for all $r \in (0,s]$ and $a \in \Re$ we have the *Lyapunov inequality*

$$(E(|X-a|^r))^{\frac{1}{r}} \leq (E(|X-a|^s))^{\frac{1}{s}} . \tag{1.3.56}$$

The Lyapunov inequality shows that, if the $s$th moment of a random variable is finite, then any $r$th moment for $r \in (0, s]$ is also finite. For any random variable $X$ we have the *Markov inequality*

$$P(X \geq a) \leq \frac{1}{a} E(|X|) \tag{1.3.57}$$

for all $a > 0$. From this we can deduce the widely used *Chebyshev inequality*

$$P(|X - E(X)| \geq a) \leq \frac{1}{a^2} \operatorname{Var}(X) \tag{1.3.58}$$

for all $a > 0$. For two random variables $X_1$ and $X_2$ the *Cauchy-Schwartz inequality* provides the estimate

$$|E(|X_1 X_2|)| \leq \sqrt{E((X_1)^2)\, E((X_2)^2)}. \tag{1.3.59}$$

Further properties of moments can be found at the end of the following section.

### Conditional Expectations

The notion of conditional expectation is central to many of the ideas that arise in probability theory and stochastic calculus. The mean value or expectation $E(X)$ is the coarsest estimate that we have for an integrable random variable $X$, that is, for which $E(|X|) < \infty$, see (1.3.6). If we know that some event $A$ has occurred we may be able to improve on this estimate. For instance, suppose that the event $A = \{\omega \in \Omega : X(\omega) \in [a, b]\}$ has occurred. Then in evaluating our estimate of the value of $X$ we need only to consider corresponding values of $X$ and weight them according to their likelihood of occurrence, which is now the conditional probability, see (1.1.13), given this event. The resulting estimate is called the *conditional expectation* of $X$ given event $A$ and is denoted by $E(X| A)$.

For a discrete random variable $X$ with possible values in a set of real numbers $\mathcal{X} = \{\ldots, x_{-1}, x_0, x_1, \ldots\}$ the conditional probability for the outcome $x_i$ given the event $A = \{\omega \in \Omega : X(\omega) \in [a, b]\}$ satisfies

$$P(X = x_i \,|\, A) = \begin{cases} 0 & \text{for} \quad x_i \notin [a, b] \\ \frac{p_i}{\sum_{a \leq x_j \leq b} p_j} & \text{for} \quad x_i \in [a, b] \end{cases} \tag{1.3.60}$$

and so the conditional expectation is given by

$$E(X \,|\, A) = \sum_{x_i \in \mathcal{X}} x_i \, P(X = x_i \,|\, A) = \frac{\sum_{a \leq x_i \leq b} x_i \, p_i}{\sum_{a \leq x_j \leq b} p_j}. \tag{1.3.61}$$

More generally, for an integrable random variable $X$ and an event $A \in \mathcal{A}$ the *conditional expectation* $E(X \,|\, A)$ is given by

$$E(X \mid A) = \frac{\int_A X(\omega)\, dP(\omega)}{P(A)}. \tag{1.3.62}$$

For a continuous random variable $X$ with a density function $f_X$ the corresponding *conditional density* is

$$f_X(x \mid A) = \begin{cases} 0 & \text{for} \quad x < a \ \text{ or } \ b < x \\ \frac{f_X(x)}{\int_a^b f_X(s)\, ds} & \text{for} \quad x \in [a, b] \end{cases}$$

with the *conditional expectation*

$$E(X \mid A) = \int_{-\infty}^{\infty} x\, f_X(x \mid A)\, dx = \frac{\int_a^b x\, f_X(x)\, dx}{\int_a^b f_X(x)\, dx}, \tag{1.3.63}$$

which is conditioned on the event $A$ and is thus a number.

More generally let $(\Omega, \mathcal{A}, P)$ be a given probability space with an integrable, see (1.3.6), random variable $X$. We denote by $\mathcal{S}$ a sub-sigma-algebra of $\mathcal{A}$, thus representing a coarser type of information than is given by $\mathcal{A}$. We then define the *conditional expectation* of $X$ with respect to the sub-sigma-algebra $\mathcal{S}$, which we denote by $E(X \mid \mathcal{S})$, as an $\mathcal{S}$-measurable function satisfying

$$\int_S E(X \mid \mathcal{S})(\omega)\, dP(\omega) = \int_S X(\omega)\, dP(\omega), \tag{1.3.64}$$

see Sect. 1.1, for all $S \in \mathcal{S}$. The Radon-Nikodym theorem, see Shiryaev (1984), guarantees the existence and uniqueness of the random variable $E\,(X \mid \mathcal{S})$ a.s. Note that $E\,(X \mid \mathcal{S})$ is a random variable defined on the coarser probability space $(\Omega, \mathcal{S}, P)$ and thus on $(\Omega, \mathcal{A}, P)$. However, $X$ is usually not a random variable on $(\Omega, \mathcal{S}, P)$, but when it is we have

$$E\left(X \,\middle|\, \mathcal{S}\right) = X, \tag{1.3.65}$$

which is the case when $X$ is $\mathcal{S}$-measurable.

Let us consider an example with a random variable $X(\omega) = \omega$ for $\omega \in [0, 1]$ with probability density $f_X(x) = 2x$ for $x \in [0, 1]$. We define the sigma-algebra $\mathcal{S}$ generated by the event $A = \{\omega \in [0, 0.5]\}$. It is then an easy calculation by using (1.3.63) to obtain the conditional expectation

$$E\,(X \mid \mathcal{S})\,(\omega) = \begin{cases} E(X \mid A) = \frac{1}{3} & \text{for} \quad \omega \in [0, 0.5] \\ E(X \mid A^c) = \frac{7}{9} & \text{for} \quad \omega \notin [0, 0.5], \end{cases}$$

where $P(A) = \frac{1}{4}$, $P(A^c) = \frac{3}{4}$ and $E(X) = \frac{2}{3}$.

For nested sigma-algebras $\mathcal{S} \subset \mathcal{T} \subset \mathcal{A}$ and an integrable random variable $X$ we have the *law of iterated conditional expectations*

$$E\left(E\left(X \,\middle|\, \mathcal{T}\right) \,\middle|\, \mathcal{S}\right) = E\left(X \,\middle|\, \mathcal{S}\right) \tag{1.3.66}$$

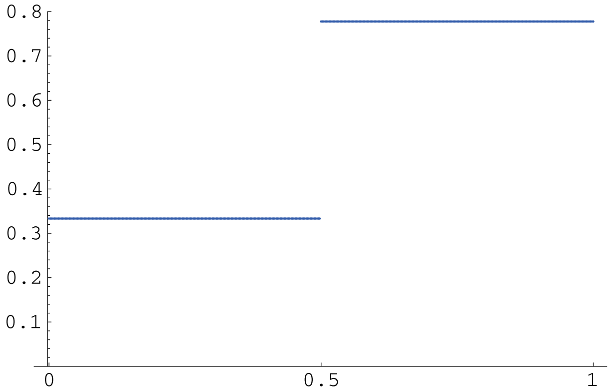



**Fig. 1.3.2.** Conditional expectation

a.s. and when $X$ is independent of the events in $\mathcal{S}$, see (1.1.16), we have

$$E\left(X \,\middle|\, \mathcal{S}\right) = E(X), \tag{1.3.67}$$

a.s. Setting $\mathcal{S} = \{\emptyset, \Omega\}$ it can be seen that

$$E(E\,(X \,|\, \mathcal{S})) = E(X). \tag{1.3.68}$$

This property is easy to check for the above example, where

$$E(X) = \frac{1}{4}\,\frac{1}{3} + \frac{3}{4}\,\frac{7}{9} = \frac{2}{3}.$$

Conditional expectations have similar properties to those of ordinary integrals such as linearity

$$E\left(\alpha\, X + \beta\, Y \,\middle|\, \mathcal{S}\right) = \alpha\, E\left(X \,\middle|\, \mathcal{S}\right) + \beta\, E\left(Y \,\middle|\, \mathcal{S}\right), \tag{1.3.69}$$

where $X$ and $Y$ are integrable random variables and $\alpha, \beta \in \Re$ are deterministic constants. In addition, if $X$ is $\mathcal{S}$-measurable, then

$$E(X\,Y \,|\, \mathcal{S}) = X\,E(Y \,|\, \mathcal{S}). \tag{1.3.70}$$

Furthermore, we have the order preserving property

$$E\left(X \,\middle|\, \mathcal{S}\right) \le E\left(Y \,\middle|\, \mathcal{S}\right) \tag{1.3.71}$$

if $X \le Y$ a.s.

The conditional expectation $E(X \,|\, \mathcal{S})$ is in some sense obtained by smoothing $X$ over the events in $\mathcal{S}$. Thus the finer the information set $\mathcal{S}$, the more $E(X \,|\, \mathcal{S})$ resembles the random variable $X$.

## Least-Squares Estimate

Let $\mathcal{S} \subset \mathcal{A}$ be a given sigma-algebra and $X$ a square integrable random variable on $(\Omega, \mathcal{A}, P)$. We shall show below that

$$E\left(\left(X - E\left(X \,\middle|\, \mathcal{S}\right)\right)^2\right) \leq E\left((X - Y)^2\right) \tag{1.3.72}$$

for all $\mathcal{S}$-measurable, square integrable random variables $Y$. Consequently, $E(X \,|\, \mathcal{S})$ is the *least-squares estimate* or best forecast for $X$ amongst the random variables $Y$ which are $\mathcal{S}$-measurable.

The conditional expectation $E(X \,|\, \mathcal{S})$ can therefore be interpreted as the best estimate, in a least-squares sense, for $X$ under the information given by the events in $\mathcal{S}$. In the case where $\mathcal{S}$ is the sigma-algebra of events generated by a random variable $Y$ we may also write $E(X \,|\, Y)$ for the conditional expectation $E(X \,|\, \mathcal{S})$. This notion of a least-squares estimate, or best forecast, is central to many ideas that arise in stochastic calculus and quantitative finance.

Since the inequality (1.3.72) has fundamental importance we derive it in the following few lines:

Let $Y$ be any square integrable $\mathcal{S}$-measurable random variable and $X$ be a square integrable random variable. Then with $Z = E(X \,\big|\, \mathcal{S})$ we obtain

$$\begin{aligned} E((X - Y)^2) &= E((X - Z + Z - Y)^2) \\ &= E((X - Z)^2) + 2\,E((X - Z)\,(Z - Y)) + E((Z - Y)^2). \end{aligned} \tag{1.3.73}$$

Using the above described properties of conditional expectations it follows that

$$\begin{aligned} E((X - Z)\,(Z - Y)) &= E\left(E\left((X - Z)\,(Z - Y) \,\middle|\, \mathcal{S}\right)\right) \\ &= E\left(E(X - Z \,\big|\, \mathcal{S})\,(Z - Y)\right) = E((Z - Z)\,(Z - Y)) = 0. \end{aligned}$$

Consequently, (1.3.73) is minimized by choosing $Y = Z = E(X \,\big|\, \mathcal{S})$, which proves (1.3.72).

## Moment Generating Functions (*)

The *cumulants* $k_1, k_2, \ldots$ of a random variable $X$ appear as coefficients of the power series expansion of its *Laplace transform* $\psi_X$, which is also called the *moment generating function*, and has the form

$$\psi_X(\lambda) = E(\exp\{\lambda\, X\}) = 1 + k_1\,\lambda + k_2\,\frac{\lambda^2}{2} + k_3\,\frac{\lambda^3}{3!} + k_4\,\frac{\lambda^4}{4!} + \ldots \tag{1.3.74}$$

for $\lambda \in \Re$ if $\psi_X(\lambda)$ is finite. Note that $\psi_X(\lambda)$ is always finite for $\lambda = 0$ but may be infinite for other values of $\lambda$. The derivatives of the Laplace transform with respect to $\lambda$ can be used to find the moments. The first four cumulants are related to the first moment and the central moments up to order four, see (1.3.45), by the equations

$$k_1 = \alpha_1 = \mu_X, \quad k_2 = m_2, \quad k_3 = m_3, \quad k_4 = m_4 - 3m_2^2. \tag{1.3.75}$$

The Laplace transform of an $N(\mu, \sigma^2)$ Gaussian distributed random variable $X$ is given by

$$\psi_X(\lambda) = E\left(\exp\{\lambda\, X\}\right) = \exp\left\{\lambda\,\mu + \frac{\lambda^2\,\sigma^2}{2}\right\} \tag{1.3.76}$$

for $\lambda \in \Re$. This Laplace transform can be used to obtain expectations for asset prices under the standard market model, which is the *lognormal* or *Black-Scholes model*. Under this model returns are normalized increments of exponentials of Gaussian random variables or, equivalently, the log-returns are Gaussian.

## Characteristic Functions (*)

Another important functional of a random variable $X$ is its *characteristic function* $\phi_X$, which is defined as the expectation

$$\phi_X(\theta) = E(\exp\{\imath\,\theta\, X\}), \tag{1.3.77}$$

for all $\theta \in \Re$, where $\imath$ denotes the *imaginary unit*, that is $\imath = \sqrt{-1}$. This function always exists and its absolute value is less than or equal to one, that is

$$|\phi_X(\theta)| \leq 1. \tag{1.3.78}$$

It can be used to identify uniquely the distribution of a given random variable. In this sense the characteristic function encapsulates all of the information content of the distribution of a random variable. For instance, the $p$th moment of $X$, if it exists, can be obtained by the formula

$$\alpha_p = E(X^p) = (-\imath)^p\, \frac{d^p}{(d\theta)^p}\, \phi_X(0). \tag{1.3.79}$$

The mean, variance, skewness and kurtosis can then be derived from these moments according to (1.3.45). For example, the characteristic function of the Poisson distribution with intensity $\lambda$ is from (1.1.31) given by

$$\begin{aligned} \phi_X(\theta) &= \sum_{n=0}^{\infty} \exp\{\imath\,\theta\, n\}\, \frac{\lambda^n}{n!}\, \exp\{-\lambda\} \\ &= \exp\{-\lambda\,(1 - \exp\{\imath\,\theta\})\} \end{aligned} \tag{1.3.80}$$

for $\theta \in \Re$. By using (1.2.6) the characteristic function of an $N(\mu, \sigma^2)$ Gaussian distributed random variable $X$ takes the form

$$\phi_X(\theta) = \int_{-\infty}^{\infty} \exp\{\imath\,\theta\,x\}\,\frac{1}{\sqrt{2\,\pi}\,\sigma}\,\exp\left\{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right\} dx$$

$$= \exp\left\{\imath\,\theta\,\mu - \theta^2\,\frac{\sigma^2}{2}\right\} \qquad (1.3.81)$$

for $\theta \in \Re$. Note the similarity with the Laplace transform (1.3.76). For a $G(p,\alpha)$ gamma distributed random variable, see (1.2.9), we obtain the expression

$$\phi_X(\theta) = \int_0^{\infty} \exp\{\imath\,\theta\,x\}\,\frac{\alpha^p}{\Gamma(p)}\,\exp\{-\alpha\,x\}\,x^{p-1}\,dx$$

$$= \left(\frac{\alpha}{\alpha - \imath\,\theta}\right)^p \qquad (1.3.82)$$

for $\theta \in \Re$. For $p = 1$ and $\alpha = \lambda$ this is the characteristic function of an exponential distributed random variable $X \sim Exp(\lambda)$, see (1.2.5). Characteristic functions are often used to analyze and characterize properties of random variables. They are closely related to Fourier transforms of the corresponding density function. A characteristic function $\phi_X(\theta)$ uniquely determines the density function $f_X(x)$ of a continuous random variable $X$. Indeed, the corresponding density function can be found by the *inverse Fourier transform*

$$f_X(x) = \frac{1}{2\,\pi}\int_{-\infty}^{\infty} \exp\{-\imath\,x\,\theta\}\,\phi_X(\theta)\,d\theta, \qquad (1.3.83)$$

see Lukacs (1960).

We mentioned at the end of Sect. 1.2 that the SGH distribution covers log-return distributions for a number of important asset price models. Using the notation and parametrization given there one obtains for the SGH distribution the characteristic function

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\,\frac{K_\lambda(\sqrt{(\alpha\,\delta)^2 + \delta^2\,\Delta\,\theta^2})\,(\alpha\,\delta)^\lambda}{K_\lambda(\alpha\,\delta)\,((\alpha\,\delta)^2 + \delta^2\,\Delta\,\theta^2)^{\frac{1}{2}\lambda}} \qquad (1.3.84)$$

for $\theta \in \Re$. Recall that $K_\lambda$ is the modified Bessel function of the third kind with index $\lambda$.

If one searches in probability or statistics textbooks and encyclopedias, then the characteristic function of the Student $t$ distribution is notably absent or erroneous. However, a simple closed form solution has been found in Hurst (1997) that is given by the formula

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\,\frac{K_{\frac{1}{2}n}(\varepsilon\,\sqrt{n\,\Delta}\,|\theta|)\,(\varepsilon\,\sqrt{n\,\Delta}\,|\theta|)^{\frac{1}{2}n}}{\Gamma(\frac{1}{2}\,n)\,2^{\frac{1}{2}\,n-1}} \qquad (1.3.85)$$

for all degrees of freedom $n > 0$ and $\theta \in \Re$. For the normal-inverse Gaussian distribution the characteristic function is

$$\phi_X(\theta) = \exp\left\{\imath\,\mu\,\Delta\,\theta + \bar{\alpha}\left(1 - \sqrt{1 + \frac{c^2\,\Delta\,\theta^2}{\bar{\alpha}}}\right)\right\} \tag{1.3.86}$$

for $\theta \in \Re$, where the parametrization is again as in (1.2.29). Furthermore, the hyperbolic distribution has the characteristic function

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\,\frac{\bar{\alpha}\,K_1\,(\sqrt{\bar{\alpha}^2 + \delta^2\,\Delta\,\theta^2})}{K_1(\bar{\alpha})\,\sqrt{\bar{\alpha}^2 + \delta^2\,\Delta\,\theta^2}} \tag{1.3.87}$$

for $\theta \in \Re$. Finally, a variance gamma distributed random variable $X$ has the characteristic function

$$\phi_X(\theta) = \exp\{\imath\,\mu\,\Delta\,\theta\}\left(1 + \frac{c^2\,\Delta\,\theta^2}{2\,\lambda}\right)^{-\lambda} \tag{1.3.88}$$

for $\theta \in \Re$. A convenient proof for the above results can be obtained by interpreting the above distributions as normal mixture distributions. This means that the random variable is assumed to be conditionally Gaussian distributed with independent random variance. For instance, a Student $t$ distribution with $n$ degrees of freedom is obtained when the inverse of the variance is chi-square distributed with $n$ degrees of freedom. If instead the variance is chi-square distributed, then a variance gamma distribution arises.

### Gaussian Shift (*)

In the context of option pricing, see Buchen & Konstandatos (2005), and other applications it can be useful to apply the following basic relation for *shifted Gaussian random variables*. Let $X \sim N(0,1)$ denote a standard Gaussian random variable, $\theta \in \Re$ a real valued constant and $H(\cdot)$ a real valued function of $x \in \Re$ with $|E(H(X+\theta))| < \infty$. Then it can be shown by exploiting the structure of the Gaussian density that the expectation of a shifted standard Gaussian random variable is of the form

$$E(H(X+\theta)) = E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\,X\right\}H(X)\right). \tag{1.3.89}$$

Interestingly, this allows one also to include the case of more general Gaussian random variables $Y = a + bX$ for $a, b \in \Re$ with mean $E(Y) = a$ and variance $\text{Var}(Y) = b^2$, where we derive the following relation from (1.3.89) for a real valued function $G(y) = G(a + b\,x)$

$$E(G(Y+\theta)) = E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\,X\right\}G(Y)\right). \tag{1.3.90}$$

This is an important relation because the function $G$ can be freely chosen. We shall see later on that the Gaussian shift forms, in principle, the basis for the probability measure transformation that is used in standard derivative pricing.

## 1.4 Joint Distributions and Random Vectors

For many practical applications we need to consider several random variables $X_1$, $X_2$, ..., $X_n$. For instance, these may represent the daily log-returns of all stocks in a market. This leads us to the introduction of joint distributions. The random variables may sometimes be interpreted as components of a vector-valued random variable, which is then called a random vector.

### Joint Distributions

As in the case of a single random variable, we can similarly form a distribution function for $n$ random variables $X_1$, $X_2$, ..., $X_n$, which are defined on the same probability space. The distribution function $F_{X_1,X_2,\cdots,X_n} : \Re^n \to [0,1]$ is called the *joint distribution function* and is defined by the relation

$$F_{X_1,X_2,\cdots,X_n}(x_1, x_2, \ldots, x_n) = P\left(X_i \leq x_i,\ i \in \{1, 2, \ldots, n\}\right). \tag{1.4.1}$$

Its properties can be illustrated by considering the case of two random variables $X_1$ and $X_2$. Then $F_{X_1,X_2}(x_1, x_2)$ satisfies the limit condition

$$\lim_{x_i \to -\infty} F_{X_1,X_2}(x_1, x_2) = 0 \tag{1.4.2}$$

for $i = 1$ and fixed $x_2 \in \Re$ or $i = 2$ and fixed $x_1 \in \Re$, and also the limit condition

$$\lim_{x_1,x_2 \to \infty} F_{X_1,X_2}(x_1, x_2) = 1. \tag{1.4.3}$$

Furthermore, $F_{X_1,X_2}(x_1, x_2)$ is nondecreasing and continuous from the right in $x_1$ and $x_2$. Additionally, it can be seen that

$$F_{X_1,X_2}(x_1, x_2) = F_{X_2,X_1}(x_2, x_1) \tag{1.4.4}$$

for $(x_1, x_2) \in \Re^2$. The *marginal distribution* $F_{X_1}$ satisfies

$$F_{X_1}(x_1) = \lim_{x_2 \to \infty} F_{X_1,X_2}(x_1, x_2). \tag{1.4.5}$$

for $x_1 \in \Re$.

For continuous random variables the joint distribution function is often differentiable, except possibly at some isolated or boundary points. For a wide class of continuous random variables there is a density function $f_{X_1,X_2} : \Re^2 \to \Re^+ = [0, \infty)$ given by

$$f_{X_1,X_2}(x_1, x_2) = \frac{\partial^2 F_{X_1,X_2}(x_1, x_2)}{\partial x_1\, \partial x_2}, \tag{1.4.6}$$

satisfying

$$F_{X_1,X_2}(x_1, x_2) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} f_{X_1,X_2}(s_1, s_2)\, ds_1\, ds_2. \tag{1.4.7}$$

Here $\frac{\partial}{\partial x_1}$ and $\frac{\partial^2}{\partial x_1\, \partial x_2}$ denote first and second order partial derivatives.

## Correlated Random Variables

Let us consider two random variables $X_1$ and $X_2$ with means $\mu_{X_1}$ and $\mu_{X_2}$ and variances $\sigma^2_{X_1}$ and $\sigma^2_{X_2}$, respectively. Their *covariance* is then defined as

$$\mathrm{Cov}(X_1, X_2) = E((X_1 - \mu_{X_1})\,(X_2 - \mu_{X_2})). \tag{1.4.8}$$

Obviously, we have for two random variables $X_1$ and $X_2$

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Cov}(X_2, X_1) \tag{1.4.9}$$

and for any constant $a_i \in \Re$, $i \in \{1, 2, 3, 4\}$,

$$\mathrm{Cov}(a_1\,X_1 + a_2, a_3\,X_2 + a_4) = a_1\,a_3\,\mathrm{Cov}(X_1, X_2). \tag{1.4.10}$$

If $X_1$ and $X_2$ are independent, then

$$\mathrm{Cov}(X_1, X_2) = 0. \tag{1.4.11}$$

If $X_1 = X_2$, then

$$\mathrm{Cov}(X_1, X_2) = \mathrm{Var}(X_1). \tag{1.4.12}$$

We define the *correlation* $\varrho_{X_1,X_2}$ of $X_1$ and $X_2$ in the form

$$\varrho_{X_1,X_2} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)\,\mathrm{Var}(X_2)}}. \tag{1.4.13}$$

By the Cauchy-Schwartz inequality (1.3.59) it follows that

$$-\,1 \le \varrho_{X_1,X_2} \le 1. \tag{1.4.14}$$

If $X_2 = a_1\,X_1 + a_2$ and $a_1 > 0$, then by (1.4.13) and (1.4.8) we have the correlation

$$\varrho_{X_1,X_2} = 1. \tag{1.4.15}$$

The correlation $\varrho_{X_1,X_2}$ provides a measure of the degrees of linear dependence between $X_1$ and $X_2$ using second moments. If $\varrho_{X_1,X_2} \neq 0$, then we call $X_1$ and $X_2$ *correlated*. Two independent random variables are always uncorrelated. For Gaussian random variables also the converse is true, that is, two uncorrelated Gaussian random variables are independent. Note however, in general, two uncorrelated random variables can be still dependent. This is important for log-returns. These can be highly dependent even if they are uncorrelated. This point is often missed in practice. A simple example is given when $X_1$ is $N(0,1)$ Gaussian distributed and $X_2 = \frac{1}{\sqrt{2}}\,((X_1)^2 - 1)$. Obviously, by (1.3.47) the correlation is zero. However, both random variables $X_1$ and $X_2$ are strongly dependent.

### Bivariate Gaussian Density

Let $\boldsymbol{A}^\top$ denote the *transpose* of the vector or matrix $\boldsymbol{A}$. A matrix $\boldsymbol{A}$ is *regular* if it is invertible. This is the case if its *determinant* $\det(\boldsymbol{A})$ is not equal to zero.

An important example of a two-dimensional density function is the *bivariate Gaussian density* given by

$$f_{X_1,X_2}(x_1,x_2) = \frac{1}{2\pi\sqrt{\det(\boldsymbol{D})}} \exp\left\{ -\frac{1}{2} \sum_{i,j=1}^{2} C^{i,j}(x_i-\mu_i)(x_j-\mu_j) \right\} \tag{1.4.16}$$

for $(x_1,x_2)^\top \in \Re^2$, with *mean vector* $\boldsymbol{\mu} = (\mu_1,\, \mu_2)^\top \in \Re^2$, *covariance matrix* $\boldsymbol{D} = [D^{i,j}]_{i,j=1}^2$, with components $D^{i,j} = E((X_i-\mu_i)(X_j-\mu_j)),\ i,j \in \{1,2\}$, which is here a $2\times 2$ regular matrix, and the *inverse* of the *matrix* $\boldsymbol{C} = [C^{i,j}]_{i,j=1}^2$. We say that two random variables $X_1$ and $X_2$ having the density (1.4.16) are *jointly Gaussian* distributed with *mean vector* $\boldsymbol{\mu}$ and *covariance matrix* $\boldsymbol{D}$.

If the random vector $\boldsymbol{Z} = (Z_1, Z_2)^\top$ has independent standard Gaussian components $Z_1$ and $Z_2$, then there exists an upper triangular, invertible $2\times 2$ matrix $\boldsymbol{S}$ such that $\boldsymbol{D} = \boldsymbol{S}^\top\boldsymbol{S}$ and the vector

$$\boldsymbol{X} = (X_1, X_2)^\top = \boldsymbol{S}^\top\boldsymbol{Z} + \boldsymbol{\mu} \tag{1.4.17}$$

is jointly Gaussian with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{D}$. $\boldsymbol{S}$ is sometimes called the *Cholesky decomposition* of the covariance matrix $\boldsymbol{D}$.

As an example let us construct pairs of correlated Gaussian random variables $X_1$, $X_2$ with means $\mu_1 = E(X_1) = 0$, $\mu_2 = E(X_2) = 0$ and variances $E(X_1^2) = 1$, $E(X_2^2) = \frac{1}{3}$ and covariance $E(X_1X_2) = \frac{1}{2}$ out of independent standard Gaussian distributed random variables $Z_1$ and $Z_2 \sim N(0,1)$. Some *Value at Risk* (VaR) evaluations are based on constructions of this type.

We note that for

$$X_1 = S^{1,1}\, Z_1 + S^{2,1}\, Z_2 \quad \text{and} \quad X_2 = S^{1,2}\, Z_1 + S^{2,2}\, Z_2 \tag{1.4.18}$$

with

$$\boldsymbol{S} = \begin{pmatrix} S^{1,1} & S^{1,2} \\ S^{2,1} & S^{2,2} \end{pmatrix} = \begin{pmatrix} 1 & \frac{1}{2} \\ 0 & \frac{1}{\sqrt{12}} \end{pmatrix}$$

we have

$$\boldsymbol{D} = \boldsymbol{S}^\top\,\boldsymbol{S} = \begin{pmatrix} 1 & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{3} \end{pmatrix} \tag{1.4.19}$$

and

$$\boldsymbol{C} = \boldsymbol{D}^{-1} = \begin{pmatrix} 4 & -6 \\ -6 & 12 \end{pmatrix},$$

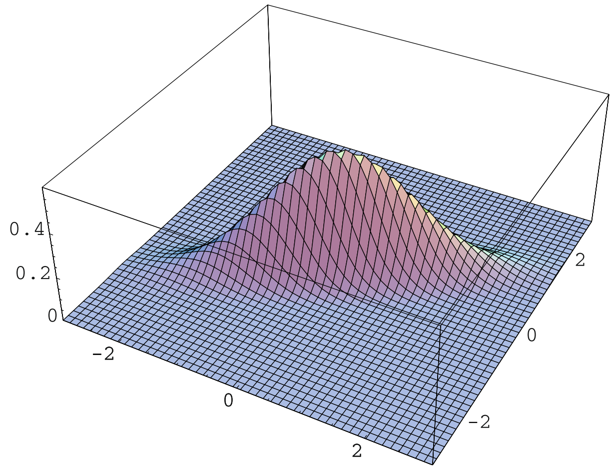



**Fig. 1.4.1.** Bivariate Gaussian density

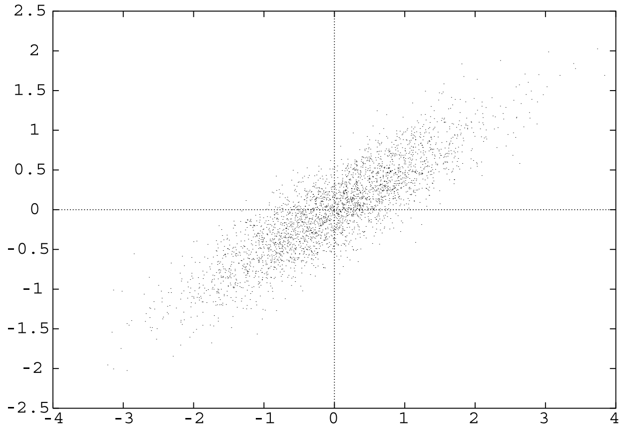



**Fig. 1.4.2.** Points with correlated Gaussian coordinates

where

$$\det(\boldsymbol{C}) = C^{1,1}\, C^{2,2} - C^{1,2}\, C^{2,1} = 12.$$

In Fig. 1.4.1 we show the two-dimensional Gaussian joint density for $X_1$ and $X_2$ with mean vector zero and covariance matrix $\boldsymbol{D}$ given in (1.4.19). We remark that the lines of the Gaussian density that have the same level have an elliptic shape.

Figure 1.4.2 shows 3000 simulated realizations of pairs $(X_1, X_2)$ of these Gaussian random variables using $X_1$ as the $x$-coordinate and $X_2$ as the $y$-coordinate. Note that the points are concentrated mostly in the area where the density given in Fig. 1.4.1 is largest.

For the bivariate Gaussian density with covariance matrix (1.4.19) the correlation coefficient is according to (1.4.18) and (1.4.13) given by

$$\varrho_{X_1,X_2} = \frac{S^{1,1}\,S^{1,2}}{\sigma_{X_1}\,\sigma_{X_2}} = \frac{1}{2}\sqrt{3} \approx 0.866.$$

This means that Fig. 1.4.2 displays a set of 3000 outcomes of correlated Gaussian random variables with the above correlation coefficient.

### Conditional Expectation for the Bivariate Gaussian Case

For given random variables $X_1$ and $X_2$ with bivariate Gaussian distribution one can prove that if $\text{Cov}(X_1, X_2) = 0$, then $X_1$ and $X_2$ are independent. Furthermore, if $\text{Var}(X_2) > 0$, then

$$E(X_1 \,\big|\, X_2) = E(X_1) + \frac{\text{Cov}(X_1, X_2)}{\text{Var}(X_1)}\,(X_2 - E(X_2)) \tag{1.4.20}$$

and

$$E\left((X_1 - E(X_1 \,\big|\, X_2))^2\right) = \text{Var}(X_1) - \frac{(\text{Cov}(X_1, X_2))^2}{\text{Var}(X_2)}. \tag{1.4.21}$$

Here $E(X_1 \,|\, X_2)$ denotes the conditional expectation of $X_1$ given the information generated by $X_2$.

We emphasize that the above constructions use jointly Gaussian distributed random variables. Now consider two independent $N(0,1)$ standard Gaussian random variables $Y_1$ and $Y_2$. From these we construct $X_1 = |Y_2|\,\text{sgn}(Y_1)$ and $X_2 = Y_2$. Using these definitions it can be shown that $X_1 \sim N(0,1)$ and $X_2 \sim N(0,1)$ with

$$\begin{aligned}\text{Cov}(X_1, X_2) &= E(X_1\,Y_2) - E(X_1)\,E(Y_2) = E(Y_2\,|Y_2|\,\text{sgn}(Y_1)) \\ &= E(Y_2\,|Y_2|)\,E(\text{sgn}(Y_1)) = 0,\end{aligned}$$

but $X_1$ and $X_2$ are dependent random variables. As a consequence, $X_1$ and $X_2$ are not jointly Gaussian distributed and

$$\text{Cov}(|X_1|, |X_2|) = E(|Y_2|^2) - (E(|Y_2|))^2 > 0.$$

Note that these types of effects need to be taken into account if one is modeling log-returns of securities.

### Properties of Independent Random Variables

Recall the definition of independent random variables in Sect. 1.1. It can be shown that two random variables $X_1$ and $X_2$ are independent if their joint and marginal distribution functions satisfy the relation

$$F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1)\,F_{X_2}(x_2) \tag{1.4.22}$$

for all $x_1$, $x_2 \in \Re$. This is equivalent to saying that

$$E(g_1(X_1)\, g_2(X_2)) = E(g_1(X_1))\, E(g_2(X_2)) \tag{1.4.23}$$

for all measurable functions $g_1$, $g_2$ for which the above expectations exist. If both $F_{X_1}$ and $F_{X_2}$ have density functions $f_{X_1}$ and $f_{X_2}$, respectively, and if $X_1$ and $X_2$ are independent, then their joint distribution function $F_{X_1,X_2}$ has a density function $f_{X_1,X_2}$ which satisfies the equation

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1)\, f_{X_2}(x_2). \tag{1.4.24}$$

Moreover, choosing $g_1$ and $g_2$ to equal the identity function in (1.4.23) it can be seen that for two independent random variables $X_1$ and $X_2$ the product $X_1 X_2$ has an expectation given by

$$E(X_1 X_2) = E(X_1)\, E(X_2), \tag{1.4.25}$$

and the sum $X_1 + X_2$ has a variance satisfying the *additivity property*

$$\mathrm{Var}(X_1 + X_2) = \mathrm{Var}(X_1) + \mathrm{Var}(X_2). \tag{1.4.26}$$

The Gaussian random variables $X_1$ and $X_2$ obtained from (1.4.18) in the corresponding example are by (1.4.19) not independent since $E(X_1 X_2) = \frac{1}{2}$ but $E(X_1) = E(X_2) = 0$. They are correlated, as will be shown in the next subsection.

### First and Second Moments of Random Vectors

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ denote a *random vector*. Then the expectation is taken componentwise and we obtain

$$E(\boldsymbol{X}) = (E(X_1), E(X_2), \ldots, E(X_n))^\top. \tag{1.4.27}$$

In the case when $\boldsymbol{B} = [B^{i,j}]_{i,j=1}^{n,m}$ is an $n \times m$ *random matrix*, where $B^{i,j}$ is some random variable we obtain its expectation as the $n \times m$ matrix

$$E(\boldsymbol{B}) = [E(B^{i,j})]_{i,j=1}^{n,m}. \tag{1.4.28}$$

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ and $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_m)^\top$ with $n, m \in \mathcal{N}$ denote two random vectors. Their *covariance matrix* $\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y})$ is defined as

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{Y}) &= E\left( (\boldsymbol{X} - E(\boldsymbol{X}))\, (\boldsymbol{Y} - E(\boldsymbol{Y}))^\top \right) \\ &= \Big[ E\big( (X_i - E(X_i))\, (Y_j - E(Y_j)) \big) \Big]_{i,j=1}^{n,m} \end{aligned} \tag{1.4.29}$$

The matrix $\mathrm{Cov}(\boldsymbol{X}) = \mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X})$ is called the *autocovariance matrix* of the vector $\boldsymbol{X}$.

If $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ is an $n$-dimensional vector, $\boldsymbol{A} = [A^{i,j}]_{i,j=1}^{n,m}$ a deterministic $n \times m$ matrix and $\boldsymbol{b} = (b_1, b_2, \ldots, b_m)^\top$ a deterministic $m$-dimensional vector, then for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ it is straightforward to show that

$$E(\boldsymbol{Y}) = E(\boldsymbol{A}\,\boldsymbol{X} + \boldsymbol{b}) = \boldsymbol{A}\,E(\boldsymbol{X}) + \boldsymbol{b} \tag{1.4.30}$$

and

$$\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y}) = \boldsymbol{A}\,\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X})\,\boldsymbol{A}^\top. \tag{1.4.31}$$

For example, if $\boldsymbol{X}$ is a vector of $n$ independent random variables with variance $\mathrm{Var}(X_i) = 1,\ i \in \{1, 2, \ldots, n\}$, then

$$\mathrm{Cov}(\boldsymbol{X}, \boldsymbol{X}) = \boldsymbol{I}, \tag{1.4.32}$$

where $\boldsymbol{I}$ is the identity matrix or unit matrix and we have for $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ the autocovariance matrix

$$\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y}) = \boldsymbol{A}\,\boldsymbol{A}^\top. \tag{1.4.33}$$

To construct from such a vector $\boldsymbol{X}$ an $n$-dimensional vector $\boldsymbol{Y}$ with given autocovariance matrix $\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y})$ it is sufficient to find an upper triangular $n \times n$-matrix $\boldsymbol{A}$ that satisfies (1.4.33). This matrix is then the Cholesky decomposition of $\mathrm{Cov}(\boldsymbol{Y}, \boldsymbol{Y})$, see (1.4.17).

For any $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ one has the equality

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i,j=1}^{n} \mathrm{Cov}(X_i, X_j), \tag{1.4.34}$$

and if $\mathrm{Cov}(X_i, X_j) = 0$ for $i \neq j$, then

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i). \tag{1.4.35}$$

### Multivariate Joint Distributions

The properties (1.4.2)–(1.4.5) of joint distribution functions generalize to any number $n \geq 2$ of random variables $X_1$, $X_2$, …, $X_n$. With the notation introduced in (1.4.1) the joint distributions $F_{X_1, X_2, \cdots, X_n}$ satisfy

$$\lim_{x_i \to -\infty} F_{X_1, X_2, \cdots, X_n}(x_1, x_2, \ldots, x_n) = 0 \tag{1.4.36}$$

for $i \in \{1, 2, \ldots, n\}$ and fixed $x_j,\ j \in \{1, 2, \ldots, i-1, i+1, \ldots, n\}$. We also have the limit condition

$$\lim_{x_1, \ldots, x_n \to +\infty} F_{X_1, X_2, \cdots, X_n}(x_1, x_2, \ldots, x_n) = 1. \tag{1.4.37}$$

In addition, $F_{X_1,X_2,\cdots,X_n}$ is non-decreasing and continuous from the right in $x_i$ for $i \in \{1,2,\ldots,n\}$. For any permutation $\{i_1,i_2,\ldots,i_n\}$ of the set $\{1,2,\ldots,n\}$ we have

$$F_{X_{i_1},X_{i_2},\cdots,X_{i_n}}(x_{i_1},x_{i_2},\ldots,x_{i_n}) = F_{X_1,X_2,\cdots,X_n}(x_1,x_2,\ldots,x_n). \quad (1.4.38)$$

Furthermore, if $\{i_1,i_2,\ldots,i_k\}$ is any subset of the set $\{1,2,\ldots,n\}$, then the marginal distribution $F_{X_{i_1},X_{i_2},\cdots,X_{i_k}}$ for $k \in \{1,2,\ldots,n\}$ satisfies

$$F_{X_{i_1},X_{i_2},\cdots,X_{i_k}}(x_{i_1},x_{i_2},\ldots,x_{i_k}) = \lim_{x_i \to +\infty} F_{X_1,X_2,\cdots,X_n}(x_1,x_2,\ldots,x_n), \quad (1.4.39)$$

where this limit has to be taken for all $i \notin \{i_1,i_2,\ldots,i_k\}$.

The properties (1.4.22)–(1.4.26) can also be generalized to $n$ random variables. Thus, the random variables $X_1$, $X_2$, …, $X_n$ are independent if their joint distribution satisfies the equation

$$F_{X_1,X_2,\cdots,X_k}(x_1,x_2,\ldots,x_k) = F_{X_1}(x_1)\,F_{X_2}(x_2)\,\cdots\,F_{X_k}(x_k) \quad (1.4.40)$$

for all $k \in \{1,2,\ldots,n\}$. If in this case each $F_{X_i}$ has a density function $f_{X_i}$, then $F_{X_1,X_2,\cdots,X_n}$ has a joint density function $f_{X_1,X_2,\ldots,X_n}$ that takes the form

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = f_{X_1}(x_1)\,f_{X_2}(x_2)\,\cdots\,f_{X_n}(x_n). \quad (1.4.41)$$

In addition, for $n$ independent random variables $X_1, X_2, \cdots, X_n$ the product $g_1(X_1)\ g_2(X_2)\ \cdots g_n(X_n)$ involving measurable functions $g_1, g_2, \ldots, g_n$ has expectation

$$E(g_1(X_1)g_2(X_2)\cdots g_n(X_n)) = E(g_1(X_1))\,E(g_2(X_2))\,\cdots\,E(g_n(X_n)), \quad (1.4.42)$$

whereas their sum has variance

$$\mathrm{Var}\left(\sum_{i=1}^{n} g_i(X_i)\right) = \sum_{i=1}^{n} \mathrm{Var}(g_i(X_i)). \quad (1.4.43)$$

### Multivariate Gaussian Density

Consider a *random vector* $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ with *mean vector*

$$\boldsymbol{\mu} = (\mu_1,\mu_2,\ldots,\mu_n)^\top = (E(X_1),E(X_2),\ldots,E(X_n))^\top \quad (1.4.44)$$

and an $n\times n$ *autocovariance matrix* $\boldsymbol{D} = \mathrm{Cov}(\boldsymbol{X},\boldsymbol{X}) = \left[D^{\ell,m}\right]_{\ell,m=1}^{n}$, where

$$D^{\ell,m} = E\left((X_\ell - \mu_\ell)\,(X_m - \mu_m)\right) = E(X_\ell\,X_m) - E(X_\ell)\,E(X_m). \quad (1.4.45)$$

If $\boldsymbol{D}$ is regular, that is $\det(\boldsymbol{D}) \neq 0$, and its density is for $\boldsymbol{x} = (x_1,x_2,\ldots,x_n)^\top \in \Re^n$ given by

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n)$$
$$= \frac{\exp\left\{-\frac{1}{2}\,(\boldsymbol{x}-\boldsymbol{\mu})^\top\,\boldsymbol{D}^{-1}\,(\boldsymbol{x}-\boldsymbol{\mu})\right\}}{\sqrt{(2\,\pi)^n\,\det(\boldsymbol{D})}}, \qquad (1.4.46)$$

then $\boldsymbol{X}$ has an $n$-dimensional Gaussian density. The components of a Gaussian distributed random vector are *independent* if and only if they are pairwise *uncorrelated*. Furthermore, if $\boldsymbol{X}$ is an $n$-dimensional Gaussian random vector, $\boldsymbol{A}$ a deterministic matrix with $m$ rows and $n$ columns and $\boldsymbol{b}$ a deterministic $m$-dimensional vector, then $\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} + \boldsymbol{b}$ is an $m$-dimensional Gaussian random vector with mean $\boldsymbol{A}\boldsymbol{\mu} + \boldsymbol{b}$ and covariance matrix $\boldsymbol{A}\boldsymbol{D}\boldsymbol{A}^\top$.

### Conditional Expectation for Multivariate Gaussian Case (*)

We can generalize the relationships (1.4.20) and (1.4.21) on conditional expectations for bivariate Gaussian random variables to the case where $X_1$ is a scalar random variable and $\boldsymbol{X_2} = (X_2^1, X_2^2, \ldots, X_2^n)^\top$ is an $n$-dimensional random vector such that $X_1$ and the components of $\boldsymbol{X_2}$ are jointly Gaussian distributed. One can prove that if $\mathrm{Cov}(X_1, X_2^i) = 0$ for all $i \in \{1,2,\ldots,n\}$, then the random variable $X_1$ and the components of the random vector $\boldsymbol{X_2}$ are independent. In the case when the autocovariance matrix of $\boldsymbol{X_2}$ is invertible, that is $\mathrm{Cov}(\boldsymbol{X_2},\boldsymbol{X_2})^{-1}$ exists, then one has the following conditional expectations

$$E(X_1 \,\big|\, \boldsymbol{X_2}) = E(X_1) + \mathrm{Cov}(X_1,\boldsymbol{X_2})\,(\mathrm{Cov}(\boldsymbol{X_2},\boldsymbol{X_2}))^{-1}\,(\boldsymbol{X_2} - E(\boldsymbol{X_2})) \qquad (1.4.47)$$

and

$$E\left((X_1 - E(X_1 \,\big|\, \boldsymbol{X_2}))^2\right) = \mathrm{Var}(X_1) - \mathrm{Cov}(X_1,\boldsymbol{X_2})\,(\mathrm{Cov}(\boldsymbol{X_2},\boldsymbol{X_2}))^{-1} \times \mathrm{Cov}(X_1,\boldsymbol{X_2})^\top. \qquad (1.4.48)$$

These relationships are quite helpful in statistical analysis and for the pricing of derivatives for multiple securities.

### Multivariate Gaussian Shift (*)

The following relationships can be used for Value at Risk calculations and also in multi-asset option pricing. Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top \in \Re$ denote an $n$-dimensional vector with correlated $N(0,1)$ distributed components. The correlation matrix equals the covariance matrix $\boldsymbol{D}$ with components

$$D^{\ell,m} = \varrho_{X_\ell,X_m},$$

see (1.4.13). We denote according to (1.4.46) the corresponding joint density by

$$f_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) = N'_{n,\boldsymbol{D}}(\boldsymbol{x}) = \frac{\exp\{-\frac{1}{2}\,\boldsymbol{x}^\top \boldsymbol{D}^{-1}\,\boldsymbol{x}\}}{\sqrt{(2\pi)^n\,\det(\boldsymbol{D})}} \tag{1.4.49}$$

for $\boldsymbol{x} \in \Re^n$. The associated Gaussian distribution function for $X$ is given by

$$\begin{aligned} F_{X_1,X_2,\ldots,X_n}(x_1,x_2,\ldots,x_n) &= N_{n,\boldsymbol{D}}(\boldsymbol{x}) \\ &= P(X_i < x_i,\, i \in \{1,2,\ldots,n\}) \\ &= E\left(\prod_{i=1}^n \mathbf{1}_{\{X_i<x_i\}}\right) \\ &= \int_{-\infty}^{x_n} \cdots \int_{-\infty}^{x_1} N'_{n,\boldsymbol{D}}(\boldsymbol{y})\, dy_1 \cdots dy_n \end{aligned} \tag{1.4.50}$$

for $\boldsymbol{x} = (x_1,x_2,\ldots,x_n)^\top \in \Re^n$. We say that the $n$-dimensional vector $\boldsymbol{x} \sim N_n(\mathbf{0},\boldsymbol{D})$ is Gaussian distributed with mean vector $\boldsymbol{\mu} = (0,\ldots,0)^\top$ and covariance matrix $\boldsymbol{D}$.

Let $\boldsymbol{X} \sim N_n(\mathbf{0},\boldsymbol{D})$ and $\boldsymbol{b} = (b^1,b^2,\ldots,b^n)^\top \in \Re^n$ be an $n$-dimensional deterministic vector, then the scalar random variable

$$Z = \boldsymbol{b}^\top \boldsymbol{X}$$

is Gaussian with

$$Z \sim N(0, \boldsymbol{b}^\top \boldsymbol{D}\,\boldsymbol{b}). \tag{1.4.51}$$

More generally, let $\boldsymbol{B} = [B^{i,j}]_{i,j=1}^{m,n}$ be a deterministic $m \times n$ matrix, then we obtain

$$\boldsymbol{Y} = \boldsymbol{B}\,\boldsymbol{X} \sim N_m(\mathbf{0}, \boldsymbol{B}\,\boldsymbol{D}\,\boldsymbol{B}^\top), \tag{1.4.52}$$

where the mean vector is a vector of zeros and the covariance matrix $\boldsymbol{BDB}^\top$ is an $m \times m$ matrix. Additionally, let us normalize the vector $\boldsymbol{Y}$ by using the diagonal matrix $\boldsymbol{A} = [A^{i,j}]_{i,j=1}^m$, where $A^{i,i} = \sqrt{(\boldsymbol{B}\,\boldsymbol{D}\,\boldsymbol{B}^\top)^{i,i}}$ and $A^{i,j} = 0$ for $i \neq j$. We set

$$\tilde{\boldsymbol{Y}} = \boldsymbol{A}^{-1}\,\boldsymbol{Y} = \boldsymbol{A}^{-1}\,\boldsymbol{B}\,\boldsymbol{X},$$

where $\tilde{\boldsymbol{Y}} = (\tilde{Y}_1,\tilde{Y}_2,\ldots,\tilde{Y}_m)^\top \sim N_m(\mathbf{0}, \boldsymbol{A}^{-1}\,\boldsymbol{B}\,\boldsymbol{D}\,\boldsymbol{B}^\top(\boldsymbol{A}^{-1})^\top)$ turns out to be an $m$-dimensional Gaussian vector with zero mean vector and standard variances for its components. Therefore, it follows for $\tilde{\boldsymbol{y}} = (\tilde{y}_1,\tilde{y}_2,\ldots,\tilde{y}_m)^\top = \boldsymbol{A}^{-1}\boldsymbol{y}$ that

$$\begin{aligned} P\left(\tilde{Y}_i < \tilde{y}_i,\, i \in \{1,2,\ldots,m\}\right) &= P(Y_i < y_i,\, i \in \{1,2,\ldots,m\}) \\ &= N_{m,\boldsymbol{A}^{-1}\boldsymbol{BDB}^\top(\boldsymbol{A}^{-1})^\top}(\boldsymbol{A}^{-1}\,\boldsymbol{y}), \end{aligned} \tag{1.4.53}$$

where the multivariate Gaussian distribution function is given in (1.4.50).

From the properties of the probability density $N'_{n,\boldsymbol{D}}(\boldsymbol{x})$ of an $n$-dimensional vector $\boldsymbol{X}$ of standard Gaussian random variables with covariance matrix $\boldsymbol{D}$, see (1.4.49), we have the relation

$$N'_{n,\boldsymbol{D}}(\boldsymbol{x}) = \exp\left\{\boldsymbol{\theta}^\top \boldsymbol{x} - \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{D}\,\boldsymbol{\theta}\right\} N'_{n,\boldsymbol{D}}(\boldsymbol{x} - \boldsymbol{D}\,\boldsymbol{\theta}) \tag{1.4.54}$$

for any vectors $\boldsymbol{\theta}, \boldsymbol{x} \in \Re^n$. This yields the *multivariate Gaussian shift* property for $\boldsymbol{X} \sim N_n(\mathbf{0}, \boldsymbol{D})$, a deterministic vector $\boldsymbol{\theta} = (\theta^1, \theta^2, \ldots, \theta^n)^\top$ and a scalar function $H(\boldsymbol{x})$ of an $n$-dimensional vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^\top$ in the form

$$E(H(\boldsymbol{X} + \boldsymbol{\theta})) = E\left(\exp\left\{-\frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{D}\,\boldsymbol{\theta} + \boldsymbol{\theta}^\top \boldsymbol{X}\right\} H(\boldsymbol{X})\right). \tag{1.4.55}$$

This result can be employed in the pricing of derivatives involving several securities, see Buchen (2004) and Buchen & Konstandatos (2005).

### Multivariate Characteristic Functions (*)

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\top$ be a random vector. The *characteristic function* $\phi_{\boldsymbol{X}}(\boldsymbol{\theta})$ with $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top$ is defined for all values of $\boldsymbol{\theta} \in \Re^p$ by

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = E(\exp\{\imath\,\boldsymbol{\theta}^\top \boldsymbol{X}\}), \tag{1.4.56}$$

where $\imath$ is the imaginary unit. Note that

$$|\phi_{\boldsymbol{X}}(\boldsymbol{\theta})| \leq 1 \tag{1.4.57}$$

for all $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_p)^\top \in \Re^p$. This characteristic function uniquely identifies the distribution of the corresponding random vector. For a continuous $n$-dimensional random vector we have

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \exp\left\{\imath \sum_{k=1}^{p} \theta_k\, x_k\right\} f_{\boldsymbol{X}}(x_1, \ldots, x_p)\, dx_1, \ldots, dx_p. \tag{1.4.58}$$

The characteristic function $\phi_{\boldsymbol{X}}(\boldsymbol{\theta})$ of a $p$-dimensional jointly Gaussian distributed random vector $\boldsymbol{X}$ with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{D}$ is of the form

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = \exp\left\{\imath\,\boldsymbol{\mu}^\top \boldsymbol{\theta} - \frac{1}{2}\boldsymbol{\theta}^\top \boldsymbol{D}\,\boldsymbol{\theta}\right\}. \tag{1.4.59}$$

for all $\boldsymbol{\theta} \in \Re^p$.

Let us give another example using a $p$-dimensional Student $t$ distributed random variable $\boldsymbol{X} = (X_1, X_2, \ldots, X_p)^\top$ with $n > 0$ degrees of freedom, zero mean vector $\boldsymbol{\mu} = (0, \ldots, 0)^\top$ and regular covariance matrix $\boldsymbol{D}$. This random variable can be obtained from a multivariate Gaussian vector $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_p)^\top$, with mean vector $\boldsymbol{\mu}_{\boldsymbol{Y}} = (0, \ldots, 0)^\top$ and covariance matrix $\boldsymbol{D}$, scaled by the inverse of the square root of an independent scalar $\chi^2(n)$ distributed random variable $Z \in (0, \infty)$ such that

$$\boldsymbol{X} = \frac{\boldsymbol{Y}}{\sqrt{\frac{Z}{n}}}, \tag{1.4.60}$$

see (1.2.16). If $\boldsymbol{Y}$ is a vector of independent standard Gaussian random variables, then $\boldsymbol{X}$ has the characteristic function

$$\phi_{\boldsymbol{X}}(\boldsymbol{\theta}) = E\left(e^{\imath\,\boldsymbol{\theta}^\top \boldsymbol{X}}\right) = \frac{K_{\frac{n}{2}}\left(\{n\,\boldsymbol{\theta}^\top\boldsymbol{\theta}\}^{\frac{1}{2}}\right)\,(n\,\boldsymbol{\theta}^\top\boldsymbol{\theta})^{\frac{n}{4}}}{\Gamma(\frac{n}{2})\,2^{\frac{n}{2}-1}} \tag{1.4.61}$$

for $n > 0$ and $\boldsymbol{\theta} = (\theta_1, \dots, \theta_p)^\top \in \Re^p$, where $K_\lambda(\cdot)$ is again the modified Bessel function of the third kind with index $\lambda$. Its probability density function is then of the form

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{\Gamma(\frac{1}{2}(n+p))}{(\pi\,n)^{\frac{p}{2}}\,\Gamma(\frac{n}{2})}\left(1 + \frac{\boldsymbol{x}^\top\boldsymbol{x}}{n}\right)^{-\frac{1}{2}(n+p)} \tag{1.4.62}$$

for $\boldsymbol{x} = (x_1, x_2, \dots, x_p)^\top \in \Re^p$.

### Further Properties of Moments (*)

When we are considering $n$ different random variables $X_1, X_2, \dots, X_n$, then it is often convenient to use vector notation. For vectors $\boldsymbol{x} = (x_1, x_2, \dots, x_n)^\top$ and $\boldsymbol{y} = (y_1, y_2, \dots, y_n)^\top$ in $\Re^n$ recall that the *inner product* $(\boldsymbol{x}, \boldsymbol{y})$ and the *Euclidean norm* $|\boldsymbol{x}|$ are defined by

$$(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^\top\,\boldsymbol{y} = \sum_{i=1}^n x_i\,y_i \quad \text{and} \quad |\boldsymbol{x}| = \sqrt{\boldsymbol{x}^\top\,\boldsymbol{x}} = \sqrt{\sum_{i=1}^n (x_i)^2}, \tag{1.4.63}$$

respectively. Note that for $n = 1$ the Euclidean norm coincides with the absolute value operator.

The following moment inequalities are often useful and follow from more general inequalities for integrals, see Shiryaev (1984). Let $\boldsymbol{X} = (X_1, X_2, \dots, X_n)^\top$ and $\boldsymbol{Y} = (Y_1, Y_2, \dots, Y_n)^\top$ be random vectors, then

$$E(|\boldsymbol{X} + \boldsymbol{Y}|^r) \le c_r\,(E(|\boldsymbol{X}|^r) + E(|\boldsymbol{Y}|^r)) \tag{1.4.64}$$

with $c_r = 1$ for $r \le 1$ and $c_r = 2^{r-1}$ for $r \ge 1$. Furthermore,

$$(E(|\boldsymbol{X} + \boldsymbol{Y}|^r))^{\frac{1}{r}} \le (E(|\boldsymbol{X}|^r))^{\frac{1}{r}} + (E(|\boldsymbol{Y}|^r))^{\frac{1}{r}} \tag{1.4.65}$$

for $r \ge 1$, and

$$E(|(\boldsymbol{X}, \boldsymbol{Y})|) \le (E(|\boldsymbol{X}|^p))^{\frac{1}{p}}\,(E(|\boldsymbol{Y}|^q))^{\frac{1}{q}} \tag{1.4.66}$$

for $p$, $q > 1$ with $\frac{1}{p} + \frac{1}{q} = 1$.

## 1.5 Copulas (*)

*Copulas* play an important role in the analysis and modeling of the dependence structures of financial random variables. They are used, for instance, in *Value at Risk* (VaR) and *credit risk* modeling applications. Since they are widely used in different areas in quantitative finance we summarize below a few basic facts on copulas.

### Copula Function (*)

A copula function can be considered as a basic building block for constructing multivariate densities and distributions, see Nelsen (1999). A *copula function* $C : [0,1]^n \to [0,1]$ in $\Re^n$, $n \in \{2, 3, \ldots\}$, is a multivariate distribution function with the property that its marginal distributions are standard uniform distributions.

By this definition a copula has the $U(0,1)$ uniform density as the density for all of its marginal distributions, see (1.4.5). The following theorem by Sklar (1959) makes clear that copulas are universal tools for analyzing multivariate distributions.

**Theorem 1.5.1.** (Sklar) *Let $F_{X_1,X_2,\ldots,X_n} : \Re^n \to [0,1]$ be a multivariate $n$-dimensional distribution function with marginal distributions $F_{X_i} : \Re \to [0,1]$, $i \in \{1, 2, \ldots, n\}$, then there exists a copula $C : [0,1]^n \to [0,1]$ such that*

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = C\left(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\right) \quad (1.5.1)$$

*for $(x_1, x_2, \ldots, x_n)^\top \in \Re^n$. Moreover, if the marginal distributions have a density, then the copula is unique.*

The proof of this important result exploits the essential fact that one has for $(u_1, u_2, \ldots, u_n)^\top \in [0,1]^n$ the relation

$$C(u_1, u_2, \ldots, u_n) = F_{X_1,X_2,\ldots,X_n}\left(F_{X_1}^{-1}(u_1), F_{X_2}^{-1}(u_2), \ldots, F_{X_n}^{-1}(u_n)\right). \quad (1.5.2)$$

**Corollary 1.5.2.** *For any copula $C : [0,1]^n \to [0,1]$ in $\Re^n$, $n \in \{2, 3, \ldots\}$, and distribution functions $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$ the function*

$$F_{X_1,X_2,\ldots,X_n}(x_1, x_2, \ldots, x_n) = C\left(F_{X_1}(x_1), F_{X_2}(x_2), \ldots, F_{X_n}(x_n)\right) \quad (1.5.3)$$

*for $(x_1, x_2, \ldots, x_n)^\top \in \Re^n$ defines a multivariate distribution function with marginal distributions $F_{X_1}, F_{X_2}, \ldots, F_{X_n}$.*

This means that every multivariate distribution with continuous marginal distribution function admits a unique copula representation. Furthermore, the above result shows that copulas and marginal distribution functions are the building blocks for general multivariate distributions.

### Gaussian Copula (*)

One of the most common copulas that arise in finance is the *Gaussian copula* $C_{N,\boldsymbol{D}}$, which is defined as

$$C_{N,\boldsymbol{D}}(u_1, u_2, \ldots, u_n) = N_{n,\boldsymbol{D}}\left(N_{X_1}^{-1}(u_1), N_{X_2}^{-1}(u_2), \ldots, N_{X_n}^{-1}(u_n)\right) \quad (1.5.4)$$

for $(u_1, u_2, \ldots, u_n)^\top \in [0,1]^n$. Here $\boldsymbol{D}$ is the regular $n \times n$ covariance matrix of the multivariate Gaussian random variable $\boldsymbol{X}$, see (1.4.45). It is common in standard VaR calculations to use the Gaussian copula if one has to deduce from the log-returns of the constituents of a portfolio the VaR number of the portfolio.

As an example, let us consider points $(X_1, X_2)^\top$ with Gaussian marginals that have the bivariate Gaussian copula. Here we set

$$X_i = \varrho\, Z_0 + \sqrt{1-\varrho^2}\, Z_i \tag{1.5.5}$$

for $i \in \{1, 2\}$, where $Z_0$, $Z_1$, $Z_2$ are independent standard Gaussian random variables. The parameter $\varrho \in [-1, 1]$ measures the correlation between $X_i$ and $Z_0$ for $i \in \{1, 2\}$. In Fig. 1.4.2 we have plotted 3000 of such points that relate to a bivariate Gaussian distribution with correlation $\varrho \approx 0.866$. The corresponding bivariate Gaussian copula is then

$$C_{N,\boldsymbol{D}}(u_1, u_2) = N_{2,\boldsymbol{D}}\left(N_{X_1}^{-1}(u_1), N_{X_2}^{-1}(u_2)\right) \tag{1.5.6}$$

for $(u_1, u_2)^\top \in [0, 1]^2$. Although the Gaussian copula is widely used in VaR calculations, it usually provides a poor fit to multivariate log-return data.

### Student $t$ Copula (*)

It has been reported in Breymann, Dias & Embrechts (2003) that a good fit for multivariate log-returns of currencies is obtained by the *Student $t$ copula* $C_{t,\boldsymbol{D},\delta}$ with approximately $\delta \approx 4$ degrees of freedom. This copula is defined by the function

$$C_{t,\boldsymbol{D},\delta}(u_1, u_2, \ldots, u_n) = t_{\delta,\boldsymbol{D}}\left(t_{X_1}^{-1}(u_1), t_{X_2}^{-1}(u_2), \ldots, t_{X_n}^{-1}(u_n)\right) \tag{1.5.7}$$

for $(u_1, u_2, \ldots, u_n)^\top \in [0, 1]^n$. Here $t_{\delta,\boldsymbol{D}}$ is the Student $t$ distribution with $\delta > 2$ degrees of freedom and $\boldsymbol{D}$ as the covariance matrix of the components $(X_1, X_2, \ldots, X_n)^\top$, see (1.4.60) and (1.4.62). For currency log-returns Breymann et al. (2003) identified a Student $t$ copula with approximately four degrees of freedom.

The isolines of the bivariate $t$ density have an elliptical shape as is the case for the Gaussian density. This is not surprising due to the representation (1.4.60) of multivariate Student $t$ distributed random variables as multivariate Gaussian random variables with independent inverse chi-square distributed variance.

According to (1.5.7) the bivariate $t$ copula with covariance matrix $\boldsymbol{D}$ and $\delta$ degrees of freedom is obtained from the expression

$$C_{t,\boldsymbol{D},\delta}(u_1, u_2) = t_{\delta,\boldsymbol{D}}\left(t_{X_1}^{-1}(u_1), t_{X_2}^{-1}(u_2)\right) \tag{1.5.8}$$

for $(u_1, u_2)^\top \in [0, 1]^2$.

## 1.6 Exercises for Chapter 1

**1.1.** Show that $\mathrm{Var}(X) = E(X^2) - (E(X))^2$.

**1.2.** Calculate the first and second moments and the variance for a Poisson random variable with intensity $\lambda > 0$.

**1.3.** Calculate the first and second moments and the variance for a $U(a, b)$ uniformly distributed random variable.

**1.4.** Determine for an exponentially distributed random variable with intensity parameter $\lambda > 0$ the first and second moments and the variance.

**1.5.** Calculate the first and second moments and the variance for an $N(0, 1)$ standard Gaussian distributed random variable.

**1.6.** Determine the even moments for a standard Gaussian distributed random variable.

**1.7.** If a random variable $Y$ is $N(\mu, \sigma^2)$ Gaussian distributed show that $X = \frac{Y-\mu}{\sigma}$ is $N(0, 1)$ distributed.

**1.8.** If a random variable $Y$ is $N(0, 1)$ Gaussian distributed what is the distribution of $Y^2$?

**1.9.** Compute the expectation of the exponential $Y = \exp\{X\}$ of a Gaussian $N(\mu, \sigma^2)$ distributed random variable.

**1.10.** (*) Show for a standard Gaussian random variable $X \sim N(0, 1)$, a deterministic constant $\theta \in \Re$ and a real valued function $H(x)$ for $x \in \Re$ with $|E(H(X + \theta))| < \infty$ that

$$E(H(X+\theta)) = E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\, X\right\} H(X)\right).$$

**1.11.** (*) Prove that for a correlated pair of Gaussian random variables the corresponding joint density is, in general, not the product of their marginal densities. When are these random variables independent?

**1.12.** (*) Compute the mean for the Cauchy distribution with density $p(x) = [\pi\,(1 + x^2)]^{-1}$. Is this mean finite?

**1.13.** (*) Compute the conditional expectation $E(X|A)$ for a random variable $X(\omega) = \omega \in [0, 1]$ with density $f_X(x) = x$ with respect to the event $A = \{\omega \in [0, 0.5]\}$.

# 2

# Statistical Methods

We introduce in this chapter further fundamental results from probability theory and statistics which are important in quantitative finance. They are highly relevant for the empirical analysis of financial data. In particular, limit theorems are presented and confidence intervals constructed. Furthermore, the log-returns of a world stock index will be estimated pointing at a stylized empirical fact.

## 2.1 Limit Theorems

In this section some fundamental limit theorems are summarized. These include the Law of Large Numbers and the Central Limit Theorem.

### Law of Large Numbers

In Sect. 1.1 we mentioned the intuitive idea of defining probabilities as limits of relative frequencies determined from many independent repetitions of a given probabilistic experiment. This idea can be given some theoretical justification from an asymptotic analysis of sequences of *independent* and *identically distributed* (i.i.d.) random variables $X_1$, $X_2$, .... An example would be a sequence of daily log-returns. Let us assume for the moment that these random variables have the same distribution as some random variable $X$ with finite second moments. We then write for their mean

$$\mu = E(X_n) \tag{2.1.1}$$

and for their variance

$$\sigma^2 = \text{Var}(X_n), \tag{2.1.2}$$

$n \in \mathcal{N}$. Since the random variables $X_1, X_2, \ldots$ are independent it follows that the *sample mean*

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

$$\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{2.1.3}$$

has the mean

$$E(\hat{\mu}_n) = \mu \tag{2.1.4}$$

and the variance

$$\mathrm{Var}(\hat{\mu}_n) = \frac{\sigma^2}{n}. \tag{2.1.5}$$

Note that one does not need for (2.1.3) the independence of the random variables. The *Law of Large Numbers* (LLN) is one of the fundamental results of probability theory and statistics. To formulate this law we say, that a sequence of random variables $Y_1, Y_2, \ldots$ *converges in the mean square sense* to a random variable $Y$ if

$$\lim_{n\to\infty} E\left(|Y_n - Y|^2\right) = 0. \tag{2.1.6}$$

In this case we write

$$Y \overset{\text{m.s.}}{=} \lim_{n\to\infty} Y_n. \tag{2.1.7}$$

The mean square version of the LLN using this mode of convergence is stated by the following result.

**Theorem 2.1.1.** (Mean-square LLN) *If the independent random variables $X_1$, $X_2, \ldots$ have the same finite first and second moments, then the sample mean $\hat{\mu}_n$ converges in the mean square sense to the mean $\mu$, that is*

$$\mu \overset{m.s.}{=} \lim_{n\to\infty} \hat{\mu}_n. \tag{2.1.8}$$

To see this we can write, using the independence property of $X_1, X_2, \ldots$ and equations (2.1.3), (2.1.1) and (2.1.5), the relation

$$\begin{aligned} E\left((\hat{\mu}_n - \mu)^2\right) &= E\left(\left(\frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)\right)^2\right) = \frac{1}{n^2}\sum_{i=1}^{n} E\left((X_i - \mu)^2\right) \\ &= \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^{n}\sigma^2\right) = \frac{1}{n}\,\sigma^2. \end{aligned} \tag{2.1.9}$$

Using this formula we see by (2.1.6) and (2.1.7) that (2.1.8) is established.

There exists also a strong LLN, which goes back to Kolmogorov. Here the sample mean converges *almost surely* (a.s.) and we write

$$\mu \overset{\text{a.s.}}{=} \lim_{n\to\infty} \hat{\mu}_n \tag{2.1.10}$$

for

$$P\left(\lim_{n\to\infty} \hat{\mu}_n = \mu\right) = 1. \tag{2.1.11}$$

**Theorem 2.1.2.** (Strong LLN, Kolmogorov) *For a sequence of independent random variables* $X_1, X_2, \ldots$ *with mean* $\mu$ *and*

$$\sum_{i=1}^{\infty} \frac{\text{Var}(X_i)}{i^2} < \infty \tag{2.1.12}$$

*it holds that*

$$\mu \overset{a.s.}{=} \lim_{n\to\infty} \hat{\mu}_n. \tag{2.1.13}$$

To underline that there are different types of convergence let us also state a weak LLN, which is due to Markov. For this purpose we say that a sequence of random variables $Y_1, Y_2, \ldots$ *converges in probability* to a random variable $Y$ if for any $\varepsilon > 0$

$$\lim_{n\to\infty} P(|Y_n - Y| > \varepsilon) = 0 \tag{2.1.14}$$

and we write

$$Y \overset{P}{=} \lim_{n\to\infty} Y_n. \tag{2.1.15}$$

**Theorem 2.1.3.** (Weak LLN, Markov) *For a sequence of uncorrelated random variables* $X_1, X_2, \ldots$ *with mean* $E(X_i) = \mu,\ i \in \mathcal{N}$, *and*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum_{i=1}^{n} \text{Var}(X_i) = 0 \tag{2.1.16}$$

*one has*

$$\mu \overset{P}{=} \lim_{n\to\infty} \hat{\mu}_n. \tag{2.1.17}$$

In the case of an i.i.d. sequence of log-returns one can therefore estimate via the sample mean the mean of the log-returns. In Fig. 2.1.2 we plot the sample mean for the log-returns of the S&P500 as it evolves for an increasing number of observations.

The link between relative frequencies, as discussed in Sect. 1.1, and corresponding probabilities can now be directly established by using the weak LLN. If $A$ is an event and $\frac{N(A)}{N}$ the relative frequency of $A$ occurring in $N \in \mathcal{N}$ independent, identical observations of $A$, then

$$P(A) \overset{P}{=} \lim_{N\to\infty} \frac{N(A)}{N}. \tag{2.1.18}$$

This is a fundamental result, which supports our empirical analysis and stochastic modeling in finance.

**Empirical Moments**

We have shown by the weak LLN under appropriate conditions that the sample mean $\hat{\mu}_n$, which is also the *first empirical moment*, approaches the true mean $\mu$ of uncorrelated random variables $X_1, X_2, \ldots, X_n$ in probability for increasing $n$. The sample mean $\hat{\mu}_n$ is therefore a reasonable estimate for the mean $\mu$. This provides a method for estimating the mean of a sequence of uncorrelated random variables.

Consider i.i.d. random variables $X_1, X_2, \ldots$ under the conditions $E(X_i^2) < \infty$, $i \in \mathcal{N}$, one can also show that the *sample variance*

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \hat{\mu}_n)^2, \tag{2.1.19}$$

which we define as the *second empirical central moment*, converges almost surely to the variance $\sigma^2$ of the i.i.d. random variables $X_1, X_2, \ldots$.

Similarly, under the conditions $E(X_i^3) < \infty$ and $\text{Var}(X_i) > 0$, $i \in \mathcal{N}$, the *sample skewness*

$$\hat{\beta}_n = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{X_i - \hat{\mu}_n}{\hat{\sigma}_n} \right)^3 \tag{2.1.20}$$

approaches almost surely the skewness $\beta_X$, see (1.3.22). The *sample kurtosis*

$$\hat{\kappa}_n = \frac{1}{n} \sum_{i=1}^{n} \left( \frac{X_i - \hat{\mu}_n}{\hat{\sigma}_n} \right)^4 \tag{2.1.21}$$

provides under the conditions $E(X_i^4) < \infty$ and $\text{Var}(X) > 0$, $i \in \mathcal{N}$, an a.s. converging estimate for the kurtosis $\kappa_X$, see (1.3.30), for i.i.d. random variables $X_1, X_2, \ldots$. To obtain useful estimates for these empirical moments one has, therefore, only to ensure that the corresponding moments are finite if one has i.i.d. observations.

Let us consider a simulated sequence of independent identically $N(-1, 1)$ Gaussian distributed random variables. These have by Table 1.3.1 mean $\mu = -1$, variance $\sigma^2 = 1$, skewness $\beta = 0$ and kurtosis $\kappa = 3$. Figure 2.1.1 displays linearly interpolated graphs for the resulting sample mean, sample variance, sample skewness and sample kurtosis for increasing values of the sample size $n \in \{10, 11, \ldots, 1000\}$.

As suggested by the weak LLN, for increasing sample sizes we see that the empirical moments appear to converge towards the respective values of the moments; in this case the mean $\mu = -1$, variance $\sigma^2 = 1$, skewness $\beta = 0$ and kurtosis $\kappa = 3$. One notes that for higher order moments one needs more observations to stabilize the corresponding empirical sample moment.

As another illustration, let us calculate the empirical moments from observations of daily log-returns of the S&P500 index covering the twenty year period from 1977 until 1997. For these S&P500 log-returns we obtain the empirical moments

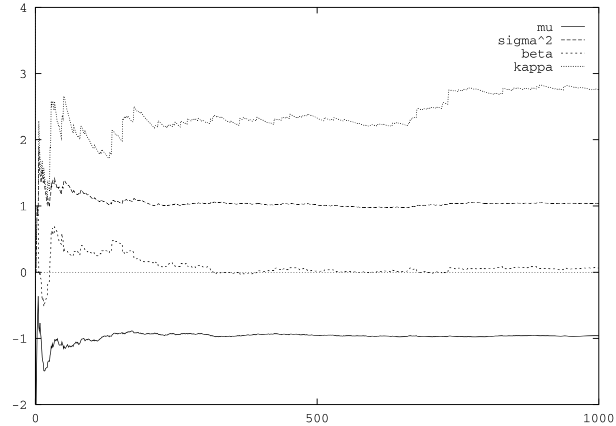



**Fig. 2.1.1.** Empirical moments from a simulation

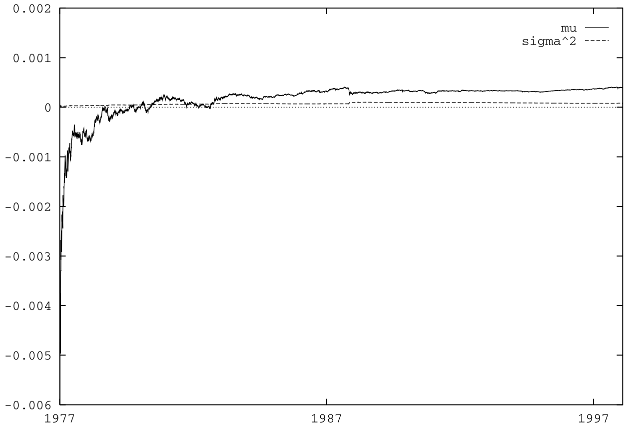



**Fig. 2.1.2.** Sample mean and variance for S&P500 log-returns

$$\hat{\mu}_n = 0.00040, \quad \hat{\sigma}_n^2 = 0.000082, \quad \hat{\beta}_n = -2.22, \quad \hat{\kappa}_n = 58.43 \tag{2.1.22}$$

for sample size $n = 5478$. In Fig. 2.1.2, we show the corresponding sample mean and sample variance as they evolve over time in dependence on time. These converge reasonably well towards the corresponding values shown in (2.1.22). We then display the resulting evolution of the sample skewness and sample kurtosis in Fig. 2.1.3. It is apparent that these are not very stable estimates. In particular, the values jump considerably at the 1987 stock market crash. If we remove from our sample the largest absolute log-return that occurred at the October 1987 market crash, then we obtain with the remaining $n = 5477$ observations the empirical moments

$$\hat{\mu}_n = 0.00044, \quad \hat{\sigma}_n^2 = 0.000074, \quad \hat{\beta}_n = -0.098, \quad \hat{\kappa}_n = 11.06. \tag{2.1.23}$$

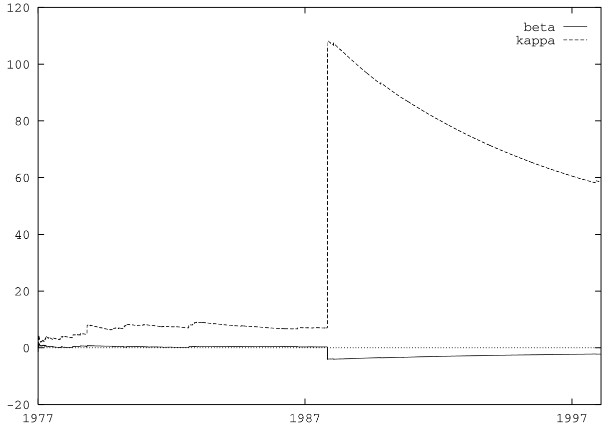



**Fig. 2.1.3.** Sample skewness and kurtosis for S&P500 log-returns

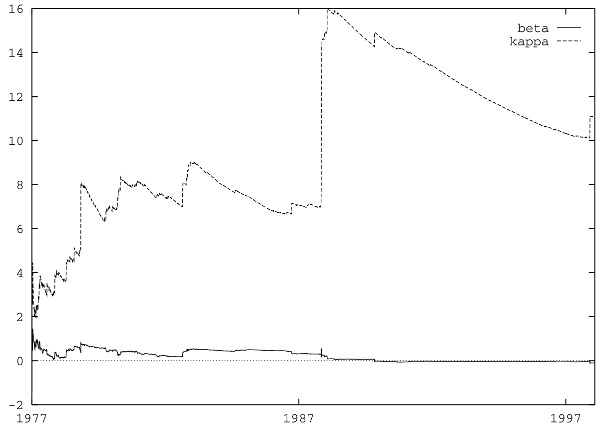



**Fig. 2.1.4.** Sample skewness and kurtosis for S&P500 log-returns without 1987 crash

This calculation shows that the estimated sample kurtosis changed dramatically after removing the most extreme log-return. In Fig. 2.1.4 we show the corresponding empirical skewness and kurtosis for the reduced sample. A comparison of Fig. 2.1.3 and Fig. 2.1.4 indicates that the fourth empirical moment is extremely sensitive with respect to this data set. The kurtosis $\kappa$ might not even be finite for log-returns of the S&P500 index if fitted to a reasonable class of models. We shall show later in this chapter that typical parameter estimates of stock market index log-returns in the class of symmetric generalized distributions imply infinite kurtosis. For this reason, when estimating log-returns, it is recommended one uses a statistical approach that exploits the entire distribution and does not depend on any higher order empirical moments, such as the sample kurtosis.

### Central Limit Theorem

To obtain more information regarding the asymptotics of the sample mean $\hat{\mu}_n$ one needs another fundamental result. To prepare its formulation we say that a sequence of random variables $X_1, X_2, \ldots$ *converges in distribution* to a random variable $X$ if the distribution function $F_{X_n}(x)$ converges at each point $x$ of continuity of $F_X(x)$, and we write

$$X \stackrel{\mathrm{d}}{=} \lim_{n\to\infty} X_n. \tag{2.1.24}$$

The Central Limit Theorem CLT states the following result.

**Theorem 2.1.4.** (CLT) *A standardized sample average*

$$\hat{Z}_n = \sqrt{n}\,\frac{(\hat{\mu}_n - \mu)}{\sigma}, \tag{2.1.25}$$

*for a sequence $X_1$, $X_2$, $\ldots$ of i.i.d. random variables with mean $\mu \in \Re$ and variance $\sigma^2 \in (0,\infty)$ converges in distribution, as $n \to \infty$, to a standard Gaussian random variable $Z \sim N(0,1)$, that is*

$$\lim_{n\to\infty} \hat{Z}_n \stackrel{d}{=} Z. \tag{2.1.26}$$

This fundamental theorem states that

$$\lim_{n\to\infty} F_{\hat{Z}_n}(z) = F_Z(z) = N(z), \tag{2.1.27}$$

for all $z \in \Re$, where $N(\cdot)$ denotes the standard Gaussian distribution function, see (1.2.7). One can show that one only needs the existence of the third absolute moment $E(|X_i|^3) < \infty,\ i \in \mathcal{N}$, of the i.i.d. distributed random variables to achieve Gaussianity for the standardized sample average together with the *Berry-Esseen inequality*

$$\sup_{x\in\Re} \left| P\left(\hat{Z}_n < x\right) - N(x) \right| \le n^{-\frac{1}{2}}\,\frac{0.8}{\sigma^2}\,E\left(|X_1 - E(X_1)|^3\right). \tag{2.1.28}$$

As a consequence of the CLT the independence of the random variables $X_1, X_2, \ldots$ and the existence of second moments guarantee a Gaussian limit for $\hat{Z}_n$. This provides an explanation for the dominant role of the Gaussian distribution in probability and statistics and many areas of application including quantitative finance. For instance, one observes for most financial securities that for increasing periods of time the corresponding long term log-returns seem to approach Gaussian random variables. In view of the CLT this is not a surprising observation if one interprets short term log-returns as i.i.d. random variables.

### Bernoulli Trials

Let us provide a simple illustration of the Law of Large Numbers and also the Central Limit Theorem. *Bernoulli trials* are independent repetitions of an experiment with two basic outcomes which might occur with probabilities $p$ and $1-p$, respectively. If we set $X_n = 1$ for a positive log-return and $X_n = 0$ for a non-positive log-return, then we can model these log-returns by using an i.i.d. sequence of random variables $X_1, X_2, \ldots$ with mean

$$\mu = E(X_n) = p \tag{2.1.29}$$

and variance

$$\sigma^2 = \mathrm{Var}(X_n) = p\,(1-p). \tag{2.1.30}$$

The sum

$$H_n = X_1 + X_2 + \cdots + X_n = n\,\hat{\mu}_n, \tag{2.1.31}$$

see (2.1.3), counts the number of positive log-returns occurring out of $n$ observed trials. In a Bernoulli trial one is typically interested in the number of outcomes that correspond to a given specific event. Furthermore, $\hat{\mu}_n = \frac{H_n}{n}$ measures the relative frequency of observing such an event, in our example the occurrence of positive log-returns. The LLNs tell us, as $n \to \infty$, that the random variables $\hat{\mu}_n$ converge in a meaningful sense to the value $p$, which in our example is the probability of having a positive log-return for a single observation. Additionally, by the CLT we know that, as $n \to \infty$, the standardized sample mean

$$\hat{Z}_n = \sqrt{n}\,\frac{\left(\frac{H_n}{n} - p\right)}{\sqrt{p\,(1-p)}},$$

see (2.1.25)–(2.1.31), converges in distribution to a standard Gaussian random variable $Z$.

### Binomial Distribution

We remark that the probability for the event $H_n = m$ is the same as the *binomial probability* for $m$ successes out of $n$ trials, that is

$$P(H_n = m) = p^m\,(1-p)^{n-m}\,\frac{n!}{(n-m)!\,m!}, \tag{2.1.32}$$

where we recall that $k! = 1 \cdot 2 \cdot \ldots \cdot k$ and $0! = 1$. Figure 2.1.5 shows for $p = 0.5$ and $n = 10$ the resulting binomial probabilities, when these are interpolated. These probabilities resemble the corresponding bell shaped Gaussian density function, as indicated by the CLT, which is also included in Fig. 2.1.5 for comparison. The Gaussian density is the curve with the slightly larger value at the mean. This means, for $p$ asymptotically not vanishing for large $n$ that the binomial probabilities tend asymptotically to the values of the Gaussian density.

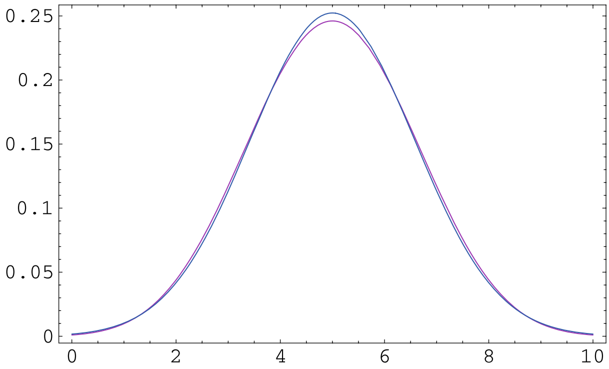



**Fig. 2.1.5.** Binomial probabilities for $p = 0.5$ and $n = 10$ and Gaussian density

Binomial probabilities appear in finance, for example, in random walk and binomial tree approximations of continuous time asset price models, as will be discussed later. However, binomial distributions are also linked to the Poisson distribution as the following statement shows. If $n$ is large and $p$ is small in (2.1.32), such that $\lambda = np > 0$, then it holds asymptotically for $n \to \infty$ that

$$P(H_n = m) \approx \exp\{-\lambda\} \frac{\lambda^m}{m!} \tag{2.1.33}$$

for $m \in \{0, 1, \ldots\}$. This means that for large $n$ and $\lambda = np$ the binomial probabilities approach Poisson probabilities, see (1.1.30).

## 2.2 Confidence Intervals

To analyze empirically market data one needs sophisticated statistical tools. With the sample mean that we introduced in the previous section we have an estimate for the mean. We may ask, what is the number of observations needed to obtain a reasonably accurate estimate and how correct is this estimate? This question can be answered in different ways, as we shall see below.

### Basic Confidence Intervals

Let us again consider the Bernoulli trials introduced previously. These are formed by a sequence of i.i.d. random variables $X_1, X_2, \ldots$ taking the value 1 with probability $p$ and the value 0 with probability $1 - p$, where we now assume that $p$ is unknown to us. We recall that $\mu = E(X_n) = p$ and $\sigma^2 = \text{Var}(X_n) = p\,(1 - p)$ and so it follows

$$E(\hat{\mu}_n) = p$$

and by (1.4.26) and (2.1.30)

$$\mathrm{Var}(\hat{\mu}_n) = \frac{p\,(1-p)}{n}.$$

As outlined previously, the sample mean $\hat{\mu}_n$ converges by the LLN, see (2.1.8), in a mean square sense to $p$. We can apply the Chebyshev inequality (1.3.58) to the random variable $\hat{\mu}_n - p$ and may then use the inequality $\sigma^2 = p\,(1-p) \le \frac{1}{4}$ for $p \in [0,1]$ to obtain

$$P\left(|\hat{\mu}_n - p| \ge a\right) = P(|\hat{\mu}_n - \mu| \ge a) \le \frac{\sigma^2}{n\,a^2} = \frac{p\,(1-p)}{n\,a^2} \le \frac{1}{4\,n\,a^2} \tag{2.2.1}$$

for $a > 0$ and so

$$P\left(|\hat{\mu}_n - p| < a\right) = 1 - P\left(|\hat{\mu}_n - p| \ge a\right) \ge 1 - \frac{1}{4\,n\,a^2}. \tag{2.2.2}$$

Thus, for any $0 < \alpha < 1$ and $a > 0$ we can conclude that the unknown mean $p$ lies in the interval $(\hat{\mu}_n - a,\, \hat{\mu}_n + a)$ with at least probability $1 - \alpha$ when

$$n \ge n(a, \alpha) = \frac{1}{4\,\alpha\,a^2}. \tag{2.2.3}$$

In statistical terminology we say that the hypothesis that $p$ belongs to the interval $(\hat{\mu}_n - a,\, \hat{\mu}_n + a)$ is acceptable at a $100(1-\alpha)\%$ *level of confidence* if

$$P(|\hat{\mu}_n - p| < a) = 1 - \alpha \tag{2.2.4}$$

and call

$$(\hat{\mu}_n - a, \hat{\mu}_n + a)$$

the $100(1-\alpha)\%$ *confidence interval*. In our example, $(\hat{\mu}_n - 0.1,\, \hat{\mu}_n + 0.1)$ is at least a 95% confidence interval when $n \ge n(0.1, 0.05) = 500$. If $n$ were fixed, then we would obtain from (2.2.3) in this example the inequality

$$a \ge \frac{1}{2\sqrt{\alpha\,n}}. \tag{2.2.5}$$

We note that the length of the confidence interval only decreases proportionally to $n^{-\frac{1}{2}}$ for increasing number of observations $n$. This is a general phenomenon, as we shall see later on.

### Gaussian Confidence Interval and VaR

For a Gaussian random variable $X$ with known mean $\mu$ and known variance $\sigma^2$ the $100(1-\alpha)\%$ confidence interval with

$$P\left(\left|\frac{X-\mu}{\sigma}\right| < p_{1-\frac{\alpha}{2}}\right) = 1 - \alpha \tag{2.2.6}$$

is given in the form

$$(\mu - \sigma\, p_{1-\frac{\alpha}{2}}, \mu + \sigma\, p_{1-\frac{\alpha}{2}}). \tag{2.2.7}$$

Here $p_{1-\alpha}$ is the $100(1-\alpha)\%$ quantile of the standard Gaussian distribution. For instance, a 99% confidence interval requires one to choose $p_{1-\alpha} \approx 2.58$.

The confidence interval (2.2.7) is a *two-sided confidence interval.* Sometimes however, in particular, in the computation of *Value at Risk* (VaR), one is interested in determining a critical maximum loss $\mathrm{VaR}((1-\alpha)\%)$ so that one can assert with $(1-\alpha)\%$ confidence that $X$ is at least as large as $-\mathrm{VaR}((1-\alpha)\%)$, that is

$$P\left(X \geq -\mathrm{VaR}((1-\alpha)\%)\right) = 1 - \alpha \tag{2.2.8}$$

or equivalently

$$P\left(\frac{X-\mu}{\sigma} < -z_\alpha\right) = \alpha, \tag{2.2.9}$$

with

$$z_\alpha = \frac{\mathrm{VaR}((1-\alpha)\%) + \mu}{\sigma} \tag{2.2.10}$$

Then (2.2.9) corresponds to the *one sided confidence interval*

$$(-\infty, -z_\alpha), \tag{2.2.11}$$

where the random variable $\frac{X-\mu}{\sigma}$ can be found with $\alpha\%$ probability in that interval. According to (2.2.8) and (2.2.10) $X$ will thus be with $(1-\alpha)\%$ probability above the level

$$-\,\mathrm{VaR}((1-\alpha)\%) = -(z_\alpha\, \sigma - \mu). \tag{2.2.12}$$

For example, for a Gaussian random variable $X$ and $\alpha = 0.01$ we have the $\alpha\%$ percentile with value $z_{0.01} \approx 2.35$, which determines according to (2.2.12) the corresponding maximum critical loss $\mathrm{VaR}(99\%) = z_{0.01}\, \sigma - \mu$.

### Student $t$ Confidence Interval

Often we do not know the variance $\sigma^2$ or do not have a reasonable estimate for it. However, in these cases we can use the sample variance $\hat{\sigma}_n^2$, see (2.1.19), to construct appropriate confidence intervals.

Let $X_1$, $X_2$, …, $X_n$ denote $n$ i.i.d. $N(\mu, \sigma^2)$ Gaussian random variables with known mean $\mu$ and unknown variance $\sigma^2$. As described in (2.1.3) and (2.1.19) we have the sample mean

$$\hat{\mu}_n = \frac{1}{n} \sum_{j=1}^{n} X_j \tag{2.2.13}$$

and sample variance

**Table 2.2.1.** Quantiles for the Student $t$ distribution

| n | 10 | 20 | 30 | 40 | 60 | 100 | 200 |
|---|---|---|---|---|---|---|---|
| $t_{0.9,n-1}$ | 1.83 | 1.73 | 1.70 | 1.68 | 1.67 | 1.66 | 1.65 |
| $t_{0.99,n-1}$ | 3.25 | 2.86 | 2.76 | 2.70 | 2.66 | 2.62 | 2.58 |

$$\hat{\sigma}_n^2 = \frac{1}{n-1} \sum_{j=1}^{n} \left(X_j - \hat{\mu}_n\right)^2 . \tag{2.2.14}$$

Then for $n > 3$ it can be shown that the random variable

$$T_n = \frac{\hat{\mu}_n - \mu}{\sqrt{\frac{\hat{\sigma}_n^2}{n}}} \tag{2.2.15}$$

is Student $t$ distributed, see (1.2.17), with $n-1$ degrees of freedom, that is $T_n \sim t(n-1)$.

We have

$$P\left(|\hat{\mu}_n - \mu| < a\right) = P\left(|T_n| < t\right) = 2\left(F_{T_n}(t) - 0.5\right), \tag{2.2.16}$$

where

$$t = a\sqrt{\frac{n}{\hat{\sigma}_n^2}} \tag{2.2.17}$$

and $F_{T_n}(x)$ is the value of the Student $t$ distribution function with $n-1$ degrees of freedom evaluated at $x \in \Re$.

Thus, for a given $100\alpha\%$ confidence level, we can check whether or not the test variable

$$T_n^0 = \frac{\hat{\mu}_n - \mu_0}{\sqrt{\frac{\hat{\sigma}_n^2}{n}}}$$

with hypothesized mean $\mu_0$ satisfies the inequality

$$|T_n^0| < t_{1-\alpha,n-1}. \tag{2.2.18}$$

Here $t_{1-\alpha,n-1}$ is the $100\,(1-\alpha)\%$ *quantile* of the Student $t$ distribution with $n-1$ degrees of freedom for which

$$2\left(F_{T_n^0}(t_{1-\alpha,n-1}) - 0.5\right) = 1 - \alpha.$$

Some values of the quantile $t_{1-\alpha,n-1}$ are given in Table 2.2.1. If the relation (2.2.18) is not fulfilled, then we reject the *null hypothesis* $H_0$ that $\mu = \mu_0$. Otherwise, we accept it on the basis of this test. In addition, we can form the corresponding $100(1-\alpha)\%$ confidence interval

$$(\hat{\mu}_n - a, \hat{\mu}_n + a),$$

see (2.2.17), with

$$a = t_{1-\alpha,n-1}\sqrt{\frac{\hat{\sigma}_n^2}{n}}. \tag{2.2.19}$$

We call this interval the *Student t confidence interval.* It contains all of the values $\mu_0$ for which the null hypothesis would not be rejected by this test. Applications of this result can be found, for instance, in the treatment of errors in Monte Carlo simulations.

We note also that the length of the above confidence interval is proportional to $n^{-\frac{1}{2}}$. This means, to obtain a ten times smaller confidence interval requires approximately hundred times as many observations. We face this phenomenon in the application of Monte Carlo methods.

The above described procedure requires $X_1, X_2, \ldots$ to be Gaussian. When this is not the case we note by the CLT that sample means of sufficiently large groups or batches of these i.i.d. random variables will be approximately Gaussian. Consequently, with such a construction we can approximately apply the above methodology to these sample means. To be more precise we take $n$ batches of $m$ i.i.d. random variables $X_1^{(j)}$, $X_2^{(j)}$, ..., $X_m^{(j)}$ for $j \in \{1, 2, \ldots, n\}$. Then we form the sample means

$$\hat{\mu}_m^{(j)} = \frac{1}{m}\sum_{\ell=1}^{m} X_\ell^{(j)}$$

and apply the above Student $t$ methodology to these sample means rather than the original random variables $X_i^{(j)}$. For practical applications it has often been found that the batches should consist of at least 15 random variables to provide a reasonable approximation.

In the case when the variance $\sigma^2$ is known the following test variable

$$\bar{T}_n = \sqrt{n}\,\frac{(\hat{\mu}_n - \mu)}{\sigma}$$

is Gaussian and we can construct similar confidence intervals as above, but based on the Gaussian distribution rather than the Student $t$ distribution. Recall also that the Student $t$ distribution asymptotically approaches a Gaussian distribution as the degrees of freedom tend to infinity.

### VaR Analysis for Student $t$ Log-Returns (*)

As outlined in regulatory recommendations the modeling of, so-called, *event risk* is of increasing importance in VaR analysis.

In the following version of a Student $t$ log-return model we exploit the fact that symmetric generalized hyperbolic distributions admit a representation as a mixture of normal distributions. This means, if one chooses the variance of a conditionally Gaussian distribution as the inverse of a Gamma distributed random variable, then the resulting distribution is a Student $t$ distribution, see also Sect. 1.2. For simplicity, we neglect here the impact of any asymmetry since for the short time intervals considered this is not relevant. We shall provide later in this chapter more details on *normal variance mixture models*.

To generate Student $t$ distributed log-returns $Z^{(1)}, \ldots, Z^{(d)}$ for $d$ securities at a fixed time we set

$$Z^{(k)} = \sqrt{\tau}\, Y^{(k)} \tag{2.2.20}$$

for $k \in \{1, 2, \ldots, d\}$, where $\tau$ denotes the conditional variance with

$$\tau = \left(1 - \frac{2}{n}\right) \left(\frac{1}{n} \sum_{\ell=1}^{n} \left(\psi^{(\ell)}\right)^2\right)^{-1} \tag{2.2.21}$$

and $n \in \{3, 4, \ldots\}$. Additionally to the independent standard Gaussian distributed random variables $Y^{(k)}$ that appear in (2.2.20) we employ further independent standard Gaussian random variables $\psi^{(\ell)}$. Hence, the random variable $\tau$ is chi-square distributed with $n$ degrees of freedom, see Sect. 1.2. Consequently, the random variables $Z^{(k)}$, $k \in \{1, 2, \ldots, d\}$, are Student $t$ distributed with unit variance and $n$ degrees of freedom. The conditional variance $\tau$ can be interpreted as a measure of the random activity of the market during the time period of interest. Note that the conditional variance converges to one as the degrees of freedom $n$ tend to infinity, which yields asymptotically normal log-returns.

In addition to the typical parameters of the lognormal model we have used here only the extra parameter $n$, which is sufficient to characterize the leptokurtosis of the Student $t$ distribution. As will be shown later, a typical parameter choice for $n$ is about four. Smaller degrees of freedom generate log-returns with more extreme movements.

An important feature of the resulting multivariate Student $t$ distribution for log-returns is its copula, see Sect. 1.5. It realistically captures the estimated dependence of extreme asset price movements, as shown in Embrechts, McNeil & Straumann (2002) and McNeil, Frey & Embrechts (2005). Let $\boldsymbol{Y} = (Y^{(1)}, \ldots, Y^{(d)})^\top$ denote a vector of independent standard Gaussian distributed random variables and $\tau$ be an independent chi-square random variable. Note that the joint distribution of the random vector $\boldsymbol{X} = (X^{(1)}, \ldots, X^{(d)})^\top$ with

$$\boldsymbol{X} = \sqrt{\tau}\, \boldsymbol{D}\, \boldsymbol{Y}$$

is a multivariate Student $t$ distribution with $n$ degrees of freedom. Here $\boldsymbol{D}$ is the Cholesky decomposition of the covariance matrix $\mathrm{Cov}_{\boldsymbol{X}}$ of $\boldsymbol{X}$, see Sect. 1.4. Since $\boldsymbol{Y}$ is Gaussian and $\frac{1}{\tau}$ is independent chi-square distributed, the resulting multivariate Student $t$ distribution of $\boldsymbol{X}$ belongs to the class of *elliptic distributions*. One can show that the calculation of VaR numbers is for this class of distributions analytically tractable, see Platen & Stahl (2003). More precisely, a theorem in Fang, Kotz & Ng (1990) yields the representation

$$\boldsymbol{a}^\top \boldsymbol{X} = |\boldsymbol{a}^\top \boldsymbol{D}|\, \zeta \tag{2.2.22}$$

for any given weight vector $\boldsymbol{a} = (a^1, a^2, \ldots, a^d)^\top$, where $|\cdot|$ is the Euclidean norm, $\boldsymbol{a} \in \Re^d$, $\boldsymbol{D}^\top \boldsymbol{D} = \text{Cov}_{\boldsymbol{X}}$ and $\zeta$ denotes a Student $t$ distributed scalar random variable with $n$ degrees of freedom. The representation (2.2.22) significantly simplifies the VaR calculation for portfolios even if these have an extremely large number of constituents.

Since the multivariate Student $t$ distribution is an elliptical distribution, it follows from Embrechts et al. (2002), that VaR is in this case a, so-called, *coherent risk measure*, see Artzner, Delbaen, Eber & Heath (1997). This fact is highly important for the consistent use of VaR as a risk measure for internal capital allocation to particular business lines. The property of coherent risk measures that sometimes creates problems for VaR is the additivity, where the risk measure for the sum of two risky securities should never be greater than the sum of their risk measures, see Föllmer & Schiedt (2002).

In order to calculate VaR for the given short term horizon we apply, the so-called, square root time rule, which is in line with regulatory recommendations. From (2.2.22) we obtain then the following formula for the VaR number of a given portfolio at the given time:

$$\text{VaR}_h(V, \alpha) \approx V \sqrt{\boldsymbol{a}^\top \, \text{Cov}_{\boldsymbol{X}} \, \boldsymbol{a}} \, \sqrt{h \, \varDelta} \; \tilde{t}_\alpha(n). \tag{2.2.23}$$

Here $\sqrt{\boldsymbol{a}^\top \text{Cov}_{\boldsymbol{X}} \boldsymbol{a}}$ characterizes the total volatility of the portfolio, $V$ denotes the market value of the portfolio at the given time, $\varDelta$ is the time step size for a trading day, $h$ the number of trading days and $\tilde{t}_{1-\alpha}(n)$ the $100(1-\alpha)\%$-quantile of the Student $t$ distribution with $n$ degrees of freedom.

The product (2.2.23) generalizes a short hand formula, used in practice, to calculate VaR by including the *event factor*

$$\varphi = \frac{\tilde{t}_{1-\alpha}(n)}{p_{1-\alpha}}, \tag{2.2.24}$$

that is

$$\text{VaR}_h(V, \alpha) \approx V \sqrt{\boldsymbol{a}^\top \, \text{Cov}_{\boldsymbol{X}} \, \boldsymbol{a}} \, \sqrt{h \, \varDelta} \; p_{1-\alpha} \, \varphi. \tag{2.2.25}$$

Here $p_{1-\alpha}$ is the $100(1-\alpha)\%$-quantile of the standard Gaussian distribution. Consequently, the event factor $\varphi$ adjusts the standard VaR formula to a level that captures the, so-called, event risk when one uses Student $t$ log-returns. According to the quantiles of the Gaussian and Student $t$ distribution one obtains by (2.2.23) the event factors shown in Table 2.2.2. Even for rather small degrees of freedom, say $n \approx 2$, the additional regulatory capital will not surpass 16%.

**Table 2.2.2.** Event factor $\varphi$ in dependence on degrees of freedom $n$

| $n$ | $\infty$ | 10 | 5 | 4 | 3 | 2 |
|---|---|---|---|---|---|---|
| $\varphi$ | 1 | 1.06 | 1.11 | 1.12 | 1.14 | 1.16 |

Gibson (2001) performed an extensive study using an extremely large set of representative portfolios of US institutions, where he identified empirically an event factor of about $\hat{\varphi} \approx 1.12$. One notes that this is exactly the value of the event factor that matches in Table 2.2.2 the one for the degrees of freedom $n = 4$. We shall see later that this finding supports a model proposed in Platen (2002), the minimal market model, which will be derived later in Chap. 13. Also our inference later in this chapter will suggest a Student $t$ distribution with four degrees of freedom as a realistic estimate for the log-return distribution of indices.

## 2.3 Estimation Methods

There exists a wide range of estimation techniques developed for various inference problems that have major importance in quantitative finance. The choice of a suitable estimation method depends on the available data and the assumed model. In this section we concentrate mainly on linear models.

### Estimators

Assume that there are $n \in \mathcal{N}$ real valued observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n} \in \Re$ of, say, log-returns. These observations contain information about the parameters $\theta_1, \theta_2, \ldots, \theta_q \in \Re$ that we wish to estimate. The observations can be represented as *observation vector* $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top \in \Re^n$ and the parameters as *parameter vector* $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Theta \subseteq \Re^q$, where $\Theta$ specifies the set of allowable values for the parameters, $q \in \mathcal{N}$.

Generally, an *estimator* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q)^\top \in \Theta$ is a function $\hat{\boldsymbol{\theta}} : \Re^n \to \Theta$ by which the parameters can be approximately identified from the observations, that is, by the *estimate*

$$\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(R_{t_1}, R_{t_2}, \ldots, R_{t_n}). \tag{2.3.1}$$

For example, as discussed previously, two typical parameters that are often needed are the mean $\theta_1 = E(X)$ and the variance $\theta_2 = E((X-\mu)^2)$ of a Gaussian random variable $X \sim N(\theta_1, \theta_2)$, say Gaussian daily log-returns. Given an observation vector $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top$ with components that consist of i.i.d. observations of $X$, these parameters can be estimated. According to (2.2.13) the sample mean

$$\hat{\theta}_1 = \frac{1}{n} \sum_{j=1}^{n} R_{t_j} \tag{2.3.2}$$

estimates the value of $\theta_1$. By (2.2.14) the sample variance

$$\hat{\theta}_2 = \frac{1}{n-1} \sum_{j=1}^{n} (R_{t_j} - \hat{\theta}_1)^2 \tag{2.3.3}$$

provides an estimator for $\theta_2$. In this example we have $q = 2$ and $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2)^\top \in \Re^2$. For Gaussian daily log-returns, $\hat{\theta}_1$ would be the estimate of the expected daily growth rate and $\hat{\theta}_2$ the estimate of the variance of the daily log-returns.

### Unbiasedness and Consistency

The assessment of the quality of an estimate $\hat{\theta}_i$, $i \in \{1, 2, \ldots, q\}$, of the $i$th parameter $\theta_i$ can be based on the *estimation error*

$$\tilde{\theta}_i(n) = \theta_i - \hat{\theta}_i. \tag{2.3.4}$$

Ideally, the estimation error should be close to zero almost surely. However, this is difficult to achieve for a finite set of observations and a general model. Therefore, less stringent requirements are typically used.

The first requirement, which is often formulated, is that the expected value of the estimation error should be zero. That is, by taking expectations we obtain from (2.3.4) the condition

$$0 = E(\tilde{\theta}_i(n)) = E(\theta_i) - E(\hat{\theta}_i) = \theta_i - E(\hat{\theta}_i) \tag{2.3.5}$$

for $i \in \{1, 2, \ldots, q\}$. Estimators that satisfy relation (2.3.5) are called *unbiased* and it follows by (2.3.5) that

$$E(\hat{\theta}_i) = \theta_i \tag{2.3.6}$$

for $i \in \{1, 2, \ldots, q\}$.

In the case when the estimator $\hat{\theta}_i$ does not meet the unbiasedness condition (2.3.5), then $\hat{\theta}_i$ is said to be *biased*. The *bias* $E(\tilde{\theta}_i(n))$ is defined as the expected value of the estimation error (2.3.4). If the bias tends to zero as the number $n$ of observations increases, then the estimator is called *asymptotically unbiased*, that is

$$\lim_{n \to \infty} E(\tilde{\theta}_i(n)) = 0, \tag{2.3.7}$$

$i \in \{1, 2, \ldots, q\}$. A reasonable requirement for an unbiased estimator $\hat{\theta}_i$ is that its estimation error should, for increasing number $n$ of observations, converge in probability to zero, that is

$$\lim_{n \to \infty} \tilde{\theta}_i(n) \stackrel{P}{=} 0 \tag{2.3.8}$$

for $i \in \{1, 2, \ldots, q\}$. An estimator satisfying the property (2.3.8) is called *consistent*.

In our previous example, where $X \sim N(\theta_1, \theta_2)$ is a Gaussian random variable and $R_{t_1}, R_{t_2}, \ldots$ are independent observations of $X$, the expected value of the sample mean $\hat{\theta}_1$ given in (2.3.2) is

$$E(\hat{\theta}_1) = \frac{1}{n}\sum_{j=1}^{n} E(R_{t_j}) = \frac{1}{n}\, n\, \theta_1 = \theta_1. \tag{2.3.9}$$

Consequently, in this case the sample mean $\hat{\theta}_1$ is an unbiased estimator of the mean $\theta_1$. We obtain for the estimation error $\tilde{\theta}_1(n)$ the variance

$$E(\tilde{\theta}_1(n)^2) = E\left(\left(\frac{1}{n}\sum_{j=1}^{n}(R_{t_j} - \theta_1)\right)^2\right) = \frac{1}{n^2}\sum_{j=1}^{n} E\left((R_{t_j} - \theta_1)^2\right) = \frac{1}{n}\,\theta_2. \tag{2.3.10}$$

The variance in (2.3.10) converges to zero as $n \to \infty$. This implies by the Chebyshev inequality (1.3.58) and equation (2.3.10) that $\tilde{\theta}_1(n)$ converges in probability to zero, that is

$$P(\tilde{\theta}_1(n) > \varepsilon) \leq \frac{1}{\varepsilon^2}\, E\left(\tilde{\theta}_1(n)^2\right) = \frac{\theta_2}{n\,\varepsilon^2}. \tag{2.3.11}$$

Therefore, the estimator $\hat{\theta}_1$ is by (2.3.8) consistent.

In (2.1.22) we have estimated for the S&P500 daily log-returns with mean $\theta_1 \approx 0.0004$ and variance $\theta_2 \approx 0.00008$. By (2.3.11) to obtain with about only a probability of $\frac{\theta_2}{n\,\varepsilon^2} \approx 0.1$ an estimation error $\tilde{\theta}_1(n) > \varepsilon \approx 0.1\theta_1$ one needs more than $n \approx \frac{\theta_2}{0.1\,\varepsilon^2} \approx 500,000$ observations. This is an enormous number of daily observations that is needed to get any rough idea about the daily expected growth of an underlying security, as the S&P500. It requires far more data than market history offers. Therefore, without the availability of any extra structure it is highly unrealistic to expect any reliably estimates of trend, drift or growth parameters for financial securities. In particular, it is unlikely that with the available data one can estimate equity risk premia realistically.

### Efficiency

One calls an estimator, which yields the lowest variance estimate an *efficient estimator*. It uses optimally, in a least-square sense, the information contained in the observations. Therefore, a useful measure of the quality of an estimator with estimation error vector $\tilde{\boldsymbol{\theta}}(n) = (\tilde{\theta}_1(n), \ldots, \tilde{\theta}_q(n))^\top$ is given by the *error covariance matrix*

$$\mathrm{Cov}_{\tilde{\boldsymbol{\theta}}(n)} = E\left(\tilde{\theta}(n)\,\tilde{\theta}(n)^\top\right). \tag{2.3.12}$$

This matrix measures the errors of individual estimators also in relation to each other. One obtains a scalar error measure for the $i$th estimator by considering the $i$th diagonal element in (2.3.12), which is the $i$th *mean-square error*

$$\mathrm{Cov}^{i,i}_{\tilde{\boldsymbol{\theta}}(n)} = E\left((\tilde{\theta}_i(n))^2\right), \tag{2.3.13}$$

$i \in \{1, 2, \ldots, q\}$. An overall scalar error measure is obtained by summing up all the diagonal elements of $\mathrm{Cov}_{\tilde{\boldsymbol{\theta}}(n)}$, which yields the *mean-square error*

$$\mathrm{Mse}_{\tilde{\boldsymbol{\theta}}(n)} = E\left(\tilde{\boldsymbol{\theta}}(n)^\top \, \tilde{\boldsymbol{\theta}}(n)\right). \tag{2.3.14}$$

One calls a symmetric $q \times q$ matrix $\boldsymbol{B}$ *positive definite* if

$$\boldsymbol{a}^\top \boldsymbol{B}\, \boldsymbol{a} > 0 \tag{2.3.15}$$

for all $q$-vectors $\boldsymbol{a}$. A symmetric matrix $\boldsymbol{C}$ is said to be *smaller* than another symmetric matrix $\boldsymbol{A}$, or $\boldsymbol{C} < \boldsymbol{A}$, if the matrix $\boldsymbol{A} - \boldsymbol{C}$ is positive definite. This allows us to state that an estimator $\hat{\boldsymbol{\theta}}$, which provides the smallest error covariance matrix among all unbiased estimators, is the best estimator in the mean-square sense. Such an estimator is called an *efficient estimator*.

### Fisher Information

It can be shown that there exists a lower bound for the error covariance matrix given in (2.3.12), see Mendel (1995). This bound involves the *Fisher information matrix* $\boldsymbol{J} = [J^{i,j}]_{i,j=1}^{q}$, where

$$J^{i,j} = E\left(\frac{\partial}{\partial \theta_i} \ln(F_{\boldsymbol{X}}(\boldsymbol{R}))\, \frac{\partial}{\partial \theta_j} \ln(F_{\boldsymbol{X}}(\boldsymbol{R}))\right) \tag{2.3.16}$$

and $F_{\boldsymbol{X}}(\boldsymbol{R})$ is the joint distribution of the vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)^\top$ with the given parameter vector $\boldsymbol{\theta} \in \Theta$, when taken under the information given by the observations $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top$. The term $\frac{\partial}{\partial \theta_i} \ln(F_{\boldsymbol{X}}(\boldsymbol{R}))$ is the partial derivative with respect to the parameter $\theta_i$ of the natural logarithm of the joint distribution $F_{\boldsymbol{X}}(\cdot)$ of the observed quantities $\boldsymbol{R}$. We assume that the partial derivatives exist and are absolutely integrable. If $\hat{\boldsymbol{\theta}}$ is any unbiased estimator of $\boldsymbol{\theta}$, then the error covariance matrix is bounded from below by the inverse of the Fisher information matrix $\boldsymbol{J}$, that is

$$\mathrm{Cov}_{\tilde{\boldsymbol{\theta}}(n)} \geq \boldsymbol{J}^{-1}. \tag{2.3.17}$$

The lower bound is called the *Cramér-Rao lower bound* and provides a useful measure for testing the efficiency of specific estimation methods.

In the context of estimating trend, drift or growth parameters in log-returns of financial securities it tells us that there is an objective lower bound for the error covariance matrix. As already indicated previously, this bound is so high that there are not enough data available to accurately estimate trend and growth parameters in security prices.

### Method of Moments

A rather obvious and simple estimation method is the *method of moments*, see Hansen (1982) or Cochrane (2001). It often leads to computationally simple estimators and is intuitively satisfying. However, it has also some weaknesses, in particular, when moments that are involved in the derivation of the estimators are not certain to exist in reality. For instance, as we shall see later, it seems that empirical studies on log-return data indicate that the existence of the fourth moment could be questionable, see also Dacorogna, Müller, Pictet & De Vries (2001).

Assume that there are $n$ i.i.d. observations $R_{t_1}, R_{t_2}, \dots, R_{t_n} \in \Re$ that have the probability distribution function $F_{R,\boldsymbol{\theta}}$, which depends on the parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_q)^\top \in \Re^q$ for $q \in \mathcal{N}$. We know from Sect. 1.3 that the $j$th moment $m_j = m_j(\theta_1, \theta_2, \dots, \theta_q)$ of the random variable $R$ is obtained by the integral

$$m_j(\theta_1, \theta_2, \dots, \theta_q) = E((R)^j) = \int_{-\infty}^{\infty} (r)^j \, dF_{R,\boldsymbol{\theta}}(r) \tag{2.3.18}$$

for $j \in \mathcal{N}$ as long as $m_j = m_j(\theta_1, \theta_2, \dots, \theta_q) < \infty$. It is obvious that the moments $m_j$, $j \in \mathcal{N}$, are functions of the parameters $\theta_1, \theta_2, \dots, \theta_q$.

By application of the strong Law of Large Numbers, see Sect. 2.1, one can estimate under appropriate assumptions the respective moments using the given observations $R_{t_1}$, $R_{t_2}$, $\dots, R_{t_n}$. Let us denote by $\hat{m}_j$ the $j$th sample moment

$$\hat{m}_j = \frac{1}{n} \sum_{i=1}^{n} (R_{t_i})^j \tag{2.3.19}$$

for $j \in \mathcal{N}$. Typically, $q$ equations for the first $q$ moments are sufficient for identifying estimators for the $q$ unknown parameters $\theta_1, \theta_2, \dots, \theta_q$. The basic idea for the method of moments is therefore to equate the theoretical moments $m_j$ with corresponding sample moments $\hat{m}_j$, that is

$$m_j(\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q) = \hat{m}_j \tag{2.3.20}$$

for $j \in \{1, 2, \dots, q\}$, with $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ denoting the resulting parameter estimators.

If the system of equations (2.3.20) has a solution that is acceptable, then we call $\hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_q$ the *method of moments estimators*.

Sometimes, it is recommended to use the $j$th central theoretical moments

$$\mu_j = \mu_j(\theta_1, \theta_2, \dots, \theta_q) = E\left((R - m_1)^j\right) \tag{2.3.21}$$

and the respective $j$th *central sample moments*

$$\hat{\mu}_j = \frac{1}{n-1} \sum_{i=1}^{n} (R_{t_i} - m_1)^j \tag{2.3.22}$$

to form the $q$ equations

$$\mu_j(\theta_1, \theta_2, \ldots, \theta_q) = \hat{\mu}_j \tag{2.3.23}$$

for $j \in \{1, 2, \ldots, q\}$. By solving the system of equations (2.3.23) one may obtain slightly different method of moment estimators $\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q$.

The theoretical justification for the method of moments relies on the fact that under appropriate assumptions the sample moments are consistent estimators of the respective theoretical moments, see (2.3.8). It is well-known that the method of moments is sometimes not very efficient. Furthermore, the unbiasedness of the method of moment estimators cannot be easily guaranteed.

### Linear Least-Squares Estimation

The following well-known estimation method does, in principle, not require any information about the structure of the underlying distribution function of the observations. It uses only first and second order moments. In its basic form the *least-squares estimation method* assumes that the $n$-dimensional observation vector $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top \in \Re^n$ satisfies the following linear model

$$\boldsymbol{R} = \boldsymbol{B}\,\boldsymbol{\theta} + \boldsymbol{\varepsilon}. \tag{2.3.24}$$

Here $\boldsymbol{B} = [b^{i,j}]_{i,j=1}^{n,q}$ is the $(n,q)$-*observation matrix*, $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^\top \in \Re^n$ is the $n$-dimensional *observation error vector* and $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Theta$ denotes the $q$-dimensional parameter vector, $n, q \in \mathcal{N}$ with $q < n$. The observation matrix $\boldsymbol{B}$ is assumed to be known and to be of maximum rank $q$.

Note that for $\boldsymbol{\varepsilon} = (0, 0, \ldots, 0)^\top$ equation (2.3.24) has no solution. However, in reality the observation errors $\varepsilon$ are random and unknown. Therefore, the best that one can achieve is to find an estimator $\hat{\boldsymbol{\theta}}$ that minimizes in a reasonable sense the effect of the observation errors. From a mathematical viewpoint it is convenient to use a *least-squares criterion* of the form

$$U_{\mathrm{LS}}(\boldsymbol{\theta}) = \frac{1}{2}\,\boldsymbol{\varepsilon}^\top\,\boldsymbol{\varepsilon} = \frac{1}{2}\,(\boldsymbol{R} - \boldsymbol{B}\,\boldsymbol{\theta})^\top\,(\boldsymbol{R} - \boldsymbol{B}\,\boldsymbol{\theta}). \tag{2.3.25}$$

One notes that no expectation is taken in the least-squares criterion (2.3.25). This criterion simply minimizes the observation error $\boldsymbol{\varepsilon}$. It does not directly minimize the absolute value of the estimation error $\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}$.

Minimizing the quadratic form (2.3.25) with respect to the unknown parameter vector $\boldsymbol{\theta}$ yields by the corresponding first order conditions the, so-called, *normal equations* in the form

$$(\boldsymbol{B}^\top\,\boldsymbol{B})\,\hat{\boldsymbol{\theta}} = \boldsymbol{B}^\top\,\boldsymbol{R}. \tag{2.3.26}$$

This allows one to determine the least-squares estimator $\hat{\boldsymbol{\theta}}$, since we assumed that the matrix $\boldsymbol{B}$ has maximum rank $q < n$. Then we obtain

$$\hat{\boldsymbol{\theta}} = \boldsymbol{B}^{+} \boldsymbol{R} \tag{2.3.27}$$

with

$$\boldsymbol{B}^{+} = (\boldsymbol{B}^{\top} \boldsymbol{B})^{-1} \boldsymbol{B}^{\top}. \tag{2.3.28}$$

Statistically one can analyze the least-squares estimator by assuming that the measurement errors have zero mean, that is

$$E(\varepsilon_{t_j}) = 0 \tag{2.3.29}$$

for all $i \in \{1, 2, \ldots, n\}$. Obviously, the least-squares estimator is unbiased, since by (2.3.27) and (2.3.24) we obtain

$$E(\hat{\boldsymbol{\theta}}) = \boldsymbol{B}^{+} (\boldsymbol{B} \boldsymbol{\theta} + E(\boldsymbol{\varepsilon})) = \boldsymbol{\theta}. \tag{2.3.30}$$

Of great interest is the covariance matrix of the observation error vector, which has the form

$$\mathrm{Cov}(\boldsymbol{\varepsilon}) = E(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^{\top}). \tag{2.3.31}$$

If this matrix is known, then one can use the relation

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \boldsymbol{B}^{+} \boldsymbol{R} - \boldsymbol{B}^{+} \boldsymbol{B} \boldsymbol{\theta} = \boldsymbol{B}^{+} (\boldsymbol{R} - \boldsymbol{B} \boldsymbol{\theta}) = \boldsymbol{B}^{+} \boldsymbol{\varepsilon}$$

to calculate the covariance matrix

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) &= E\left((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\,(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^{\top}\right) = E\left(\boldsymbol{B}^{+} \boldsymbol{\varepsilon}\,(\boldsymbol{B}^{+} \boldsymbol{\varepsilon})^{\top}\right) \\ &= \boldsymbol{B}^{+} E\left(\boldsymbol{\varepsilon} \boldsymbol{\varepsilon}^{\top}\right) (\boldsymbol{B}^{+})^{\top} = \boldsymbol{B}^{+} \mathrm{Cov}(\boldsymbol{\varepsilon})\,(\boldsymbol{B}^{+})^{\top} \end{aligned} \tag{2.3.32}$$

of the estimation error. One notes that this covariance matrix depends on the second moments of the observation errors and the observation matrix.

### Curve Fitting

The linear least-squares estimation method is widely used, for instance, in linear regression analysis, that is, linear curve fitting. Linear and nonlinear curve fitting are common tasks in quantitative finance. Let us fit to given observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$ the model

$$R_{t_k} = \sum_{i=1}^{q} \theta_i \, \phi_i(t_k) + \varepsilon_k \tag{2.3.33}$$

for $t_k \in \{t_1, t_2, \ldots, t_n\}$. Here $\phi_i : [0, \infty) \to \Re$ is the given $i$th basis function with $i \in \{1, 2, \ldots, q\}$, which can be a nonlinear function of the variable $t \in [0, \infty)$. The unknown parameter vector is again $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^{\top}$. The observation error vector $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)^{\top}$ is as before.

If we now assume that the observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$ are available at the arguments $t_1, t_2, \ldots, t_n$, then by (2.3.33) the observation matrix $\boldsymbol{B} = [B_{\ell,i}]_{\ell,i=1}^{n,q}$ has the elements

$$B_{\ell,i} = \phi_i(t_\ell) \tag{2.3.34}$$

for $\ell \in \{1, 2, \ldots, n\}$ and $i \in \{1, 2, \ldots, q\}$.

By inserting the known values of the functions $\phi_i$, $i \in \{1, 2, \ldots, q\}$ into the observation matrix and using the observation vector one obtains directly the least-squares estimate (2.3.27). Often it is convenient to choose the basis functions such that they are orthogonal, that is

$$\sum_{\ell=1}^{n} \phi_i(t_\ell)\, \phi_k(t_\ell) = \begin{cases} 1 & \text{for} \quad i = k \\ 0 & \text{otherwise.} \end{cases} \tag{2.3.35}$$

In this case one obtains $\boldsymbol{B}^\top \boldsymbol{B} = \boldsymbol{I}$, where $\boldsymbol{I}$ is the identity matrix. This simplifies the least-squares estimator (2.3.27) yielding the simple expression

$$\hat{\boldsymbol{\theta}} = \boldsymbol{B}^\top \boldsymbol{R}. \tag{2.3.36}$$

In this case, one obtains for the least-squares estimator of the $i$th parameter the formula

$$\hat{\theta}_i = \sum_{\ell=1}^{n} \phi_i(t_\ell)\, R_{t_\ell} \tag{2.3.37}$$

for $i \in \{1, 2, \ldots, q\}$. The covariance matrix of the estimation error reduces here to the matrix

$$\operatorname{Cov}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \boldsymbol{B}^\top E(\boldsymbol{\varepsilon}\, \boldsymbol{\varepsilon}^\top)\, \boldsymbol{B} = \boldsymbol{B}^\top \operatorname{Cov}(\boldsymbol{\varepsilon})\, \boldsymbol{B}, \tag{2.3.38}$$

which does not depend on the true parameter vector $\boldsymbol{\theta}$. The linear least-squares method is widely used because of its simplicity. Its application can be largely successful if the chosen model is reasonably accurate for the data.

#### Generalized Least-Squares Estimators

One can refine the previously given linear least-squares problem by adding a symmetric positive weighting matrix $\boldsymbol{W}$ into the least-squares criterion (2.3.25). The *generalized least-squares* criterion is then of the form

$$U_{\text{GLS}}(\boldsymbol{\theta}) = \frac{1}{2}\, \boldsymbol{\varepsilon}^\top \boldsymbol{W}\, \boldsymbol{\varepsilon} = \frac{1}{2}\, (\boldsymbol{R} - \boldsymbol{B}\, \boldsymbol{\theta})^\top \boldsymbol{W}\, (\boldsymbol{R} - \boldsymbol{B}\, \boldsymbol{\theta}). \tag{2.3.39}$$

The optimal choice for the weighting matrix $\boldsymbol{W}$ is the inverse of the covariance matrix of the observation error, that is

$$\boldsymbol{W} = (\operatorname{Cov}(\boldsymbol{\varepsilon}))^{-1}. \tag{2.3.40}$$

This choice follows from the fact that the resulting generalized least-squares estimator

$$\hat{\boldsymbol{\theta}} = \left(\boldsymbol{B}^\top (\operatorname{Cov}(\boldsymbol{\varepsilon}))^{-1} \boldsymbol{B}\right)^{-1} \boldsymbol{B}^\top (\operatorname{Cov}(\boldsymbol{\varepsilon}))^{-1}\, \boldsymbol{R} \tag{2.3.41}$$

minimizes also the mean square error criterion

$$U_{\mathrm{MSE}}(\boldsymbol{\theta}) = E\left((\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})^\top (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right). \tag{2.3.42}$$

The estimator (2.3.41) is often called the *best linear unbiased estimator*.

In some applications the generalized linear least-squares method is not sufficient for capturing the dependence between the observations $R_{t_1}, R_{t_2}, \ldots, R_{t_n}$ and the parameter vector $\boldsymbol{\theta}$. In such case one can consider a nonlinear model of the form

$$\boldsymbol{R} = \boldsymbol{G}(\boldsymbol{\theta}) + \boldsymbol{\varepsilon} \tag{2.3.43}$$

with a given nonlinear vector valued function $\boldsymbol{G} : \Theta \to \Re^n$.

Similarly as before, one can in this case minimize the observation error to obtain the criterion

$$U_{\mathrm{NLS}}(\boldsymbol{\theta}) = \frac{1}{2}\,(\boldsymbol{R} - \boldsymbol{G}(\boldsymbol{\theta}))^\top\,(\boldsymbol{R} - \boldsymbol{G}(\boldsymbol{\theta})). \tag{2.3.44}$$

By minimizing the criterion (2.3.44) one typically obtains a *nonlinear least-squares estimator*. However, one must note that this involves a nonlinear optimization, which can only be performed numerically and may not always yield unique estimators.

## 2.4 Maximum Likelihood Estimation

### Maximum Likelihood Method

For proper financial modeling, it is essential to use an objective and reliable statistical methodology to distinguish between competing models. A key problem is the identification of a typical distribution for log-returns. For instance, one can try to use some moment based methods, as described previously. However, these may not say enough about the shape of the distribution. Alternatively, one could use the following *maximum likelihood methodology*, which appears to be reasonably objective. It is based on some hypothesized family of probability densities and does not require the use of higher order empirical moments.

In the following the maximum likelihood methodology will be explained in the context of observed sequences of i.i.d. log-returns. This framework can be used to identify a best fit of log-return distributions, for instance, for stock market index data as will be discussed later in detail. However, one must be aware of the fact that there is never a "true" distribution behind the random variables that one observes in practice. More realistic estimation techniques are, for instance, provided by the quasi-likelihood theory, as presented in Heyde (1997).

The maximum likelihood estimation method assumes that there is no prior information available on the parameters $\theta_1, \theta_2, \ldots, \theta_q$. What is needed for the maximum likelihood method is the probability density function $f_R$ of the independent identically distributed observations $R_{t_1}$, $R_{t_2}$, …, $R_{t_n}$. The *maximum likelihood estimators* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_n)^\top$ have several theoretically highly desirable asymptotic optimality properties when the sample size $n$ is large. For instance, in the case when there exists an estimator which satisfies the Cramer-Rao lower bound (2.3.17), then it can be constructed by using the maximum likelihood method. The maximum likelihood estimator $\hat{\theta}$ is consistent and asymptotically efficient, which means that it achieves asymptotically the Cramer-Rao lower bound for the estimation error.

### Likelihood Function

We assume that the observed log-returns are denoted by $R_{t_1}$, $R_{t_2}$, … and form a sequence of i.i.d. random variables with some hypothesized, parameterized density $f_{\boldsymbol{R}}$. Note that the joint probability density function $f_{\boldsymbol{R}}$ of the random vector $\boldsymbol{R} = (R_{t_1}, R_{t_2}, \ldots, R_{t_n})^\top$ of log-returns can be written for these i.i.d. random variables, see (1.4.41), in the form

$$\mathcal{L}(\boldsymbol{\theta}) = f_{R_{t_1}, R_{t_2}, \ldots, R_{t_n}}(R_{t_1}, R_{t_2}, \ldots, R_{t_n}, \boldsymbol{\theta}) = \prod_{i=1}^{n} f_{\boldsymbol{R}}(R_{t_i}, \boldsymbol{\theta}). \tag{2.4.1}$$

Here $f_{\boldsymbol{R}}(\cdot, \boldsymbol{\theta})$ is the density of $R_{t_i}$, $i \in \{1, 2, \ldots, n\}$, given the parameter values $\boldsymbol{\theta} = (\theta_1, \theta_2, \ldots, \theta_q)^\top \in \Theta \subseteq \Re^q$, $q \in \mathcal{N}$. The set $\Theta$ specifies again the set of allowable values that the parameters can take. Our aim will be to find some best parameter estimates that fit the data. We call the above function (2.4.1) the *likelihood function* for the parameter $\boldsymbol{\theta} \in \Theta$.

### Maximum Likelihood Estimate

To be able to optimize the choice of the parameter we need a criterion to identify a best fit. The maximum likelihood methodology uses the *maximum likelihood estimator* $\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\theta}_2, \ldots, \hat{\theta}_q)^\top$ as a best estimate of $\boldsymbol{\theta} \in \Theta$, where

$$\mathcal{L}^* = \mathcal{L}(\hat{\boldsymbol{\theta}}) = \sup_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta}). \tag{2.4.2}$$

Here $\mathcal{L}^* = \sup_{\boldsymbol{\theta} \in \Theta} \mathcal{L}(\boldsymbol{\theta})$ denotes the *supremum* of $\mathcal{L}(\boldsymbol{\theta})$, that is in our case the least upper bound of $\mathcal{L}(\boldsymbol{\theta})$ over all $\boldsymbol{\theta} \in \Theta$. This means, $\hat{\boldsymbol{\theta}}$ is the parameter that maximizes the likelihood function with respect to the set of permitted parameter values $\boldsymbol{\theta} \in \Theta$. Intuitively, this choice yields the parameter $\hat{\boldsymbol{\theta}}$ for which the observed log-returns are most likely chosen from the hypothesized density in the given parameterized family of probability densities. This means that $f_{\boldsymbol{R}}(\cdot, \hat{\boldsymbol{\theta}})$ represents the most probable density from the given class of densities having observed the log-returns $R_{t_1}, R_{t_2} \ldots, R_{t_n}$.

In practice, it is convenient to work with the *log-likelihood function*

$$\ell(\boldsymbol{\theta}) = \ln(\mathcal{L}(\boldsymbol{\theta})). \tag{2.4.3}$$

Under suitable conditions, when the true parameter is an interior point of $\Theta$, the maximum likelihood estimate $\hat{\boldsymbol{\theta}}$ can be obtained as a root of the first order conditions

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \theta_i} = 0 \tag{2.4.4}$$

for all $i \in \{1, 2, \ldots, q\}$, where $\frac{\partial}{\partial \theta_i}$ denotes the partial derivative with respect to $\theta_i$. The system of equations (2.4.4) is called the system of *maximum likelihood equations*.

If one cannot explicitly solve the maximum likelihood equations, then a root finding method, for instance, a multi-dimensional Newton method, can be applied to solve the system of maximum likelihood equations. The above maximum likelihood approach only yields reliable estimates if the hypothesized model is suitable and a sufficient number of observations is available. For the identification of log-return distributions this means that the hypothesized distribution must be reasonably close to the true distribution and one needs a large number of observed log-returns that can be interpreted as being independent and identically distributed.

### Likelihood Ratio Test

Now, let us suppose that we have a class of models that corresponds to a class of parameters characterizing some hypothesized density. Our goal will be to identify the parameters of the density which best fits our data set of observed log-returns using the maximum likelihood approach.

This can be achieved by the *likelihood ratio test* which is due to Neyman & Pearson (1928). We emphasize, that the maximum likelihood approach does not rely on certain higher moments, for instance the kurtosis, that might not even exist in a given situation, as we shall see later.

We define the *likelihood ratio* in the form

$$\Lambda = \frac{\mathcal{L}^*_{\text{model}}}{\mathcal{L}^*_{\text{general model}}}. \tag{2.4.5}$$

Here $\mathcal{L}^*_{\text{model}}$ represents the maximized likelihood function of a hypothesized model density, say, with $q$ parameters. On the other hand, $\mathcal{L}^*_{\text{general model}}$ denotes the maximized likelihood function for the density of a more general model that has, say $q + \nu$ parameters and nests the hypothesized model density $q, \nu \in \mathcal{N}$.

Under appropriate conditions it can be shown, see Rao (1973), that the density of the *test statistic*

$$L_n = -2 \ln(\Lambda) \tag{2.4.6}$$

is asymptotically a chi-square density, or more generally a gamma density, see (1.2.9), for increasing number of observations $n \to \infty$. Here the degrees

**Table 2.4.1.** Quantiles for the chi-square-distribution

| $\nu$ | 1 | 2 | 30 |
|---|---|---|---|
| $\chi^2_{0.99,\nu}$ | 6.635 | 9.210 | 50.9 |
| $\chi^2_{0.95,\nu}$ | 3.841 | 5.991 | 43.8 |
| $\chi^2_{0.90,\nu}$ | 2.706 | 4.605 | 40.3 |
| $\chi^2_{0.20,\nu}$ | 0.064 | 0.446 | 23.4 |
| $\chi^2_{0.10,\nu}$ | 0.0158 | 0.211 | 20.6 |
| $\chi^2_{0.05,\nu}$ | 0.0039 | 0.103 | 18.5 |
| $\chi^2_{0.01,\nu}$ | 0.000157 | 0.020 | 15.0 |
| $\chi^2_{0.001,\nu}$ | 0.000002 | 0.002 | 11.6 |

of freedom $\nu$ equal the difference between the number of parameters in the general model density and the hypothesized model density. It can then be shown that as $n \to \infty$

$$P\left(L_n < \chi^2_{1-\alpha,\nu}\right) \approx F_{\chi^2(\nu)}\left(\chi^2_{1-\alpha,\nu}\right) = 1 - \alpha, \tag{2.4.7}$$

where $F_{\chi^2(\nu)}$ denotes the chi-square distribution with $\nu$ degrees of freedom and $\chi^2_{1-\alpha,\nu}$ is its $100(1-\alpha)\%$ quantile. In Table 2.4.1 we summarize some quantiles of the chi-square distribution.

As in Sect. 2.2 we can similarly check for a given $100\alpha\%$ confidence level whether or not the test statistic $L_n$ is in the $100(1-\alpha)\%$ quantile of the chi-square distribution with $\nu$ degrees of freedom. If the relation

$$L_n < \chi^2_{1-\alpha,\nu} \tag{2.4.8}$$

is satisfied, then we cannot reject at the $100\alpha\%$ significance level the hypothesis that the suggested model is the true underlying model. Otherwise, we reject this hypothesis on the basis of this likelihood ratio test.

## 2.5 Normal Variance Mixture Models

### Subordination

It is well-known that the log-return distributions of security prices are strongly leptokurtic. This means that they have larger kurtosis than the Gaussian distribution provides. The following simple modeling approach is called *subordination* and goes back to Bochner (1955) and Clark (1973). We used already some kind of subordination when generating Student $t$ log-returns in (2.2.20). For capturing typical features of log-return distributions one can simply make the conditional variances themselves independent random variables. This yields a class of models with normal-variance mixture distributed log-returns, see Feller (1968). This kind of models allows us to keep the empirical analysis fairly simple. For simplicity we shall only consider here symmetric log-returns since any log-return mean is in reality extremely small and can be easily added to the model.

We now assume a normal-variance mixture density for the $i$th log-return $Z_i$ by setting

$$Z_i = \sqrt{m_i}\,\xi_i. \tag{2.5.1}$$

Here we use for all $i \in \{0, 1, \ldots, n-1\}$ an independent identically distributed nonnegative *conditional variance* $m_i$, together with some independent, standard Gaussian distributed random variable $\xi_i \sim N(0, 1)$. Note that each log-return $Z_i$ can here be linked to a corresponding conditional variance $m_i$. The conditional variance $m_i$ for the log-return at time $t_i$, $i \in \{0, 1, \ldots\}$, is assumed to be distributed according to a given density $f_m$. The generality of the resulting class of normal-variance mixture densities for log-returns follows from the freedom to adjust the density $f_m$. One obtains here the normal-variance mixture density function of the log-return $Z_i$, in the form

$$f_Z(x) = \frac{1}{\sqrt{2\,\pi}} \int_0^\infty \frac{1}{\sqrt{u}} \exp\left\{-\frac{x^2}{2\,u}\right\} f_m(u)\,du \tag{2.5.2}$$

for $x \in \Re$, as long as this integral exists. The $i$th log-return $Z_i$ has then mean zero, variance $v_Z = v_m$, skewness zero and kurtosis

$$\kappa_Z = 3\left(1 + \frac{v_m}{(\mu_m)^2}\right). \tag{2.5.3}$$

Here the $i$th conditional variance $m_i$, $i \in \{0, 1, \ldots, n-1\}$ has mean $\mu_m$ and variance $v_m$. One notes for the case of nonzero variance of the conditional variance $m_i$ that it follows by (2.5.3) that any normal-variance mixture density has kurtosis greater than three and is, therefore, *leptokurtic*. This can be used as an explanation for widely observed leptokurtic log-returns in practice. The *lognormal model* with constant $m_i$ is an extremely useful modeling attempt that has as its justification mainly its mathematical simplicity.

Samuelson (1957), Osborne (1959) and subsequently many other authors have modeled asset price increments by lognormal random variables, where the resulting log-returns are Gaussian random variables. This has been an extremely important first step in quantitative finance. The corresponding Gaussian density for the lognormal model results from (2.5.2) when the density of the conditional variance degenerates to that of a constant $m_i = 1$ with $v_m > 0$ for $i \in \{0, 1, \ldots, n-1\}$. We emphasize that the Gaussian density has been clearly rejected in many studies as a suitable log-return density for most securities.

Let us remark that in a wider range of models, beyond the models that we consider here, one does not need the conditional variance for the log-returns to exist. This allows one to cover *logstable models* as suggested in Mandelbrot (1963, 1967), Mandelbrot & Taylor (1967), Fama (1965) and Hurst, Platen & Rachev (1999). These models have typically one additional parameter when compared to the lognormal model and generate log-returns that may have no conditional variance. We are not studying here any of these models since most log-returns in financial markets seem to have finite conditional variance.

### SGH Models

We shall not consider any further the case of a Gaussian log-return density but study instead a rich class of analytically tractable densities that include the Gaussian one as a limit. Our aim will be to discriminate between a wide range of possible leptokurtic densities. The densities that we shall analyze can be classified as a class of normal-variance mixture densities. The Gaussian log-return density arises simply as limiting case for certain extreme parameters. As described in Sect. 1.2, it is noticeable that a large group of authors have proposed important models with log-returns that relate to the class of generalized hyperbolic densities. This class of densities was extensively examined by Barndorff-Nielsen (1977, 1978) and Barndorff-Nielsen & Blaesild (1981). We assume, for simplicity, zero skewness and consider in the following the *symmetric generalized hyperbolic* (SGH) density as a possible density for log-returns. This density results when the density of the conditional variance $m_i$, $i \in \{0, 1, \ldots\}$, is a *generalized inverse Gaussian density*. We call the resulting discrete time log-return models *SGH models*.

By (1.2.24) the SGH density function of a log-return $Z_i$ is of the form

$$f_Z(x) = \frac{1}{\delta\, K_\lambda(\alpha\, \delta)} \sqrt{\frac{\alpha\, \delta}{2\, \pi}} \left(1 + \frac{x^2}{\delta^2}\right)^{\frac{1}{2}\left(\lambda - \frac{1}{2}\right)} K_{\lambda - \frac{1}{2}}\!\left(\alpha\, \delta \sqrt{1 + \frac{x^2}{\delta^2}}\right) \tag{2.5.4}$$

for $x \in \Re$, where $\lambda \in \Re$ and $\alpha, \delta \geq 0$. We set $\alpha \neq 0$ if $\lambda \geq 0$ and $\delta \neq 0$ if $\lambda \leq 0$. We remark that the corresponding probability density function of a conditional variance $m_i$ in the normal-variance mixture density (2.5.2) is here the *generalized inverse Gaussian density* of the form

$$f_m(x) = \frac{\alpha^\lambda}{2\, \delta^\lambda K_\lambda(\alpha\, \delta)}\, x^{\lambda - 1} \exp\left\{-\frac{1}{2}\left(\frac{\delta^2}{x} + \alpha^2 x\right)\right\}, \tag{2.5.5}$$

where $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index $\lambda$, see (1.2.25).

The SGH density is a four parameter density. The two *shape parameters* are $\lambda$ and $\bar{\alpha} = \alpha\, \delta$, defined so that they are invariant under scale transformations. The other parameters contribute to the scaling of the density. We define as in (1.2.27) the parameter $c$ as the *scale parameter* such that $v_m = v_Z = c^2$, that is

$$c^2 = \begin{cases} \frac{2\,\lambda}{\alpha^2} & \text{if } \delta = 0 \text{ for } \lambda > 0,\ \bar{\alpha} = 0, \\ \frac{\delta^2\, K_{\lambda+1}(\bar{\alpha})}{\bar{\alpha}\, K_\lambda(\bar{\alpha})} & \text{otherwise.} \end{cases} \tag{2.5.6}$$

The variance of $m_i$ is

$$v_m = c^4 \left( \frac{K_\lambda(\bar{\alpha})\, K_{\lambda+2}(\bar{\alpha})}{K_{\lambda+1}(\bar{\alpha})^2} - 1 \right). \tag{2.5.7}$$

Consequently, the log-return $Z_i$ has mean zero, variance $v_Z = c^2$, skewness zero and kurtosis

$$\kappa_Z = \frac{3\, K_\lambda(\bar{\alpha})\, K_{\lambda+2}(\bar{\alpha})^2}{K_{\lambda+1}(\bar{\alpha})}. \tag{2.5.8}$$

Furthermore, it can be shown that as $\lambda \to \pm\infty$ or $\bar{\alpha} \to \infty$ the SGH density asymptotically approaches the Gaussian density.

To illustrate certain candidate densities for the log-returns within the class of SGH densities we recall in the following four special cases of the SGH density that coincide with the log-return densities of important asset price models, which we previously mentioned in Sect. 1.2.

Praetz (1972) and Blattberg & Gonedes (1974) suggested for log-returns a *Student t density* with degrees of freedom $n > 0$. This density follows from the above SGH density for the shape parameters $\lambda = -\frac{1}{2}\, n < 0$ and $\bar{\alpha} = 0$, that is $\alpha = 0$ and $\delta = \varepsilon\, \sqrt{n}$. Using these parameter values the Student $t$ density function has then the form (1.2.28). Here the variance mixture density is an inverse gamma density.

Barndorff-Nielsen (1995) considered log-returns to follow a *normal-inverse Gaussian mixture distribution*. The corresponding density is obtained from the SGH density when the shape parameter $\lambda = -\frac{1}{2}$ is chosen. For this parameter value the conditional variance $m_i$ is inverse Gaussian distributed and it follows by (2.5.4) that the probability density function (1.2.29) for $Z_i$.

Eberlein & Keller (1995) and Küchler et al. (1999) suggested models, where log-returns appear to be *hyperbolicly distributed*. This type of models is covered by the choice of the shape parameter $\lambda = 1$ in the SGH density. The probability density function of the $Z_i$ is then of the form (1.2.30).

Madan & Seneta (1990), Geman, Madan & Yor (2001) and Carr, Geman, Madan & Yor (2003) assumed log-returns to be distributed with a *normal-variance gamma mixture distribution*. This particular case occurs when the shape parameters are such that $\lambda > 0$ and $\bar{\alpha} = 0$, that is, $\delta = 0$ and $\alpha = \frac{\sqrt{2\,\lambda}}{c}$. In this case the conditional variance $m_i$ is gamma distributed and the probability density function of $Z_i$ is given by (1.2.31)

This means the variance mixture density is here a gamma density. The model is also known as *variance gamma* (VG) model. Below we shall investigate which kinds of densities best fit observed log-returns.

## 2.6 Distribution of Index Log-Returns

### Estimation of Log-Returns

In the literature a vast amount of empirical work has been accomplished estimating the distributions of log-returns of financial securities. Starting with papers by Mandelbrot (1963) and Fama (1963), and in a subsequent stream of literature, it became clear that the standard assumption that log-returns are Gaussian distributed is very crude and for certain risk management tasks even a dangerous assumption. It is now widely recognized that, in reality, extreme log-returns are more likely than suggested by the Gaussian distribution.

From the perspective of the benchmark approach, which we present later in this book, it is not surprising that empirical studies on log-returns of exchange rates and equity prices have so far not identified a particular distribution that fits the observed data well. The reason is that an exchange rate involves at least two major factors. One of these relates to the domestic economy and the second reflects the foreign economy. As we shall explain later, the benchmark approach suggests we analyze and model the denominations of the market portfolio in different currencies. Intuitively, one separates other impacts from that of the currency of interest. The market portfolio acts as the reference unit that is practically least disturbed by any particular security. The benchmark approach will suggest that one may more easily find some testable empirical evidence about the log-return distribution of an index than for an exchange rate or equity price.

In two papers, the Nobel Laureate Harry Markowitz together with Usmen; see Markowitz & Usmen (1996a, 1996b), analyzed daily S&P500 stock index log-returns for the period from 1963 until 1983 in a Baysian framework. Within the wide family of Pearson distributions, see Stuart & Ord (1994), they identified the Student $t$ distribution with about $n = 4.5$ degrees of freedom as the best fit.

In an independent empirical analysis of stock index log-returns observed from 1982 until 1996 of the S&P500 and other stock indices, Hurst & Platen (1997) demonstrated that the Student $t$ distribution provides the best fit in a range of normal-variance mixture distributions when employing a maximum likelihood methodology. In a recent study on log-returns of a world stock index, when denominated in different currencies, Fergusson & Platen (2006) showed via likelihood ratio tests by using daily data for the period from 1970 until 2004 that the Student $t$ distribution provides clearly the best fit in the class of SGH distributions. In the following we provide more details on these findings.

### A World Stock Index

We apply the previously described SGH model, where we use as security a *world stock index* (WSI), which we construct as a self-financing portfolio of stock market accumulation indices. The weights for the market capitalization that were chosen at the end of the observation period in 2004 are given in the last column of Table 2.6.1. Weights for earlier years reflect the market capitalization as obtainable from Thomson Financial.

We cover with this index more than 95% of the world stock market capitalization for the period from 1970 until 2004. We exclude in our study weekends

**Table 2.6.1.** Empirical moments for log-returns of WSI

| Country | Currency | $\hat{m}_y$ | $\hat{\sigma}_y$ | $\hat{\beta}_y$ | $\hat{\kappa}_y$ | Weights |
|---|---|---|---|---|---|---|
| Argentina | ARS | 0.000495 | 0.009705 | 4.971980 | 170.049000 | 0.0023 |
| Australia | AUD | 0.000403 | 0.008815 | -0.372951 | 14.051762 | 0.0017 |
| Austria | ATS, EUR | 0.000367 | 0.008716 | -0.489599 | 13.406629 | 0.0161 |
| Belgium | BEF, EUR | 0.000332 | 0.009387 | -0.443605 | 13.231579 | 0.0061 |
| Brazil | BRL | 0.001392 | 0.011284 | 2.330662 | 79.025855 | 0.0013 |
| Canada | CAD | 0.000404 | 0.007831 | -0.579593 | 18.075123 | 0.0251 |
| Denmark | DKK | 0.000356 | 0.009526 | -0.394520 | 14.431344 | 0.0037 |
| Finland | FIM, EUR | 0.000396 | 0.010612 | 0.238200 | 70.189255 | 0.0032 |
| France | FRF, EUR | 0.000376 | 0.009332 | -0.416397 | 13.655084 | 0.0302 |
| Germany | DEM, EUR | 0.000285 | 0.009417 | -0.483043 | 13.632198 | 0.0342 |
| Greece | GRD, EUR | 0.000569 | 0.009454 | 0.841563 | 35.815507 | 0.0012 |
| Hong Kong | HKD | 0.000421 | 0.008000 | -0.446156 | 17.986618 | 0.0231 |
| Hungary | HUF | 0.000476 | 0.008578 | -0.184216 | 17.086200 | 0.0023 |
| India | INR | 0.000542 | 0.013561 | -0.131696 | 78.892868 | 0.0047 |
| Indonesia | IDR | 0.000571 | 0.008906 | -0.118789 | 19.917163 | 0.0201 |
| Ireland | IRP, EUR | 0.000373 | 0.009033 | -0.507152 | 13.844893 | 0.0055 |
| Italy | ITL, EUR | 0.000373 | 0.008957 | -0.535339 | 16.379118 | 0.0132 |
| Japan | JPY | 0.000238 | 0.009033 | -0.638245 | 14.427542 | 0.1550 |
| Korea | KRW | 0.000439 | 0.009636 | -0.139465 | 57.088739 | 0.0072 |
| Malaysia | MYR | 0.000398 | 0.008500 | -0.616024 | 18.996454 | 0.0158 |
| Mexico | MXN | 0.001152 | 0.020686 | 3.860894 | 278.762521 | 0.0016 |
| Netherlands | NLG, EUR | 0.000300 | 0.009358 | -0.472830 | 13.914414 | 0.0193 |
| Norway | NOK | 0.000373 | 0.009196 | -0.340799 | 14.541824 | 0.0029 |
| Philippines | PHP | 0.000460 | 0.009009 | -0.354508 | 19.339190 | 0.0041 |
| Portugal | PTE, EUR | 0.000391 | 0.008810 | -0.452567 | 13.152174 | 0.0013 |
| Singapore | SGD | 0.000353 | 0.007991 | -0.554040 | 17.215279 | 0.0079 |
| Spain | ESP, EUR | 0.000464 | 0.010334 | 1.079021 | 48.935156 | 0.0124 |
| Sweden | SEK | 0.000421 | 0.009195 | 0.214018 | 19.355551 | 0.0124 |
| Switzerland | CHF | 0.000239 | 0.010097 | -0.407918 | 11.406303 | 0.0206 |
| Taiwan | TWD | 0.000357 | 0.007900 | -0.558432 | 17.784127 | 0.0141 |
| Thailand | THB | 0.000450 | 0.009285 | 0.628255 | 36.231337 | 0.0049 |
| Turkey | TRL | 0.001029 | 0.011354 | 5.010563 | 155.199258 | 0.0018 |
| UK | GBP | 0.000414 | 0.009141 | -0.482534 | 13.541676 | 0.0846 |
| US | USD | 0.000374 | 0.007724 | -0.613548 | 18.808550 | 0.4301 |

and other nontrading days at the US and European exchanges. As shown in Platen (2004c, 2005b) and as will be discussed later in Sect. 10.6 under the benchmark approach, such a diversified portfolio is robust against variations in weightings as long as the weight of each contributing security remains reasonably small. The constructed portfolio is a proxy for the world stock portfolio or market portfolio.

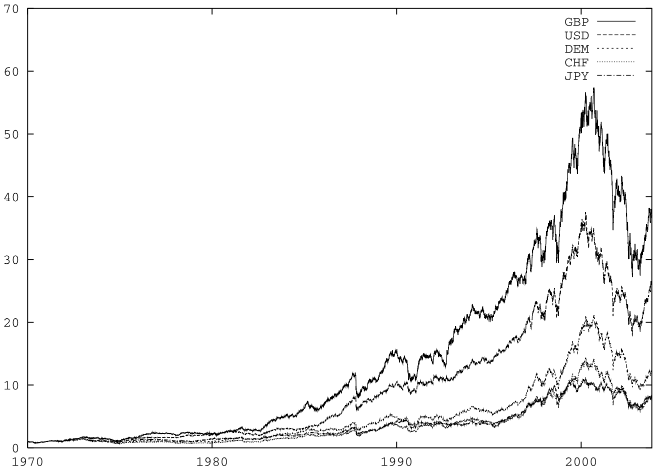



**Fig. 2.6.1.** WSI in units of different currencies

In Fig. 2.6.1 we plot the resulting WSI for the observation period when denominated in units of British Pound, US dollar, Deutsche Mark, Swiss Franc and Japanese Yen. For convenience we normalized the initial values to one.

In the following we shall study the distribution of log-returns of the WSI when denominated in 34 major currencies. This will provide some distributional characterization of the general market risk for the respective markets. The findings will be quite important for supporting the appropriateness of theoretically suggested financial market models.

We deliberately do not adjust for any changes over time, market crashes or other influences that may have affected the data. Some methods of data analysis discard extreme values of observations as outliers. But this would be inappropriate in a financial context because it is most important for risk management to capture the probability for extreme log-returns. For *daily log-returns* of the WSI for the period from 1970 until 2004 in 34 currency denominations, the first four empirical moments yield the average empirical mean $\hat{m}_y = 0.000486$, average standard deviation $\hat{\sigma}_y = 0.009789$, average skewness $\hat{\beta}_y = 0.460316$ and average kurtosis $\hat{\kappa}_y = 44.485182$.

For each currency let us centralize the log-returns of WSI denominations with respect to the mean. Furthermore, we scale these to obtain unit variance. The resulting transformed log-returns have then an estimated zero mean and unit variance. Now, we combine all observed centralized and scaled daily log-returns in one large sample. We show in Fig. 2.6.2 the logarithm of the resulting log-return histogram, that is, the relative frequencies. This figure shows also the logarithm of the Student $t$ density with degrees of freedom 3.64, which appears to fit the data extremely well already by visual inspection. When comparing the shapes of the other densities covered in Fig. 1.2.8–1.2.10, then none of these match the shape shown in Fig. 2.6.2.

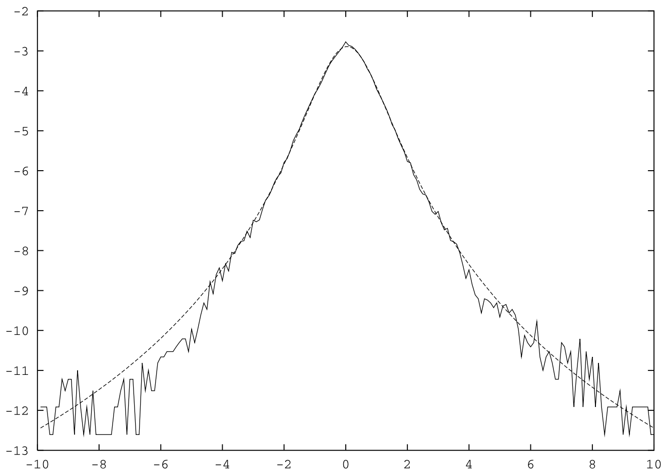



**Fig. 2.6.2.** Logarithm of WSI log-return histogram

By considering the above fit and the relatively small empirical skewness the empirical log-return density appears to be fairly symmetrical. This is also consistent with the findings in Markowitz & Usmen (1996a, 1996b) for the case of S&P500 log-returns and those in Hurst & Platen (1997) for stock index log-returns. Therefore, to simplify our analysis and to focus on the identification of the tail properties of log-return densities, we assume that the densities that we shall consider are symmetric, that is we assume zero mean and zero skewness. This assumption does not very much influence the empirical results that we obtain and allows us to use the SGH models described in Sect. 2.5. We emphasize that the following study focuses on the shape of the log-return densities. It avoids relying on any particular moment properties. Note that certain higher order moments may not exist. In particular, the kurtosis of Student $t$ distributed log-returns with less than four degrees of freedom is *infinite*, see (1.3.37). This could become a problem in reality when taking into account that in Fig. 2.6.2 we fitted a Student $t$ density with about 3.64 degrees of freedom.

### Maximum Likelihood Estimation for SGH Densities

Let us now identify the typical distribution for log-returns of the WSI in the class of SGH distributions.

The class of SGH densities that we introduced in Sect. 1.2 represents a rich class of leptokurtic densities. To reject, on a given significance level, the assumption that a hypothetical SGH density is not the true underlying density we apply the maximum likelihood ratio test, as described in Sect. 2.4. We define the *likelihood ratio* in the form

$$\Lambda = \frac{\mathcal{L}^*_{\text{model}}}{\mathcal{L}^*_{\text{SGH}}}. \tag{2.6.1}$$

Here $\mathcal{L}^*_{\text{model}}$ represents the maximized likelihood function of a given specific, nested log-return density, for instance, the Student $t$ density. With respect to this density the maximum likelihood estimate for the parameters is calculated and then used to obtain the corresponding *likelihood function* $\mathcal{L}^*_{\text{model}}$. On the other hand, $\mathcal{L}^*_{\text{SGH}}$ denotes the maximized likelihood function for the SGH density, which is the nesting density that has been similarly obtained. We then calculate according to (2.4.6) the test statistic $L_n = -2\ln(\Lambda)$, which is for increasing number $n \to \infty$ of observations asymptotically chi-square distributed. Here the degrees of freedom equal the difference between the number of parameters of the nesting density and the nested density. The nesting density, which is the SGH density, is a four-parameter density and our nested densities are the Student $t$, normal-inverse Gaussian, hyperbolic and variance-gamma density, see Sect. 1.2. Each of these is a three-parameter density. Therefore, in the cases considered, the test statistic $L_n$ is, for $n \to \infty$, asymptotically chi-square distributed with one degree of freedom.

According to (2.4.7) we have asymptotically as $n \to \infty$ that

$$P\left(L_n < \chi^2_{1-\alpha,1}\right) \approx 1 - \alpha, \tag{2.6.2}$$

where $\chi^2_{1-\alpha,1}$ is the $100(1-\alpha)\%$ quantile of the chi-square distribution $F_{\chi^2(1)}$ with one degree of freedom. One can then check, say, for a 99% *significance level* whether or not the test statistic $L_n$ is in the 1% quantile of the chi-square distribution with one degree of freedom. By Table 2.4.1 it follows that if the relation

$$L_n < \chi^2_{0.01,1} \approx 0.000157 \tag{2.6.3}$$

is satisfied, then we cannot reject on a 99% significance level the hypothesis that the suggested density is the true underlying density. Similarly, if

$$L_n < \chi^2_{0.001,1} \approx 0.000002 \tag{2.6.4}$$

holds, then we cannot reject the hypothesis that the given density is the true underlying density on a 99.9% significance level.

The above maximum likelihood methodology offers a natural definition of a *best fit*. We call the density with the smallest test statistic $L_n$ the best fit in the given class of SGH densities. This density maximizes the likelihood ratio $\Lambda$ given in (2.6.1).

### Log-Returns of World Stock Indices

Now let us study the log-returns of the WSI when denominated in units of each of the 34 currencies. By the above formulas together with the different SGH densities described in Sects. 1.2 and 2.5 one can construct the corresponding maximum likelihood estimators and test statistics. In Table 2.6.2 we display the test statistics $L_n$ for log-returns of the WSI in the 34 different currency denominations, as given in Fergusson & Platen (2006). It is apparent that the

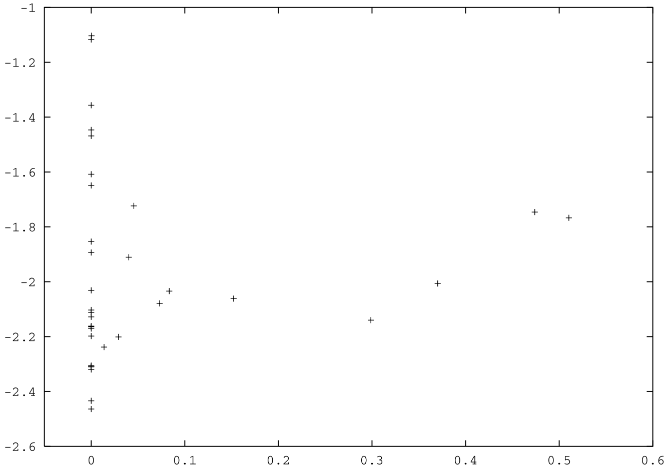



**Fig. 2.6.3.** $(\bar{\alpha}, \lambda)$-plot for log-returns of WSI in different currencies

Student $t$ density shows in all cases the smallest test statistic. For 25 of the 34 currencies one cannot reject on a 99.9% significance level the hypothesis that the Student $t$ density is the true density. The inverse Gaussian density seems to be the second best choice but can be for all stocks rejected on any reasonable significance level. In the last column of Table 2.6.2 one finds the estimated degrees of freedom for the Student $t$ density for each of the currency denominations. One notes that the degrees of freedom are in the range of 2.2 to 4.9. We emphasize that the resulting arithmetic average of 3.9899 of the estimated degrees of freedom for log-returns of WSI currency denominations is very close to the number four. We shall find towards the end of the book a natural explanation for this empirical result.

Let us visualize the estimated shape parameters $\bar{\alpha}$ and $\lambda$ of the SGH density in Fig. 2.6.3 in an $(\bar{\alpha}, \lambda)$-scatter plot for the log-returns of the 34 WSI currency denominations.

Interestingly, the estimated parameter points are scattered near the negative $\lambda$-axis with values between $-2.5$ and $-1.0$. It is the Student $t$ density that is characterized by points directly located on the $\lambda$-axis with $n = -2\lambda$ degrees of freedom. This means that the large number of points near $\lambda = -2$ on the $\lambda$-axis indicate a *Student t density of about four degrees of freedom.* For comparison, the variance-gamma density, which would favor the Madan-Seneta variance gamma model, see Madan & Seneta (1990) and Geman et al. (2001), would need to have points on the positive $\lambda$-axis in Fig. 2.6.3. The hyperbolic density, suggested by Eberlein & Keller (1995) as log-return density corresponds to the shape parameter $\lambda = 1$ and any $\bar{\alpha} > 0$. Our findings in Fig. 2.6.3 do not support this type of model. The normal-inverse Gaussian distributed log-returns of a model proposed in Barndorff-Nielsen (1995) would generate points in the scatter plot of Fig. 2.6.3 for $\lambda = -\frac{1}{2}$ and $\alpha > 0$. However, there are also no points that come close to the line $\lambda = -\frac{1}{2}$. One

**Table 2.6.2.** The $L_n$ test statistic for log-returns of the WSI in different currencies

| Country | Student t | Inverse Gaussian | Hyperbolic | Variance Gamma | Degrees of Freedom |
|---|---|---|---|---|---|
| Argentina | 0.000000 | 137.566726 | 377.265316 | 414.030946 | 3.215953 |
| Australia | 0.000000 | 24.403096 | 54.527486 | 73.900060 | 4.618187 |
| Austria | 1.272478 | 10.533450 | 43.923596 | 63.696828 | 4.177998 |
| Belgian | 0.000000 | 15.755730 | 41.664530 | 59.922270 | 4.639238 |
| Brazil | 0.000000 | 132.348986 | 430.843576 | 444.117392 | 2.937178 |
| Canada | 0.000000 | 42.829996 | 80.784298 | 110.357598 | 4.927523 |
| Denmark | 0.000000 | 29.607334 | 72.807340 | 96.852868 | 4.328256 |
| Finland | 0.000000 | 130.807532 | 286.692740 | 326.546792 | 3.707350 |
| France | 0.303708 | 12.578282 | 42.737860 | 61.753332 | 4.329017 |
| Germany | 0.000000 | 17.945998 | 45.546996 | 64.739532 | 4.620013 |
| Greece | 0.000000 | 60.072066 | 120.786456 | 150.578024 | 4.340990 |
| Hong Kong | 0.000000 | 31.399542 | 84.191740 | 111.611116 | 4.080899 |
| Hungary | 0.002812 | 33.488642 | 102.407366 | 132.556744 | 3.827598 |
| India | 0.000000 | 218.752194 | 1096.862148 | 962.798700 | 2.283747 |
| Indonesia | 0.000000 | 54.595328 | 121.131360 | 148.098694 | 4.062652 |
| Ireland | 0.031796 | 16.660834 | 53.818850 | 76.062630 | 4.187040 |
| Italy | 0.002606 | 19.207820 | 60.267448 | 83.332082 | 4.172936 |
| Japan | 0.000000 | 24.017652 | 60.214094 | 81.351358 | 4.392711 |
| Korea S. | 0.000000 | 129.955438 | 386.626152 | 425.311040 | 3.265655 |
| Malaysia | 0.000000 | 56.525498 | 149.299592 | 189.659002 | 3.786499 |
| Mexico | 0.000000 | 440.818300 | 2132.850298 | 1746.118774 | 2.207160 |
| Netherlands | 0.000000 | 15.802518 | 41.848016 | 60.873070 | 4.611005 |
| Norway | 0.000000 | 27.920608 | 71.785758 | 96.835862 | 4.256095 |
| Philippines | 0.017290 | 52.048754 | 167.407546 | 199.781080 | 3.458544 |
| Portugal | 1.582056 | 13.129484 | 54.154638 | 76.914946 | 4.071114 |
| Singapore | 0.000000 | 30.656496 | 73.326620 | 99.354034 | 4.396040 |
| Spain | 0.000000 | 70.602362 | 139.884600 | 165.163122 | 4.206288 |
| Sweden | 0.000000 | 66.852560 | 130.642934 | 166.827332 | 4.468983 |
| Switzerland | 0.144462 | 15.390592 | 46.179620 | 67.726852 | 4.471172 |
| Taiwan | 0.000000 | 33.290522 | 82.900518 | 110.434774 | 4.224246 |
| Thailand | 0.000000 | 100.851126 | 282.814096 | 314.709524 | 3.298124 |
| Turkey | 0.000000 | 152.625500 | 493.285862 | 506.466162 | 2.893205 |
| UK | 0.000000 | 21.124390 | 47.980512 | 68.613654 | 4.868241 |
| US | 0.000000 | 31.352956 | 75.539480 | 102.259640 | 4.323809 |

must emphasize that all the above discussed log-return densities are strongly leptokurtic and already rather close to the underlying type of log-return density. Still, the above analysis points clearly in the direction of a Student $t$ density as the best fitting log-return density. Given the high significance level the demonstrated Student $t$ property of daily index log-returns establishes a stylized empirical fact that has to be explained in an advanced financial market model. We shall explain later in Sect. 13.2 the findings by the minimal market model, see Platen (2001, 2002), which describes in some sense the optimal dynamics of a financial market. To do this properly, we need to apply the theory of stochastic processes and use stochastic differential equations for modeling.

## 2.7 Convergence of Random Sequences

As we have already seen in the LLN and the CLT, questions about the convergence of random sequences arise naturally in both theory and applications. This is certainly true in the area of quantitative finance, where uncertainty is the key feature that has to be modeled and large numbers of random variables arise.

### Different Types of Convergence

In contrast to a deterministic setting one faces in a stochastic environment several different types of convergence, which is sometimes confusing. For this reason, we summarize several different types of convergence that are commonly used. Some of these we have already introduced previously. In what follows we assume that the random variables $X_1$, $X_2$, ... and $X$ are all defined on the same probability space $(\Omega, \mathcal{A}, P)$.

I. The sequence $X_1$, $X_2$, ... *converges in probability* to $X$ if for every $\varepsilon > 0$

$$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0. \tag{2.7.1}$$

In this case we write $X \stackrel{P}{=} \lim_{n\to\infty} X_n$.

II. The sequence $X_1$, $X_2$, ... *converges with probability one* to $X$ if

$$P\left(\lim_{n\to\infty} X_n = X\right) = 1. \tag{2.7.2}$$

For this type of convergence we write $X \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n$ or say that the sequence converges *almost surely*, that is $P$-a.s. or a.s. to $X$.

III. For $p \in (0, \infty)$ the sequence $X_1$, $X_2$, ... *converges in mean order* $p$ to $X$ if

$$\lim_{n\to\infty} E\left(|X_n - X|^p\right) = 0. \tag{2.7.3}$$

Here we write $X \stackrel{L^p}{=} \lim_{n\to\infty} X_n$. For $p = 2$, this is convergence in the mean square sense, that is $X \stackrel{\text{m.s.}}{=} \lim_{n\to\infty} X_n$, see (2.1.7).

IV. The sequence $X_1$, $X_2$, ... *converges in distribution* to $X$ if

$$\lim_{n\to\infty} E(f(X_n)) = E(f(X)) \tag{2.7.4}$$

for every bounded continuous function $f : \Re \to \Re$. In this case we write $X \stackrel{\mathrm{d}}{=} \lim_{n\to\infty} X_n$. It can be shown, see Shiryaev (1984), that this is equivalent to

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x) \tag{2.7.5}$$

for all continuity points of $x \in \Re$.

For the above types of convergence the following implications, denoted by $\Longrightarrow$, can be shown, see Shiryaev (1984):

$$X \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n \Longrightarrow X \stackrel{P}{=} \lim_{n\to\infty} X_n$$

$$X \stackrel{L^p}{=} \lim_{n\to\infty} X_n \Longrightarrow X \stackrel{P}{=} \lim_{n\to\infty} X_n$$

$$X \stackrel{P}{=} \lim_{n\to\infty} X_n \Longrightarrow X \stackrel{\mathrm{d}}{=} \lim_{n\to\infty} X_n \tag{2.7.6}$$

for $p \in (0, \infty)$. Furthermore, it can be shown, see Shiryaev (1984), that if $X \stackrel{P}{=} \lim_{n\to\infty} X_n$, then there exists a subsequence $X_{i_1}, X_{i_2}, \ldots$ with $X \stackrel{\text{a.s.}}{=} \lim_{k\to\infty} X_{i_k}$. By the implications (2.7.6) it follows if $X \stackrel{L^p}{=} \lim_{n\to\infty} X_n$, then there exists also such a subsequence.

Note that for $X \stackrel{P}{=} \lim_{n\to\infty} X_n$ and $Y \stackrel{P}{=} \lim_{n\to\infty} Y_n$ it follows that

$$X + Y \stackrel{P}{=} \lim_{n\to\infty} (X_n + Y_n) \quad \text{and} \quad X\,Y \stackrel{P}{=} \lim_{n\to\infty} X_n\,Y_n. \tag{2.7.7}$$

Similarly, for $X \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n$ and $Y \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} Y_n$ one has

$$X + Y \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} (X_n + Y_n) \quad \text{and} \quad X\,Y \stackrel{\text{a.s.}}{=} \lim_{n\to\infty} X_n\,Y_n. \tag{2.7.8}$$

In particular, the limits (2.7.8) allow us to work with a rich variety of equations that involve almost surely asymptotically determined random variables.

The *Borel-Cantelli Lemma* states for a sequence of events $A_1, A_2, \ldots$ in $\mathcal{A}$ that

(i) if $\sum_{k=1}^{\infty} P(A_k) < \infty$, then the event that consists of the realization of infinitely many of the events $A_1, A_2, \ldots$ has probability zero, that is

$$P\big(\omega : \text{ there exists a } j \in \mathcal{N} \text{ such that } \omega \in A_i \text{ for all } i \geq j\big) = 0$$

(ii) if $\sum_{k=1}^{\infty} P(A_k) = \infty$ and $A_1, A_2, \ldots$ are independent, then the event that consists of the realization of infinitely many of the events $A_1, A_2, \ldots$ has probability one, that is

$$P\big(\omega : \text{ there exists a } j \in \mathcal{N} \text{ such that } \omega \in A_i \text{ for all } i \geq j\big) = 1.$$

**Limits under Expectation (*)**

For many results in quantitative finance one needs to deal with limits and expectations. Therefore, let us formulate some fundamental results, which allow us to interchange limits and expectations, see Shiryaev (1984). First we mention the *Monotone Convergence Theorem*.

**Theorem 2.7.1.** (Monotone Convergence) *Let $Y$, $X$, $X_1$, $X_2$, ... be random variables.*

*(i) If $X_n \geq Y$ for all $n \in \mathcal{N}$, $E(Y) > -\infty$ and the sequence $(X_n)_{n\in\mathcal{N}}$ is monotone increasing, where $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$, then*

$$\lim_{n\to\infty} E(X_n) = E(X). \tag{2.7.9}$$

*(ii) If $X_n \leq Y$ for all $n \in \mathcal{N}$, $E(Y) < \infty$ and the sequence $(X_n)_{n\in\mathcal{N}}$ is monotone decreasing, where $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$, then*

$$\lim_{n\to\infty} E(X_n) = E(X). \tag{2.7.10}$$

Let us denote by *lim inf* the lower limit and by *lim sup* the upper limit when a sequence has several limits. The following fundamental result on inequalities when interchanging limits and expectations is known as *Fatou's Lemma*.

**Lemma 2.7.2.** (Fatou) *Let $Y$, $X_1$, $X_2$, ... be random variables.*

*(i) If $X_n \geq Y$ for all $n \in \mathcal{N}$ and $E(Y) > -\infty$, then*

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} E(X_n). \tag{2.7.11}$$

*(ii) If $X_n \leq Y$ for all $n \in \mathcal{N}$ and $E(Y) < \infty$, then*

$$\limsup_{n\to\infty} E(X_n) \leq E\left(\limsup_{n\to\infty} X_n\right). \tag{2.7.12}$$

*(iii) If $|X_n| \leq Y$ for all $n \in \mathcal{N}$ and $E(Y) < \infty$, then*

$$E\left(\liminf_{n\to\infty} X_n\right) \leq \liminf_{n\to\infty} E(X_n) \leq \limsup_{n\to\infty} E(X_n) \leq E\left(\limsup_{n\to\infty} X_n\right). \tag{2.7.13}$$

We now mention *Lebesgue's Dominated Convergence Theorem* that underpins a range of practically important results in quantitative finance.

**Theorem 2.7.3.** (Lebesgue) *Let $Y$, $X$, $X_1$, $X_2$, ... be random variables such that $|X_n| \leq Y$, $E(Y) < \infty$ and $\lim_{n\to\infty} X_n \stackrel{a.s.}{=} X$, then we have*

$$E(|X|) < \infty, \tag{2.7.14}$$

$$\lim_{n\to\infty} E(|X_n - X|) = 0 \tag{2.7.15}$$

*and thus*

$$\lim_{n\to\infty} E(X_n) = E(X). \tag{2.7.16}$$

**Definition 2.7.4.** *A family $(X_n)_{n\in\mathcal{N}}$ of random variables is said to be* uniformly integrable *if*

$$\lim_{q\to\infty} \left( \sup_{n\in\mathcal{N}} E\left( |X_n| \, I_{\{|X_n|>q\}} \right) \right) = 0. \tag{2.7.17}$$

Obviously, if $|X_n| \leq Y$ for $n \in \mathcal{N}$ and $E(Y) < \infty$, then the family of random variables $(X_n)_{n\in\mathcal{N}}$ is uniformly integrable. This observation allows the proof of Fatou's Lemma and Lebesgue's Dominated Convergence Theorem by the following general result.

**Theorem 2.7.5.** *Let $X$, $X_1$, $X_2$, ... denote nonnegative random variables with $\lim_{n\to\infty} X_n \overset{a.s.}{=} X$ and $E(X_n) < \infty$ for all $n \in \mathcal{N}$, then it holds*

$$\lim_{n\to\infty} E(X_n) = E(X) \tag{2.7.18}$$

*if and only if the family of random variables $(X_n)_{n\in\mathcal{N}}$ is uniformly integrable.*

### Extreme Value Theorem (*)

The understanding of the occurrence of extreme losses is the key to many risk management problems. It is important, for instance, in Value at Risk analysis and the evaluation of catastrophic insurance claims. We call a random variable or distribution *nondegenerate* if its probability is not concentrated in one single value. Let us assume that catastrophic insurance losses, say for hurricanes, are modeled by an i.i.d. sequence of nondegenerate random variables $X_1, X_2, \ldots$ with distribution function $F_X$. Given a number $n \in \{1, 2, \ldots\}$ of loss data $X_1, X_2, \ldots, X_n$ we are interested in the maximum

$$M_n = \max(X_1, X_2, \ldots, X_n) \tag{2.7.19}$$

of these losses. The minimum can be studied using a maximum by transforming the sequence such that

$$\min(X_1, X_2, \ldots, X_n) = -\max(-X_1, -X_2, \ldots, -X_n). \tag{2.7.20}$$

The following theorem is known as the *Extreme Value Theorem* or *Fisher-Tippett Theorem*, see Embrechts, Klüppelberg & Mikosch (1997) for further details.

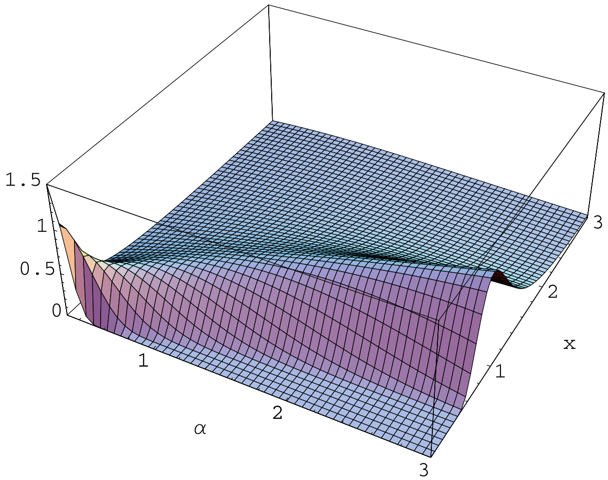



**Fig. 2.7.1.** Density of the Fréchet distribution in dependence on $\alpha$

**Theorem 2.7.6.** (Fisher-Tippett) *If there exist sequences of norming constants* $c_n > 0$*,* $d_n \in \Re$ *for* $n \in \{2, 3, \ldots\}$ *and some random variable* $H$ *with nondegenerate distribution function* $F_H$ *such that*

$$\lim_{n \to \infty} (c_n M_n + d_n) \stackrel{d}{=} H, \tag{2.7.21}$$

*then* $F_H$ *belongs to the type of one of the following three distribution functions:*

*(i)* *the* Fréchet distribution

$$F_H(x) = \begin{cases} 0 & \textit{for} \quad x \le 0, \\ \exp\{-x^{-\alpha}\} & \textit{for} \quad x > 0, \end{cases} \tag{2.7.22}$$

*with* $\alpha > 0$*,*

*(ii)* *the* Weibull distribution

$$F_H(x) = \begin{cases} \exp\{-(-x)^{\alpha}\} & \textit{for} \quad x \le 0, \\ 1 & \textit{for} \quad x > 0, \end{cases} \tag{2.7.23}$$

*with* $\alpha > 0$ *or*

*(iii)* *the* Gumbel distribution

$$F_H(x) = \exp\{-e^{-x}\} \tag{2.7.24}$$

*for* $x \in \Re$*.*

This is a remarkable and rather fundamental result. Whatever underlying loss distribution is given, there is only one of the above three limit distributions that is possible for its maxima. In Fig. 2.7.1, we show the density of the

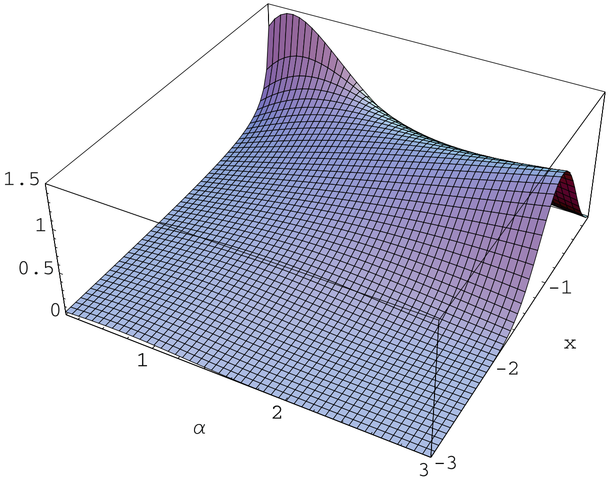



**Fig. 2.7.2.** Density of the Weibull distribution

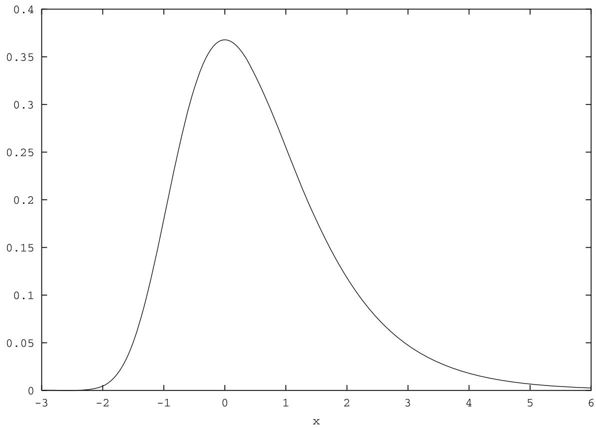



**Fig. 2.7.3.** Density of the Gumbel distribution

Fréchet distribution in dependence on the parameter $\alpha$. Note that the maxima have here only positive values. Figure 2.7.2 displays the Weibull density. This density captures only negative maxima. For the Gumbel distribution the density is plotted in Fig. 2.7.3. In this case the maxima can be positive or negative.

The Extreme Value Theorem raises the question whether a distribution $F_X$ of losses is in the domain of attraction of a given extreme value distribution. For a detailed answer we refer to Embrechts et al. (1997). However, let us mention that appropriately normalized maxima of the Cauchy distribution are, for instance, in the domain of attraction of the Fréchet distribution. Normalized minima of the Student $t$ distribution correspond to the Weibull distribution. Adequately normalized maxima of the normal, gamma, lognormal and exponential distribution are captured by the Gumbel distribution.

It is important to know for risk management purposes which extreme value distribution attracts a given loss distribution. This knowledge is extremely useful, for instance, in insurance premium calculations or in Value at Risk analysis. One knows from Theorem 2.7.6 that one needs only to consider one of the described three extreme value distributions in a specific application when one deals with extreme values.

## 2.8 Exercises for Chapter 2

**2.1.** Consider a sequence of independent random variables $X_1$, $X_2, \ldots$ with mean $\mu$ and variance $\mathrm{Var}(X_i) \leq K < \infty$. What is the almost sure limit of the sample mean $\hat{\mu}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$ of that sequence?

**2.2.** For a sequence of independent identically distributed random variables $X_1, X_2, \ldots$ with mean $\mu \in \Re$ and variance $\sigma^2 > 0$ characterize for the sequence of random variables

$$\hat{Y}_n = \sqrt{n} \left( \frac{1}{n} \sum_{i=1}^{n} X_i - \mu \right)$$

for $n \to \infty$ the limit.

**2.3.** Consider a Gaussian random variable $Z$ with known mean $E(Z)$ and known variance $\mathrm{Var}(Z)$. Provide the 99% confidence interval for $2Z$.

**2.4.** For the random variable $Z$ in Exercise 2.3 calculate the one sided confidence interval with 99% confidence such that $Z$ is at least as large as $-\mathrm{VaR}((1-\alpha)\%)$, that is

$$P(Z \geq -\mathrm{VaR}((1-\alpha)\%)) = 1 - \alpha.$$

# 3

# Modeling via Stochastic Processes

In this chapter the fundamental concept of a stochastic process is introduced. We show how stochastic processes can be applied in the context of asset price modeling. The notions of processes with independent increments, stationary processes and Markov processes are explained. Essentially, stochastic processes provide the mathematical framework that allows us to model financial quantities as families of random variables that evolve over time.

## 3.1 Introduction to Stochastic Processes

To remind us how typical asset prices evolve, we show in Fig. 3.1.1, the history of the S&P500 index from 1993 to 1997. Note that this process appears to be continuous but also seems far from being differentiable.

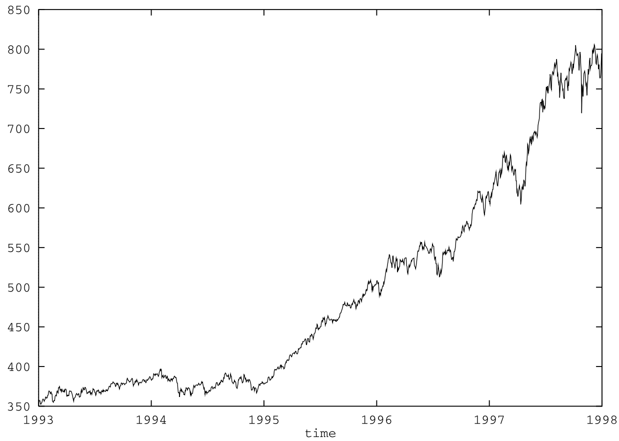



**Fig. 3.1.1.** The S&P500 index for 1993–1997

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

A natural question to ask is: What are appropriate mathematical objects that would allow us to model asset price dynamics? Since asset prices evolve randomly over time, in the early development of the theory of finance it was already realized that the best representation of price behavior would be a probabilistic one. As early as 1900 Bachelier proposed a model for the motion of stock prices. It probably was the first work that involved a mathematical object that is today known as Brownian motion. Only later Einstein, Wiener and others studied the same fundamental mathematical object to describe continuous and strongly fluctuating random dynamics. Today the theory of stochastic processes provides a general mathematical framework that allows us to build and investigate models that involve Brownian motion and more complicated families of random variables.

### Stochastic Process

Throughout the following we shall assume that there exists a common underlying probability space $(\Omega, \mathcal{A}, P)$ consisting of the sample space $\Omega$, the collection of events $\mathcal{A}$ and the probability measure $P$, as discussed in Sect. 1.1.

A collection of random variables $X_{t_0}, X_{t_1}, \ldots$ can be conveniently used to describe the evolution of an observed asset price, for instance, daily closing values for the S&P500 index over any given set of observation times $t_0 < t_1 < \ldots$.

**Definition 3.1.1.** *We call a family* $X = \{X_t,\, t \in \mathcal{T}\}$ *of random variables* $X_t \in \Re$ *a* stochastic process*, where the totality of its finite-dimensional distribution functions*

$$F_{X_{t_{i_1}},\ldots,X_{t_{i_j}}}(x_{i_1}, \ldots, x_{i_j}) = P(X_{t_{i_1}} \leq x_{i_1}, \ldots, X_{t_{i_j}} \leq x_{i_j}) \tag{3.1.1}$$

*for* $i_j \in \{0, 1, \ldots\}$*,* $j \in \mathcal{N}$*,* $x_{i_j} \in \Re$ *and* $t_{i_j} \in \mathcal{T}$ *determines its* probability law*, see* (1.4.36)–(1.4.40)*.*

The stochastic process $X$ is indexed by the time $t$ and we call $\mathcal{T}$ the *time set*. The *state space* of $X$ is here the one-dimensional Euclidean space $\Re$ or a subset of it. As in Definition 3.1.1 we define analogously a stochastic process $X = \{X_t,\, t \in \mathcal{T}\}$ where $X_t \in \Re^d$, $d \in \mathcal{N}$, $t \in \mathcal{T}$.

Since asset prices can be observed at any time instant, it is desirable that stochastic processes can be defined for all time instants in an interval which, for example, can be the finite time set $\mathcal{T} = [0, 1]$ or an infinite set such as $\mathcal{T} = [0, \infty)$. In these cases we call the process a *continuous time stochastic process*. We typically consider stochastic processes on a time set $\mathcal{T} = [0, \infty)$ or $\mathcal{T} = [0, T]$ with $T \in (0, \infty)$. The benchmark approach, which we develop later in a continuous time setting, will allow us to consider the evolution of financial markets on $[0, \infty)$.

As an example for $\mathcal{T} = [0, T]$ we can take the linearly interpolated path of the S&P500 daily prices in Fig. 3.1.1, which represents a realization of a stochastic process on the finite time set $\mathcal{T} = [1993, 1998]$ for the time period from 1993 up until 1998. We could also choose a discrete time set $\mathcal{T} = \{0, 1, \ldots\}$ to number, for instance, the observation days. In Sect. 2.5 we have used a discrete time set to model financial markets. However, time evolves continuously and we shall concentrate later on a framework for continuous time financial market models.

The most fundamental characteristic of a stochastic process $X$ is the totality of its finite-dimensional distribution functions given in (3.1.1). These have according to (1.4.39) the *consistency property*

$$F_{X_{t_{i_1}},\ldots,X_{t_{i_k}}}(x_{i_1},\ldots,x_{i_k}) = \lim_{x_i \to \infty} F_{X_{t_1},\ldots,X_{t_n}}(x_1,\ldots,x_n) \tag{3.1.2}$$

for $\{i_1, i_2, \ldots, i_k\} \subseteq \{1, 2, \ldots, n\}$ with $k \in \{1, 2, \ldots, n\}$, where the limit has to be taken for all $i \notin \{i_1, i_2, \ldots, i_k\}$. The following fundamental result, which is due to Kolmogorov, shows that under the above consistency property a given family of finite dimensional distribution functions is that of a stochastic process. This is an important statement that gives access to the direct construction of particular stochastic processes.

**Theorem 3.1.2.** (Kolmogorov) *Let there be a given family of finite-dimensional distribution functions $\{F_{X_{t_{i_1}},\ldots,X_{t_{i_k}}}(x_{i_1},\ldots,x_{i_k})\}$ satisfying the consistency property (3.1.2). Then there exists a probability space $(\Omega, \mathcal{A}, P)$ and a stochastic process $X = \{X_t,\, t \in [0, \infty)\}$ such that (3.1.1) is satisfied.*

The families of random variables that we are going to consider satisfy the consistency property (3.1.2) and are therefore stochastic processes. A trivial example of a stochastic process arises if the daily closing prices of an asset are taken to be independent and identically distributed. In such a model, what happens at present is completely unaffected by what has happened in the past. The corresponding probability law is then quite simple, since the finite-dimensional distribution function is here given by the product

$$F_{X_{t_{i_1}},X_{t_{i_2}},\ldots,X_{t_{i_j}}}(x_{i_1},x_{i_2},\ldots,x_{i_j}) = F_{X_{t_{i_1}}}(x_{i_1})\, F_{X_{t_{i_2}}}(x_{i_2}) \cdots F_{X_{t_{i_j}}}(x_{i_j}), \tag{3.1.3}$$

where $F$ is the common distribution of the independent random variables. Such an asset price model appears to be rather unrealistic since consecutive values of prices can reasonably be expected to be related. A sequence of independent random variables to be used for modeling increments of asset prices would seem to be more realistic. Since the consistency property (3.1.2) is satisfied in both cases, the corresponding models form stochastic processes on the probability space $(\Omega, \mathcal{A}, P)$. The discrete time models introduced in Sect. 2.5 provide further examples for stochastic processes.

A model that is generating a sequence of independent outcomes would much better fit, for example, daily increments or log-returns rather than actual values of asset prices themselves. Note that the variances of price increments are much larger than those of log-returns if the corresponding prices

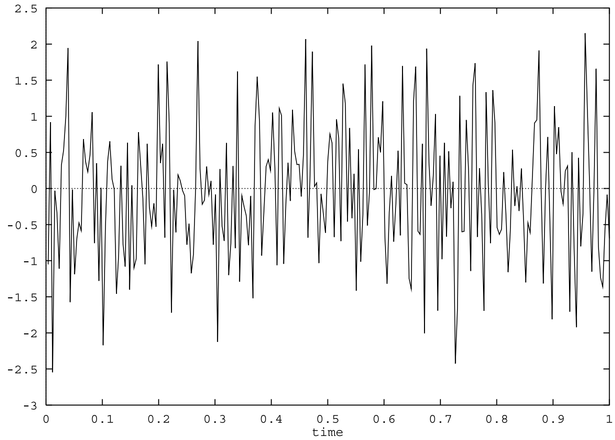



**Fig. 3.1.2.** Interpolated independent standard Gaussian outcomes

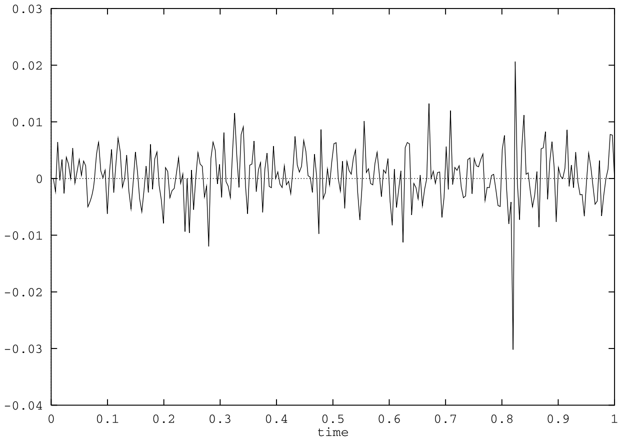



**Fig. 3.1.3.** Log-returns of the S&P500 index for 1997

are large. In Fig. 3.1.2 we show a linearly interpolated path of a sequence of 256 independent standard Gaussian distributed outcomes.

For comparison we show in Fig. 3.1.3 the linearly interpolated log-returns of the S&P500 for the year 1997, these are the increments of the logarithm of the index. Note that they show some similarity with a sequence of independent Gaussian distributed random variables as shown in Fig. 3.1.2. In both cases the values change drastically over time. However, in the case of the S&P500 we seem to observe extreme log-returns.

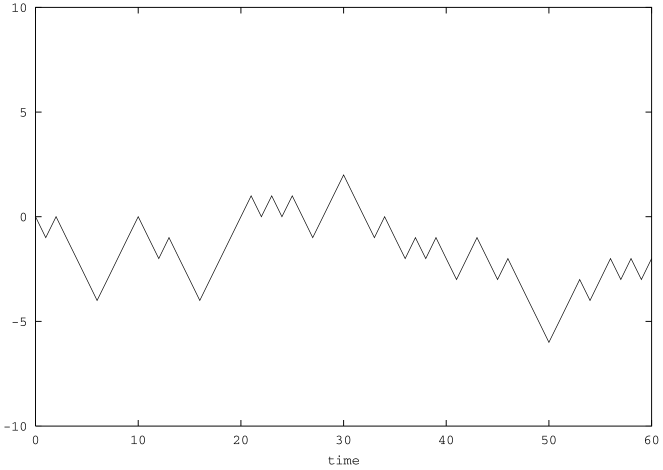



**Fig. 3.1.4.** Path of a symmetric random walk

## Symmetric Random Walk and Binomial Tree

An asset price model that has been widely used arises if we assume that the logarithm $X_{t_{n+1}}$ of the asset price of the $(n+1)$th trading day results from an equally probable up or downward move by a certain fixed amount relative to the previous logarithm $X_{t_n}$ of the price, see, for example, Cox, Ross & Rubinstein (1979). This corresponds to the well-known *symmetric random walk*, which is characterized by a first order difference equation for the log-return $R_{t_n}$ of the form

$$R_{t_n} = X_{t_{n+1}} - X_{t_n} = a\,\xi_n, \tag{3.1.4}$$

where $a > 0$ is the fixed amount for the daily absolute log-price increment. The variables $\xi_n,\ n \in \{0, 1, \ldots\}$, are independent, two-point distributed random variables with the probabilities

$$P(\xi_n = \pm 1) = \frac{1}{2}, \tag{3.1.5}$$

corresponding to either up or downward moves. Here the initial value of the log-price is of the form $X_{t_0} = ka$ for some given integer $k \in \mathcal{N}$. A linearly interpolated sample path obtained for the symmetric random walk is shown in Fig. 3.1.4 for the parameters $k = 0$ and $a = 1$. Note that such a random walk admits only states given by integer multiples of the amount $a$ and does not take other values.

A symmetric random walk is therefore still a rather restrictive log-price model since, in practice, log-returns usually vary considerably as we have seen in Sect. 2.6. Figure 3.1.5 shows the possible paths that a random walk could traverse. Note that the random walk forces the log-asset values to evolve only in the bounded cone of possible paths. For this reason the symmetric random

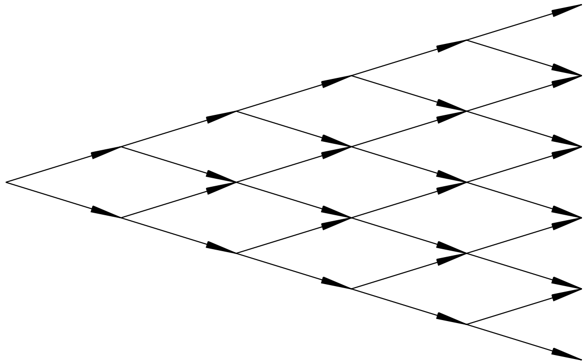


**Fig. 3.1.5.** Binomial tree

walk is sometimes referred to as a symmetric *binomial tree*. Such binomial trees are widely used in quantitative finance for pricing options and other derivatives under various models. We shall describe later examples of how this can be done.

### Gaussian and Student *t* Random Walk

We can construct a more realistic model for the logarithm of an asset price by allowing Gaussian, instead of two-point distributed independent log-returns, which then result in a *Gaussian random walk*. Here, in principle, any real value can occur. The Gaussian random walk can be formally characterized again by the difference equation (3.1.4), but the random variables $a\xi_n$ are now taken to be independent Gaussian $N(\mu, \sigma^2)$ distributed random variables for $n \in \{0, 1, \ldots\}$. Figure 3.1.6 shows a linearly interpolated sample path of a standard Gaussian random walk. Note that here the daily log-returns vary when compared with Fig. 3.1.4 for the symmetric random walk. In our example the signs of the upward and downward moves in Fig. 3.1.4 and Fig. 3.1.6 coincide. Of course, one can take any distribution for the log-returns. As we have seen in Sect. 2.6 the Student $t(4)$ distribution matches well the estimation of index log-returns. In Fig. 3.1.7 we show a random walk with Student $t(4)$ distributed increments.

Note that the Student $t(4)$ distributed random increments are generated by multiplying the independent Gaussian increments in Fig. 3.1.6 with the square root of the inverse of a chi-square distributed random variable with four degrees of freedom. Such chi-square distributed random variable is obtained as the sum of the four squared independent standard Gaussian random variables, as described in Sect. 1.2.

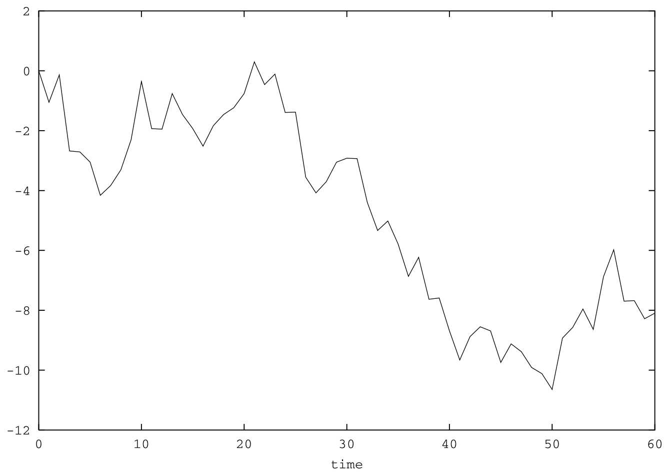



**Fig. 3.1.6.** Path of a Gaussian random walk

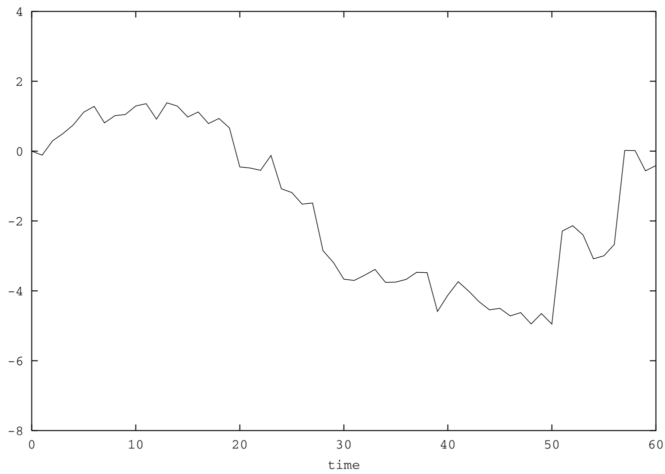



**Fig. 3.1.7.** Path of a Student $t(4)$ random walk

### Realizations of Stochastic Processes

As mentioned at the beginning of this section, a stochastic process is a family of random variables which is indexed by a time parameter and has to satisfy the consistency property (3.1.2). In principle, it is a function $X : \mathcal{T} \times \Omega \to \Re^d$ with some technical measurability constraints. For a fixed sample point $\omega \in \Omega$ we call $X_\cdot(\omega) : \mathcal{T} \to \Re^d$ a *realization*, *sample path* or *trajectory* of the stochastic process. The curves plotted in Fig. 3.1.4 and Fig. 3.1.5 or the history of the S&P500 in Fig. 3.1.1 can be interpreted as specific realizations of corresponding stochastic processes. For fixed time $t \in \mathcal{T}$ the quantity $X_t(\cdot)$ is a random variable with $X_t(\cdot) : \Omega \to \Re^d$ with all properties of random variables as discussed in Chap. 1.

Recall by (3.1.2) that the function $X : \mathcal{T} \times \Omega \to \Re^d$ cannot be completely arbitrary. If we do not state anything else, then we shall consider stochastic processes with right-continuous sample paths that equal at any time $t \in \mathcal{T}$ their right hand limits, that is

$$X_t = \lim_{\varepsilon \to 0+} X_{t+\varepsilon},$$

having left hand limits

$$X_{t-} = \lim_{\varepsilon \to 0-} X_{t-\varepsilon},$$

where both limits are finite for each $\omega \in \Omega$ and $t \in \mathcal{T}$ with $\varepsilon > 0$.

To be mathematically precise we would have to refer to some technically demanding notation and results from measure theory, see Doob (1994), which are similar to those already applied to random variables in Sect. 1.1. Similarly to random variables, which we defined in Chap. 1, we assume that a stochastic process is *measurable*. We do not further specify this here and refer instead to Karatzas & Shreve (1991). All stochastic processes that we consider have this measurability property.

The question of when two stochastic processes model the same dynamics is an important one for modeling in quantitative finance. Any ambiguity could lead to potential problems for practical applications. Let us suppose that there are two stochastic processes $X$ and $Y$ defined on the same probability space $(\Omega, \mathcal{A}, P)$ taking values in $\Re^d$, $d \in \mathcal{N}$. In the given probabilistic setting $X$ and $Y$ can be distinguished in a number of different ways. One possibility is to say that the processes $X$ and $Y$ are *indistinguishable* for the time set $\mathcal{T}$ if

$$X_t(\omega) = Y_t(\omega) \tag{3.1.6}$$

a.s. for all $t \in \mathcal{T}$, that is, all their sample paths coincide a.s.

Let us note that if two stochastic processes $X$ and $Y$ are equal a.s. for all $t \in [0, \infty)$, then if they are right continuous they are indistinguishable. This is a useful fact that we exploit because we consider typically only right-continuous stochastic processes with left hand limits.

## 3.2 Certain Classes of Stochastic Processes

It is useful to distinguish between various classes of stochastic processes according to their specific temporal relationships. This is important for applications in finance because we need to find classes of stochastic processes that we can use as basic building blocks for realistic market models. Furthermore, it is essential that we find classes of stochastic processes that are for constant parameters analytically tractable. This allows us to establish fast and efficient numerical methods to calculate important characteristics as moments, probabilities, option prices or other important financial quantities.

**Moments of Stochastic Processes**

Some important information about a stochastic process can be provided by the *mean*

$$\mu(t) = E(X_t) \tag{3.2.1}$$

and the *variance*

$$v(t) = \mathrm{Var}(X_t) = E\left((X_t - \mu(t))^2\right) \tag{3.2.2}$$

for $t \in \mathcal{T}$ and the *covariance*

$$C(s,t) = \mathrm{Cov}(X_s, X_t) = E((X_s - \mu(s))(X_t - \mu(t))) \tag{3.2.3}$$

for $s, t \in \mathcal{T}$. Obviously, these quantities only make sense if they are finite. For instance, the fourth moment of the increments of the Student $t(4)$ random walk in Fig. 3.1.7 do not theoretically exist. The stochastic process formed by the independent standard Gaussian outcomes plotted in Fig. 3.1.2 has mean $\mu(t_i) = 0$, variance $v(t_i) = 1$ and covariances $C(t_i, t_k) = 0$ for $i \neq k$.

**Stationary Processes**

An important class of stochastic processes is that of *stationary processes* since they represent a form of probabilistic equilibrium. This property is appealing and is often useful in a financial modeling context. For instance, interest rates, dividend rates, inflation rates, volatilities and credit spreads are likely to be modeled by stationary processes since they typically exhibit some equilibrium type dynamics.

**Definition 3.2.1.** *We say that a stochastic process* $X = \{X_t,\, t \in \mathcal{T}\}$ *is* stationary *if its joint distributions are all invariant under time displacements, that is if*

$$F_{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}} = F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}} \tag{3.2.4}$$

*for all* $h > 0$, $t_i \in \mathcal{T}$, $i \in \{1, 2, \ldots, n\}$ *and* $n \in \mathcal{N}$.

In particular, the random variables $X_t$ have the same distribution for all $t \in \mathcal{T}$ and thus, means, variances and covariances satisfy the equations

$$\mu(t) = \mu(0), \qquad v(t) = v(0) \qquad \text{and} \qquad C(s,t) = c(t-s) \tag{3.2.5}$$

for all $s, t \in \mathcal{T}$, where $c : \Re \to \Re$ is a function.

For example, a sequence of i.i.d. random variables forms a stationary stochastic process on a discrete time set $\mathcal{T} = \{0, 1, \ldots\}$.

For comparison, in Fig. 3.2.1 we show a trajectory of a stationary continuous Gaussian process with mean $\mu(t) = 0.05$. It can be interpreted as a sample path of interest rates obtained from the Vasicek model, see Vasicek (1977). We describe the Vasicek model in the next chapter. Note that the

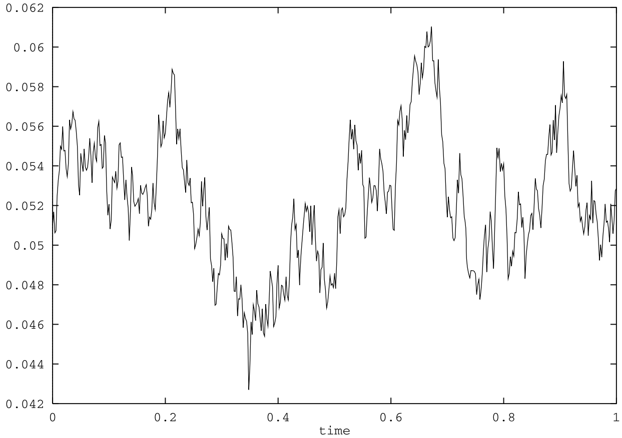



**Fig. 3.2.1.** Sample path for the Vasicek interest rate model

process appears to be oscillating around a reference level over time. Its key feature as a stationary process is that we could continue to observe its movements over longer and longer time horizons but its distributional properties and thus its mean, variance and covariance would not change. More precisely, its probabilistic features would be time shift invariant.

### Processes with Stationary Independent Increments

*Processes with stationary independent increments* form another important class of stochastic processes. These processes have outstanding mathematical properties that make them suitable as fundamental building blocks in financial modeling. For these processes the random increments $X_{t_{j+1}} - X_{t_j}$, $j \in \{0, 1, \ldots, n-1\}$, are independent for any combination of time instants $t_0 < t_1 < \ldots < t_n$ in $\mathcal{T}$ for all $n \in \mathcal{N}$. If $t_0$ is the smallest time instant in $\mathcal{T}$, then the initial value $X_{t_0}$ and the random increments $X_{t_j} - X_{t_0}$ for any other $t_j \in \mathcal{T}$ are also required to be independent. The increments are assumed to be stationary, that is $X_{t+h} - X_t$ has the same distribution as $X_h - X_0$ for all $h > 0$ and $t \geq 0$. On the basis of Kolmogorov's theorem, that is Theorem 3.1.2, it follows that the resulting model forms a stochastic process on the given probability space $(\Omega, \mathcal{A}, P)$.

### Wiener Process

The so-called Brownian motion was discovered by Robert Brown in the early 19th century when observing, under a microscope, the motion of pollen grains that were subject to a large number of small random molecular collisions. The displacement of such a particle over time, say in a north-south direction, resembles much that of the paths of the Wiener process as is given in

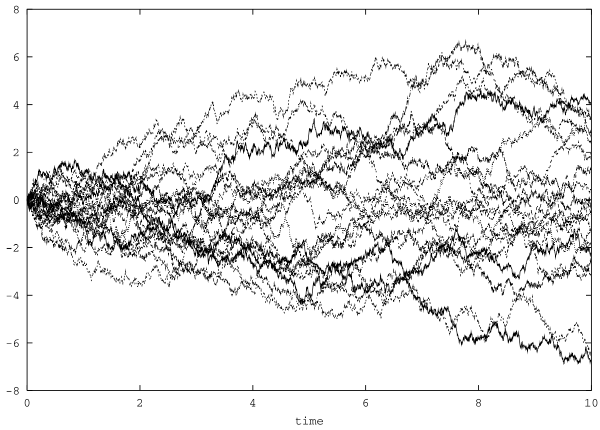



**Fig. 3.2.2.** Sample paths of the Wiener process

Fig. 3.2.2. For this reason the Wiener process that we introduce now is also called Brownian motion.

The most important continuous process with independent increments is the *Wiener process*. It is a continuous time stochastic process with independent Gaussian distributed increments and continuous sample paths.

**Definition 3.2.2.** *We define the* standard Wiener process $W = \{W_t, t \in [0, \infty)\}$ *as a process with Gaussian stationary independent increments and continuous sample paths for which*

$$W_0 = 0, \qquad \mu(t) = E(W_t) = 0, \qquad \text{Var}(W_t - W_s) = t - s \tag{3.2.6}$$

*for all* $t \in [0, \infty)$ *and* $s \in [0, t]$.

In Fig. 3.2.2 we display 20 trajectories of a Wiener process. Each trajectory is highly erratic but always forms a continuous path. In Bachelier (1900), a linearly *transformed Wiener process* of the type

$$Y_t = Y_0 + b\, W_t \tag{3.2.7}$$

was employed to model stock prices observed at the Paris Bourse. Only later an object like the Wiener process was used in physics by Einstein (1905). In Fig. 3.2.3 a sample path for such a transformed Wiener process with $Y_0 = 100$ and $b = 20$ for a period of $T = 1$ year is displayed.

The Wiener process has fundamental mathematical properties and is used in many applications in finance. Often one uses also the name Brownian motion to characterize it. Note that the Wiener process is *not* a stationary process, as can be seen from its increasing variance, see (3.2.6).

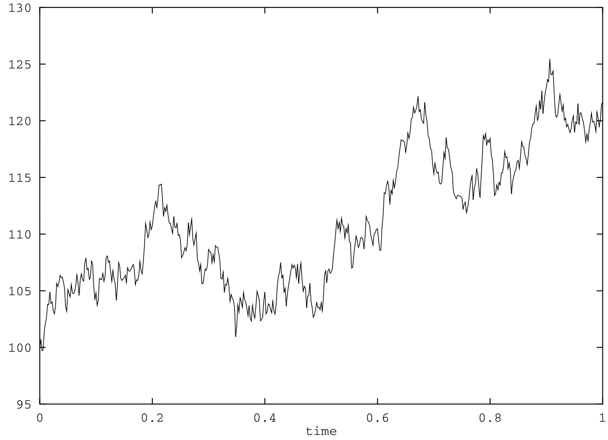



**Fig. 3.2.3.** Sample path of a transformed Wiener process

## 3.3 Discrete Time Markov Chains

### Markov Processes

Very suitable processes for financial modeling are the, so-called, *Markov processes*. Some classes of Markov processes will be considered in the following section. As we shall see, they provide a high degree of tractability. In the context of stochastic processes it is the *Markov property* that allows us to specify the *future* evolution of an asset price only in dependence on its *present* value and, therefore, not depending on any *past* information. In the following we shall discuss different classes of Markov processes including discrete and continuous time Markov chains and later also diffusion processes. The class of Markov processes has great significance for financial modeling. These processes offer an efficient way of characterizing properties of state variables without the need to consider or memorize their entire past history. For the implementation of tractable quantitative methods this is a crucial advantage.

We first consider the Markov property in a discrete time setting because this can be more easily explained.

### Discrete Time Markov Property

To analyze the relationships between the random variables of a stochastic process at different time instants, the corresponding conditional probabilities are often used. We consider a sequence of random variables $X_{t_0}, X_{t_1}, \ldots$ taking values in a discrete *set of states*

$$\mathcal{X} = \{\ldots, y_{-j}, y_{-(j-1)}, \ldots, y_{-1}, y_0, y_1, \ldots, y_{j-1}, y_j, \ldots\}. \tag{3.3.1}$$

To illustrate this, let us consider the symmetric random walk as given in (3.1.4). Here we have $y_j = ja$ for $j \in \{\ldots, -1, 0, 1, \ldots\} = \bar{\mathcal{N}}$. Suppose that we observe at a time instant $t_n$ the present outcome $x_n$, that is the value of $X_{t_n} = x_n$, as well as the past outcomes $X_{t_0} = x_0$, $X_{t_1} = x_1$, ..., $X_{t_{n-1}} = x_{n-1}$ for $x_0, x_1, \ldots, x_n \in \mathcal{X}$. By using this information we want to describe the probabilistic properties of the future values, say log-asset prices $X_{t_{n+1}}$, $X_{t_{n+2}}, \ldots$. The value of the next immediate future log-asset price $X_{t_{n+1}}$ is of particular interest. We can write the corresponding conditional probabilities in the form

$$P\left(X_{t_{n+1}} = x_{n+1} \,\middle|\, X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\right) = \frac{P\left(X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n, X_{t_{n+1}} = x_{n+1}\right)}{P\left(X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\right)} \tag{3.3.2}$$

for each $x_0, x_1, \ldots, x_{n+1} \in \mathcal{X}$ and $n \in \{0, 1, \ldots\}$. For our example, the symmetric random walk, we obtain due to the difference equation (3.1.4) and the independence property a *one-step transition probability*

$$P\Big(X_{t_{n+1}} = x_{n+1} \,\Big|\, X_{t_n} = x_n\Big) = P\Big(X_{t_{n+1}} = x_{n+1} \,\Big|\, X_{t_0} = x_0, \ldots, X_{t_n} = x_n\Big) \tag{3.3.3}$$

for each $n \in \{0, 1, \ldots\}$ and $x_0, x_1, \ldots, x_{n+1} \in \mathcal{X}$. That is, the future log-asset price value $X_{t_{n+1}}$ does not depend on the past if the present log-asset price value $X_{t_n}$ is known. We call property (3.3.3) the *discrete time Markov property.*

More generally, a discrete time stochastic process satisfying condition (3.3.3) with a discrete set $\mathcal{X}$ of possible states, is called a *discrete time Markov chain.* The set $\mathcal{X}$ of states can be finite or infinite and quite different from what is used for our symmetric random walk.

## Probabilities of a Markov Chain

For a discrete time Markov chain let us denote for each $n \in \{0, 1, \ldots\}$ the transition probability at time $t_n$ by

$$p^{i,j}(n) = P\left(X_{t_{n+1}} = y_j \,\middle|\, X_{t_n} = y_i\right) \tag{3.3.4}$$

for integers $i$ and $j$, where $p^{i,j}(n)$ denotes the probability of moving from state $y_i$ to state $y_j$ at time $t_n$. In our example of a symmetric random walk only the one step upward or downward transition probabilities, $p^{i,i+1}(n)$ and $p^{i,i-1}(n)$ respectively, are not zero. Obviously, we have

$$0 \le p^{i,j}(n) \le 1 \tag{3.3.5}$$

for all integers $i$ and $j$ and $n \in \{0, 1, \ldots\}$. Since $X_{t_{n+1}}$ can only attain states in $\mathcal{X}$ we also have

$$\sum_{j \in \bar{\mathcal{N}}} p^{i,j}(n) = 1 \tag{3.3.6}$$

for each integer $i$ and $n \in \{0, 1, \ldots\}$.

For $n \in \{0, 1, \ldots\}$ let

$$p_i(n) = P(X_{t_n} = y_i) \tag{3.3.7}$$

be the probability that the log-asset price, or more generally the value $X_{t_n}$, of the Markov chain $X$ at time $t_n$ equals $y_i$. Then the probability $p_j(n+1)$, that at the next time instant $t_{n+1}$ the Markov chain is in the state $y_j$, is related to $p_i(n)$ through the equation

$$p_j(n+1) = \sum_{i \in \bar{\mathcal{N}}} p_i(n)\, p^{i,j}(n) \tag{3.3.8}$$

for all integers $j$. Hence, if we know the initial probabilities $p_i(0)$, for all integers $i$, then we obtain by application of (3.3.8) in several steps

$$p_j(n+1) = \sum_{i_0 \in \bar{\mathcal{N}}} p_{i_0}(0) \sum_{i_1 \in \bar{\mathcal{N}}} p^{i_0,i_1}(0) \cdots \sum_{i_{n-1} \in \bar{\mathcal{N}}} p^{i_{n-2},i_{n-1}}(n-2)$$
$$\cdot \sum_{i_n \in \bar{\mathcal{N}}} p^{i_{n-1},i_n}(n-1)\, p^{i_n,j}(n) \tag{3.3.9}$$

for all integers $j$ with

$$\sum_{j \in \bar{\mathcal{N}}} p_j(n+1) = 1 \tag{3.3.10}$$

for $n \in \{0, 1, \ldots\}$. The relations (3.3.5)–(3.3.10) permit, in principle, a rather efficient computation of probabilities and related functionals for a discrete time Markov chain. In the case of the symmetric random walk this will be extremely simple as we shall see below.

### Probabilities of a Symmetric Binomial Tree

The symmetric random walk is an important example of a discrete time Markov chain that is frequently applied in finance, see Cox, Ingersoll & Ross (1985), mainly because of its simplicity. We have seen in Fig. 3.1.5 the range of possible paths that can be traversed. This simple log-asset price model provides easy access to the computation of a wide range of derivative prices as we shall see later. However, due to the simplicity of the model these prices can be rather inaccurate.

The one step transition probability for a symmetric random walk or equivalently a symmetric binomial tree, see (3.1.5), is given by

$$p^{i,j}(n) = \begin{cases} \frac{1}{2} \text{ for } & j \in \{i-1, i+1\} \\ 0 \text{ otherwise} \end{cases} \tag{3.3.11}$$

for $n \in \{0, 1, \ldots\}$.

Let us now compute the probabilities that correspond to specific log-asset price values in the binomial tree using (3.3.9) and (3.3.11). We assume that $X$ starts at time $t_0 = 1$, with initial value $X_{t_0} = ka \in \mathcal{X}$ with probability $p_k(0) = 1$ and $p_i(0) = 0$ for $i \neq k$.

The total number of upward plus downward moves until time $t_n$ equals the number $n$. Obviously, the value of the log-asset price at time $t_n$ is a function of the number of upward moves of the random walk. Since a jump at any time step is assumed to be independent from the other jumps we can compute the corresponding probability for the log-asset price value being in a particular state. This probability follows from the binomial probabilities given in (2.1.32) with $p = 0.5$. More precisely, the log-asset price takes at time $t_n$ the value $X_{t_n} = aj$ with probability

$$p_j(n) = \frac{n!}{\left(\frac{j-(k-n)}{2}\right)! \left(n - \frac{j-(k-n)}{2}\right)!} \left(\frac{1}{2}\right)^n \tag{3.3.12}$$

for $n \in \{0, 1, \ldots\}$ and

$$j \in \{k-n, k-(n-2), k-(n-4), \ldots, k+(n-4), k+(n-2), k+n\},$$

where, as mentioned earlier, $i! = 1 \cdot 2 \cdot \ldots \cdot (i-1) \cdot i$ and $0! = 1$. Recall that some binomial probabilities are displayed in Fig. 2.1.5, which indicates that after only a few time steps binomial probabilities resemble Gaussian ones. This is a consequence of the Central Limit Theorem, as pointed out in Sect. 2.1. In this sense the symmetric random walk can also be interpreted as an approximation of a transformed Wiener process.

As previously noted, the symmetric binomial tree or random walk is a powerful but simple model for log-asset prices. Its probabilities are easily computed and the possible log-price movements are kept to a minimum as these can occur in only a fixed up or down jump. In the context of derivative pricing and hedging, binomial, trinomial and other types of discrete time Markov chains are often used as approximations for log-asset price or asset price models. Later we shall give examples for such discrete time Markov chains tailored to suit particular models.

## 3.4 Continuous Time Markov Chains

Let us now study the Markov property in a continuous time setting where only discrete levels can be reached.

### Continuous Time Markov Property

A continuous time Markov chain is a continuous time stochastic process $X = \{X_t, t \in [0, \infty)\}$, which takes values in a finite set of states $\mathcal{X} =$

$\{y_1, y_2, \ldots, y_N\}$ such that its values jump at random time instants from one state to another. For instance, this could be a model for the short term interest rate that is set by a Central Bank. It could also model the credit ratings for a company. The random jump times are not known in advance. In this context the *continuous time Markov property* takes the form

$$P\left(X_{t_{n+1}} = x_{n+1} \,\middle|\, X_{t_0} = x_0, \ldots, X_{t_n} = x_n\right) = P\left(X_{t_{n+1}} = x_{n+1} \,\middle|\, X_{t_n} = x_n\right) \tag{3.4.1}$$

for all $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1} < \infty$ and $x_0, x_1, \ldots, x_{n+1} \in \mathcal{X}$, where $n \in \{0, 1, \ldots\}$. Note that in (3.4.1), given the present, the past does not affect the transition probability that there is a jump from state $x_n$ at time $t_n$ to state $x_{n+1}$ at time $t_{n+1}$.

## Transition Probability Matrix

To make this more precise let us define for $0 \leq t_0 < t_1 < \infty$ an $N \times N$ *transition probability matrix* $\boldsymbol{P}(t_0; t_1) = [p^{i,j}(t_0; t_1)]_{i,j=1}^N$ componentwise by

$$p^{i,j}(t_0; t_1) = P\left(X_{t_1} = y_j \,|\, X_{t_0} = y_i\right)$$

for $i$, $j \in \{1, 2, \ldots, N\}$. Obviously, $\boldsymbol{P}(t_0; t_0) = \boldsymbol{I}$ is the unit matrix. Let us introduce for $t \geq 0$ and $y_i \in \mathcal{X}$ the probability

$$p_i(t) = P(X_t = y_i) \tag{3.4.2}$$

and form the probability vector

$$\boldsymbol{p}(t) = (p_1(t), p_2(t), \ldots, p_N(t))^\top. \tag{3.4.3}$$

The probability vectors $\boldsymbol{p}(t_0) = (p_1(t_0), p_2(t_0), \ldots, p_N(t_0))^\top$ and $\boldsymbol{p}(t_1)$ for $0 \leq t_0 < t_1 < \infty$ are related by the equation

$$\boldsymbol{p}(t_1)^\top = \boldsymbol{p}(t_0)^\top \, \boldsymbol{P}(t_0; t_1).$$

Using this result for $t_0 < t_1 < t_2$ we have $\boldsymbol{p}(t_2)^\top = \boldsymbol{p}(t_0)^\top \boldsymbol{P}(t_0; t_2)$ and

$$\boldsymbol{p}(t_2)^\top = \boldsymbol{p}(t_1)^\top \, \boldsymbol{P}(t_1; t_2) = \boldsymbol{p}(t_0)^\top \, \boldsymbol{P}(t_0; t_1) \, \boldsymbol{P}(t_1; t_2) \tag{3.4.4}$$

for any initial probability vector $\boldsymbol{p}(t_0)$. It can be shown that the transition probability matrices satisfy the relationship

$$\boldsymbol{P}(t_0; t_2) = \boldsymbol{P}(t_0; t_1) \, \boldsymbol{P}(t_1; t_2) \tag{3.4.5}$$

for all $t_0 < t_1 < t_2$. In the special case where the transition matrices $\boldsymbol{P}(t_0; t_1)$ depend only on the time difference $t_1 - t_0$, that is $\boldsymbol{P}(t_0; t_1) = \boldsymbol{P}(0; t_1 - t_0)$ for all $0 \leq t_0 < t_1 < \infty$, we say that the continuous time Markov chain is *homogeneous* and write $\boldsymbol{P}(t)$ for $\boldsymbol{P}(0; t)$. Then relation (3.4.5) can be reduced to the matrix equation

$$\boldsymbol{P}(s+t) = \boldsymbol{P}(s)\,\boldsymbol{P}(t) = \boldsymbol{P}(t)\,\boldsymbol{P}(s) \tag{3.4.6}$$

for all $s, s+t \in [0, \infty)$. We note that the transition probability matrix of a continuous time Markov chain satisfies a particular product relationship, which is a reflection of its Markov property. It leads to the following system of ODEs.

### Kolmogorov Equations

For a homogeneous continuous time Markov chain it can be shown that there exists an $N \times N$ *intensity matrix* $\boldsymbol{A} = [a^{i,j}]_{i,j=1}^N$ with

$$a^{i,j} = \begin{cases} \lim_{t\to 0} \dfrac{p^{i,j}(t)}{t} & \text{for} \quad i \neq j \\ \lim_{t\to 0} \dfrac{p^{i,i}(t) - 1}{t} & \text{for} \quad i = j \end{cases} \tag{3.4.7}$$

for $i, j \in \{1, 2, \ldots, N\}$ which, together with the initial probability vector $\boldsymbol{p}(0)$, completely characterizes its random behavior. Note that for each $i \in \{1, 2, \ldots, N\}$ the transition intensities always add up to zero, that is

$$\sum_{j=1}^{N} a^{i,j} = 0. \tag{3.4.8}$$

If the diagonal components $a^{i,i}$ of the intensity matrix $\boldsymbol{A}$ are finite for each $i \in \{1, 2, \ldots, N\}$, then the transition probabilities satisfy the *Kolmogorov forward equation*

$$\frac{dp^{i,j}(t)}{dt} - \sum_{k=1}^{N} p^{i,k}(t)\, a^{k,j} = 0 \tag{3.4.9}$$

and the *Kolmogorov backward equation*

$$\frac{dp^{i,j}(t)}{dt} - \sum_{k=1}^{N} a^{i,k}\, p^{k,j}(t) = 0 \tag{3.4.10}$$

for all $i, j \in \{1, 2, \ldots, N\}$ and $t \in [0, \infty)$. We can then write by (3.4.9)

$$\frac{d\boldsymbol{P}(t)}{dt} = \boldsymbol{P}(t)\,\boldsymbol{A}$$

and, thus, by (3.4.4)

$$\frac{d\boldsymbol{p}(t)^\top}{dt} = \frac{d(\boldsymbol{p}(0)^\top\,\boldsymbol{P}(t))}{dt} = \boldsymbol{p}(0)^\top\,\boldsymbol{P}(t)\,\boldsymbol{A} = \boldsymbol{p}(t)^\top\,\boldsymbol{A}.$$

This yields for the probability vector $\boldsymbol{p}(t)$ the vector ODE

$$\frac{d\boldsymbol{p}(t)}{dt} = \boldsymbol{A}^\top \boldsymbol{p}(t) \tag{3.4.11}$$

for $t \in [0, \infty)$ with $\boldsymbol{p}(0) \in [0, 1]^N$.

Another consequence of the Markov property is that the *waiting times* of a homogeneous continuous time Markov chain, defined as the times between transitions from a given level $y_i$ to any other level, are exponentially distributed, see (1.2.4), with intensity parameter

$$\lambda_i = \sum_{j \neq i} a^{i,j} \tag{3.4.12}$$

for $i \in \{1, 2, \ldots, N\}$.

### An Interest Rate Example

Let us consider an interest rate example by assuming that the interest rate process $X = \{X_t,\, t \in [0, \infty)\}$ takes only the two values $y_1 = 0.05$ and $y_2 = 0.06$ with probabilities $(p_1(t), p_2(t))^\top = \boldsymbol{p}(t)$. Switching between these levels proceeds according to the time homogeneous transition matrix

$$\boldsymbol{P}(t) = \begin{bmatrix} \frac{1+e^{-10t}}{2} & \frac{1-e^{-10t}}{2} \\ \frac{1-e^{-10t}}{2} & \frac{1+e^{-10t}}{2} \end{bmatrix} \tag{3.4.13}$$

for $t \in [0, \infty)$. Using the definition (3.4.7) of the intensity matrix and L'Hôpital's rule, one obtains the intensity matrix in this case in the form

$$\boldsymbol{A} = \begin{bmatrix} -5 & 5 \\ 5 & -5 \end{bmatrix}. \tag{3.4.14}$$

For the special initial probability vector $\boldsymbol{p}(0) = \bar{\boldsymbol{p}} = (0.5, 0.5)^\top$, that is, the initial interest rate takes each of both possible levels with equal probability, we see that

$$\bar{\boldsymbol{p}}^\top \boldsymbol{P}(t) = \boldsymbol{p}(t)^\top = \bar{\boldsymbol{p}}^\top \tag{3.4.15}$$

for all $t \in [0, \infty)$. A probability vector $\bar{\boldsymbol{p}}$, which solves (3.4.15), is called a *stationary probability vector*. This means that the above interest rate example forms a stationary process with stationary probability vector $\bar{\boldsymbol{p}} = (0.5, 0.5)^\top$. Note that stationarity requires here exactly this choice for the initial probability vector.

In Fig. 3.4.1 we plot a typical path for this simple continuous time Markov chain. As previously mentioned, the waiting times between jumps are exponentially distributed with an intensity of $\lambda_1 = \lambda_2 = 5$ jumps per year, that is, with a mean waiting time of $\frac{1}{\lambda_1} = \frac{1}{\lambda_2} = \frac{1}{5} = 0.2$ years, see Table 1.3.1.

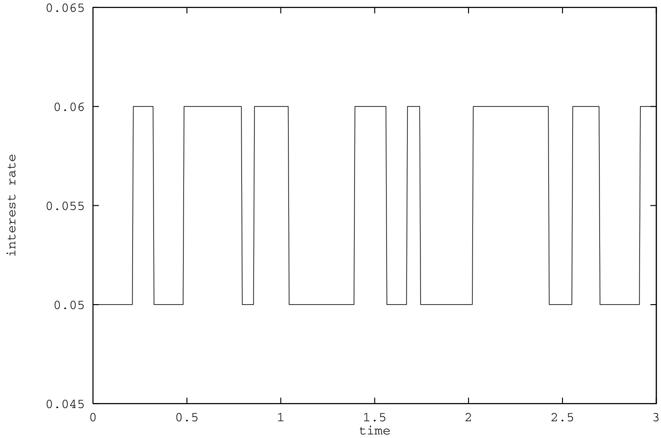



**Fig. 3.4.1.** Sample path of the continuous time Markov chain

## A Model for the Default of a Company

Let us mention also another example. A simple way of modeling the default risk of a company is to assume that the company might default according to a given intensity $a^{1,2}$, the default intensity. Let us denote by $X = \{X_t,\, t \in [0,\infty)\}$ the default process, which has only two possible values. The value $X_t = 1 = y_1$ describes the state when there has been no default up to time $t$. However, from the time of default $\tau$ onwards we set $X_t = 0 = y_2$ for all times $t \geq \tau$. This means, $y_2 = 0$ is the second possible state for $X$. One can say that $X_t$ is the indicator for the event that there is no default up until time $t$. Then $X$ can be shown to form a two-state homogeneous continuous time Markov chain with initial probability vector

$$\boldsymbol{p}(0) = (p_1(0), p_2(0))^\top = (1,0)^\top \tag{3.4.16}$$

and intensity matrix

$$\boldsymbol{A} = \begin{bmatrix} -a^{1,2} & a^{1,2} \\ 0 & 0 \end{bmatrix}. \tag{3.4.17}$$

This means, the state 0, that is default, is an absorbing state. Once it is reached, the process $X$ will never leave this state as follows from the zero intensity $a^{2,1} = 0$ in (3.4.17). According to the Kolmogorov forward equation (3.4.9) we obtain

$$\frac{dp^{1,1}}{dt}(t) = -p^{1,1}(t)\, a^{1,2}. \tag{3.4.18}$$

That is, we obtain from (3.4.17) and (3.4.18) the explicit transition probability for not having a default until time $t$ in the form

$$p^{1,1}(t) = \exp\{-a^{1,2}\, t\}. \tag{3.4.19}$$

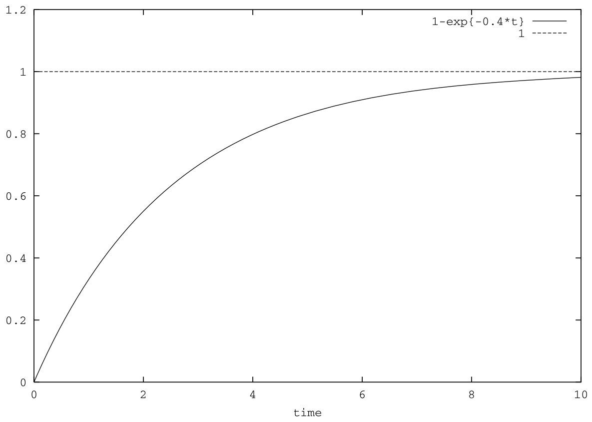



**Fig. 3.4.2.** Default probability for $a^{1,2} = 0.4$

Thus, the probability for a default until time $t$ is

$$p^{1,2}(t) = 1 - p^{1,1}(t) = 1 - \exp\{-a^{1,2}\, t\} \tag{3.4.20}$$

for $t \in [0, \infty)$. Fig. 3.4.2 displays the probability of default for a low rated company with default intensity $a^{1,2} = 0.4$ over a period of 10 years. One notes that in this case the probability of default is rather high after several years.

### Changes in the Credit Rating of a Firm

Rating agencies publish, at regular intervals, ratings for the credit worthiness of a firm. These ratings are important for the pricing of debt securities. The credit rating usually changes over time according to the performance and management of the firm. These changes can be interpreted as jumps that occur with certain intensities. Again, a homogeneous continuous time Markov chain can be used to model these changes. As a simple example let us consider the following four categories for the possible credit ratings of a firm, $y_1 = AA$, $y_2 = B$, $y_3 = C$, $y_4 = D$ with $D$ denoting default.

The intensity matrix $\boldsymbol{A}$, see (3.4.7), of the Markov chain could be, for instance, of the form given in Table 3.4.1. Note that the elements of each row add up to 0 and $y_4 = D$ is obviously the absorbing state. Assuming an initial probability vector $\boldsymbol{p}(0)$ one can apply the Kolmogorov forward equation (3.4.9) to obtain the transition probabilities for this Markov chain.

Figure 3.4.3 displays a typical sample path for this Markov chain with the levels 3, 2, 1 and 0 corresponding to the ratings $AA$, $B$, $C$ and $D$, respectively. Note that after about 17 years the rating process reaches in this figure the state 0, that is the state $D$, which corresponds to default.

**Table 3.4.1.** Credit rating transition matrix

| | $AA$ | $B$ | $C$ | $D$ |
|---|---|---|---|---|
| $AA$ | -0.06 | 0.03 | 0.02 | 0.01 |
| $B$ | 0.1 | -0.4 | 0.2 | 0.1 |
| $C$ | 0.2 | 0.4 | -1.0 | 0.4 |
| $D$ | 0 | 0 | 0 | 0 |

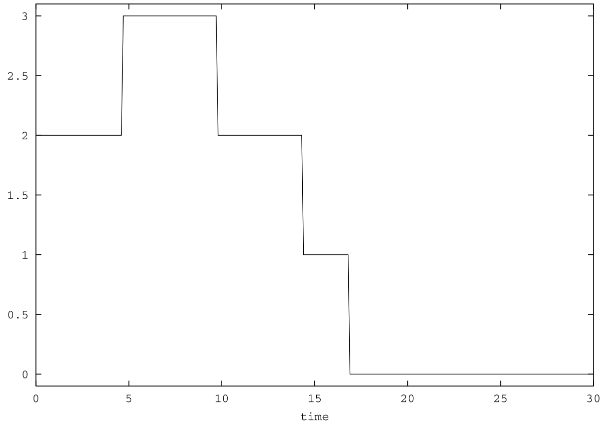



**Fig. 3.4.3.** Credit ratings for a firm

## Ergodicity of Continuous Time Markov Chains (*)

A stochastic process $X$ that describes some equilibrium often enjoys a powerful and important property called *ergodicity*, which we mention briefly. For a continuous time Markov chain $X$, a time $T \in (0, \infty)$ and any bounded, measurable function $f : \mathcal{X} \rightarrow \Re$ we can define the corresponding *time average* over the interval $(0, T]$ as

$$A_f(T) = \frac{1}{T} \int_0^T f(X_t)\, dt. \tag{3.4.21}$$

Such time averages are random and often refer to important quantities in economic or financial applications. For instance, certain statistics describe the average growth rate, average interest rate or the average inflation rate over longer time periods. An important theoretical and practical question is: What happens with these averages if the length $T$ of the observation interval $(0, T]$ becomes large, that is as $T \rightarrow \infty$? This question can be answered for those stochastic processes that satisfy the following ergodicity property:

A continuous time Markov chain $X$ taking values in the set $\mathcal{X} = \{y_1, y_2, \ldots, y_N\}$ is called *ergodic* if

$$\lim_{T\to\infty} A_f(T) = \lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_t)\, dt = \sum_{i=1}^{N} f(y_i)\, \bar{p}_i \tag{3.4.22}$$

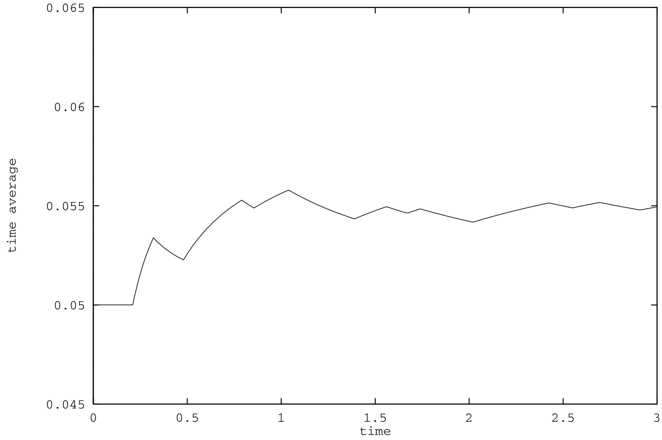



**Fig. 3.4.4.** Time average of an ergodic interest rate process

a.s. for all bounded, measurable functions $f : \mathcal{X} \rightarrow \Re$. Here $\bar{\boldsymbol{p}}$ denotes the stationary probability vector, which satisfies the relation

$$\bar{\boldsymbol{p}}\,\boldsymbol{P}(t) = \bar{\boldsymbol{p}} \tag{3.4.23}$$

for all $t \in [0, \infty)$. That is, the random time average $A_f(T)$ converges as $T \rightarrow \infty$, towards the spatial average appearing on the right hand side of equation (3.4.22). This spatial average on the right hand side of equation (3.4.22) is the expectation of a random variable $f(X)$, where $X$ has as probabilities those of the stationary probability vector $\bar{\boldsymbol{p}}$.

Let us show for $T \rightarrow \infty$ a typical convergence pattern of the time average (3.4.22) for the interest rate example that was given in Fig. 3.4.1. Figure 3.4.4 displays a corresponding trajectory of the time average $A_f(T)$ for the mean, that is $f(y) = y$, for a three year time interval. We see that the long-term time average converges to the stationary mean of the interest rate process, which has according to (3.4.22), the theoretical value $\mu = 0.055$.

## 3.5 Poisson Processes

The Wiener process is the fundamental mathematical object to model continuous uncertainty with independent increments, see Sect. 3.2. As we have discussed already in previous sections, we need to be able to model in finance various random events, such as defaults, operational failures, insurance claims and trading times.

A stochastic process with independent increments, see Sect. 3.2, that counts events can be used to model event driven uncertainty. We call this fundamental process a *Poisson process*.

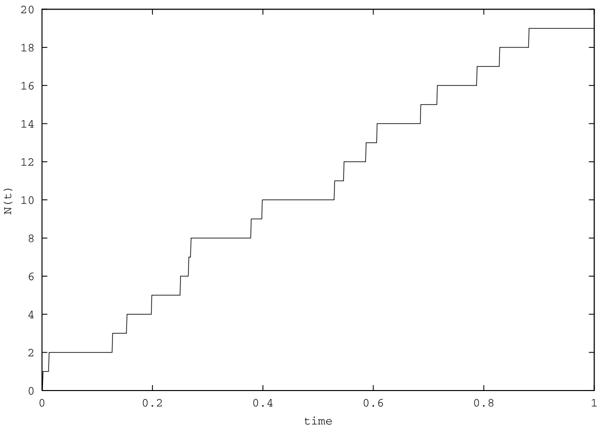



**Fig. 3.5.1.** Standard Poisson process with intensity $\lambda = 20$

## Standard Poisson Process

**Definition 3.5.1.** *A* standard Poisson process $N = \{N_t,\, t \in [0, \infty)\}$ *with* intensity $\lambda > 0$ *is a process with stationary independent increments with initial value* $N_0 = 0$ *such that* $N_t - N_s$ *is Poisson distributed with intensity* $\lambda(t-s)$*, see* (1.1.30)*, that is with probability*

$$P(N_t - N_s = k) = \frac{e^{-\lambda\,(t-s)}\,(\lambda\,(t-s))^k}{k!} \tag{3.5.1}$$

*for* $k \in \{0, 1, \ldots\}$*,* $t \in [0, \infty)$ *and* $s \in [0, t]$*.*

For the Poisson process $N$ with intensity $\lambda$ we have the mean

$$\mu(t) = E(N_t) = \lambda\, t, \tag{3.5.2}$$

see (1.3.3), and the variance

$$v(t) = \mathrm{Var}(N_t) = E\left((N_t - \mu(t))^2\right) = \lambda\, t, \tag{3.5.3}$$

see (1.3.16), for $t \in [0, \infty)$.

In Fig. 3.5.1 we plot a graph for a standard Poisson process with intensity $\lambda = 20$. That means, according to (3.5.2), we should expect on average 20 events to happen during the time period $[0, 1]$, which is almost the case for this trajectory.

A standard Poisson process $N$ is a counting process. It generates an increasing sequence of *jump times* $\tau_1, \tau_2, \ldots$ with each event that it counts. Thus, $N_t$ equals the number of events that occurred up until time $t \in [0, \infty)$. Here for $t \in [0, \infty)$ the time $\tau_{N_t}$ denotes the last time that $N$ made a jump such that

$$\tau_k = \inf\{\, t \in [0, \infty) :\, N_t \geq k\} \tag{3.5.4}$$

for $k \in \mathcal{N}$.

### Transformed Poisson Process

In some applications, as in the modeling of defaults, the intensity $\lambda(t)$ that a certain type of event occurs may dependent on the time $t \in [0, \infty)$. This leads to a time *transformed Poisson process* $N = \{N_t,\, t \in [0, \infty)\}$, where

$$P(N_t - N_s = k) = \frac{\exp\left\{-\int_s^t \lambda(z)\, dz\right\} \left(\int_s^t \lambda(z)\, dz\right)^k}{k!} \tag{3.5.5}$$

for $k \in \{0, 1, \ldots\},\ t \in [0, \infty)$ and $s \in [0, t]$.

Additionally, the *mark* $\xi_k$ of the $k$th event, for instance, the *recovery rate* of the $k$th default, may also depend on the number of events that occurred. For the moment let us assume that $\xi_k$ is deterministic, $k \in \{1, 2, \ldots\}$. We can then consider the transformed Poisson process $Y = \{Y_t,\, t \in [0, \infty)\}$ with

$$Y_t = \sum_{k=1}^{N_t} \xi_k \tag{3.5.6}$$

for $t \in [0, \infty)$. If the intensity process $\lambda = \{\lambda(t),\, t \in [0, \infty)\}$ is deterministic and the $k$th mark $\xi_k$ is deterministic for each $k \in \{1, 2, \ldots\}$, then it follows from (3.5.5) that the mean $\mu(t)$ of $Y_t$ is given by the expression

$$\mu(t) = E(Y_t) = \sum_{k=1}^{\infty} \xi_k\, P(N_t = k), \tag{3.5.7}$$

where the above probabilities are expressed in (3.5.5). In this case $Y$ is a process with independent increments. If one chooses $\xi_1 = 1$ and $\xi_k = 0$ for $k \in \{2, 3, \ldots\}$, then this allows us to model some kind of credit worthiness $C(t)$ of a company at time $t$ by setting

$$C(t) = 1 - Y_t.$$

The credit worthiness may start at time zero with $C(0) = 1$ and declines then to zero at the time when the first default arises. In this simple example the expected value $E(C(t))$ of credit worthiness at time $t$ equals by (3.5.7) and (3.5.5) the probability

$$E(Y_t) = P(N_t = 0) = \exp\left\{-\int_0^t \lambda(z)\, dz\right\}$$

for $t \in [0, \infty)$.

### Compound Poisson Process

It is often important to differentiate in a sequence of events various types of events, for instance, the different recovery rates of a default. One also needs to

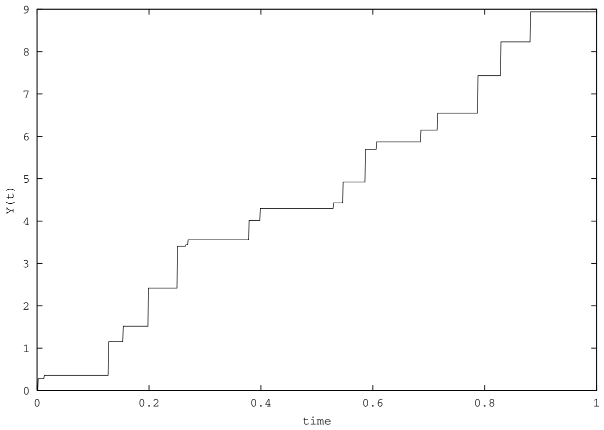



**Fig. 3.5.2.** Compound Poisson process

allocate a corresponding intensity to the occurrence of each type of possible event. Changes in credit ratings could be interpreted as such types of events. Similarly, one could also model operational failures of different kinds, which may occur in a company or institution, as particular types of events. Another example is the varying severity of certain damages in insurance claims.

To construct a process that models sequences of different types of events, let us consider a Poisson process $N$ with intensity $\lambda > 0$, together with a sequence of i.i.d. random variables $\xi_1, \xi_2, \ldots$ that are independent of $N$. Here

$$P(\xi_1 \leq z) = F_{\xi_1}(z) \tag{3.5.8}$$

denotes for $z \in \Re$ the given distribution function of $\xi_1$. We now construct the *compound Poisson process* $Y = \{Y_t,\, t \in [0, \infty)\}$, where $Y_0 = 0$ and

$$Y_t = \sum_{k=1}^{N_t} \xi_k \tag{3.5.9}$$

for $t \in [0, \infty)$. A compound Poisson process generates a sequence of pairs $(\tau_k, \xi_k)_{k \in \mathcal{N}}$ of *jump times* $\tau_k$ and *marks* $\xi_k$. In Fig. 3.5.2 we show the trajectory of a compound Poisson process $Y$ where the i.i.d. random variables are uniformly $U(0, 1)$ distributed, see (1.2.3), and $N$ is as in Fig. 3.5.1.

In insurance, a simple version of the *Cramér-Lundberg model* for the *risk reserve* or *surplus* $X_t$ at time $t$ of an insurance company can be described in the form

$$X_t = X_0 + c\,t - Y_t, \tag{3.5.10}$$

where the *claim process* $Y = \{Y_t,\, t \in [0, \infty)\}$ is a compound Poisson process with i.i.d. claim sizes $\xi_k > 0$, $k \in \mathcal{N}$. Here $c$ is the *premium rate* that describes the premium payments per unit of time that the insurance company collects. We show in Fig. 3.5.3 the trajectory of a risk reserve process $X$ when setting $c = 10$ and $X_0 = 100$.

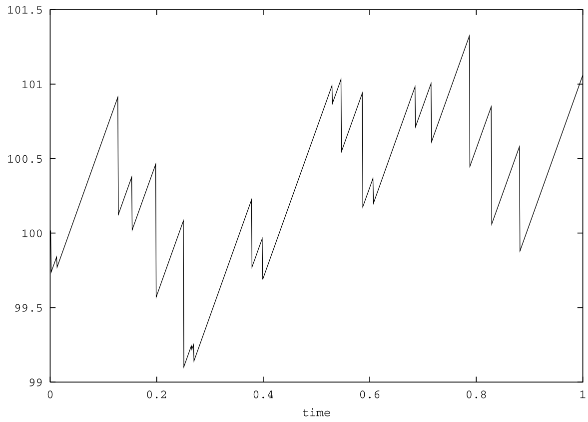



**Fig. 3.5.3.** Risk reserve process

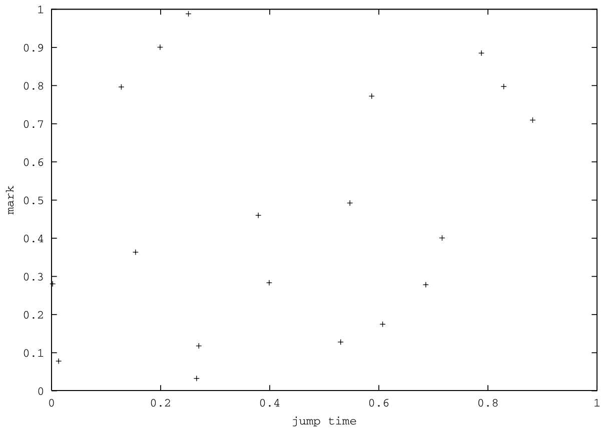



**Fig. 3.5.4.** Pairs of jump times and marks

In Fig. 3.5.4 we plot the points for the pairs of jump times and marks of the path of the compound Poisson process $Y$ shown in Fig. 3.5.2. These correspond also to the risk reserve process $X$ in Fig. 3.5.3.

A compound Poisson process is an example of a so-called *marked point process*. Its trajectory is fully characterized by the sequence of pairs $(\tau_k, \xi_k)_{k \in \mathcal{N}}$. As already mentioned, one could think of accumulated insurance claims that are presented to an insurance company. Here the marks are the claim sizes and the jump times the instances when claims arise. In Sect. 3.7 we shall present some mathematical results for the Cramér-Lundberg model that allow to calculate quantities of interest in insurance or operational risk modeling.

## Poisson Measure (*)

In the remainder of this section we list briefly some notions and results that prepare the ground for introducing general event driven processes that become increasingly common in quantitative finance. To deal properly with the modeling of many events arising with high intensity the notion of a *random measure* is needed. For this purpose let us extend our example of a compound Poisson process.

It is a fundamental feature of the Poisson process, given in Definition 3.5.1, that due to the independence of its increments the location of the set of points in the time interval $[0, 1]$, see Fig. 3.5.4, is such as if one has generated $N_1$ independent uniformly $U(0, 1)$ distributed random variables. On the other hand, the marks $\xi_k$ of the compound Poisson process $Y$ used in Fig. 3.5.4 are also uniformly distributed on $[0, 1]$, that is $\xi_k \in U(0, 1)$. Consequently, the pairs $(\tau_k, \xi_k)$, for $k \in \mathcal{N}$, which are generated by the compound Poisson process $Y$ during the time period $[0, 1]$, shown in Fig. 3.5.4, are uniformly distributed in the square $[0, 1] \times [0, 1]$. Such a graph can be interpreted as the trajectory of some *Poisson measure*.

The notion of a Poisson measure will appear to be quite technical. However, this notion is essential if one aims to consider the important class of stochastic processes that will be introduced in the next section.

We now introduce the *mark set*

$$\mathcal{E} = \Re \backslash \{0\}. \tag{3.5.11}$$

Here the element $\{0\}$ is excluded to avoid in some modeling applications jumps of size zero. Let $\mathcal{B}(\Gamma)$ denote the smallest sigma-algebra containing all open sets of a set $\Gamma$. Now, we construct a *Poisson measure* $p_\varphi(dv \times dt)$ on $\mathcal{E} \times [0, \infty)$ with *intensity measure*

$$\nu_\varphi(dv \times dt) = \varphi(dv)\, dt. \tag{3.5.12}$$

Here $\varphi(\cdot)$ is a measure on $\mathcal{B}(\mathcal{E})$ with

$$\int_{\mathcal{E}} \min(1, v^2)\, \varphi(dv) < \infty. \tag{3.5.13}$$

The random Poisson measure $p_\varphi(\cdot)$ is assumed to be such that for each set $A$ from the product-sigma-algebra of $\mathcal{B}(\mathcal{E})$ and $\mathcal{B}([0, T])$, see Protter (2004), the random variable $p_\varphi(A)$, which counts the number of points in $A$, is Poisson distributed with intensity

$$\nu_\varphi(A) = \int_0^T \int_{\mathcal{E}} \mathbf{1}_{\{(v,t) \in A\}}\, \varphi(dv)\, dt, \tag{3.5.14}$$

that is, one has

$$P(p_\varphi(A) = \ell) = \frac{\nu_\varphi(A)^\ell}{\ell\,!}\, e^{-\nu_\varphi(A)} \tag{3.5.15}$$

for $\ell \in \{0, 1, \ldots\}$ and each $T \in [0, \infty)$, see (1.1.30). For disjoint sets $A_1, \ldots, A_r$, $r \in \mathcal{N}$, the random variables $p_\varphi(A_1), \ldots, p_\varphi(A_r)$ are here assumed to be independent.

For example, the points in Fig. 3.5.4 can be interpreted as a realization of a Poisson measure on $[0,1] \times [0,1]$ with $\varphi(dv) = \lambda dv$, where $\frac{\varphi(dv)}{dv} = \lambda = 20$ and $T = 1$. We shall use throughout the following the notations $p(dv, dt) = p_\varphi(dv, dt) = p_\varphi(dv \times dt)$ for a Poisson measure and $\nu_\varphi(dv, dt) = \nu_\varphi(dv \times dt)$ for the corresponding intensity measure, whenever it is convenient.

## 3.6 Lévy Processes (*)

We now introduce the class of Lévy processes, which has been used by a number of researchers in financial modeling, see, for instance, Barndorff-Nielsen & Shephard (2001), Geman et al. (2001) and Eberlein (2002). It includes as special cases, for instance, the Poisson process and the Wiener process. Lévy processes are processes with stationary independent increments which enjoy a number of elegant mathematical properties. In the market microstructure of tick by tick intraday data the piecewise constant path of an intraday stock price resembles that of paths of certain Lévy processes. The following definition refers to Sect. 3.2 and the concept (2.7.1) of convergence in probability.

**Definition 3.6.1.** *A stochastic process* $X = \{X_t,\, t \in [0, \infty)\}$ *with* $X_0 = 0$ *a.s. is called a* Lévy process *if*

1. $X$ *is a process with* independent increments, *where* $X_t - X_s$ *for* $0 \le s < t < \infty$ *is independent of the past, that is the* $X_r$ *with* $0 \le r \le s$*;*
2. $X$ *has* stationary increments, *which means that* $X_t - X_s$ *has for* $0 \le s < t < \infty$ *the same distribution as* $X_{t-s}$*;*
3. $X$ *is* continuous in probability, *that is* $X_s \stackrel{P}{=} \lim_{t \to s} X_t$ *for* $s \in [0, \infty)$.

This definition generalizes that of the Wiener process, see Definition 3.2.2, and also that of the standard Poisson process, see Definition 3.5.1.

A popular asset price model is that of an *exponential Lévy model* $S = \{S_t,\, t \in [0, \infty)\}$, where

$$S_t = S_0 \exp\{X_t\} \tag{3.6.1}$$

for $t \in [0, \infty)$. This kind of model has been studied, for instance, in Madan & Seneta (1990) and Eberlein & Keller (1995) among others. By taking the exponential in (3.6.1) the asset price is guaranteed to stay positive.

### Lévy Decomposition (*)

Let $X$ be a Lévy process. Then it can be shown that $X$ has a decomposition of the form

$$X_t = \alpha\, t + \beta\, W_t + \int_0^t \int_{|v|<1} v\, (p_\varphi(dv, ds) - \varphi(dv)\, ds) + \int_0^t \int_{|v| \ge 1} v\, p_\varphi(dv, ds) \tag{3.6.2}$$

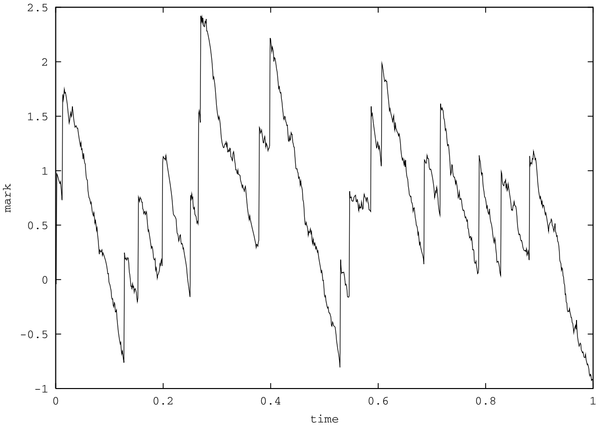



**Fig. 3.6.1.** A linearly interpolated sample path of a Lévy process

for $t \in [0, \infty)$, see, for instance, Protter (2004). Here $W = \{W_t, \, t \in [0, \infty)\}$ is a standard Wiener process. Furthermore, $p_\varphi$ is a *Poisson measure* as defined at the end of the previous section. Therefore, for any set $A \in \mathcal{E}$ the process $p_\varphi(A) = \{p_\varphi(A, [0, t]), \, t \in [0, \infty)\}$ is a Poisson process independent of $W$ with intensity $\varphi(A)$, where $\varphi(dv)$ is called the *Lévy measure*. It is defined on $\mathcal{E}$ assuming that (3.5.13) holds. Note that for disjoint sets $A, B \in \mathcal{E}$ the corresponding Poisson processes $p_\varphi(A)$ and $p_\varphi(B)$ are independent. Equation (3.6.2) shows that a Lévy process is a superposition of a constant trend process, a multiple of a Wiener process and some generalized marked point process.

A Lévy process is fully characterized by the parameters $\alpha, \beta \in \Re$ together with the Lévy measure $\varphi$. In the case $\beta = 0$ and $\varphi(A) = 0$ for all $A \in \mathcal{E}$, the Lévy process value $X_t$ at time $t$ equals the linear trend function $X_t = \alpha \, t$. If $\varphi(A) = 0$ for all $A \in \mathcal{E}$, then the Lévy process is a transformed Wiener process of the form $X_t = \alpha \, t + \beta \, W_t$. This is the only possible form that a continuous Lévy process can have. As an example, for $\alpha = \beta = 0$ and $\varphi(\mathcal{E}) = \varphi(\{1\}) \in (0, \infty)$ the Lévy process $X = p_\varphi(\{1\}) = \{p_\varphi(\{1\}, [0, t]), \, t \in [0, \infty)\}$ is a Poisson process with intensity $\varphi(\{1\})$. In Fig. 3.6.1 the path of a Lévy process is shown for the case $\alpha = 0$, $\beta = 1$ and

$$\varphi(dv) = \begin{cases} \lambda \, dv \text{ for } & v \in (0, 1) \\ 0 & \text{otherwise}, \end{cases}$$

where $\lambda = 20$. Note that we used in this example the jump times and marks of the compound Poisson process exhibited in Figs. 3.5.2 and 3.5.4. This means that a compound Poisson process is a Lévy process. A compound Poisson process assumes finite intensity for the underlying Poisson process that generates the events. A Lévy process can also model the case where certain extremely small jumps may arise with unlimited intensity. This is potentially the case for price processes observed from tick by tick data if one tries to avoid a Wiener process for modeling. Such prices can be modeled via the use of a Lévy measure. A simple asset price model is obtained by the exponential Lévy process $Y = \{Y_t,\, t \in [0, \infty)\}$ with

$$Y_t = Y_0 \exp\{X_t\}, \tag{3.6.3}$$

where $Y_0 > 0$ and $X_t$ is the value of a Lévy process at time $t \in [0, \infty)$, see (3.6.1). Such exponential Lévy processes have been used in asset price modeling, for instance, in Kou (2002).

Lévy processes have a range of elegant mathematical properties, which make these processes theoretically attractable for financial modeling. A disadvantage of Lévy processes arises from the fact that the Lévy measure is difficult to estimate in practice. Furthermore, financial models that are based on Lévy processes can become mathematically very challenging when used for derivative pricing and portfolio optimization.

Special Lévy processes, as the Wiener process, Poisson process and compound Poisson process, seem to be often sufficient as building blocks for sophisticated models in many areas of finance and insurance. These particular Lévy processes often allow us to construct fast and accurate computational methods. In some cases they provide analytic solutions.

It is advisable to model in finance important single events that arise with finite intensity together with continuous noise, which aggregates most of the many small events in a compact manner. This way of modeling reduces strongly the complexity of a model and, potentially, secures its tractability.

## 3.7 Insurance Risk Modeling (*)

In this section we examine actuarial quantities which are important to insurers under the classical insurance risk model. Complex quantities and functionals can be accessed under the given framework. It is worthwhile mentioning that the methods and techniques employed in the following can also be applied to quantify some types of *operational risk* and *credit risk*.

### Classical Insurance Risk Model (*)

Let $(\Omega, \mathcal{A}, P)$ be a probability space underlying all mathematical objects that will be introduced in the following. We study the classical model of insurance risk theory, the Cramér-Lundberg model, see (3.5.10). Here $X_a = u \geq 0$ is the insurer's *initial capital*. The *premiums* are received continuously at a constant rate $c \geq 0$ per unit time. The aggregate claims constitute a compound Poisson process $Y = \{Y_t,\, t \in [0, \infty)\}$ characterized by the Poisson intensity parameter $\lambda > 0$ and the individual *claim amount distribution* function $F_\xi(z)$, $z \in (0, \infty)$ with $F_\xi(0) = 0$ and mean value $\mu$. That is,

$$Y_t = \sum_{k=1}^{N_t} \xi_k \tag{3.7.1}$$

for $t \in [0, \infty)$, where $N = \{N_t,\, t \in [0, \infty)\}$ is a Poisson process with intensity $\lambda$. The claim amounts $(\xi_k)_{k \ge 1}$ are independent, identically distributed random variables with common distribution function $F_\xi$. Then for $t \in [0, \infty)$, the *surplus* or *risk reserve* at time $t$ for the insurer is given in the form

$$X_t = u + c\,t - Y_t, \tag{3.7.2}$$

which represents according to (3.5.10) the *Cramér-Lundberg model*, see Gerber (1979) or Grandell (1991).

For simplicity we assume that the distribution function $F_\xi$ is differentiable with $f_\xi(z) = \frac{dF_\xi(z)}{dz}$ being the individual *claim amount probability density*. Consequently, for every $t \in [0, \infty)$ and $y > 0$ the compound Poisson process value $Y_t$ has the probability density

$$f_{Y_t}(y) = e^{-\lambda t} \sum_{n=0}^{\infty} \frac{(\lambda t)^n}{n\,!}\, f_\xi^{n*}(y), \tag{3.7.3}$$

where $f^{n*}(y)$ denotes the $n$-fold convolution of the claim amount probability density $f_\xi(y)$ with itself. By convention, we define $f_\xi^{0*}(y) = 1$. Similarly, we shall use "*" more generally to denote the convolution operation in the following. Throughout the section we assume that $N = \{N_t,\, t \in [0, \infty)\}$ and $(\xi_k)_{k \ge 1}$ are independent and that

$$c > \lambda\,\mu. \tag{3.7.4}$$

Under this assumption it follows for $X = \{X_t,\, t \in [0, \infty)\}$ from the strong Law of Large Numbers, see (2.1.13), that

$$P\left(\lim_{t \to \infty} X_t = \infty\right) = 1. \tag{3.7.5}$$

In our modeling we allow the surplus process $X$ still to continue even if the surplus reaches a negative value. A typical realization of the surplus or risk reserve process is given in Fig. 3.5.3.

Let $\tau$ denote the *time of ruin* and $\varrho$ be the time of the surplus process $X = \{X_t,\, t \in [0, \infty)\}$ leaving zero ultimately, which is called the *ultimate leaving-time*, that is,

$$\tau = \inf\{t \ge 0 :\, X_t < 0\}, \tag{3.7.6}$$

where $\tau = \infty$ if the set is empty, and

$$\varrho = \sup\{t \ge 0 :\, X_t < 0\} \tag{3.7.7}$$

with $\varrho = 0$ if the set is empty.

Let us abbreviate the phrase "there exists" by $\exists$. For $X_0 = u \geq 0$ the quantity

$$\Psi(u) = P\left(\exists t > 0, X_t < 0 \,\middle|\, X_0 = u\right) = P\left(\tau < \infty \,\middle|\, X_0 = u\right) \tag{3.7.8}$$

denotes the *probability of ruin* with initial capital $u$. Then

$$\Phi(u) = 1 - \Psi(u) \tag{3.7.9}$$

is the *non-ruin probability*, also called *survival probability*. It follows from Grandell (1991) that

$$\Phi(0) = \frac{c - \lambda\,\mu}{c} \quad \text{and} \quad \Psi(0) = \frac{\lambda\,\mu}{c}. \tag{3.7.10}$$

### Actuarial Diagnostics (*)

In the given classical insurance risk model it is possible to calculate a number of important quantities, which are called *actuarial diagnostics*. Let us consider the random variable $X_{\tau-}$, which is the *surplus immediately prior to ruin*. Furthermore, $|X_\tau|$ denotes the *deficit at ruin*. We denote by $\sup_{0 \leq t < \tau} X_t$ the *maximum profit before ruin*. Similarly, $\inf_{0 \leq t < \tau} X_t$ is the *minimum profit before ruin*. Furthermore, $\sup_{0 \leq t < \varrho} X_t$ denotes the *maximum profit after ruin*. Finally, $\inf_{0 \leq t < \varrho} X_t$ is the *maximum loss after ruin and before the ultimate leaving time*. For $x > u = X_0$, let $\pi(u; x, t)$ denote the probability density that the surplus process crosses upward the level $x$ at $t$ for the first time. According to Gerber (1979) it follows that

$$\pi(u; x, t) = \pi(u; x - u, t) = \frac{x - u}{t}\, f_{X_t}\,(u + c\,t - x). \tag{3.7.11}$$

Let

$$\tau_1 = \inf\{t \geq 0 :\, X_t < 0\}$$

with $\tau_1 = \infty$ if the set is empty, and

$$\tau_1^0 = \inf\{t > 0 :\, X_t = 0\},$$

where $\tau_1^0 = \infty$ if the set is empty. Obviously, we have $\tau_1 = \tau$. In general, for $k \geq 2$, define recursively

$$\tau_k = \inf\{t > \tau_{k-1}^0 :\, X_t < 0\}$$

with $\tau_k = \infty$ if the set is empty, and

$$\tau_k^0 = \inf\{t > \tau_{k-1}^0 :\, X_t = 0\},$$

where $\tau_k^0 = \infty$ if the set is empty. We consider

$$\kappa = \sup\{k : \tau_k^0 < \infty\}$$

with $\kappa = 0$ if the set is empty, as the total number of zeros of the surplus process $X = \{X_t,\, t \in [0,\infty)\}$. It follows from (3.7.5) that $P(\kappa < \infty) = 1$. Furthermore, we have $\tau_\kappa^0 < \infty$, $\tau_{\kappa+1}^0 = \infty$ when $\kappa > 0$ and $\varrho = \tau_\kappa^0$ when $\varrho > 0$. With any initial value $u \in \Re$ we have

$$P\left(\tau_k < \tau_k^0 \,\middle|\, \tau_k < \infty\right) = 1,$$

for all $k \geq 1$.

In insurance risk analysis, there is an increasing interest in the distributions of functionals related to the surplus process for providing information about an insurer's risk. These include the distribution of the surplus immediately prior to ruin, the distribution of the deficit at ruin and the distribution of the maximum profits before ruin. References on this topic include Dickson (1992, 1993), Gerber & Shiu (1997) and Picard (1994). Using techniques described, for instance, in Wei & Wu (2002) one can derive explicit expressions for the joint distributions of almost all of these actuarial diagnostics and can give the exact results when the individual claim amounts are exponentially distributed. As already indicated, these techniques appear to have also relevance for operational risk and credit risk evaluations.

## 3.8 Exercises for Chapter 3

**3.1.** Show that a standard Wiener process $W$ has covariance $C(s,t) = \min\{s,t\}$.

**3.2.** Is the Wiener process stationary?

**3.3.** Derive the probabilities for the symmetric random walk.

**3.4.** Compute the probability for having $j \in \{0, 1, \ldots, n\}$ upward moves in a *non-symmetric* random walk after $n$ time steps with probability $p \in (0,1)$ for an upward move.

**3.5.** Determine the mean and variance for the stationary continuous time Markov chain interest rate example.

**3.6.** What is the long term time average of the squared interest rate in the interest rate example?

**3.7.** For a Poisson process with intensity $\lambda > 0$ derive a formula for its second moment at time $t > 0$.

**3.8.** Compute the first moment for a compound Poisson process with intensity $\lambda > 0$ and $U(0,1)$ distributed i.i.d. marks.

**3.9.** What is the probability for a compound Poisson process with intensity $\lambda > 0$ and $U(0,1)$ distributed i.i.d. marks to have no jump until time $t > 0$?

**3.10.** (*) For a Lévy process with $\alpha = 1$, $\beta = 1$ and

$$\varphi(dv) = \begin{cases} \lambda \text{ for } \quad v = \frac{1}{2} \\ 0 \text{ otherwise} \end{cases}$$

with $\lambda > 0$ calculate the mean and the variance.

# 4

# Diffusion Processes

In this chapter diffusion processes are introduced. These are potential candidates for the modeling of asset prices, interest rates and other financial quantities. We cover examples on geometric Brownian motion, Ornstein-Uhlenbeck and square root processes.

## 4.1 Continuous Markov Processes

A Markov process that evolves in continuous time and has continuous trajectories is called a *continuous Markov process*. This type of process would appear to be well suited for the modeling of a range of financial quantities such as stock prices, exchange rates and interest rates. Unlike Markov chains, that have discontinuous paths, it allows us to model continuous random movements of stock prices. The typical trajectory of a transformed Wiener process, as given in Fig. 3.2.3, would seem to be a reasonable candidate for the representation of asset price dynamics, for example, the path of the S&P500 index that was displayed in Fig. 3.1.1.

The Wiener process evolves in continuous time and has continuous trajectories. That is, it has paths without any jumps. Since it has independent increments it is also a Markov process. However, the transformed Wiener process given in Sect. 3.2 can take negative values. To see this better we plot in Fig. 4.1.1 the Gaussian transition densities for the standard Wiener process for the time interval $[0.1, 3.0]$. The figure shows that for negative values to be obtained there is a positive probability. This observation also applies to a transformed Wiener process. It indicates that the Wiener process or a transformed Wiener process would not be suitable for the modeling of asset price dynamics.

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

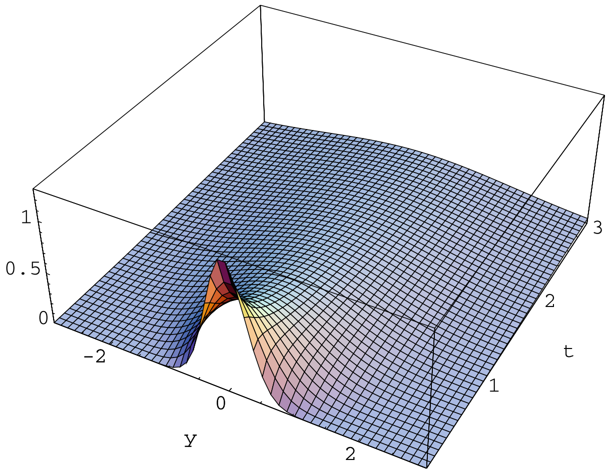



**Fig. 4.1.1.** Probability densities for the standard Wiener process

### Black-Scholes Model

It is intuitively appealing to assume that asset prices can be modeled using some positive process which changes its value proportionally to its current value. On the basis of this assumption it makes sense to exponentially transform the Wiener process $W$ to ensure positive asset price values. That is, we consider the random variable

$$X_t = \exp\{g\,t + b\,W_t\} \tag{4.1.1}$$

for $t \in [0, \infty)$. Here $g$ denotes the *growth rate* and $b$ is known as the *volatility* of the asset price process $X$. In Samuelson (1955, 1965a) this model was suggested for asset prices. Later, it was used in Merton (1973b) and Black & Scholes (1973) as a stock price model in their Nobel prize winning work on option pricing. The stochastic process given in (4.1.1) is called *geometric Brownian motion*. The corresponding asset price model is the *lognormal* or *Black-Scholes model*.

In Fig. 4.1.2 we show a path for geometric Brownian motion over a period of ten years with growth rate $g = 0.05$ and volatility $b = 0.2$. Note that the fluctuations become larger for larger values of the asset price.

To have flexibility in using different initial values in the lognormal model we define *geometric Brownian motion* more generally by the expression

$$X_t = X_{t_0}\,\exp\{g\,(t - t_0) + b\,(W_t - W_{t_0})\} \tag{4.1.2}$$

for $t \in [t_0, \infty)$ with initial asset price $X_{t_0} > 0$, growth rate $g$ and volatility $b$, where $W$ denotes a standard Wiener process.

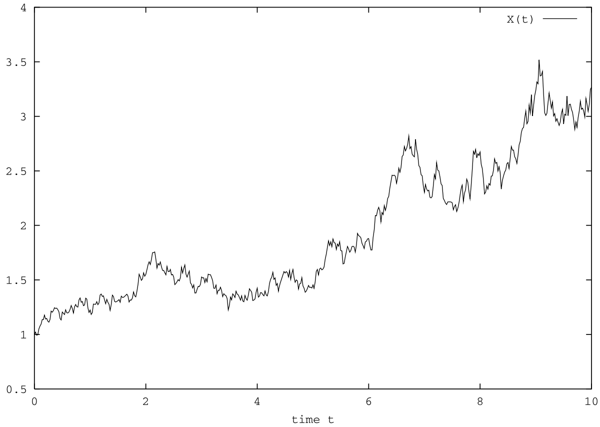



**Fig. 4.1.2.** A path of geometric Brownian motion

## Markov Property

Geometric Brownian motion is an example of a *continuous Markov process*, a class of stochastic processes widely used for asset price modeling. Let us suppose that the share price of a stock is at present \$1 and follows a continuous Markov process. Then it is reasonable to assume that predictions of future stock price values should only depend on the present share price and be unaffected by the price one year, one month or one week ago. The only relevant information is that the price at present is \$1. Any predictions of future prices are uncertain, however, they can be expressed in terms of a probability distribution. The Markov property then implies that the probability distribution of the stock price at a particular future time depends only on the current stock price. This simplifies considerably the modeling, statistical inference and numerical analysis that typically arise.

The Markov property has a natural economic interpretation in the modeling of asset prices: The present price of a stock encapsulates all of the information contained in the knowledge of past prices. This does not exclude the possibility of using certain statistical properties of the stock price history to determine, that is calibrate, model parameters, for instance, the growth rate or the volatility of the lognormal model.

In what follows we shall suppose that for $k \in \{0, 1, \ldots\}$ every joint distribution $F_{X_{t_0}, X_{t_1}, \ldots, X_{t_k}}(x_0, x_1, \ldots, x_k)$ of the process $X = \{X_t,\ t \in [0, \infty)\}$ under consideration has a density $p(t_0, x_0;\ t_1, x_1;\ \ldots;\ t_k, x_k)$. This allows us to define the conditional probability distribution in the form

$$P\left(X_{t_{n+1}} < x_{n+1} \,|\, X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\right)$$
$$= \frac{\displaystyle\int_{-\infty}^{x_{n+1}} p(t_0, x_0; t_1, x_1; \ldots; t_n, x_n; t_{n+1}, y)\, dy}{\displaystyle\int_{-\infty}^{\infty} p(t_0, x_0; t_1, x_1; \ldots; t_n, x_n; t_{n+1}, y)\, dy} \tag{4.1.3}$$

for all time instants $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1} < \infty,\ n \in \{0, 1, \ldots\}$, and all states $x_0, x_1, \ldots, x_{n+1} \in \Re$, provided the denominator is nonzero. Now, the *Markov property* can be formulated in the form

$$P\left(X_{t_{n+1}} < x_{n+1} \,|\, X_{t_0} = x_0, X_{t_1} = x_1, \ldots, X_{t_n} = x_n\right) \\ = P\left(X_{t_{n+1}} < x_{n+1} \,|\, X_{t_n} = x_n\right) \tag{4.1.4}$$

for all time instants $0 \leq t_0 < t_1 < \ldots < t_n < t_{n+1} < \infty,\ n \in \{0, 1, \ldots\}$ and all states $x_0, x_1, \ldots, x_{n+1} \in \Re$ for which the conditional probabilities are defined.

For a continuous Markov process $X$ we write its transition probability distribution in the form

$$P(s, x; t, (-\infty, y)) = P(X_t < y \,|\, X_s = x),$$

for $t \in [0, \infty),\ s \in [0, t]$ and $x, y \in \Re$. If for $s$, $x$ and $t$ the probability distribution function $P(s, x; t, \cdot)$ has a probability density $p(s, x; t, \cdot)$, called the *transition density*, then it holds

$$P(s, x; t, (-\infty, y)) = \int_{-\infty}^{y} p(s, x; t, u)\, du \tag{4.1.5}$$

for all $y \in \Re,\ t \in [0, \infty)$ and $s \in [0, t]$.

**Chapman-Kolmogorov Equation**

The transition matrix equation (3.4.5) for continuous time Markov chains has a counterpart for the transition densities of continuous Markov processes. This continuous version is called the *Chapman-Kolmogorov equation* and has the form

$$p(s, x; t, y) = \int_{-\infty}^{\infty} p(s, x; \tau, z)\, p(\tau, z; t, y)\, dz \tag{4.1.6}$$

for $0 \leq s \leq \tau \leq t < \infty$ and $x, y \in \Re$, which follows directly from the Markov property. The Chapman-Kolmogorov equation is a fundamental relation that is used to derive important properties of continuous Markov processes.

## 4.2 Examples for Continuous Markov Processes

Let us discuss some examples of continuous Markov processes that, as we shall see later, are diffusion processes and play a role in financial modeling.

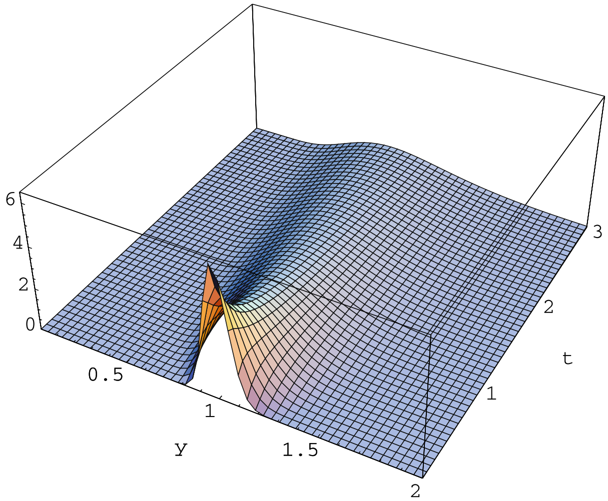



**Fig. 4.2.1.** Transition density for geometric Brownian motion

## Wiener Process

An example of a continuous Markov process is given by the standard Wiener process defined in (3.2.6). The Wiener process obtains the Markov property from its independent increments. It has the Gaussian transition density

$$p(s,x;t,y) = \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{-\frac{(y-x)^2}{2(t-s)}\right\}, \tag{4.2.1}$$

for $t \in [0,\infty)$, $s \in [0,t]$ and $x, y \in \Re$. Figure 4.1.1 shows the transition density for a Wiener process that starts at time 0 with the initial value 0.

## Geometric Brownian Motion

Geometric Brownian motion, see (4.1.2), is also a continuous Markov process. As we see later, it can be expressed as an exponential of a linearly transformed Wiener process, which gives it its Markov property. It has the transition density

$$p(s,x;t,y) = \frac{1}{\sqrt{2\,\pi(t-s)}\,b\,y} \exp\left\{-\frac{(\ln(y)-\ln(x)-g(t-s))^2}{2\,b^2(t-s)}\right\}, \tag{4.2.2}$$

for $t \in [0,\infty)$, $s \in [0,t]$ and $x, y \in (0,\infty)$. Figure 4.2.1 shows the transition density for a geometric Brownian motion with growth rate $g = 0.05$, volatility $b = 0.2$ and initial value $x = 1$ at time $s = 0$ for the period from 0.1 to 3 years.

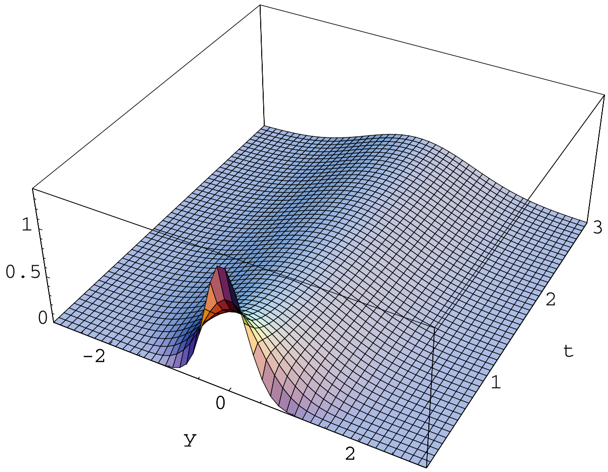



**Fig. 4.2.2.** Transition density of standard OU process starting at $(s, x) = (0, 0)$

## Standard Ornstein-Uhlenbeck Process

Let us consider an example of another continuous Markov process which is also a Gaussian process. This is the *standard Ornstein-Uhlenbeck (OU) process* $X = \{X_t,\, t \in [0, \infty)\}$, where we start from an initial value $X_0$. Since it is a Gaussian process it can be characterized by the mean and the variance of its increments. More precisely, its Gaussian transition density is defined in the form

$$p(s, x; t, y) = \frac{1}{\sqrt{2\pi \left(1 - e^{-2(t-s)}\right)}} \exp\left\{ -\frac{\left(y - x e^{-(t-s)}\right)^2}{2\left(1 - e^{-2(t-s)}\right)} \right\}, \tag{4.2.3}$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $x, y \in \Re$, with mean $x\, e^{-(t-s)}$ and variance $1 - e^{-2(t-s)}$.

To illustrate the stochastic dynamic of this process we show in Fig. 4.2.2 the transition density of a standard OU process for the period from 0.1 to 3 years with initial value $x = 0$ at time $s = 0$. As can be seen from Fig. 4.2.2 that the transition densities for the standard OU process seem to stabilize after a period of about one year. In fact, as can be seen from (4.2.3) these transition densities asymptotically approach, as $t \to \infty$, a standard Gaussian density. This is in contrast, for example, to transition densities for the Wiener process, which do not converge to a *stationary density*, see (4.2.1) and Fig. 4.1.1. For illustration, we plot in Fig. 4.2.3 the transition density for a standard OU process that starts at the initial value $x = 2$ at time $t = 0$. Note how the transition density evolves towards a median that is close to 0.

In Fig. 4.2.4 a path of a standard OU process is shown. It can be observed that this trajectory fluctuates around some reference level. Indeed, as already indicated, the standard OU process has a stationary density. This can be

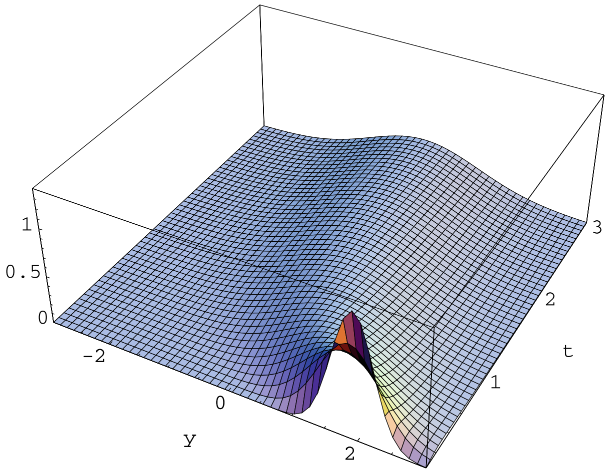



**Fig. 4.2.3.** Transition density of standard OU process starting at $(s, x) = (0, 2)$

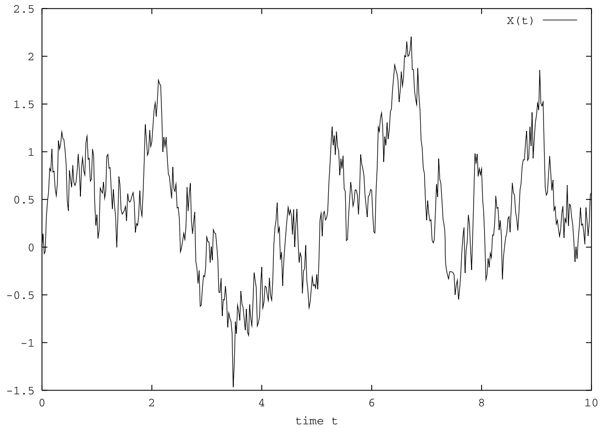



**Fig. 4.2.4.** Path of a standard Ornstein-Uhlenbeck process

seen from (4.2.3) when $t \to \infty$. Note also that the Gaussian property of the standard OU process means that even a scaled and shifted OU process may become negative.

More generally, as we shall describe later in Sect. 7.2, an *Ornstein-Uhlenbeck* (OU) *process* is a Gaussian process that is mean reverting to a reference level and its fluctuations can be more or less intense than that of a standard OU process. Such a model is suitable, for instance, for an inflation rate or a real interest rate. The fact that the OU process leads into an equilibrium dynamics is important for such modeling purposes.

### Geometric Ornstein-Uhlenbeck Process

An asset price model that both has a stationary density and is positive is obtained by the *geometric Ornstein-Uhlenbeck process*. It is expressed as the

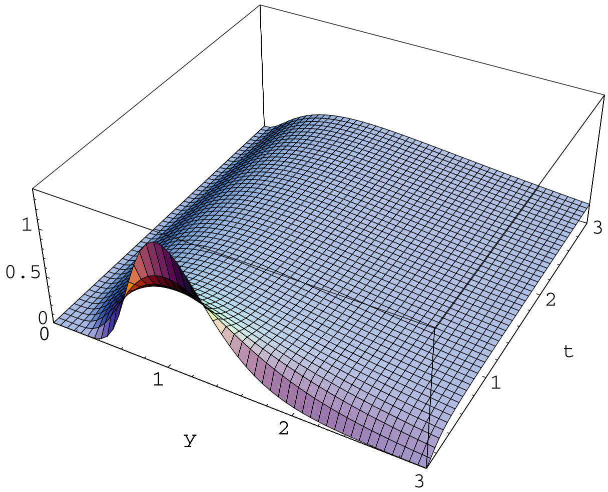



**Fig. 4.2.5.** Transition density of the geometric Ornstein-Uhlenbeck process

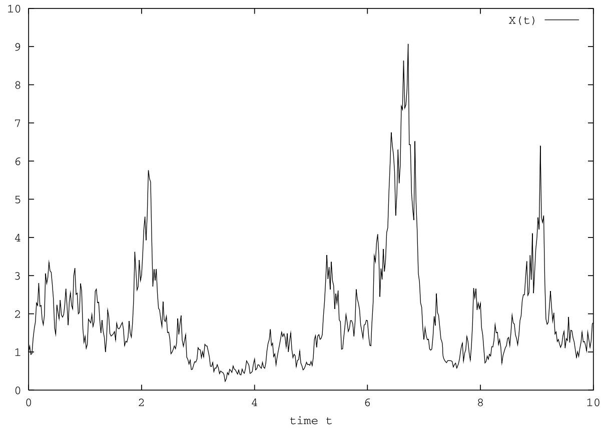



**Fig. 4.2.6.** Path of a geometric Ornstein-Uhlenbeck process

exponential of a standard OU process, that is, it has the *lognormal transition density*

$$p(s,x;t,y) = \frac{1}{y\sqrt{2\,\pi\,(1-e^{-2(t-s)})}}\,\exp\left\{-\frac{\big(\ln(y)-\ln(x)\,e^{-(t-s)}\big)^2}{2\left(1-e^{-2(t-s)}\right)}\right\}, \tag{4.2.4}$$

for $t \in [0,\infty)$, $s \in [0,t]$ and $x, y \in (0,\infty)$. In Fig. 4.2.5 we display the corresponding probability densities for the time period from 0.1 to 3 years with initial value $x = 1$ at time $s = 0$. In this case the transition density converges over time to a limiting lognormal density as stationary density, as can be seen from (4.2.4). Figure 4.2.6 shows a trajectory for the geometric OU process. We note that it stays positive and shows large fluctuations for large values.

This process was, for instance, interpreted in Föllmer & Schweizer (1993) as an asset price model. However, it is still somewhat restrictive in that it is not possible to model changes in the trend or volatility of the asset price. This will be conveniently achieved in the context of more general diffusion processes, which form a class of special continuous Markov processes and will be considered below.

## 4.3 Diffusion Processes

It is not surprising that the Wiener process serves as a prototype example of a diffusion process since it can model the diffusive motion of Brownian particles. As we attempt to show, diffusion processes form a powerful class of stochastic processes that can be applied to a range of financial modeling problems.

### Characterization of Diffusion Processes

**Definition 4.3.1.** *A continuous time Markov process with* transition density $p(s,x;t,\,y)$ *is called a* diffusion process *if the following three limits exist for all* $\varepsilon > 0$, $s \in [0,\infty)$ *and* $x \in \Re$ *:*

$$\lim_{t\downarrow s} \frac{1}{t-s} \int_{|y-x|>\varepsilon} p(s,x;t,y)\,dy = 0, \tag{4.3.1}$$

$$\lim_{t\downarrow s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x)p(s,x;t,y)\,dy = a(s,x) \tag{4.3.2}$$

*and*

$$\lim_{t\downarrow s} \frac{1}{t-s} \int_{|y-x|<\varepsilon} (y-x)^2 p(s,x;t,y)\,dy = b^2(s,x), \tag{4.3.3}$$

*where* $a$ *and* $b^2$ *are integrable functions.*

The condition (4.3.1) prevents the diffusion process from having jumps. At time $s$ and position $x$ the quantity $a(s,x)$ in (4.3.2) is called the *drift coefficient* and $b(s,x)$ in (4.3.3) the *diffusion coefficient*. Condition (4.3.2) implies that the drift coefficient is given by the limit of the conditional expectation

$$a(s,x) = \lim_{t\downarrow s} \frac{1}{t-s} E\left(X_t - X_s \,\Big|\, X_s = x\right). \tag{4.3.4}$$

This means that the drift $a(s,x)$ is the instantaneous rate of change in the conditional mean of the diffusion process given that $X_s = x$.

Similarly, it follows from (4.3.3) that

$$b^2(s,x) = \lim_{t\downarrow s} \frac{1}{t-s} E\left((X_t - X_s)^2 \,\Big|\, X_s = x\right), \tag{4.3.5}$$

which denotes the limit of the second moment of the increments of the diffusion process normalized by the time $t-s$, given that $X_s = x$. Thus $b(s, x)$ measures the average size of the fluctuations of the diffusion process. In fact it can be shown that $b^2(s, x)$ is approximately the normalized variance of its increments $X_t - X_s$ as $t \to s$. Furthermore, it can be shown under fairly general conditions that for a given initial value $X_0$, drift $a(\cdot, \cdot)$ and diffusion coefficient $b(\cdot, \cdot)$ the diffusion process $X$ is uniquely determined, for instance, in a mean square sense as will be discussed later.

Roughly speaking, the increment $X_t - X_s$ of a diffusion process over a small time interval of length $h = t - s$ can be interpreted approximately as a conditionally Gaussian random variable with mean $a(s, X_s)h$ and variance $b^2(s, X_s)h$. This can be expressed as

$$X_t - X_s \approx a(s, X_s)\, h + b(s, X_s)\, \sqrt{h}\, \xi, \tag{4.3.6}$$

where $\xi$ is an independent standard Gaussian random variable. This equation is useful as a first approximation of increments of diffusion processes, to guide intuition and to indicate a relationship with the classical Taylor series expansion. Note however, in this simplified form no information is given about the corresponding error term.

### Examples of One-Factor Asset Price Models

Let us list together with the already mentioned examples a few additional one-dimensional diffusion processes that have been applied in asset price modeling: The *linearly transformed Wiener process*, see Fig. 3.2.3, is an example of a diffusion process with drift $a(s, x) = 0$ and diffusion coefficient $b(s, x) = b$. As previously mentioned, it was used in Bachelier (1900) for stock price modeling. One of the disadvantages of this asset price model is given by the fact that it generates negative asset prices. In a very simplistic way it is sometimes argued that if one freezes the trajectory of this *Bachelier model* when it first hits zero, then one obtains a very basic asset price model. This model has many deficiencies. In particular, the asset price will hit zero with positive probability. This is usually not intended when modeling asset prices.

It can be shown that *geometric Brownian motion* or the *Black-Scholes* (BS) *model* as given in (4.1.2) is a diffusion process with drift

$$a(s, x) = x \left( g + \frac{1}{2}\, b^2 \right) \tag{4.3.7}$$

and diffusion coefficient

$$b(s, x) = x\, b. \tag{4.3.8}$$

As previously mentioned, the BS model was suggested in Samuelson (1955, 1965a) and used in Black & Scholes (1973). Despite the fact that this model became the standard financial market model, it does not generate a random, fluctuating volatility, which is usually observed in practice.

The *geometric Ornstein-Uhlenbeck* (GOU) *model*, see (4.2.4), can be shown to have drift coefficient

$$a(s, x) = x\,(1 - \ln(x)) \tag{4.3.9}$$

and diffusion coefficient

$$b(s, x) = \sqrt{2}\,x. \tag{4.3.10}$$

This asset price model permits an equilibrium type dynamics. It was used, as already mentioned, in Föllmer & Schweizer (1993), Platen & Rebolledo (1996) and Fleming & Sheu (1999). A disadvantage is again that it does not generate a fluctuating volatility.

The *constant elasticity of variance* (CEV) *model* introduced in Cox (1975), see also Schroder (1989), has drift

$$a(s, x) = x\,r \tag{4.3.11}$$

and diffusion coefficient

$$b(s, x) = \sigma\,x^{\alpha} \tag{4.3.12}$$

with constants $r$, $\sigma$ and $\alpha \in (0, 1)$. It does generate a fluctuating volatility. The elasticity of the changes of the variance of log-returns can be shown to be constant due to the power structure of the diffusion coefficient. However, as shown in Delbaen & Shirakawa (2002), the model has a deficiency since the asset price will hit zero with positive probability in finite time, which is not what one usually intends to model.

The *minimal market model* (MMM) introduced in Platen (2001, 2002) has in its stylized version the drift

$$a(s, x) = \alpha_s \tag{4.3.13}$$

and the diffusion coefficient

$$b(s, x) = \sqrt{\alpha_s\,x} \tag{4.3.14}$$

with $\alpha_s = \alpha_0 \exp\{\eta\,s\}$, for initial trend $\alpha_0 > 0$ and net growth rate $\eta > 0$. This model generates a realistic, fluctuating volatility and does not hit zero. In particular, its volatility dynamics match closely that of observed index volatility and yields realistic option prices, as we shall see later.

### Examples of One-Factor Short Rate Models

A large variety of short rate models has been developed that are formed by diffusion processes. In the following we shall mention several one-factor short rate models by specifying their drift and diffusion coefficients.

One of the simplest stochastic short rate models arises if the Wiener process is linearly transformed by assuming a deterministic drift coefficient $a(s, x) = a_s$ and a deterministic diffusion coefficient $b(s, x) = b_s$. This leads to the *Merton model*, see Merton (1973a), or to some specification of the continuous time version of the *Ho-Lee model*, see Ho & Lee (1986). Here the short rate does not remain positive, as one would expect.

A widely used short rate model is the *Vasicek model*, see Vasicek (1977), or the *extended Vasicek model*, which is an Ornstein-Uhlenbeck process with linear drift coefficient $a(s,x) = \gamma_s\,(\bar{x}_s - x)$ and deterministic diffusion coefficient $b(s,x) = b_s$. Also this model has a Gaussian transition density and, thus, allows negative interest rates.

In Black (1995) it was suggested that one considers the nonnegative value of a short rate like an option value, which only takes the positive part of an underlying quantity. This *Black model* results, when using an Ornstein-Uhlenbeck process $u = \{u_t,\, t \in [0,T]\}$ as underlying shadow short rate and a short rate of the form $x_s = (u_s)^+ = \max(0, u_s)$. Such type of short rate models, which allow the consideration of low interest rate regimes, have been studied, for instance, in Gorovoi & Linetsky (2004) and Miller & Platen (2005).

Cox et al. (1985) suggested the *CIR model*, which uses a square root process, see (4.4.6) below. Its drift coefficient $a(s,x) = \gamma_s\,(\bar{x}_s - x)$ is *affine*, which means that it is linear, and its diffusion coefficient is of the form $b(s,x) = b_s\,\sqrt{x}$. In the next section we shall describe the transition densities of the CIR model. This model has the desirable feature that it excludes negative interest rates. Furthermore, it yields an equilibrium dynamics. Unfortunately, when calibrated to market data, it shows a number of deficiencies which concern the possible shapes of the, so-called, *forward rate* or *yield curves*.

A translated or extended model of the CIR type is the *Pearson-Sun model*, see Pearson & Sun (1989), which assumes $a(s,x) = \gamma\,(\bar{x}_s - x)$ and $b(s,x) = \sqrt{b_1 + b_2\,x}$. Here the parameters are usually assumed to be constants which fulfill certain conditions, such as $\gamma(\bar{x} + \frac{b_1}{b_2}) > 0$. These ensure that the solution is contained in a certain region. Duffie & Kan (1994) generalized this model, which belongs to the affine class of diffusion processes, because the drift $a$ and squared diffusion coefficient $b^2$ are affine. This model is therefore often called an *affine model*.

Marsh & Rosenfeld (1983) and also Dothan (1978) considered a short rate model with $a(s,x) = a_s x$ and $b(s,x) = b_s x$. This specification is known as the *lognormal model*. Here the short rate remains positive, however, it does not admit a stationary regime.

A generalized lognormal model, also called the *Black-Karasinski model*, see Black & Karasinski (1991), is obtained by setting $a(s,x) = x\,(a_s + g_s\,\ln(x))$ and $b(s,x) = b_s x$. This generates a geometric Ornstein-Uhlenbeck process, see Sect. 4.2. If $g_s = -\frac{b_s'}{b_s}$, then the above model is also called the continuous-time version of the *Black-Derman-Toy model*, see Black, Derman & Toy (1990). This type of model keeps interest rates positive and allows them to have an equilibrium.

Another model arises if one sets $a(s,x) = \gamma_s\,(\bar{x}_s - x)$ and $b(s,x) = b_s x$. In the case of constant parameters this formulation is known as the *Courtadon model*, see Courtadon (1982). The *Longstaff model*, see Longstaff (1989) is obtained by setting $a(s,x) = \gamma_s\,(\sqrt{\bar{x}_s} - \sqrt{x})$ and $b(s,x) = b_s\sqrt{x}$.

A rather general model is the *Hull-White model*, see Hull & White (1990). It has linear mean-reverting drift $a(s,x) = \gamma_s\,(\bar{x}_s - x)$ and diffusion coefficient $b(s,x) = b_s x^q$ for some choice of exponent $q \geq 0$. Obviously, this structure includes several of the above models. In the case $q = 0$ the Hull-White model is also called the extended Vasicek model, as already mentioned above.

The *Sandmann-Sondermann model*, see Sandmann & Sondermann (1994), was motivated by the aim to consider annual, continuously compounded interest rates. It has drift $a(s,x) = (1 - e^{-x})(a_s - \frac{1}{2}(1 - e^{-x})\,b_s^2)$ and diffusion coefficient $b(s,x) = (1 - e^{-x})c_s$.

An alternative short rate model was proposed in Platen (1999), which suggests a drift $a(s,x) = \gamma(x - a_s)(c_s - x)$ and a diffusion coefficient of the type $b(s,x) = b_s|x - c_s|^{\frac{3}{2}}$. The *Platen model* provides a reasonably accurate reflection of the short rate drift and diffusion coefficient as estimated from market data in Ait-Sahalia (1996).

As can be seen by these examples one can, in principle, choose quite general functions for the drift and diffusion coefficients to form meaningful diffusion models of asset prices, short rates and other financial quantities. These functions then characterize, together with the initial conditions, the dynamics of the diffusion process in an elegant and efficient way. This characterization is more compact than, for instance, that given by a transition matrix of a discrete or continuous time Markov chain. As we shall see, it also allows the exploitation of smoothness and other regularity properties of the transition densities for functionals of diffusions. We shall later aim to identify an optimal diffusion type dynamics of a financial market that takes advantage of these powerful mathematical features.

## 4.4 Kolmogorov Equations

In this section we describe some important results, which show that the transition densities for diffusion processes satisfy certain *partial differential equations* (PDEs).

### Kolmogorov Equations

When the drift coefficient $a(\cdot)$ and diffusion coefficient $b(\cdot)$ of a diffusion process are appropriate functions, as will be discussed later, then its transition density $p(s,x;t,y)$ satisfies certain PDEs. These are the *Kolmogorov forward equation* or *Fokker-Planck equation*

$$\frac{\partial p(s,x;t,y)}{\partial t} + \frac{\partial}{\partial y}\left(a(t,y)\,p(s,x;t,y)\right) - \frac{1}{2}\frac{\partial^2}{\partial y^2}\left(b^2(t,y)\,p(s,x;t,y)\right) = 0, \tag{4.4.1}$$

for $(s,x)$ fixed, and the *Kolmogorov backward equation*

$$\frac{\partial p(s,x;t,y)}{\partial s} + a(s,x)\frac{\partial p(s,x;t,y)}{\partial x} + \frac{1}{2}\,b^2(s,x)\frac{\partial^2 p(s,x;t,y)}{\partial x^2} = 0, \tag{4.4.2}$$

for $(t,y)$ fixed. Obviously, the *initial* or *terminal condition* for this PDE equals the Dirac delta function

$$p(s,x;s,y) = \delta(y-x) = \begin{cases} \infty & \text{for} \quad y = x \\ 0 & \text{for} \quad y \neq x, \end{cases} \tag{4.4.3}$$

where

$$\int_{-\infty}^{\infty} \delta(y-x)\,dy = 1 \tag{4.4.4}$$

for given $x$.

The first PDE (4.4.1) describes the forward evolution of the transition density with respect to the final time and state $(t,y)$ and the second provides the backward evolution with respect to the initial time and position $(s,x)$. The forward equation (4.4.1) is commonly called the *Fokker-Planck equation*. Both Kolmogorov equations follow from the Chapman-Kolmogorov equation (4.1.6) and the conditions (4.3.1)–(4.3.3). The Kolmogorov backward equation plays, in an extended form with other boundary conditions, an essential role in derivative pricing.

A few diffusion processes, for instance, those that arise from transformations of either Gaussian or square root processes have known transition densities. It is convenient to use such transformed diffusions with explicitly known transition densities to model financial quantities. As long as one is able to stay in such a framework the resulting quantitative methods are usually superior to numerical methods for solving PDEs. However, when the drift or diffusion coefficients become more complex or time dependent, then numerical methods have to be employed to approximate the solutions of the PDEs.

**Transition Densities for the Square Root Process**

Let us consider the *square root* (SR) *process* that appears in the CIR model mentioned in the previous section. Here we use the specification of the drift coefficient $a(s,x) = \gamma(\bar{x} - x)$ and the diffusion coefficient $b(s,x) = \beta\sqrt{x}$ for $s \geq 0$, $x > 0$, with constant reference level $\bar{x} > 0$, speed of adjustment $\gamma > 0$ and scaling parameter $\beta > 0$. A key feature of the SR process is that it is linear mean-reverting and can be shown to be nonnegative, see Borodin & Salminen (2002). For a value $x$ above the reference level $\bar{x}$ the drift coefficient is negative and drives the process back to $\bar{x}$. For a value $x = 0$, the diffusion coefficient is zero and the drift coefficient is positive. Intuitively,

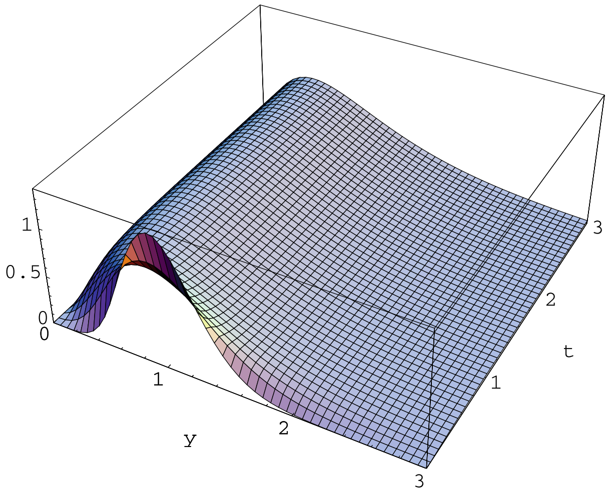



**Fig. 4.4.1.** Transition density of a square root process

the process is then driven back to its reference level $\bar{x}$. One can show that for $\frac{\gamma\,\bar{x}}{\beta^2} \geq \frac{1}{2}$ the SR process remains strictly positive, see Revuz & Yor (1999) or Borodin & Salminen (2002).

The quantity

$$n = \frac{4\,\gamma\,\bar{x}}{\beta^2} \tag{4.4.5}$$

is referred to as the *dimension* of the SR process. For the SR process with $\frac{\gamma\,\bar{x}}{\beta^2} \geq \frac{1}{2}$ the corresponding transition density $p(s, x; t, y)$ is available in analytic form. In fact, it is given by

$$p(s, x; t, y) = \frac{1}{2(\tau(t) - \tau(s))} \exp\left\{\gamma\,t - \frac{x\,\exp\{\gamma\,s\} + y\,\exp\{\gamma\,t\}}{2(\tau(t) - \tau(s))}\right\}$$

$$\times \left(\frac{y\,\exp\{\gamma\,(t-s)\}}{x}\right)^{\frac{\nu}{2}} I_\nu\left(\frac{\sqrt{x\,y\,\exp\{\gamma\,(t+s)\}}}{\tau(t) - \tau(s)}\right) \tag{4.4.6}$$

with

$$\tau(t) = \frac{(\exp\{\gamma\,t\} - 1)\,\beta^2}{4\,\gamma} \tag{4.4.7}$$

for $t \in [0, \infty)$, $s \in [0, t]$, $x > 0$, $y > 0$ and *modified Bessel function of the first kind* $I_\nu(z)$ with index

$$\nu = \frac{2}{\beta^2}\,\gamma\,\bar{x} - 1 = \frac{n}{2} - 1, \tag{4.4.8}$$

see (1.2.15). Here $\Gamma(\cdot)$ is the gamma function, see (1.2.10). One can show that the above transition density satisfies the Kolmogorov equations (4.4.1)–(4.4.4).

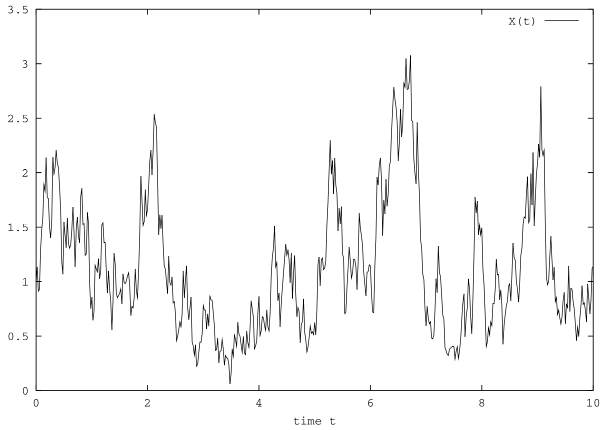



**Fig. 4.4.2.** Sample path of a square root process of dimension four

Figure 4.4.1 shows the transition density of an SR process for the period from 0.1 to 3.0 years, with initial value $X_0 = 1.0$, reference level $\bar{x} = 1.0$ and parameters $\gamma = 2$ and $\beta = \sqrt{2}$. By (4.4.5) this means that we consider an SR process of dimension $n = 4$. Figure 4.4.2 displays a sample path for the SR process.

### Generalized Square Root Processes (*)

As we have seen above, for asset price modeling and short rate modeling but also for squared volatility modeling, positive diffusion processes, which potentially allow some equilibrium, have a great appeal. Therefore, we add the following explicit transition densities for generalized square root processes. Some of these have been recently derived in Craddock & Platen (2004) by symmetry group methods. Such transition densities can be potentially rather useful in quantitative finance.

Let us consider a *generalized square root process*, which is a diffusion process $X = \{X_t,\, t \in [0, \infty)\}$ with a square root function as diffusion coefficient of the form

$$b(t, x) = \sqrt{2\,x} \tag{4.4.9}$$

for all $t \geq 0$ and $x \in [0, \infty)$. Here the drift function $a(t, x) = a(x)$ is time homogeneous but otherwise rather flexible. This drift will be specified below for certain cases.

It is of interest to identify those drift functions $a(\cdot)$, where one has an analytic solution of the Kolmogorov backward PDE for the corresponding time homogeneous transition density $p(0, x; t, y)$, which can be written as

$$-\frac{\partial p(0, x; t, y)}{\partial t} + x\,\frac{\partial^2 p(0, x; t, y)}{\partial x^2} + a(x)\,\frac{\partial p(0, x; t, y)}{\partial x} = 0 \tag{4.4.10}$$

for $t \in (0, \infty)$ with

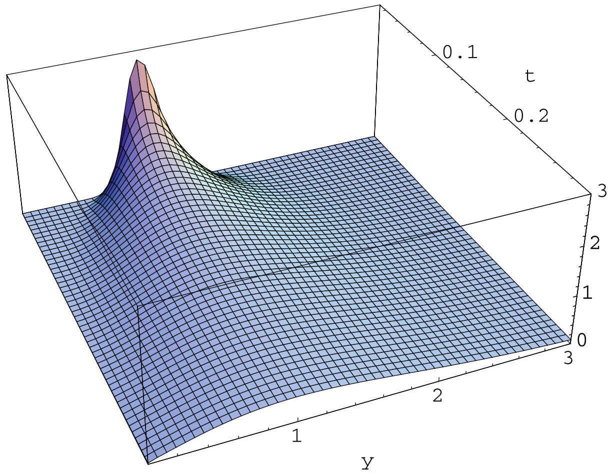



**Fig. 4.4.3.** Transition density for a squared Bessel process, case (i)

$$p(0, x, y) = \delta(x - y) \tag{4.4.11}$$

for $x, y \in (0, \infty)$. In Craddock & Platen (2004) for the following ten particular cases analytical solutions have been identified:

(i) When the drift function is a constant

$$a(x) = \alpha > 0, \tag{4.4.12}$$

then we have a, so-called, squared Bessel process of dimension $n = 2\alpha$ with transition density

$$p(0, x; t, y) = \frac{1}{t}\left(\frac{x}{y}\right)^{\frac{1-\alpha}{2}} I_{\alpha-1}\left(\frac{2\sqrt{x\,y}}{t}\right) \exp\left\{-\frac{(x+y)}{t}\right\}. \tag{4.4.13}$$

Here $I_{\alpha-1}$ is again the modified Bessel function of the first kind with index $\alpha - 1$, see (1.2.15). In Fig. 4.4.3 we plot the transition density $p(0, x; t, y)$ for $x = 1$ and $\alpha = \frac{3}{2}$, that is, for a squared Bessel process of dimension $n = 3$.

(ii) When we set the drift function to

$$a(x) = \frac{\mu\, x}{1 + \frac{\mu}{2}\, x} \tag{4.4.14}$$

for $\mu > 0$, then we obtain the transition density

$$p(0, x; t, y) = \frac{\exp\left\{-\frac{(x+y)}{t}\right\}}{\left(1 + \frac{\mu}{2}\, x\right) t} \left[\left(\sqrt{\frac{x}{y}} + \frac{\mu\sqrt{x\,y}}{2}\right) I_1\left(\frac{2\sqrt{x\,y}}{t}\right) + t\,\delta(y)\right] \tag{4.4.15}$$

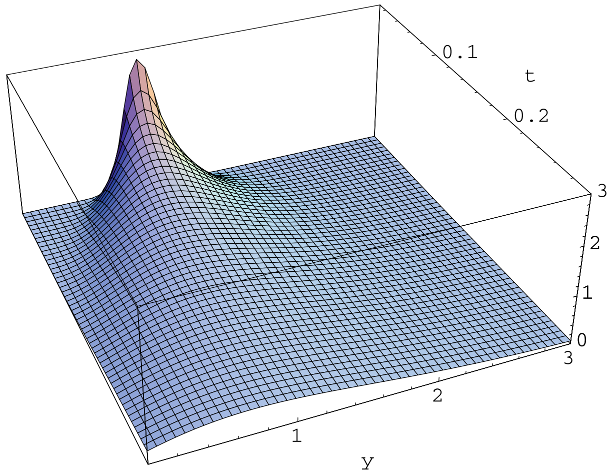



**Fig. 4.4.4.** Transition density for case (ii)

with $\delta(\cdot)$ denoting the Dirac delta function. For $y = 0$ one can interpret $\frac{\exp\{-\frac{x}{t}\}}{\left(1+\frac{\mu}{2}\,x\right)}$ as the probability of absorption at zero. In Fig. 4.4.4 we show the above transition density for $x = 1$ and $\mu = 1$.

(iii) In the case of the drift function

$$a(x) = \frac{1 + 3\sqrt{x}}{2\,(1 + \sqrt{x})}, \tag{4.4.16}$$

one obtains the transition density

$$p(0, x; t, y) = \frac{\cosh\left(\frac{2\sqrt{x\,y}}{t}\right)}{\sqrt{\pi\,y\,t}\,(1 + \sqrt{x})} \left(1 + \sqrt{y}\,\tanh\left(\frac{2\sqrt{x\,y}}{t}\right)\right) \times \exp\left\{-\frac{(x + y)}{t}\right\}. \tag{4.4.17}$$

In Fig. 4.4.5 we display the corresponding transition density for $x = 1$.

(iv) When we choose as drift function

$$a(x) = 1 + \mu\,\tanh\left(\mu + \frac{1}{2}\,\mu\,\ln(x)\right) \tag{4.4.18}$$

for $\mu = \frac{1}{2}\sqrt{\frac{5}{2}}$, then we obtain the transition density

$$p(0, x; t, y) = \left(\frac{x}{y}\right)^{\frac{\mu}{2}} \left[I_{-\mu}\left(\frac{2\sqrt{x\,y}}{t}\right) + e^{2\,\mu}\,y^{\mu}\,I_{\mu}\left(\frac{2\sqrt{x\,y}}{t}\right)\right] \times \frac{\exp\{-\frac{x+y}{t}\}}{(1 + \exp\{2\,\mu\}\,x^{\mu})\,t}. \tag{4.4.19}$$

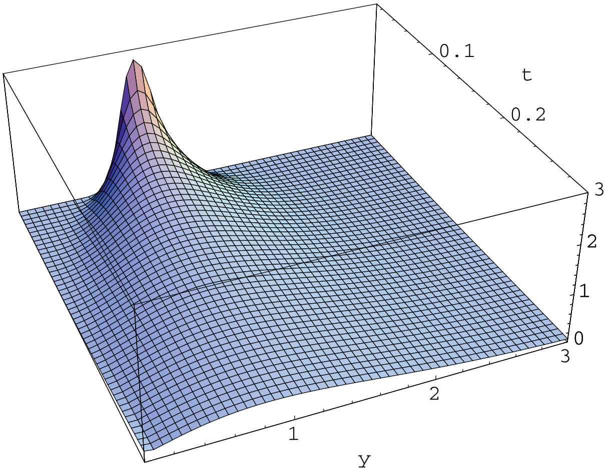



**Fig. 4.4.5.** Transition density for case (iii)

The shape of the density (4.4.19) for $x = 1$ looks quite similar to that in Fig. 4.4.5.

(v) For the drift function

$$a(x) = \frac{1}{2} + \sqrt{x}, \tag{4.4.20}$$

one obtains the transition density

$$p(0, x; t, y) = \cosh\left(\frac{(t + 2\sqrt{x})\sqrt{y}}{t}\right) \frac{\exp\{-\sqrt{x}\}}{\sqrt{\pi\, y\, t}} \exp\left\{-\frac{(x+y)}{t} - \frac{t}{4}\right\}. \tag{4.4.21}$$

Also the transition density (4.4.21) for $x = 1$ shows a lot of similarity with that in Fig. 4.4.5.

(vi) In the case where the drift function is set to

$$a(x) = \frac{1}{2} + \sqrt{x}\, \tanh(\sqrt{x}), \tag{4.4.22}$$

we obtain the transition density

$$p(0, x; t, y) = \frac{\cosh\left(\frac{2\sqrt{x\,y}}{t}\right)}{\sqrt{\pi\, y\, t}} \frac{\cosh(\sqrt{y})}{\cosh(\sqrt{x})} \exp\left\{-\frac{(x+y)}{t} - \frac{t}{4}\right\}. \tag{4.4.23}$$

The above transition density (4.4.23) for $x = 1$ has also a similar shape as that in Fig. 4.4.5.

(vii) For the drift function

$$a(x) = \frac{1}{2} + \sqrt{x}\, \coth(\sqrt{x}) \tag{4.4.24}$$

the process has the transition density

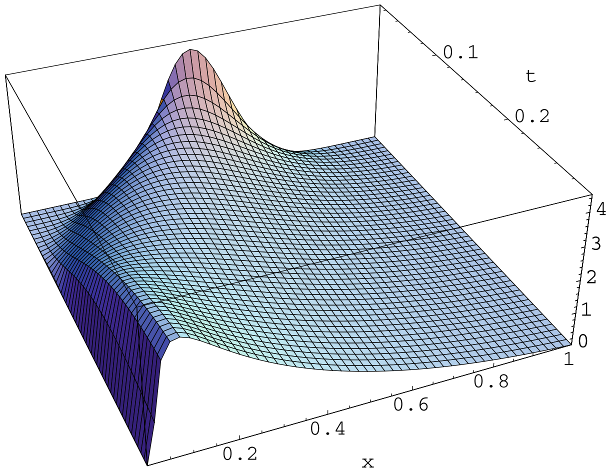



**Fig. 4.4.6.** Transition density for case (viii)

$$p(0,x;t,y) = \frac{\sinh\left(\frac{2\sqrt{x\,y}}{t}\right)}{\sqrt{\pi\,y\,t}} \, \frac{\sinh(\sqrt{y})}{\sinh(\sqrt{x})} \, \exp\left\{-\frac{(x+y)}{t} - \frac{t}{4}\right\}. \quad (4.4.25)$$

This transition density has for $x = 1$ some similarity with that shown in Fig. 4.4.3.

(viii) When we use as drift function

$$a(x) = 1 + \cot(\ln(\sqrt{x})) \quad (4.4.26)$$

for $x \in (\exp\{-2\pi\}, 1)$, then we obtain the real valued transition density

$$p(0,x;t,y) = \frac{\exp\{-\frac{(x+y)}{t}\}}{2\,\imath\,t\sin(\ln(\sqrt{x}))} \left( y^{\frac{\imath}{2}}\, I_\imath\left(\frac{2\sqrt{x\,y}}{t}\right) - y^{-\frac{\imath}{2}}\, I_{-\imath}\left(\frac{2\sqrt{x\,y}}{t}\right)\right), \quad (4.4.27)$$

where $\imath$ denotes the imaginary unit.

We plot in Fig. 4.4.6 the transition density (4.4.27) for $x = \frac{1}{2}$. Note that the process $X$ lives on the bounded interval $(\exp\{-2\pi\}, 1)$.

(ix) If we choose the drift function

$$a(x) = x \coth\left(\frac{x}{2}\right), \quad (4.4.28)$$

then we obtain the transition density

$$p(0,x;t,y) = \frac{\sinh(\frac{y}{2})}{\sinh(\frac{x}{2})} \exp\left\{-\frac{(x+y)}{2\,\tanh(\frac{t}{2})}\right\} \times \left[\frac{\exp\{\frac{t}{2}\}}{\exp\{t\}-1}\sqrt{\frac{x}{y}}\, I_1\left(\frac{\sqrt{x\,y}}{\sinh(\frac{t}{2})}\right) + \delta(y)\right], \quad (4.4.29)$$

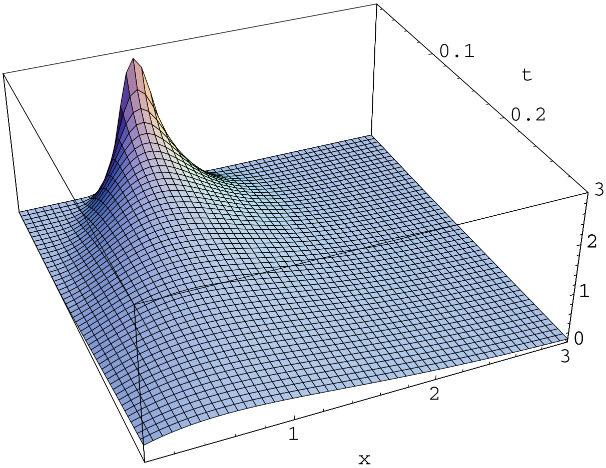



**Fig. 4.4.7.** Transition density for case (x)

where $\delta(\cdot)$ is again the Dirac delta function. In Fig. 4.4.3 we displayed a transition density of similar shape.

(x) Finally, let us set the drift function to

$$a(x) = x \tanh\left(\frac{x}{2}\right) \tag{4.4.30}$$

to obtain the transition density

$$p(0, x; t, y) = \frac{\cosh(\frac{y}{2})}{\cosh(\frac{x}{2})} \exp\left\{-\frac{(x+y)}{2 \tanh(\frac{t}{2})}\right\} \times \left[\frac{\exp\{\frac{t}{2}\}}{\exp\{t\}-1} \sqrt{\frac{x}{y}}\, I_1\left(\frac{\sqrt{x\,y}}{\sinh(\frac{t}{2})}\right) + \delta(y)\right]. \tag{4.4.31}$$

We plot in Fig. 4.4.7 the transition density for $x = 1$.

All ten cases that we described above provide examples for generalized square root processes with diffusion coefficient function $b(x) = \sqrt{2\,x}$. In all these cases we have for the prescribed drift coefficient function an explicitly known transition density. This list of explicitly known transition densities provides valuable information for a quantitative analyst when a model needs to be designed with a square root diffusion coefficient. One can try to choose one of the above models to reflect the given dynamics. As we shall see later, by applications of stochastic calculus one can describe analytically the transition densities of a much wider class of diffusion processes that arise as twice differentiable functions of the above generalized square root processes.

## 4.5 Diffusions with Stationary Densities

Let us now consider diffusion processes that can model an equilibrium. Such stationary processes are important when the probabilistic features of a diffusion process do not change after a shift in time. In finance such processes are needed to model volatilities, short rates, credit spreads, inflation rates, market activity and other key quantities.

### Stationary Density

When we use diffusion processes to provide models for financial quantities that can evolve into some equilibrium, then we restrict considerably the class of diffusion processes that we consider. For example, as previously noted, the standard and the geometric OU processes are diffusion processes with transition densities that converge over long periods of time towards corresponding stationary densities, see (4.2.3) and (4.2.4). The transition density of the standard OU process is shown in Fig. 4.2.2. In this figure we observe for increasing time the convergence of the transition density towards some stationary density, which in this case is the standard Gaussian density. Similarly, one notes in Fig. 4.2.5, the convergence of the transition density of the geometric OU process towards another stationary density, which is here the lognormal density. Also the SR process, see (4.2.4), has a stationary density.

More precisely, for a diffusion process that permits some equilibrium its *stationary density* $\bar{p}(y)$ is defined as the solution of the integral equation

$$\bar{p}(y) = \int_{-\infty}^{\infty} p(s, x; t, y)\, \bar{p}(x)\, dx$$

for $t \in [0, \infty)$, $s \in [0, t]$ and $y \in \Re$. This means, if one starts with the stationary density, then one obtains again the stationary density as the probability density of the process after any given time period. A *stationary diffusion process* is, therefore, obtained when the corresponding diffusion process starts with its stationary density. We shall not call a stationary diffusion process a diffusion process with stationary density that starts with a given fixed value. We rather say in this case that the process has a stationary density.

One can identify the stationary density $\bar{p}$ by noting that it satisfies the corresponding stationary, or time-independent, Kolmogorov forward equation, see (4.4.1). This *stationary Fokker-Planck equation* reduces to the ordinary differential equation (ODE)

$$\frac{d}{dy}\left(a(y)\,\bar{p}(y)\right) - \frac{1}{2}\,\frac{d^2}{dy^2}\left(b^2(y)\,\bar{p}(y)\right) = 0 \tag{4.5.1}$$

with drift $a(x) = a(s, x)$ and diffusion coefficient $b(x) = b(s, x)$. Consequently, it is necessary that equation (4.5.1) is satisfied to ensure that a diffusion has a stationary density. We assume in the following that a unique stationary density exists for the diffusion processes to be considered in the remainder of this section.

Note that since $\bar{p}$ is a probability density it must satisfy the relation

$$\int_{-\infty}^{\infty} \bar{p}(y)\, dy = 1. \tag{4.5.2}$$

### Analytic Stationary Densities

Fortunately, one can identify for a large class of stationary diffusion processes the analytic form of their stationary density $\bar{p}(y)$. To do this, one notes from equation (4.5.1) when setting

$$H(y) = a(y)\bar{p}(y) - \frac{1}{2}\frac{d}{dy}\left(b^2(y)\,\bar{p}(y)\right)$$

that

$$\frac{d}{dy}H(y) = 0 \tag{4.5.3}$$

for $y \in \Re$ so that

$$H(y) = H = \text{ const.} \tag{4.5.4}$$

As $y \to \infty$ then $\bar{p}(y) \to 0$ and also $\frac{d\bar{p}(y)}{dy} \to 0$. This implies that $H = 0$ and one can therefore show that the stationary density is given by the explicit expression

$$\bar{p}(y) = \frac{C}{b^2(y)} \exp\left\{2\int_{y_0}^{y} \frac{a(u)}{b^2(u)}\, du\right\}. \tag{4.5.5}$$

This density satisfies the Fokker-Planck equation (4.5.1) for $y \in \Re$ with some fixed value $y_0 \in \Re$. Here $y_0$ is some appropriate point in the interval, where the process $X$ is defined. The constant $C$ can be obtained from the normalization condition (4.5.2). The formula (4.5.5) is useful in a number of applications since it allows one to obtain explicit analytic representations for the stationary density of diffusions. Moreover, if one observes from data the stationary density of a diffusion and has either its drift or its diffusion coefficient function given, then one can deduce the form of the missing diffusion or drift coefficient function, respectively.

### Examples of Stationary Densities

Specifications for both the drift and diffusion coefficients are needed to determine the stationary density. For instance, in the case of the standard OU process, see (4.2.3), with $a(s,x) = a(x) = -x$ and $b(s,x) = b(x) = \sqrt{2}$ the stationary probability density is the standard Gaussian density

$$\bar{p}(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{y^2}{2}\right\} \tag{4.5.6}$$

for $y \in \Re$, see Fig. 1.2.3.

For the SR process, see (4.4.6), with $a(s,x) = a(x) = \gamma\,(1-x)$ and $b(s,x) = b(x) = \beta\,\sqrt{x}$ we obtain from (4.5.5) the stationary density

$$\bar{p}(y) = C\, y^{\frac{2\gamma}{\beta^2}-1}\, \exp\left\{\frac{-2\gamma}{\beta^2}\, y\right\} \tag{4.5.7}$$

for $y \in (0,\infty)$, where we assume $\frac{2\gamma}{\beta^2} > 1$. This is a gamma density, see (1.2.9), with $\alpha = p = \frac{2\gamma}{\beta^2}$.

An interesting class of diffusion processes with stationary density is obtained for a linear mean reverting drift

$$a(x) = \gamma(\bar{x} - x) \tag{4.5.8}$$

and a squared diffusion coefficient of the form

$$b^2(x) = 2(b_0 + b_1 x + b_2 x^2), \tag{4.5.9}$$

which is quadratic in $x \in \Re$. In this case, it can be shown that the corresponding stationary density $\bar{p}$ turns out to be a *Pearson type density* for an appropriate choice of constants $\gamma$, $\bar{x}$, $b_0$, $b_1$ and $b_2$. This class includes the normal, chi-square, gamma, Student $t$, uniform and exponential, but also the power exponential, beta, arcsin, Erlang and Pareto probability densities.

In Fig. 4.5.1 we show three stationary densities for specific choices of drift and diffusion coefficients. The stationary density for an Ornstein-Uhlenbeck process, labelled OU is obtained, using $\gamma = 2$ and $\bar{x} = 1$ in (4.5.8) and $b_1 = b_2 = 0$ and $b_0 = 1$ in (4.5.9). The stationary density of a square root process, labelled SR, is produced with the choices $\gamma = 2$ and $\bar{x} = 1$ in (4.5.8) and $b_0 = b_2 = 0$ and $b_1 = 1$ in (4.5.9). Finally, the stationary density of a geometric OU process, labelled GOU, is generated if we set in (4.5.5) $a(x) = x(1-\ln(x))$ and $b(x) = 2x^2$. We see in Fig. 4.5.1 the different shapes of stationary densities that can be obtained.

### Ergodicity of a Diffusion Process (*)

In Sect. 3.4 we introduced the notation of ergodicity in the context of continuous time Markov chains. This property can be analogously defined for diffusion processes with stationary densities. A diffusion process $X = \{X_t,\ t \in [0,\infty)\}$ is called *ergodic* if it has a stationary density $\bar{p}$ and

$$\lim_{T\to\infty} \frac{1}{T} \int_0^T f(X_t)\, dt = \int_{-\infty}^{\infty} f(x)\, \bar{p}(x)\, dx, \tag{4.5.10}$$

for all bounded measurable functions $f : \Re \to \Re$. That is, the limit as $T \to \infty$, of the random time average specified on the left hand side of relation (4.5.10) equals the spatial average with respect to $\bar{p}$, as given on the right hand side of (4.5.10). Ergodicity is an important property that allows us to describe and

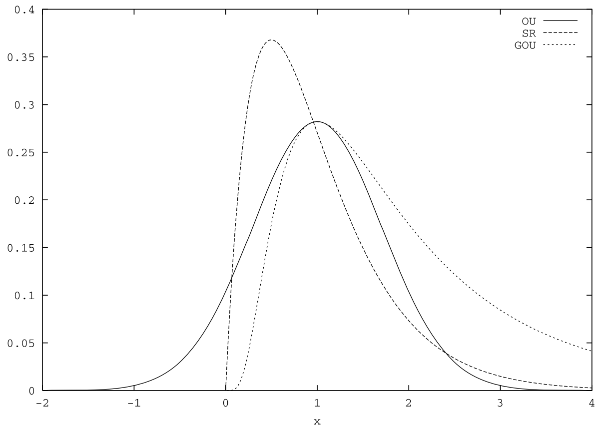



**Fig. 4.5.1.** Stationary density for OU, SR and GOU process

quantify functionals of equilibrium states of diffusion processes. It involves an expectation with respect to the stationary density. However, it does not require the diffusion process to be stationary. The process only needs to have a stationary density but it is not required to start with an initial value having the stationary density as its density.

Note that the widely used lognormal model, described in (4.1.2), does not yield an ergodic process, since it does not have a stationary density. For this reason its use and applicability, for instance, in long term short rate, volatility, credit spread or market activity modeling is limited. For instance, the geometric OU process discussed in (4.2.4) may be a better candidate for this type of modeling when aiming to use a diffusion coefficient that is multiplicative in the state variable.

Now we describe a result that permits us to identify a diffusion process with drift function $a(\cdot)$ and diffusion coefficient function $b(\cdot)$ as being ergodic. For this purpose we introduce the *scale measure* $s : \Re \to \Re^+$ given by

$$s(x) = \exp\left\{-2\int_{y_0}^{x} \frac{a(y)}{b^2(y)}\,dy\right\} \tag{4.5.11}$$

for $x \in \Re$ with $y_0$ as in (4.5.5). The following result can be found in Borodin & Salminen (2002).

**Theorem 4.5.1.** *A diffusion process with scale measure $s(\cdot)$ satisfying the following two properties:*

$$\int_{y_0}^{\infty} s(x)\,dx = \int_{-\infty}^{y_0} s(x)\,dx = \infty \tag{4.5.12}$$

*and*

$$\int_{-\infty}^{\infty} \frac{1}{s(x)\,b^2(x)}\,dx < \infty \tag{4.5.13}$$

*is ergodic and its stationary density $\bar{p}$ is given by the expression (4.5.5).*

Theorem 4.5.1 is formulated for diffusions with a state space that equals the set $\Re$ of all real numbers. In the case of diffusions that are confined to a smaller set of subintervals, the above conditions can be reformulated by including relevant boundary conditions.

### Affine Diffusions (*)

Let us now introduce the important class of *affine diffusions*. An affine function is a linear function added to some constant. Here we have the affine drift function

$$a(x) = \theta_1 + \theta_2\, x \tag{4.5.14}$$

and the affine squared diffusion function

$$b^2(x) = \theta_3 + \theta_4\, x. \tag{4.5.15}$$

The parameter vector $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3, \theta_4)^\top \in \Re^4$ is chosen so that the diffusion process $X = \{X_t,\ t \in [0, \infty)\}$ has a stationary density. In particular, we set

$$\frac{\theta_2}{\theta_4} < 0 \tag{4.5.16}$$

and

$$\eta = \frac{2}{\theta_4}\left(\theta_1 - \frac{\theta_2\, \theta_3}{\theta_4}\right) > 1 \tag{4.5.17}$$

with

$$\theta_3 \geq 0 \quad \text{and} \quad \theta_4 \geq 0. \tag{4.5.18}$$

Then it can be shown that the process $X$ is defined on the interval $(y_0, \infty)$ with $y_0 = -\frac{\theta_3}{\theta_4}$, see Borodin & Salminen (2002). One obtains from (4.5.2) and (4.5.5) the stationary density for such an affine diffusion in the form

$$\bar{p}(x) = \frac{g(x)}{\int_{y_0}^{\infty} g(y)\, dy} \tag{4.5.19}$$

with

$$g(x) = \frac{\left(\frac{-2\,\theta_2}{\theta_4}\right)^{\eta} \left(x + \frac{\theta_3}{\theta_4}\right)^{\eta - 1} \exp\left\{\frac{2\,\theta_2}{\theta_4}\left(x + \frac{\theta_3}{\theta_4}\right)\right\}}{\Gamma(\eta)} \tag{4.5.20}$$

for $x \in (y_0, \infty)$, where $\Gamma(\cdot)$ denotes the gamma function, see (1.2.10). We plot in Fig. 4.5.2 the stationary density $\bar{p}(x)$ for $\theta_1 = -\theta_2 = 1$, $\theta_4 = 1 - \theta_3$ and $\theta_3 \in [0, 0.85]$.

If we denote by $E_\infty$ the expectation under the corresponding stationary distribution, then we have the *stationary mean*

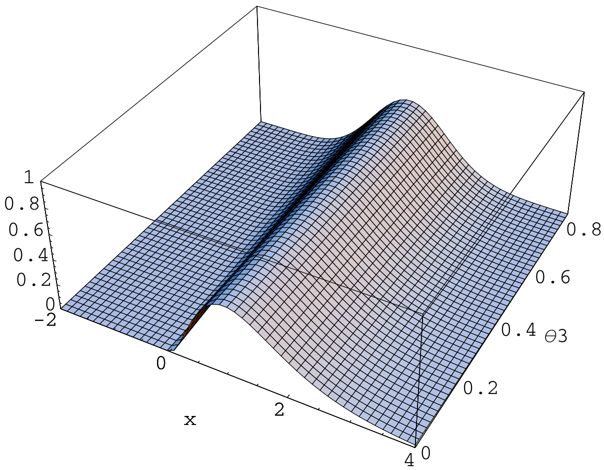



**Fig. 4.5.2.** Stationary density for $\theta_1 = -\theta_2 = 1$, $\theta_4 = 1 - \theta_3$ and $\theta_3 \in [0, 0.85]$

$$E_\infty(X_\infty) = \int_{-\infty}^{\infty} x\, \bar{p}(x)\, dx = -\frac{\theta_1}{\theta_2} \tag{4.5.21}$$

and the *stationary second moment*

$$E_\infty((X_\infty)^2) = \frac{(2\,\theta_1 + \theta_4)\,\theta_1 - \theta_3\,\theta_2}{2\,(\theta_2)^2}. \tag{4.5.22}$$

Obviously, in the case $\theta_4 = 0$ we obtain an OU process, see (4.2.3), with Gaussian stationary density. For the case when $\theta_3$ equals zero we have an SR process, see (4.4.6), of dimension

$$n = 4\,\frac{\theta_1}{\theta_4} > 2, \tag{4.5.23}$$

which has the gamma density, see (1.2.9), as stationary density. The OU and the SR process are ergodic diffusions, which have explicit expressions for their transition densities. This makes these two ergodic affine diffusion processes attractive for a wide range of applications in finance. We remark that at the end of Sect. 4.4 additional diffusion processes are mentioned that also have explicit transition densities and could be linked to ergodic diffusions.

## 4.6 Multi-Dimensional Diffusion Processes (*)

### Vector Diffusion (*)

In financial and insurance markets one observes a large number of quantities concurrently, including equity prices, exchange rates, market indices, volatilities, credit spreads and short rates. These quantities influence each other and can be modeled as a vector stochastic process because interactions need to be considered. For this type of modeling one can use a $d$-dimensional diffusion process

$$\boldsymbol{X} = \left\{ \boldsymbol{X}_t = \left(X_t^1, X_t^2, \ldots, X_t^d\right)^\top, \, t \in [0, \infty) \right\}$$

that generalizes the one-dimensional diffusion process introduced in the previous section. We call such a continuous time process with continuous paths a *vector diffusion*. Here superscripts index the components of the vector.

The transition density for the vector Markov process $\boldsymbol{X}$ to move from the state $\boldsymbol{x} \in \Re^d$ at time $s$ to the state $\boldsymbol{y} \in \Re^d$ at the later time $t$ is denoted by $p(s, \boldsymbol{x}; t, \boldsymbol{y})$. The continuous time Markov property for this vector process can be restated in a similar manner as given in (4.1.3) and we require the following limits to exist for any $\varepsilon > 0$, $s \geq 0$ and $\boldsymbol{x} \in \Re^d$, see (4.3.1)–(4.3.3):

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|\boldsymbol{y}-\boldsymbol{x}|>\varepsilon} p(s, \boldsymbol{x}; t, \boldsymbol{y}) \, d\boldsymbol{y} = 0, \tag{4.6.1}$$

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|\boldsymbol{y}-\boldsymbol{x}|\leq\varepsilon} (\boldsymbol{y} - \boldsymbol{x}) \, p(s, \boldsymbol{x}; t, \boldsymbol{y}) \, d\boldsymbol{y} = \boldsymbol{a}(s, \boldsymbol{x}) \tag{4.6.2}$$

and

$$\lim_{t \downarrow s} \frac{1}{t-s} \int_{|\boldsymbol{y}-\boldsymbol{x}|\leq\varepsilon} (\boldsymbol{y} - \boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})^\top p(s, \boldsymbol{x}; t, \boldsymbol{y}) \, d\boldsymbol{y} = \boldsymbol{S}^\top(s, \boldsymbol{x}) \, \boldsymbol{S}(s, \boldsymbol{x}). \tag{4.6.3}$$

Here $\boldsymbol{a}$ is a $d$-dimensional vector valued function and $\boldsymbol{D} = [d^{i,j}]_{i,j=1}^d = \boldsymbol{S}^\top \boldsymbol{S}$ is a symmetric $d \times d$-matrix valued function. Each component of these functions must satisfy appropriate measurability and integrability conditions, see Stroock & Varadhan (1982). We used above the Euclidean norm $|\cdot|$, see (1.4.63), and interpret the vectors as column vectors, for example, $(\boldsymbol{y} - \boldsymbol{x})(\boldsymbol{y} - \boldsymbol{x})^\top$ is a $d \times d$-matrix with $(i, j)$th component $(y_i - x_i)(y_j - x_j)$. The *drift vector* $\boldsymbol{a}$ and the *covariance matrix* $\boldsymbol{D} = \boldsymbol{S}^\top \boldsymbol{S}$ have similar interpretations to their one-dimensional counterparts in the previous section. However, we note that the components of $\boldsymbol{D}$ are the conditional covariances or variances of the increments of corresponding components of the vector diffusion, that is

$$d^{i,j}(s, x) = \lim_{t \downarrow s} \frac{1}{t-s} E\left( \left(X_t^i - X_s^i\right) \left(X_t^j - X_s^j\right) \Big| X_s = x \right),$$

where $d^{i,j}(s, x) = d^{j,i}(s, x)$. They indicate which components of the vector diffusion are correlated.

### Kolmogorov Equations (*)

For vector diffusions the transition densities satisfy the multi-dimensional *Kolmogorov forward equation*, also known as Fokker-Planck equation, given by

$$\frac{\partial p(s,\boldsymbol{x};t,\boldsymbol{y})}{\partial t} + \sum_{i=1}^{d} \frac{\partial}{\partial y_i} \left( a^i(t,\boldsymbol{y}) p(s,\boldsymbol{x};t,\boldsymbol{y}) \right)$$

$$- \frac{1}{2} \sum_{i,j=1}^{d} \frac{\partial^2}{\partial y_i \partial y_j} \left( d^{i,j}(t,\boldsymbol{y}) p(s,\boldsymbol{x};t,\boldsymbol{y}) \right) = 0 \qquad (4.6.4)$$

for $(s,\boldsymbol{x}) \in (0,\infty) \times \Re^d$ fixed and $(t,\boldsymbol{y}) \in (s,\infty) \times \Re^d$ with the initial condition

$$\lim_{t \downarrow s} p(s,\boldsymbol{x};t,\boldsymbol{y}) = \delta(\boldsymbol{x}-\boldsymbol{y})$$

for $\boldsymbol{x},\boldsymbol{y} \in \Re^d$. Here $\delta(\boldsymbol{z})$ denotes again the *Dirac delta function* but now on $\Re^d$, which defines a measure that has a mass of one concentrated at the point $(0,\ldots,0)^\top \in \Re^d$.

We can write the parabolic partial differential equation (4.6.4) more compactly in operator form as

$$\frac{\partial p(s,\boldsymbol{x};t,\boldsymbol{y})}{\partial t} - \mathcal{L}^* p(s,\boldsymbol{x};t,\boldsymbol{y}) = 0$$

for $(s,\boldsymbol{x}) \in [0,\infty) \times \Re^d$ fixed and $(t,\boldsymbol{y}) \in (s,\infty) \times \Re^d$. Here $\mathcal{L}^*$ is the formal adjoint of the operator $\mathcal{L}^0$ defined as

$$\mathcal{L}^0 u(s,\boldsymbol{x}) = \sum_{i=1}^{d} a^i(s,\boldsymbol{x}) \frac{\partial u(s,\boldsymbol{x})}{\partial x_i} + \frac{1}{2} \sum_{i,j=1}^{d} d^{i,j}(s,\boldsymbol{x}) \frac{\partial^2 u(s,\boldsymbol{x})}{\partial x_i \partial x_j} \qquad (4.6.5)$$

for $(s,\boldsymbol{x}) \in (0,\infty) \times \Re^d$. The *Kolmogorov backward equation*, which as previously mentioned plays a central role in derivative pricing, is given by

$$\frac{\partial u(s,\boldsymbol{x})}{\partial s} + \mathcal{L}^0\, u(s,\boldsymbol{x}) = 0 \qquad (4.6.6)$$

for $(s,\boldsymbol{x}) \in (0,t) \times \Re^d$ with $u(s,\boldsymbol{x}) = p(s,\boldsymbol{x};t,\boldsymbol{y})$ for fixed $t \in [0,\infty)$ and $\boldsymbol{y} \in \Re^d$.

To model and analyze the quantities in a financial market purely via corresponding partial differential equations is rather complex. A more elegant and also more general framework for modeling stochastic dynamics is provided when using stochastic calculus, which will be introduced in the following chapters.

## 4.7 Exercises for Chapter 4

**4.1.** Verify that the standard Ornstein-Uhlenbeck process is a diffusion process with stationary density and identify its mean and its variance.

**4.2.** Identify the drift and diffusion coefficient for the standard Wiener process as a specific diffusion process.

**4.3.** Compute the drift and diffusion coefficients for the standard Ornstein-Uhlenbeck process.

**4.4.** Prove that the transition density of the standard Wiener process solves the Kolmogorov forward equation and the Kolmogorov backward equation.

**4.5.** Formulate the Kolmogorov forward equation for the transition density of the standard Ornstein-Uhlenbeck process.

**4.6.** Verify that the transition density of the standard Ornstein-Uhlenbeck process satisfies the corresponding Kolmogorov backward equation.

**4.7.** Determine the stationary density for the standard Ornstein-Uhlenbeck process.

**4.8.** Does geometric Brownian motion have a stationary density?

**4.9.** Verify whether the geometric Ornstein-Uhlenbeck process has a stationary density.

**4.10.** (*) Is the geometric Brownian motion an ergodic process?

**4.11.** (*) Prove that the transition density $p(s, x; t, y)$ of the standard Wiener process satisfies the Chapman-Kolmogorov equation.

**4.12.** (*) Is the standard Ornstein-Uhlenbeck process an ergodic process?

**4.13.** (*) Show that the stationary density $\bar{p}$ of a one dimensional diffusion process solves the time-independent Kolmogorov forward equation.

**4.14.** (*) Show that a geometric Brownian motion with growth rate $g$ and volatility $b$ has the drift $a(s, x) = x(g + \frac{1}{2}b^2)$.

# 5

# Martingales and Stochastic Integrals

In this chapter we consider a class of continuous stochastic processes, called martingales, which play a central role in finance. We also define the gains realized from trading as a stochastic integral. Stochastic integration and martingales provide key tools for the analysis of the continuous time evolution of financial markets.

## 5.1 Martingales

One of the fundamental concepts in modern finance is the notion of a martingale. This is a stochastic process that, with its last observed value, provides the best forecast for its future values. Martingales exhibit the property of having no systematic trends in their dynamics. It is obvious that financial quantities, such as asset prices, are driven primarily by information. Forecasting a quantity, for example, the value of a derivative price when expressed in units of the market portfolio, is strongly dependent on the information that is available at the present time. This forces one to use a detailed notion for the information structure related to the evolution of the underlying stochastic processes.

### Information Sets and Filtrations

On a given probability space $(\Omega, \mathcal{A}, P)$, as introduced in Sect. 1.1, let us consider a financial market model that is based on the observation of a continuous time stochastic vector process $\boldsymbol{X} = \{\boldsymbol{X}_t \in \Re^n,\, t \in [0, \infty)\},\ n \in \mathcal{N}$, typically expressing asset price processes. We denote by $\hat{\mathcal{A}}_t$ the time $t$ *information set*, which is the sigma-algebra of events that are known to the market participants at time $t \in [0, \infty)$. Our interpretation of $\hat{\mathcal{A}}_t$ is that it represents the information obtained from the values of the vector process $\boldsymbol{X}$ up to time $t$. More precisely, it is the sigma-algebra

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

$$\hat{\mathcal{A}}_t = \sigma\{\boldsymbol{X}_s : s \in [0,t]\}$$

generated from all observations of $\boldsymbol{X}$ in the market up to time $t$. In a general financial market model the components of $\boldsymbol{X}$ could include diverse quantities, for instance, security prices, interest rates, indicators for certain political events, market activity, corporate data, employment figures, insurance claims, balance sheets of companies or trade balances.

Assuming that information is not lost, then the increasing family

$$\underline{\hat{\mathcal{A}}} = \{\hat{\mathcal{A}}_t,\, t \in [0,\infty)\}$$

of information sets $\hat{\mathcal{A}}_t$, which are sub-sigma-algebras of $\hat{\mathcal{A}}_\infty$ satisfy, for any sequence $0 \leq t_1 < t_2 < \ldots < \infty$ of observation times, the relation $\hat{\mathcal{A}}_{t_1} \subseteq \hat{\mathcal{A}}_{t_2} \subseteq \ldots \subseteq \hat{\mathcal{A}}_\infty = \cup_{t\in[0,\infty)}\hat{\mathcal{A}}_t$.

Furthermore, to avoid technical subtleties, we introduce the information set $\mathcal{A}_t$ as the *augmented* sigma-algebra of $\hat{\mathcal{A}}_t$ for each $t \in [0,\infty)$. This means that it is augmented by every null set in $\hat{\mathcal{A}}_\infty$ such that it belongs to $\mathcal{A}_0$, and so to each $\mathcal{A}_t$. We define $\mathcal{A}_{t+} = \cap_{\varepsilon>0}\mathcal{A}_{t+\varepsilon}$ to be the sigma-algebra of events immediately after $t \in [0,\infty)$. We say that the family $\underline{\mathcal{A}} = \{\mathcal{A}_t,\, t \in [0,\infty)\}$ is *right continuous* if $\mathcal{A}_t = \mathcal{A}_{t+}$ holds for every $t \in [0,\infty)$. Such a right-continuous family $\underline{\mathcal{A}} = \{\mathcal{A}_t,\, t \in [0,\infty)\}$ of information sets we call a *filtration*. Thus, a filtration models the evolution of information as it becomes available over time. For simplicity, we define $\mathcal{A}$ as the smallest sigma-algebra that contains $\mathcal{A}_\infty = \cup_{t\in[0,\infty)}\mathcal{A}_t$.

The above technical assumptions allow convenient mathematical derivations and do not restrict our practical modeling potential. From now on, if not stated otherwise, we shall assume a *filtered probability space* $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ to be given, where the filtration $\underline{\mathcal{A}}$ characterizes the evolution of the corresponding information. The capturing of the evolution of this information is essential for the modeling of financial markets since it is information that drives most of its dynamics.

Any given stochastic process $Y = \{Y_t,\, t \in [0,\infty)\}$ generates a filtration $\mathcal{A}^Y = \{\mathcal{A}^Y_t,\, t \in [0,\infty)\}$. Here $\mathcal{A}^Y_t = \sigma\{Y_s : s \in [0,t]\}$ is the information set, that is the sigma-algebra, generated by $Y$ up to time $t$. This information set can be interpreted as a complete record of all movements of the process $Y$ up until time $t$. $\mathcal{A}^Y$ is also called the *natural filtration* for the process $Y$. For a given model with a vector process $\boldsymbol{X}$ that describes the total evolution of the model and, thus, the corresponding increasing family of information sets, we write $\underline{\mathcal{A}} = \underline{\mathcal{A}}^{\boldsymbol{X}}$ and set $\mathcal{A}_t = \mathcal{A}^{\boldsymbol{X}}_t$, similarly as above.

If for a process $Z = \{Z_t,\, t \in [0,\infty)\}$ and each time $t \in [0,\infty)$ the random variable $Z_t$ is $\mathcal{A}^{\boldsymbol{X}}_t$-measurable, then $Z$ is called *adapted* to $\underline{\mathcal{A}}^{\boldsymbol{X}} = \{\mathcal{A}^{\boldsymbol{X}}_t,\, t \in [0,\infty)\}$. In intuitive terms this means that the history of the process $Z$ until time $t$ is covered by the information set $\mathcal{A}^{\boldsymbol{X}}_t$. As a consequence, for an $\underline{\mathcal{A}}^{\boldsymbol{X}}$-adapted process $Z$ the value $Z_t$ is known, given the information set $\mathcal{A}^{\boldsymbol{X}}_t$ up to and including time $t$. We mention that the completeness of the information set $\mathcal{A}^{\boldsymbol{X}}_t$, which includes all null events, allows us to conclude that for two random variables $Z_1$ and $Z_2$, where $Z_1 = Z_2$ a.s. and $Z_1$ is $\mathcal{A}_t^{\boldsymbol{X}}$ measurable, $Z_2$ is also $\mathcal{A}_t^{\boldsymbol{X}}$-measurable.

If the process $\boldsymbol{X}$ is Markovian, then the relevant information needed to determine properties of its future values reduces to the knowledge of the value $\boldsymbol{X}_t$ at the present time $t$. This makes it possible to express and store the relevant information in a compact form. It also highlights the importance of Markovianity for the tractability of a wide range of financial models.

In financial modeling we shall typically use later a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, where the sources of continuous uncertainty are independent standard Wiener processes $W^1, W^2, \dots, W^m$ and the sources of event driven uncertainty are independent Poisson processes $N^{m+1}, N^{m+2}, \dots, N^d$, $d \in \{1, 2, \dots\}$, $m \in \{1, 2, \dots, d\}$. We shall always assume that these Wiener and Poisson processes are $\underline{\mathcal{A}}$-adapted and that their increments $(W_t^j - W_s^j)$ are independent of $\mathcal{A}_s$, see (1.1.16), for $t \in [0, \infty)$, $s \in [0, t]$ and $j \in \{1, 2, \dots, m\}$. We call then $\boldsymbol{W} = \{\boldsymbol{W}_t = (W_t^1, W_t^2, \dots, W_t^m)^\top,\ t \in [0, \infty)\}$ an $m$-dimensional standard Wiener process on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ or an $(\underline{\mathcal{A}}, P)$-Wiener process.

### Continuous Time Martingales

In financial markets, investors have to determine best estimates for the actual value of future payoffs. If they were to use different information sets, then they might generate different value estimates. For simplicity, let us assume that they all use the same information sets. Furthermore, a value estimate needs to be based on a corresponding *benchmark* or *numeraire*, which provides the units in which the investor formulates his or her best estimates. We shall later discuss cases where one uses the savings account or the market portfolio as numeraire. Finally, an investor has also to employ a probability measure for forming some expectation when identifying the best estimate, as we shall see below. Let us use the numeraire for which it is appropriate to form an expectation under the real world probability measure when searching for the best estimate of a future payoff. We shall see later that an appropriate numeraire is the market portfolio. More generally, given an information set, a probability measure and a numeraire, we shall ask what is at present the best estimate for the value of a future cash flow or payoff.

To answer this question in a mathematically precise manner we define the quantity $F_s$ for $s \in [0, \infty)$ as the least-squares estimate, see (1.3.72), of the future value $X_t$ at the later time $t \in [s, \infty)$ under the information given by $\mathcal{A}_s$. This best estimate is $\mathcal{A}_s$-measurable and minimizes the expected least-squares error

$$\varepsilon_s = E\left((X_t - F_s)^2\right)$$

over all possible $\mathcal{A}_s$-measurable estimates. Note that we need here to assume that $X_t$ is square integrable, see (1.3.7). The random variable $F_s$ is simply the least-squares projection of $X_t$ given the information at time $s \in [0, t]$. It is obtained by the conditional expectation

$$F_s = E(X_t \,\big|\, \mathcal{A}_s), \tag{5.1.1}$$

for all $s \in [0, t]$.

In a price system a candidate for a reasonable price at time $s \in [0, t]$ for the future value $X_t$ at time $t$ is the least-squares estimate $F_s$ that can be formed on the basis of the information contained in $\mathcal{A}_s$. This means, one obtains realistic prices when setting $X_s = F_s$ by forming the price process $X = \{X_t,\, t \in [0, \infty)\}$ which satisfies the conditional expectation

$$X_s = E(X_t \,\big|\, \mathcal{A}_s) \tag{5.1.2}$$

for all $s \in [0, t]$ and $t \in [0, \infty)$.

**Definition 5.1.1.** *We call a continuous time stochastic process* $X = \{X_t,\, t \in [0, \infty)\}$*, which satisfies the property* (5.1.2) *and the integrability condition*

$$E(|X_t|) < \infty \tag{5.1.3}$$

*for all* $t \in [0, \infty)$*, a* martingale *or more precisely an* $(\underline{\mathcal{A}}, P)$-martingale*.*

If for a martingale $X$ in addition the random variable $X_t$ is square integrable for all $t \in [0, \infty)$, that is

$$E\left(|X_t|^2\right) < \infty \tag{5.1.4}$$

for all $t \in [0, \infty)$, then we call $X$ a *square integrable martingale*. Note by (5.1.1) and (5.1.2) that for a square integrable martingale the least-squares estimate of its future values is always given by its last available observation.

A martingale is defined with respect to a given filtration $\underline{\mathcal{A}}$, which denotes the family of relevant information sets, and a probability measure $P$, which expresses the likelihood of events. The conditional expectation is then taken under $P$. Since both ingredients are essential we shall call a martingale an $(\underline{\mathcal{A}}, P)$-*martingale* if it is defined with respect to the filtration $\underline{\mathcal{A}}$ and the probability measure $P$. This is sometimes important because it is not always clear from the context which filtration and probability measure are chosen. If one changes the filtration $\underline{\mathcal{A}}$ or the probability measure $P$, then a given martingale will usually no longer remain a martingale.

The martingale relation (5.1.2) is fundamental in finance, in particular, in derivative pricing. Under the benchmark approach we shall ask later derivative prices, when expressed in units of the benchmark, to form martingales. Different pricing rules are obtained by selecting different reference units or numeraires, an issue that will be discussed later in detail.

### Examples of Martingales

As an example of a continuous time martingale, let us consider a Wiener process $W = \{W_t,\, t \in [0, \infty)\}$ on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$.

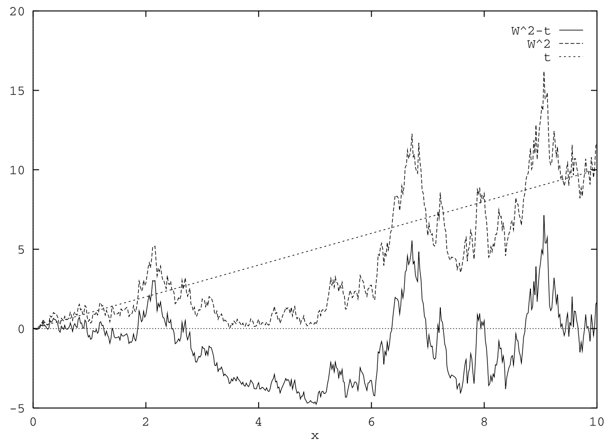



**Fig. 5.1.1.** Paths of $X_t = W_t^2 - t$, $W_t^2$ and $t$

Here, as previously mentioned, we assume that $W_{t+h} - W_t$ is independent of $\mathcal{A}_t$ for all $t \in [0, \infty)$ and $h \in [0, \infty)$. Furthermore, the natural filtration $\mathcal{A}^W$ of $W$ is such that $\mathcal{A}_t^W \subseteq \mathcal{A}_t$ for each $t \in [0, \infty)$.

Note that $W$ is $\underline{\mathcal{A}}$-adapted, which means that $W_t$ is $\mathcal{A}_t$-measurable for $t \in [0, \infty)$. We can show by the linearity and independence properties of conditional expectations, see (1.3.69) and (1.3.67), that

$$\begin{aligned} E(W_t \,|\, \mathcal{A}_s) &= E\left(W_t - W_s \,|\, \mathcal{A}_s\right) + E\left(W_s \,|\, \mathcal{A}_s\right) \\ &= E\left(W_t - W_s\right) + E\left(W_s \,|\, \mathcal{A}_s\right) \\ &= W_s \end{aligned} \tag{5.1.5}$$

for $s \in [0, \infty)$ and $t \in [s, \infty)$. From (5.1.5) it follows by Definition 5.1.1 that the above Wiener process $W$ is a martingale, or more precisely an $(\underline{\mathcal{A}}, P)$-martingale.

There are many other continuous time stochastic processes that form martingales. For example, using again the standard Wiener process $W$ it can be demonstrated that the process

$$X = \left\{X_t = W_t^2 - t,\, t \in [0, \infty)\right\} \tag{5.1.6}$$

is an $(\underline{\mathcal{A}}, P)$-martingale. In Fig. 5.1.1 we show a typical path for this process together with $W_t^2$ and $t$.

The process

$$\bar{X} = \left\{\bar{X}_t = \exp\left\{\sigma W_t - \frac{1}{2}\,\sigma^2 t\right\},\, t \in [0, \infty)\right\},$$

which is an exponential of a transformed Wiener process, is also an $(\underline{\mathcal{A}}, P)$-martingale. Note that this is a specific geometric Brownian motion with

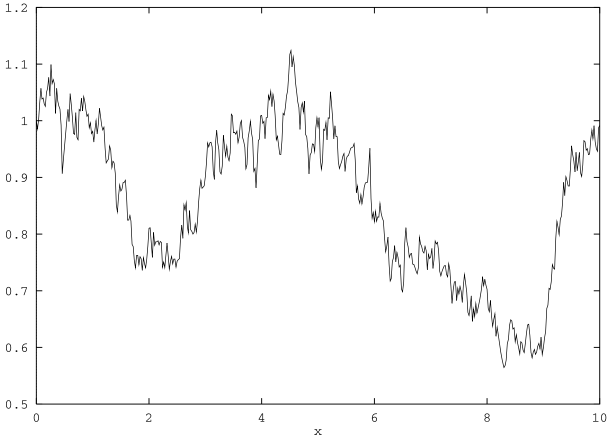



**Fig. 5.1.2.** Path of $\bar{X}_t = \exp\left\{\sigma W_t - \frac{1}{2}\sigma^2 t\right\}$

volatility $\sigma$, negative growth rate $\mu = -\frac{1}{2}\sigma^2$ and initial value $\bar{X}_0 = 1$. Figure 5.1.2 displays a sample path for this process with volatility $\sigma = 0.2$.

## Super- and Submartingales

In practice asset prices are usually not completely trendless. For instance, the price of a zero coupon bond, which pays one dollar at a fixed maturity date, increases on average over time until it reaches at maturity the value one. These types of systematically trending stochastic processes are captured by the following definition of super- and submartingales.

**Definition 5.1.2.** *One calls an $\underline{\mathcal{A}}$-adapted process $X = \{X_t,\, t \in [0,\infty)\}$ an $(\underline{\mathcal{A}}, P)$-*supermartingale (submartingale) *if*

$$X_s \overset{(\leq)}{\geq} E\left(X_t \,\middle|\, \mathcal{A}_s\right) \tag{5.1.7}$$

*and*

$$E(|X_t|) < \infty \tag{5.1.8}$$

*for $s \in [0,\infty)$ and $t \in [s,\infty)$.*

This means, on average, a supermartingale (submartingale) decreases (increases) its value over time. In comparison with a martingale the equality in (5.1.2) is replaced by the inequality (5.1.7). We call a supermartingale (submartingale) a *strict supermartingale* (submartingale) if the inequality in (5.1.7) is a strict inequality.

As an example for a submartingale we show in Fig. 5.1.3 for some geometric Brownian motion with

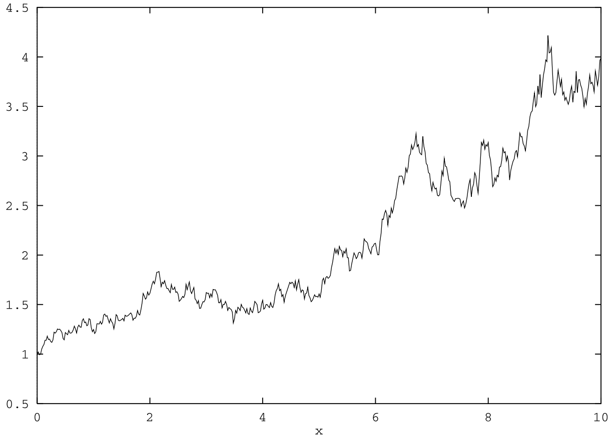



**Fig. 5.1.3.** Path of $X_t$ for a submartingale

$$X_t = \exp\left\{\left(r - \frac{1}{2}\,\sigma^2\right) t + \sigma\, W_t\right\} \tag{5.1.9}$$

a sample path over a period of ten years with expected rate of return $r = 0.05$ and volatility $\sigma = 0.2$. This example illustrates some features that are typical for asset price scenarios. They seem to exhibit larger fluctuations for larger asset price values, as is the case for the S&P500 index shown in Fig. 3.1.1. If the submartingale $X$ is discounted by the process $B = \{B_t = \exp\{r\,t\},\, t \in [0,\infty)\}$, which is simply a savings account with continuously compounding constant interest rate $r > 0$, then the discounted process $\bar{X} = \{\bar{X}_t = \frac{X_t}{B_t},\, t \in [0,\infty)\}$ is a martingale. Let us mention that Fig. 5.1.2 displays the sample path for $\bar{X}_t$, where $X_t$ is shown in Fig. 5.1.3.

As we shall see later in Chaps. 9 to 14, in financial market models supermartingales play a natural role. They appear when securities are expressed in units of a particular benchmark, which is the, so-called, *growth optimal portfolio* (GOP). This is the portfolio that maximizes the expected logarithm of its value at future dates, see Kelly (1956), Long (1990). By interpreting a diversified market index as the GOP it has been suggested in Platen (2004c) that the savings account $B$, when expressed in units of the market index should be modeled to form a strict supermartingale and not a martingale, as the classical risk neutral theory assumes, and will be explained in Chap. 9.

### Compensated Poisson Process

In Fig. 3.5.1 we plotted the path of a Poisson process $N = \{N_t,\, t \in [0,\infty)\}$ with intensity $\lambda > 0$, see Definition 3.5.1. We have assumed for any Poisson process $N$ that $N$ is $\underline{\mathcal{A}}$-adapted and such that for $t \in [0,\infty)$ and $h \in [0, T-t]$ the $\mathcal{A}_{t+h}$-measurable random variable $N_{t+h} - N_t$ is independent of $\mathcal{A}_t$. We can then show for $0 \le s < t < \infty$ by using (3.5.2) that

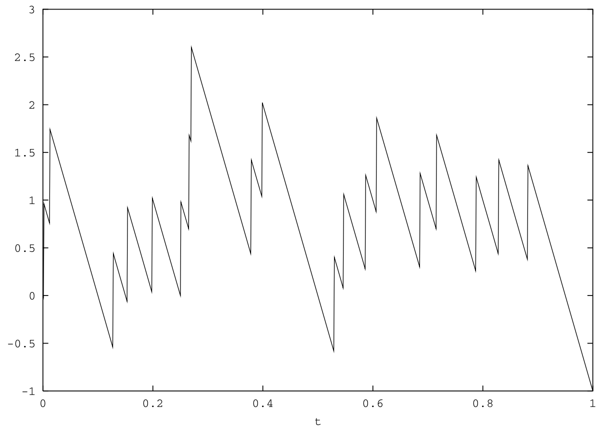



**Fig. 5.1.4.** Path of a compensated Poisson process

$$E\left(N_t \,\middle|\, \mathcal{A}_s\right) = E\left(N_t - N_s \,\middle|\, \mathcal{A}_s\right) + N_s = E\left(N_t - N_s\right) + N_s$$
$$= \lambda(t-s) + N_s \geq N_s, \tag{5.1.10}$$

which proves that the Poisson process is a submartingale.

On the other hand, the *compensated Poisson process* $q = \{q_t,\, t \in [0, \infty)\}$ with

$$q_t = N_t - \lambda\, t \tag{5.1.11}$$

is a martingale since we have by similar arguments as in (5.1.10)

$$E\left(q_t \,\middle|\, \mathcal{A}_s\right) = E\left(q_t - q_s \,\middle|\, \mathcal{A}_s\right) + q_s$$
$$= E\left(N_t - N_s\right) - \lambda\,(t-s) + q_s = q_s \tag{5.1.12}$$

for $0 \leq s \leq t < \infty$. In Fig. 5.1.4 we plot the path of a compensated Poisson process $q$ with intensity $\lambda = 20$, where Fig. 3.5.1 shows the corresponding trajectory of the Poisson process $N$.

## Stopping Times

Random times naturally appear in financial and insurance applications, for instance, as time of default of a company. We refer to Sect. 3.7 for an insurance example. Also the first hitting time of a critical barrier by an underlying asset price is a random time. Since the information structure is essential in stochastic modeling such random times have to be properly defined.

This leads us to the notion of stopping times. Let us consider a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ as introduced above.

**Definition 5.1.3.** *A random variable* $\tau : \Omega \to [0, \infty)$ *is called a* stopping time *with respect to the filtration* $\underline{\mathcal{A}}$ *if for all* $t \in [0, \infty)$

$$\{\tau \leq t\} \in \mathcal{A}_t. \tag{5.1.13}$$

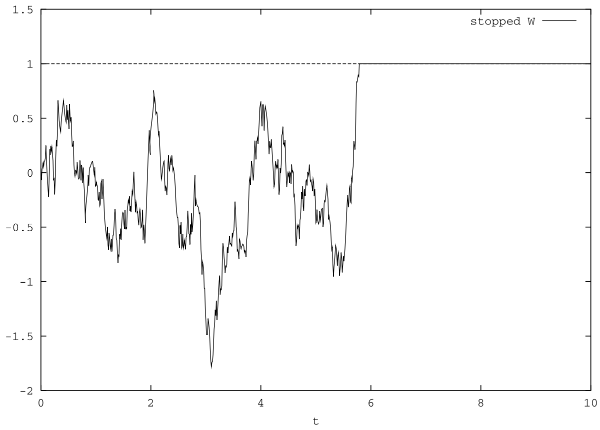



**Fig. 5.1.5.** First hitting time of a Wiener path

The relation (5.1.13) means that for all $\omega \in \Omega$ the event $\tau \leq t$ is in $\mathcal{A}_t$, which expresses the fact that it is $\mathcal{A}_t$-measurable and thus observable at time $t$. The information set, that is, the sigma-algebra associated with a stopping time $\tau$ is defined as

$$\mathcal{A}_\tau = \sigma \{A \in \mathcal{A} : \; A \cap \{\tau \leq t\} \in \mathcal{A}_t \quad \text{for} \quad t \in [0, \infty)\}. \tag{5.1.14}$$

It represents the information available before and at the stopping time $\tau$. For instance, the $k$th jump time $\tau_k$ of a Poisson process $N$, as defined in Sect. 3.5, is a stopping time. This could be the time when the $k$th company collapses in a given year. One can show that a counting process is adapted if and only if the associated jump times are stopping times.

The first time

$$\tau(a) = \inf\{t \geq 0 : \; W_t = a\} \tag{5.1.15}$$

when a Wiener process $W$ reaches a level $a \in \Re$ is a stopping time. In Fig. 5.1.5 we display the first time $\tau(1.0) \approx 5.8$ of a Wiener path hitting the level $a = 1.0$. Similarly, the default time of a company is a stopping time.

## Predictable Processes

The allocation of assets in a portfolio can, in practice, only be performed in a predictable way. That means, the investor has to decide in advance what allocation will be pursued. To make this notion of predictability precise for stopping times, we call a sigma-algebra *predictable* when it is generated by left-continuous $\underline{\mathcal{A}}$-adapted processes with right hand limits. Roughly speaking, we exclude in a predictable sigma-algebra all information about the time instant when a sudden not predictable event, like a default, occurs. Note however, immediately after the event a predictable sigma-algebra already contains also this information. A stochastic process $X = \{X_t, \, t \in [0, \infty)\}$, where $X_\tau$ is for each stopping time $\tau$ measurable with respect to a predictable sigma-algebra, is called *predictable*. For instance, all continuous stochastic processes are predictable. From a right-continuous process with left hand limits $X = \{X_t,\, t \in [0,\infty)\}$ we obtain its predictable version $\tilde{X} = \{\tilde{X}_t,\, t \in [0,\infty)\}$ by taking at each time point the left hand limit, that is

$$\tilde{X}_t = X_{t-} \tag{5.1.16}$$

for all $t \in [0,\infty)$. Later when we form stochastic integrals we shall typically request that the integrands are predictable processes. In the case when a given potential integrand is not predictable, then its left-continuous version is chosen as integrand. This is similar to the natural request that an investor has to decide about his or her portfolio allocation of stocks at the beginning of any trading period and cannot revise it afterwards.

A stopping time is called *predictable*, if $\mathcal{A}_\tau$ is predictable. This means, $\mathcal{A}_\tau$ is generated by left-continuous stochastic processes with right hand limits. A stopping time that is not predictable is called *inaccessible*. The jump times of a Poisson process are inaccessible. Here $\mathcal{A}_\tau$ cannot be generated by left-continuous processes. However, the first hitting time $\tau(a)$ of the continuous Wiener process $W$, given in (5.1.15), is predictable.

### Properties of Stopping Times (*)

For $a, b \in \Re$ we employ the notation $a \wedge b = \min(a,b)$ and $a \vee b = \max(a,b)$. One can derive the following useful properties of stopping times $\tau$ and $\tau'$, see Karatzas & Shreve (1991) and Elliott (1982).

(i) $\tau$ is $\mathcal{A}_\tau$-measurable.
(ii) For a continuous $\underline{\mathcal{A}}$-adapted process $X = \{X_t,\, t \in [0,\infty)\}$ the random variable $X_\tau$ is $\mathcal{A}_\tau$-measurable.
(iii) If $P(\tau \le \tau') = 1$, then $\mathcal{A}_\tau \subseteq \mathcal{A}_{\tau'}$.
(iv) The random variables $\tau \wedge \tau'$, $\tau \vee \tau'$ and $(\tau + \tau')$ are stopping times.
(v) If for a real valued random variable $Y$ we have $E(|Y|) < \infty$ and $P(\tau \le \tau') = 1$, then

$$E(Y \,|\, \mathcal{A}_\tau) = E(Y \,|\, \mathcal{A}_{\tau \wedge \tau'}) \tag{5.1.17}$$

and

$$E\left(E(Y \,|\, \mathcal{A}_\tau) \,\big|\, \mathcal{A}_{\tau'}\right) = E(Y \,|\, \mathcal{A}_\tau). \tag{5.1.18}$$

### Optional Sampling Theorem (*)

If $X = \{X_t,\, t \in [0,\infty)\}$ is a right continuous $(\underline{\mathcal{A}}, P)$-supermartingale, then the supermartingale property (5.1.2) is also true if the times $s$ and $t$ in (5.1.2) are stopping times. More precisely, Doob's *Optional Sampling Theorem* states the following result, see Doob (1953), Elliott (1982) or Karatzas & Shreve (1991).

**Theorem 5.1.4.** (Doob) *If $X = \{X_t,\, t \in [0,\infty)\}$ is a right continuous $(\underline{\mathcal{A}}, P)$-supermartingale on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, then it holds for two bounded stopping times $\tau$ and $\tau'$ with $\tau \le \tau'$ almost surely that*

$$E(X_{\tau'} \,|\, \mathcal{A}_\tau) \le X_\tau \tag{5.1.19}$$

*almost surely. Furthermore, if $X$ is also an $(\underline{\mathcal{A}}, P)$-martingale, then equality holds in (5.1.19).*

This theorem is important if one wants to apply a pricing rule at a stopping time or the payoff that one aims to price matures at a stopping time. Such case arises for American options that allow exercising the payoff at any time prior to maturity.

### Martingale Inequalities (*)

For a given underlying financial quantity, or more generally, a given stochastic process $X$, it is important to have some upper bounds for its maximum. If $X = \{X_t,\, t \in [0,\infty)\}$ is a right continuous supermartingale, then it can be shown, see Doob (1953) or Elliott (1982), that for any $\lambda > 0$ it holds

$$\lambda\, P\left(\sup_{t\in[0,\infty)} X_t \ge \lambda \,\Big|\, \mathcal{A}_0\right) \le E\left(X_0 \,\big|\, \mathcal{A}_0\right) + E\left(\max(0, -X_0) \,\big|\, \mathcal{A}_0\right). \tag{5.1.20}$$

By exploiting the martingale property (5.1.19) one can prove the following powerful *martingale inequalities*, see Doob (1953) or Elliott (1982). A continuous martingale $X = \{X_t,\, t \in [0,\infty)\}$ with finite $p$th moment satisfies the *maximal martingale inequality*

$$P\left(\sup_{s\in[0,t]} |X_s| > a\right) \le \frac{1}{a^p}\, E(|X_t|^p) \tag{5.1.21}$$

and the *Doob inequality*

$$E\left(\sup_{s\in[0,t]} |X_s|^p\right) \le \left(\frac{p}{p-1}\right)^p E(|X_t|^p) \tag{5.1.22}$$

for $a > 0$, $p > 1$ and $t \in [0,\infty)$. If $X$ is a continuous martingale, then the maximal martingale inequality provides an estimate for the probability that a level $a$ will be exceeded by the maximum of $X$. In particular the Doob inequality provides for $p = 2$ for the squared maximum the estimate

$$E\left(\sup_{s\in[0,t]} |X_s|^2\right) \le 4\, E\left(|X_t|^2\right)$$

for $t \in [0,\infty)$. These inequalities are important for deriving a number of fundamental results in stochastic calculus and quantitative finance.

## 5.2 Quadratic Variation and Covariation

### Quadratic Variation

The notion of the, so-called, *quadratic variation* of a given stochastic process $X$ plays a fundamental role in stochastic calculus and, therefore, in finance as well. It is a characteristic of the fluctuating part of a stochastic process and can be easily observed. In this capacity it will be useful for measuring locally in time the risk of an asset price.

To introduce this notion in a simple manner let us consider an *equidistant time discretization*

$$\{t_k = k\,h : \; k \in \{0, 1, \ldots\}\}, \tag{5.2.1}$$

with small time steps of lengths $h > 0$, such that $0 = t_0 < t_1 < t_2 < \ldots$. Thus, we have the discretization times $t_k = k\,h$ for $k \in \{0, 1, \ldots\}$. The specific structure of the time discretization is in fact not essential for the definition of the quadratic variation that we shall use, as long as the maximum time step size vanishes a.s. when approaching the limit. We employ the equidistant time discretization here to simplify our presentation. Other time discretizations with vanishing step size yield the same limit.

For a given stochastic process $X$ the *quadratic variation process* $[X] = \{[X]_t, \; t \in [0, \infty)\}$ is defined as the limit in probability, see (2.7.1), as $h \to 0$ of the sums of squared increments of the process $X$, provided this limit exists and is unique. For details we refer to Jacod & Shiryaev (2003) and Protter (2004). For instance, for semimartingales, which form a very general class of stochastic processes that we shall introduce in Sect. 5.5, the quadratic variation is uniquely defined. We have at time $t$ the *quadratic variation*

$$[X]_t \stackrel{P}{=} \lim_{h \to 0} [X]_{h,t}, \tag{5.2.2}$$

where the *approximate quadratic variation* $[X]_{h,t}$ is given by the sum

$$[X]_{h,t} = \sum_{k=1}^{i_t} (X_{t_k} - X_{t_{k-1}})^2. \tag{5.2.3}$$

Here $i_t$ denotes the integer

$$i_t = \max\{k \in \mathcal{N} : \; t_k \leq t\} \tag{5.2.4}$$

of the last discretization point before or including $t \in [0, \infty)$.

### Examples of Quadratic Variations

As an example, Fig. 5.2.1 shows for a standard Wiener process $W = \{W_t, t \in [0, \infty)\}$ a sample path and its approximate quadratic variation $[W]_{h,t}$ with time step size $h = 0.02$ on the interval $[0, 10]$. Note that the approximate

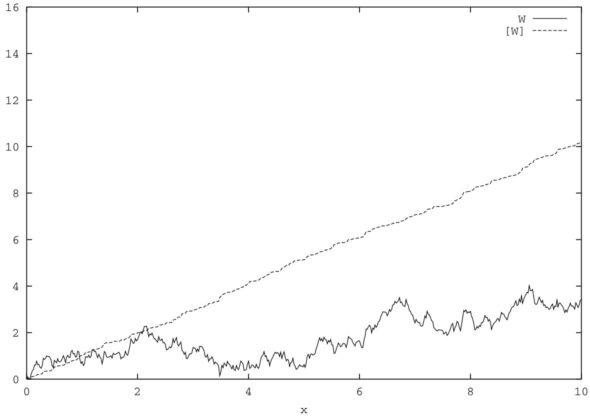



**Fig. 5.2.1.** A Wiener path $W_t$ and its approximate quadratic variation $[W]_{h,t}$

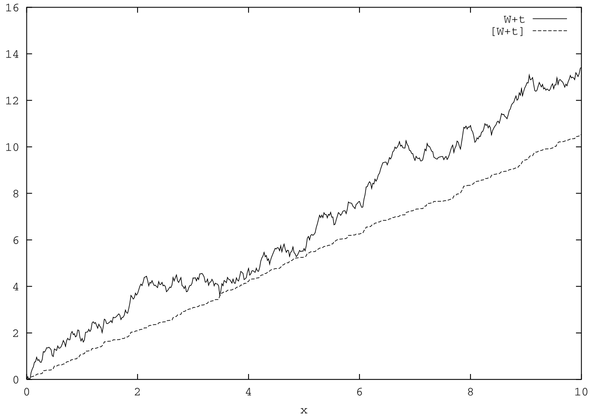



**Fig. 5.2.2.** Transformed Wiener process $Y_t$ and its approximate quadratic variation $[Y]_{h,t}$

quadratic variation in Fig. 5.2.1 forms almost a straight line with slope one. Indeed, it can be shown, see Karatzas & Shreve (1991) or Elliott (1982), that the value of the quadratic variation process $[W] = \{[W]_t,\ t \in [0,\infty)\}$ at time $t$ for a standard Wiener process $W$ is given by the relation

$$[W]_t = t \tag{5.2.5}$$

for $t \in [0,\infty)$. Thus, for finer time discretizations, the approximate quadratic variation becomes almost a perfect straight line.

In Fig. 5.2.2, a sample path of a transformed Wiener process $Y = \{Y_t,\ t \in [0,\infty)\}$ with values

$$Y_t = W_t + t$$

together with its approximate quadratic variation are displayed. Observe that the drift, which was added to the Wiener process, had practically no impact on the approximate quadratic variation, when compared to Fig. 5.2.1. This effect can be explained by noting that for a stochastic process $F = \{F_t,\, t \in [0,\infty)\}$ its, so-called, *total variation* is

$$[F]_t^{\frac{1}{2}} \stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} \left| F_{t_k} - F_{t_{k-1}} \right| \tag{5.2.6}$$

for $t \in [0,\infty)$. Note that in the case where $F_t = t$ for $t \in [0,\infty)$ the total variation $[F]_t^{\frac{1}{2}} = t$ is bounded. However, $F_t = t$ has zero quadratic variation since

$$[F]_t \stackrel{P}{=} [t]_t \stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} (t_k - t_{k-1})^2 = 0. \tag{5.2.7}$$

One notes that a differentiable function has finite total variation but zero quadratic variation. In contrast to that one can show that the strongly fluctuating Wiener process has no finite total variation but some finite quadratic variation.

It is then possible to show that the above transformed Wiener process $Y$ has the finite quadratic variation

$$\begin{aligned} [Y]_t &\stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} (Y_{t_k} - Y_{t_{k-1}})^2 \\ &\stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} \Big( (W_{t_k} - W_{t_{k-1}})^2 + 2\,(W_{t_k} - W_{t_{k-1}})\,(t_k - t_{k-1}) + \,(t_k - t_{k-1})^2 \Big) \\ &\stackrel{P}{=} [W]_t, \end{aligned} \tag{5.2.8}$$

for $t \in [0,\infty)$, which is the same as that for the Wiener process. Here only the sum of the squared Wiener process increments does not vanish asymptotically. We note that only the martingale term in the transformed Wiener process, which is in the above example the Wiener process itself, contributes to the quadratic variation.

### Another Martingale

Starting with a continuous, square integrable $(\underline{\mathcal{A}}, P)$-martingale $X$, another $(\underline{\mathcal{A}}, P)$-martingale can be constructed by using its quadratic variation $[X]$ if $E([X]_T) < \infty$ for each $T \in [0,\infty)$. More precisely, a new continuous $(\underline{\mathcal{A}}, P)$-martingale $Y = \{Y_t,\, t \in [0,\infty)\}$ is obtained by setting

$$Y_t = (X_t)^2 - [X]_t \tag{5.2.9}$$

for $t \in [0,\infty)$, see Protter (2004).

In the case of a standard Wiener process $W$, we obtain the martingale $Y = \{Y_t = (W_t)^2 - t,\, t \in [0,\infty)\}$, see (5.1.6). The type of martingale property of $Y$ given in (5.2.9) is fundamental to stochastic calculus.

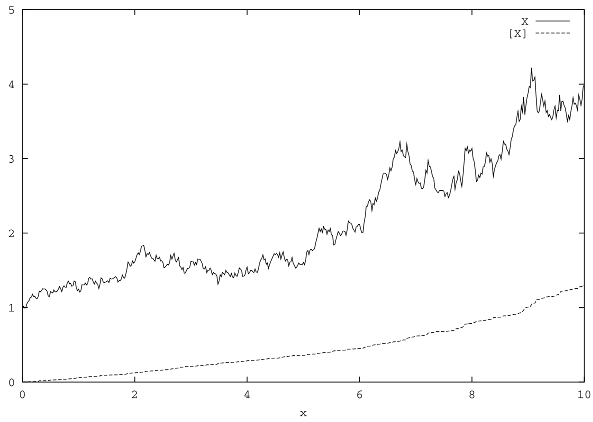



**Fig. 5.2.3.** Path of a geometric Brownian motion and its quadratic variation

### Quadratic Variation and Geometric Brownian Motion

The quadratic variation turns out to be one of the most important characteristics of a martingale. The standard market model for an asset price is the Black-Scholes (BS) model, given by a geometric Brownian motion. To highlight the usefulness of the quadratic variation in such a financial context we consider as a model for an asset price $X_t$ at time $t$ the BS model, see (4.1.2), which we write in the form

$$X_t = X_0 \exp\{L_t\}, \tag{5.2.10}$$

where

$$L_t = g\,t + \sigma\,W_t \tag{5.2.11}$$

for $t \in [0, \infty)$. Here $W = \{W_t,\, t \in [0, \infty)\}$ denotes again a standard Wiener process. With the choice of the growth rate $g = r - \frac{1}{2}\,\sigma^2$ this provides the same dynamics as was given in (5.1.9). When we use the initial value $X_0 = 1$, the expected rate of return $r = 0.05$ and the volatility $\sigma = 0.2$, then the quadratic variation $[X]$ for $X$ is shown in Fig. 5.2.3. Also displayed in Fig. 5.2.3 is the sample path for $X$, see also Fig. 5.1.3. Note that the quadratic variation is not linear. However, if we visualize the quadratic variation of the logarithm $\ln(X_t)$ of $X_t$, then we obtain, as can be seen in Fig. 5.2.4, an almost perfect straight line. The reason for this effect can be directly seen when using the following identities

$$[\ln(X)]_t = [L]_t = \sigma^2\,[W]_t = \sigma^2\,t \tag{5.2.12}$$

for $t \in [0, \infty)$. These relations hold because $L_t = \ln(X_t)$ forms a linearly transformed Wiener process and we can use the fact that $[W]_t = t$, see (5.2.5).

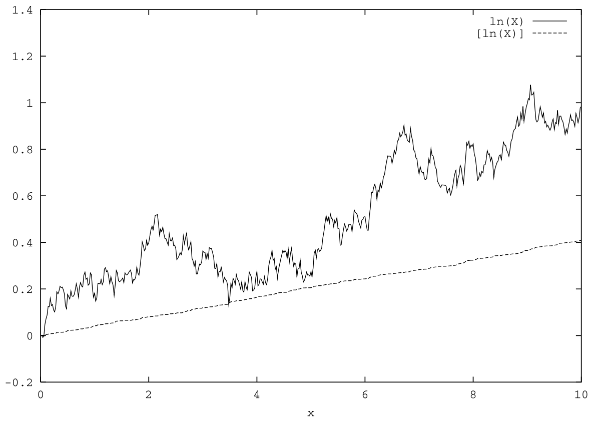



**Fig. 5.2.4.** Path of $\ln(X)$ and $[\ln(X)]$

## Volatility

The key quantity for the parametrization of the BS model, which was the standard market model for many decades, is the *volatility*. We observe in (5.2.12) that under the BS model the squared volatility is the time derivative of the quadratic variation of the logarithm of the asset price. We can express this important observation in the form

$$\sigma^2 = \frac{d}{dt}\,[\ln(X)]_t. \tag{5.2.13}$$

This relation can still be used theoretically as a definition for the volatility of a continuous asset price process, even if its dynamics is not that of a geometric Brownian motion.

To be more precise, we define the historical volatility $\mathrm{Vol}_X(t)$ at a given time $t \in [0,\infty)$ of a given continuous asset price process $X$, as the square root of the left hand derivative of the quadratic variation of the logarithm of $X$. That is, we define the historical volatility in the form

$$\mathrm{Vol}_X(t) = \sqrt{\frac{d}{dt}\,[\ln(X)]_t} \tag{5.2.14}$$

for $t \in [0,\infty)$. A common market practice for estimating squared volatility, see for instance Hull (2000), is that one estimates the sample variance of log-returns, see (2.1.19). Note that $\mathrm{Vol}_X(t)$ is by (5.2.3) and (5.2.2) asymptotically equivalent to the way that volatility is calculated in practice. However, it is well-known, see for instance, Corsi, Zumbach, Müller & Dacorogna (2001) and Barndorff-Nielsen & Shephard (2003), that the estimation of volatility is in practice a very delicate task.

The definition of volatility in (5.2.14) is quite general and can be used for all continuous asset price processes. It has the advantage that it is independent

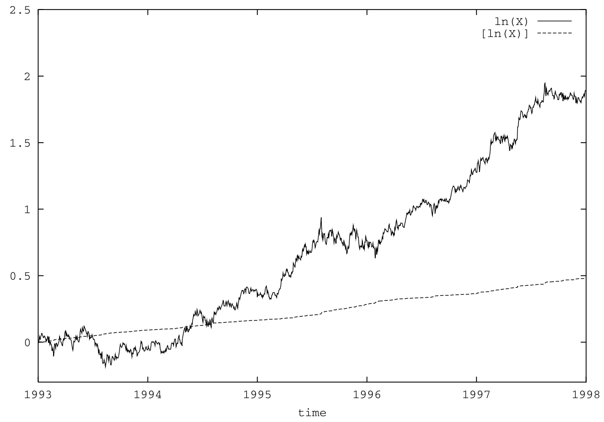



**Fig. 5.2.5.** IBM log-share price and its quadratic variation

of the specific choice of the underlying asset price model and also the time discretization employed. In the particular case of geometric Brownian motion it leads us directly to the constant volatility of the BS model, as can be seen from (5.2.12). We shall see, that the above definition of historical volatility is useful for the study of the actual volatility dynamics in asset price models. Furthermore, the approximate quadratic variation (5.2.3) can be directly used to construct a volatility estimator.

It is well-known that in reality, volatility is stochastic, as can be seen from the changing slope of the quadratic variation of the logarithm of asset prices. This indicates that the standard market model with its constant volatility can only be considered to be used as a first, rough approximation of the existing market dynamics. As another example for an application of the above definition of historical volatility, Fig. 5.2.5 shows the logarithm $\ln(\frac{X_t}{X_{t_0}})$ of the IBM share price $X_t$ from 1993 up until 1998 together with its approximate quadratic variation based on daily observations. According to the definition of historical volatility in (5.2.14) we can interpret the square root of the slope in Fig. 5.2.5 as an empirical volatility estimate of the IBM share price during the corresponding time period. By estimating the observed slope of the quadratic variation in Fig. 5.2.5, an annualized average volatility of approximately $\sqrt{\frac{0.5}{5}} = \sqrt{0.1} \approx 0.32$ is inferred.

To illustrate further the type of information that the quadratic variation provides we show in Fig. 5.2.6 the logarithm $\ln(\frac{X_t}{X_{t_0}})$ of the S&P500 index for the period from 1993 up until 1998 together with its quadratic variation. Note that the average slope of the quadratic variation in Fig. 5.2.6, that is its squared volatility, is much smaller than that for the IBM share price in Fig. 5.2.5. This is mainly due to the effect of diversification for the index. Again, an approximate estimate for the average volatility of the S&P500 index can be obtained from the square root of the slope of the quadratic variation shown in Fig. 5.2.6. Thus, we estimate an annualized average volatility of

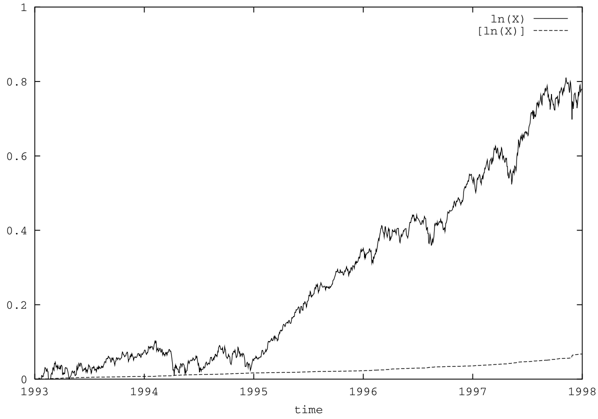



**Fig. 5.2.6.** Logarithm of S&P500 and its quadratic variation

about $\sqrt{\frac{0.05}{5}} = \sqrt{0.01} = 0.1$, which is about a third of the estimated volatility of the IBM share price.

## Covariation

In a similar manner as the quadratic variation the covariation of two continuous stochastic processes can be defined. This is another important tool which turns out to be useful for the characterization of dependencies between two stochastic processes, for instance, between asset prices. It allows the, locally in time, measurement of associations between the random fluctuations of two different continuous processes.

For the definition of covariation the same equidistant time discretization, as given in (5.2.1), is now used. That is, we set $t_k = kh$ for $k \in \{0, 1, \ldots\}$, $h > 0$. For continuous stochastic processes $Z_1$ and $Z_2$ the *covariation process* $[Z_1, Z_2] = \{[Z_1, Z_2]_t,\, t \in [0, \infty)\}$ is defined as the limit in probability, see (2.7.2), as $h \to 0$ of the values of the *approximate covariation process* $[Z_1, Z_2]_{h,\cdot}$ with

$$[Z_1, Z_2]_{h,t} = \sum_{k=1}^{i_t} (Z_1(t_k) - Z_1(t_{k-1}))(Z_2(t_k) - Z_2(t_{k-1})) \tag{5.2.15}$$

for $t \in [0, \infty)$ and $h > 0$, given by the sums of the products of the increments of the processes $Z_1$ and $Z_2$. Here the integer $i_t$ is as introduced in (5.2.4). More precisely, we define at time $t \in [0, \infty)$ the *covariation*

$$[Z_1, Z_2]_t \overset{P}{=} \lim_{h \to 0} [Z_1, Z_2]_{h,t}, \tag{5.2.16}$$

where $[Z_1, Z_2]_{h,t}$ is the approximate covariation.

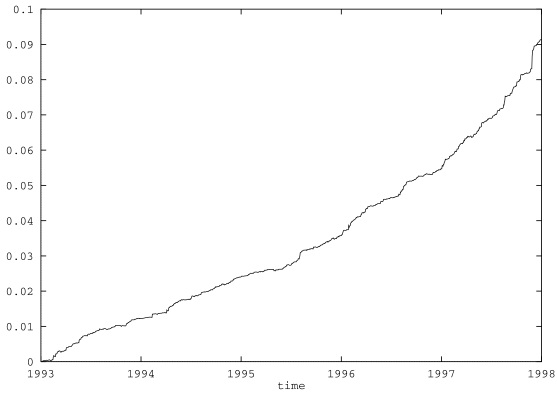



**Fig. 5.2.7.** Covariation between logarithms of S&P500 and IBM share price

As an example, we display in Fig. 5.2.7 the approximate covariation between the logarithms of the S&P500 index, see Fig. 5.2.6, and the IBM share price, see Fig. 5.2.5, for the period from 1993 up until 1998 using daily observations. Note that the average slope of the covariation seems to be here almost always positive, which indicates some association between the movements of the IBM share price and those of the S&P500 index. Summarizing these observations, it appears that the covariation provides a useful tool for measuring the degree of association of the fluctuations of two stochastic processes locally in time. Obviously, if the processes $Z_1$ and $Z_2$ are identical, then their covariation coincides with their quadratic variation.

## Covariation for Processes with Jumps (*)

For any right-continuous stochastic process $\xi = \{\xi(t),\, t \in [0,\infty)\}$ we denote by

$$\xi(t-) \overset{\text{a.s.}}{=} \lim_{h \to 0+} \xi(t-h) \tag{5.2.17}$$

the almost sure *left hand limit* of $\xi(t)$ at time $t \in (0,\infty)$. The *jump size* $\Delta\xi(t)$ at time $t$ is then defined as

$$\Delta\xi(t) = \xi(t) - \xi(t-) \tag{5.2.18}$$

for $t \in (0,\infty)$.

In the case of a pure jump process $p = \{p_t,\, t \in [0,\infty)\}$ the corresponding quadratic variation is obtained as

$$[p]_t = \sum_{0 \le s \le t} (\Delta p_s)^2 \tag{5.2.19}$$

for $t \in [0,\infty)$, where $\Delta p_s = p_s - p_{s-}$. In the case when $p$ is a Poisson process, its quadratic variation equals the process itself, that is, $[N]_t = N_t$ for all

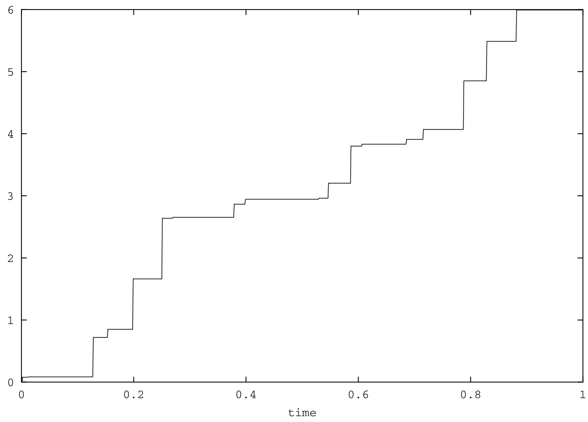



**Fig. 5.2.8.** Quadratic variation of a compound Poisson process

$t \in [0, \infty)$. We show in Fig. 5.2.8 the quadratic variation $[Y]_t$ of the trajectory of the compound Poisson process $Y$ shown in Fig. 3.5.2.

It is preferable to separate the jump part of a process when computing its quadratic variation. For a general stochastic process the quadratic variation consists of the sum of the quadratic variations of its continuous and its pure jump part. This will be made more precise below.

Let us denote by $Z_1$ and $Z_2$ two stochastic processes with continuous part

$$Z_i^c(t) = Z_i(t) - Z_i(0) - \sum_{0<s\le t} \Delta Z_i(s) \tag{5.2.20}$$

for $t \in [0, \infty)$ and $i \in \{1, 2\}$. Here the *jump size* at time $s$ is given as

$$\Delta Z_i(s) = Z_i(s) - Z_i(s-) \tag{5.2.21}$$

for $s \in [0, \infty)$ and we assume that the sum in (5.2.20) is almost surely finite. The covariation $[Z_1, Z_2]_t$ of $Z_1$ and $Z_2$ at time $t$ is then defined as

$$[Z_1, Z_2]_t = [Z_1^c, Z_2^c]_t + \sum_{0<s\le t} (\Delta Z_1(s))\,(\Delta Z_2(s)) \tag{5.2.22}$$

for $t \in [0, \infty)$, as long as the quantities involved are almost surely finite. This also means that the quadratic variation of a process $Z_1$ equals the quadratic variation $[Z_1^c]_t$ of its continuous part $Z_1^c$ plus the sum of the squares of its jumps, that is

$$[Z_1]_t = [Z_1^c]_t + \sum_{0<s\le t} (\Delta Z_1(s))^2 \tag{5.2.23}$$

for $t \in [0, \infty)$. Again, we assume that the expressions involved are almost surely finite. The above notion of covariation for processes with jumps is convenient and useful. Obviously, if the processes $Z_1$ and $Z_2$ are identical, then their quadratic variation coincides with their covariation. The quadratic variation $[q]_t$ of the compensated Poisson process $q = \{q_t, t \in [0,\infty)\}$, shown in Fig. 5.1.4, equals that of the Poisson process $N$ shown in Fig. 3.5.1, that is $[N]_t = [q]_t$ for $t \in [0,\infty)$.

We emphasize that the covariation of a process with continuous and jump part is an important characteristic in financial modeling, see Cont & Tankov (2004) and Ait-Sahalia (1996).

### *p*th Variation (*)

We call for $p > 0$ and a stochastic process $X = \{X_t, t \in [0,\infty)\}$ the process $[X]_h^{\frac{p}{2}} = \left\{[X]_{h,t}^{\frac{p}{2}}, t \in [0,\infty)\right\}$ with

$$[X]_{h,t}^{\frac{p}{2}} = \sum_{k=1}^{i_t} |X_{t_k} - X_{t_{k-1}}|^p \tag{5.2.24}$$

for $t \in [0,\infty)$ the *approximate pth variation process* of $X$. Then the $p$th *variation process* $[X]^{\frac{p}{2}} = \left\{[X]_t^{\frac{p}{2}}, t \in [0,\infty)\right\}$ is for each $t \in [0,\infty)$ defined as the limit in probability

$$[X]_t^{\frac{p}{2}} \stackrel{P}{=} \lim_{h\to 0} [X]_{h,t}^{\frac{p}{2}}, \tag{5.2.25}$$

see (2.7.1). The first order variation process $[X]^{\frac{1}{2}}$ is called *total variation*, see (5.2.6). For instance, the time $t$ with $X_t = t$ is a process with total variation $[X]_t^{\frac{1}{2}} = t < \infty$ a.s. for $t \in [0,\infty)$. Furthermore, any differentiable process can be shown to have finite total variation. Note that the Wiener process $W = \{W_t, t \in [0,\infty)\}$ does *not* have finite total variation, however, it has finite quadratic variation, as shown in (5.2.5). On the other hand, a Poisson process with finite intensity does have finite total variation.

### Local Martingales (*)

As we shall see later, in quantitative finance stochastic processes naturally appear that are not martingales but become martingales if they are properly stopped. These *local martingales* are locally in time similar to martingales.

**Definition 5.2.1.** *A stochastic process $X = \{X_t, t \in [0,\infty)\}$ is an $(\underline{\mathcal{A}}, P)$-* local martingale *if there exists an increasing sequence $(\tau_n)_{n\in\mathcal{N}}$ of stopping times, that may depend on $X$, such that $\lim_{n\to\infty} \tau_n \stackrel{a.s.}{=} \infty$ and each stopped process*

$$X^{\tau_n} = \{X_t^{\tau_n} = X_{t\wedge\tau_n}, t \in [0,\infty)\} \tag{5.2.26}$$

*is an $(\underline{\mathcal{A}}, P)$-martingale, where $t \wedge \tau_n = \min(t, \tau_n)$.*

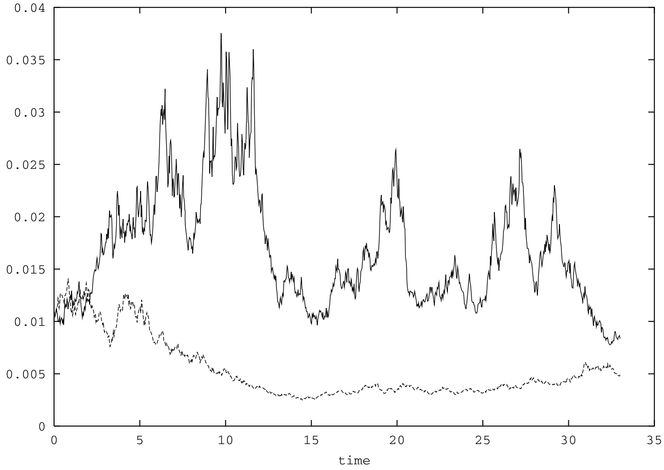



**Fig. 5.2.9.** Two trajectories of a strict local martingale

If $X$ is a local martingale, then the value $X_s$ does, in general, not equal the conditional expectation $E(X_t|\mathcal{A}_s)$ for $s \in [0,\infty)$ and $t \in [s,\infty)$. Note that an $(\underline{\mathcal{A}}, P)$-martingale is also an $(\underline{\mathcal{A}}, P)$-local martingale. However, an $(\underline{\mathcal{A}}, P)$-local martingale is not always an $(\underline{\mathcal{A}}, P)$-martingale. A local martingale that is not a martingale is called a *strict local martingale*.

To provide an example for such a strict local martingale $X = \{X_t, t \in [0,\infty)\}$ we form the sum of the squares of four independent Wiener processes $W^1$, $W^2$, $W^3$, $W^4$ which start each at the value $W_0^i = 5$, $i \in \{1,2,3,4\}$. By taking the inverse of this sum, that is

$$X_t = \left(\sum_{i=1}^{4} \left(W_t^i + 5\right)^2\right)^{-1} \tag{5.2.27}$$

for $t \in [0,\infty)$, we shall show later that this inverse of a squared Bessel process of dimension four forms a strict local martingale, see Revuz & Yor (1999). Two paths of such a process $X = \{X_t, t \in [0,\infty)\}$ are shown in Fig. 5.2.9. They both look rather different but are both constructed according to (5.2.27). It appears that they can mimic very different behaviors, in particular, over the initial time period. As we discuss later in the context of squared Bessel processes, this process has peculiar properties that differentiate it from a martingale, see Revuz & Yor (1999).

One can formulate the following statements, see Protter (2004), that will become relevant when dealing with local martingales in financial modeling under the benchmark approach.

**Lemma 5.2.2.**

*(i)* *An almost surely nonnegative (negative)* $(\underline{\mathcal{A}}, P)$*-local martingale is an* $(\underline{\mathcal{A}}, P)$*-supermartingale (submartingale).*

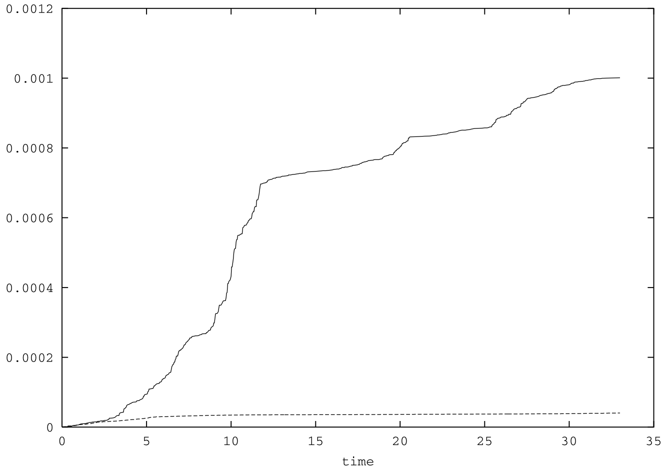



**Fig. 5.2.10.** Quadratic variations of two trajectories of a strict local martingale

*(ii) An a.s. uniformly bounded* $(\underline{\mathcal{A}}, P)$*-local martingale is an* $(\underline{\mathcal{A}}, P)$*-martingale.*

*(iii) A square integrable* $(\underline{\mathcal{A}}, P)$*-local martingale* $X$ *is a square integrable* $(\underline{\mathcal{A}}, P)$*-martingale if and only if*

$$E([X]_T) < \infty \tag{5.2.28}$$

*for all* $T \in [0, \infty)$.

We prove the assertion (i) at the end of this section. The Definition 5.2.1 of a local martingale is rather technical and somehow difficult to verify. However, the statement (iii) of the above lemma is quite useful in practice because local martingales that one typically faces in finance seem to be square integrable. The statement (iii) means that if the fluctuations of a square integrable local martingale are so strong that the mean of its quadratic variation does not exist, then it cannot be a martingale and is therefore a strict local martingale. In Fig. 5.2.10 we show the quadratic variation of the two paths of the strict local martingale shown in Fig. 5.2.9. Note that its quadratic variation appears to be highly dependent on the particular path. As we shall see later, one can show for the given example that $E([X]_t) = \infty$ for $t \in (0, \infty)$. This means that the quadratic variation of different paths varies so strongly that no finite expectation can be calculated.

We face here a subtle but important property, which will be highly relevant for the understanding of the typical dynamics of financial markets as it becomes visible under the benchmark approach.

### Nonnegative Local Martingales are Supermartingales (*)

As we shall see later the statement (i) in Lemma 5.2.2 is crucial for the benchmark approach when it establishes no-arbitrage. For completeness we provide here a proof.

**Lemma 5.2.3.** *A nonnegative* $(\underline{\mathcal{A}}, P)$*-local martingale* $X = \{X_t,\, t \in [0,\infty)\}$ *with* $E(X_t|\mathcal{A}_s) < \infty$ *for all* $0 \le s \le t < \infty$ *is an* $(\underline{\mathcal{A}}, P)$*-supermartingale.*

**Proof**: Consider a nonnegative $(\underline{\mathcal{A}}, P)$-local martingale $X = \{X_t,\, t \in [0,\infty)\}$. Then there exists an increasing sequence $(\tau_n)_{n\in\mathcal{N}}$ of stopping times, with respect to the filtration $\underline{\mathcal{A}}$, such that each stopped process $X^{\tau_n} = \{X_t^{\tau_n} = X_{t\wedge\tau_n},\, t \in [0,\infty)\}$ is an $(\underline{\mathcal{A}}, P)$-martingale and we have $\tau_n \to \infty$ almost surely. Consequently, for each $n \in \mathcal{N}$ and $0 \le s \le t < \infty$ we have

$$\begin{aligned} E(X_t \,|\, \mathcal{A}_s) &= E\left(\mathbf{1}_{\{\tau_n \ge t\}}\, X_t \,\Big|\, \mathcal{A}_s\right) + E\left(\mathbf{1}_{\{\tau_n < t\}}\, X_t \,\Big|\, \mathcal{A}_s\right) \\ &= E\left(\mathbf{1}_{\{\tau_n \ge t\}}\, X_t^{\tau_n} \,\Big|\, \mathcal{A}_s\right) + E\left(\mathbf{1}_{\{\tau_n < t\}}\, X_t \,\Big|\, \mathcal{A}_s\right) \\ &\le E\left(X_t^{\tau_n} \,\Big|\, \mathcal{A}_s\right) + E\left(\mathbf{1}_{\{\tau_n < t\}}\, X_t \,\Big|\, \mathcal{A}_s\right) \\ &= X_s^{\tau_n} + E\left(\mathbf{1}_{\{\tau_n < t\}}\, X_t \,\Big|\, \mathcal{A}_s\right). \end{aligned} \tag{5.2.29}$$

Since we have for each $t \in [0,\infty)$ by definition that $\mathbf{1}_{\{\tau_n \ge t\}} X_t$ approaches $X_t$ almost surely from below as $n \to \infty$ it follows by (5.2.29) and the Monotone Convergence Theorem, see (2.7.9), that the difference

$$E(X_t \,|\, \mathcal{A}_s) - E\left(\mathbf{1}_{\{\tau_n < t\}}\, X_t \,\Big|\, \mathcal{A}_s\right) = E\left(\mathbf{1}_{\{\tau_n \ge t\}}\, X_t \,\Big|\, \mathcal{A}_s\right)$$

approaches almost surely the conditional expectation $E(X_t \,|\, \mathcal{A}_s)$ from below as $n \to \infty$. As a consequence of that, the conditional expectation $E(\mathbf{1}_{\{\tau_n < t\}}\, X_t \,|\, \mathcal{A}_s)$ is for $n \to \infty$ decreasing and converges almost surely to zero. By using the fact that $\lim_{n\to\infty} X_s^{\tau_n} \overset{\text{a.s.}}{=} X_s$ yields in (5.2.29) the inequality $E(X_t \,|\, \mathcal{A}_s) \le X_s$ when letting $n$ tend to infinity. This proves the lemma. □

In Rogers & Williams (2000) one can find an alternative proof of this result based on Fatou's Lemma, see (2.7.11). We emphasize that it is essential in Lemma 5.2.3 that one defines the local martingale over the infinite time interval $[0,\infty)$ and not on $[0,\infty]$ or $[0,T]$ with $T \in (0,\infty)$ since the above result does not hold in these cases.


**Fig. 5.3.1.** Gains from trade of one share of IBM stock during 1993 - 1998

## 5.3 Gains from Trade as Stochastic Integral

One of the most fundamental notions in finance is that of gains from trade. In stochastic calculus this corresponds exactly to the notion of a stochastic integral, the Itô integral, which is therefore highly relevant in finance.

### Gains from Trade

Let us consider an investor who holds during the time period $[0,T]$ a constant number $\xi(0)$ of units of an asset with price process $X = \{X_t,\, t \in [0,T]\}$. The investor's allocation strategy $\xi = \{\xi(t) = \xi(0),\, t \in [0,T]\}$, characterized by the number of units of the asset held, is assumed to be constant in this case. Then the investor's *gains from trade* over the period $[0,t]$ equals

$$I_{\xi,X}(t) = \xi(0)\,\{X_t - X_0\}, \tag{5.3.1}$$

for $t \in [0,T]$. This provides the first step towards an appropriate definition of a stochastic integral, which we shall call later Itô integral. Formally, we interpret the above gains from trade $I_{\xi,X}(t)$ as an *Itô integral* of the *integrand* $\xi$ with respect to the *integrator* $X$ over the time interval $[0,t]$, and use the following notation

$$I_{\xi,X}(t) = \int_0^t \xi(s)\,dX_s. \tag{5.3.2}$$

To illustrate the above construction we show in Fig. 5.3.1 the gains from trade obtained from IBM share holdings over the period from 1993 to 1998. This refers to a constant allocation strategy which is holding $\xi(t) = 1$ share.

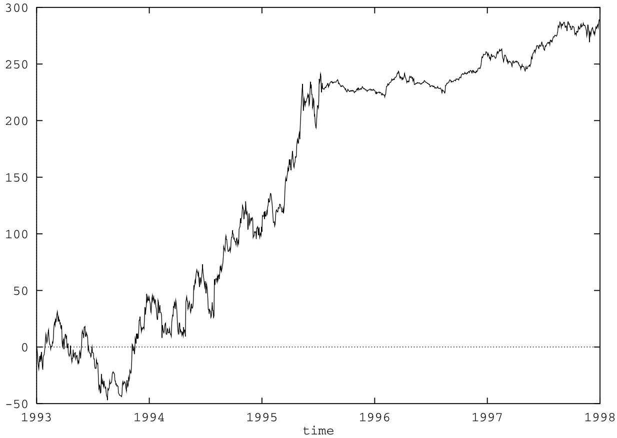



**Fig. 5.3.2.** Gains from trade of ten and later one share of IBM stock

## Piecewise Constant Allocation Strategies

Now, let us allow the investor to change his or her strategy so that it becomes a piecewise constant allocation process $\xi = \{\xi(t),\, t \in [0,T]\}$ with $\xi(t) = \xi(t_k)$ units of shares held at time $t \in [t_k, t_{k+1})$, $k \in \{0,1,\ldots\}$ and $t_k = kh$ for $h > 0$. Here the reallocation times $t_k$ form an equidistant time discretization, as given in (5.2.1). Obviously, the gains from trade over the period $[0,t]$ can be expressed in the form

$$\int_0^t \xi(s)\, dX_s = \sum_{k=1}^{i_t} \xi(t_{k-1})\, \{X_{t_k} - X_{t_{k-1}}\} + \xi(t_{i_t})\, \{X_t - X_{t_{i_t}}\}, \qquad (5.3.3)$$

where

$$i_t = \max\{k \in \mathcal{N} :\, t_k \le t\} \qquad (5.3.4)$$

is the integer index of the latest discretization time before and including $t$, see (5.2.4). Here we formally interpret the gains from trade as an Itô integral in the same form as in (5.3.2) with integrand $\xi$ and integrator $X$ covering the interval $[0,t]$.

In Fig. 5.3.2, the gains from trade are displayed when during the first half of the time period, that is until mid 1995, ten shares of IBM were held and in the second half only one share. One observes during the first period strong fluctuations of the gains from trade when compared to the second half of that time period.

## Itô Integral as a Limit

It is sufficient in many applications to use a Wiener process as integrator. Therefore, we use in a standard setting often the Itô integral with respect to the Wiener process $W = \{W_t,\, t \in [0,\infty)\}$ as integrator over the interval $[0,t]$ for a wide range of integrands $\xi = \{\xi(t),\, t \in [0,\infty)\}$.

**Definition 5.3.1.** *For a left continuous stochastic process $\xi = \{\xi(t),\, t \in [0,\infty)\}$ as integrand with*

$$\int_0^T \xi(s)^2\, ds < \infty \tag{5.3.5}$$

*for all $T \in [0,\infty)$ almost surely, the Itô integral with respect to the Wiener process $W$ is defined as the left continuous limit in probability*

$$\int_0^t \xi(s)\, dW_s \stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} \xi(t_{k-1})\, \{W_{t_k} - W_{t_{k-1}}\} \tag{5.3.6}$$

*of the sequence of corresponding approximating sums for $t \in [0,\infty)$.*

For details on the definition of Itô integrals we refer to Karatzas & Shreve (1991), Kloeden & Platen (1999) or Protter (2004). We see that the right hand sides of both (5.3.3) and (5.3.6) are very similar and coincide in the case of piecewise constant integrands. Consequently, the Itô integral can be seen as a limit in probability of gains from trade, taken over progressively finer time discretizations.

An important characteristic of the Itô integral is that the evaluation point $t_{k-1}$ for the integrand $\xi$ is always taken at the left hand side of the discretization interval $[t_{k-1}, t_k)$. This feature is natural for finance applications because an investor needs to decide at the beginning of an investment period how many units of a security he or she wants to hold. It distinguishes the Itô integral from other stochastic integrals, see Protter (2004). The choice of the evaluation point at the left hand side corresponds in finance to the economically given fact that once an allocation is made it remains constant for some period of time and cannot be changed retrospectively in a legal manner. As we shall see later, this fact is essential for establishing the martingale property for Itô integrals with respect to Wiener processes.

The above definition of Itô integrals can be extended to include more general classes of integrators rather than just the Wiener process, see Protter (2004), which will be discussed later.

### Explicit Value for an Itô Integral

To give a simple example of how the Itô integral differs from the classical, say, Riemann-Stieltjes integral, let us consider a trading strategy, where the number of shares held in an asset equals its price. For simplicity, we assume the asset price to be modeled by the Wiener process $W$. Then according to (5.3.6) we obtain

$$\int_0^t W_s\, dW_s \stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} W_{t_{k-1}} (W_{t_k} - W_{t_{k-1}}) \tag{5.3.7}$$

for $t \in [0, \infty)$. If $W$ were differentiable, then we would obtain from the deterministic integration rule the quantity $\frac{1}{2} W_t^2$ as the value of this integral at time $t$. However, in the stochastic case the correct value will be much less, as we shall see. This means that the gains from trade under this strategy do not accumulate in the same way as they would for differentiable asset prices or under the classical integration rule.

The following calculation demonstrates this important effect in more detail. By subtracting and adding $W_{t_k}^2$ and completing the square on the right hand side of (5.3.7) for each time step we see that

$$\int_0^t W_s\, dW_s \stackrel{P}{=} \lim_{h\to 0} \frac{1}{2} \sum_{k=1}^{i_t} \left\{ \left( W_{t_k}^2 - W_{t_{k-1}}^2 \right) - (W_{t_k} - W_{t_{k-1}})^2 \right\}$$

$$\stackrel{P}{=} \frac{1}{2} W_t^2 - \frac{1}{2} W_0^2 - \lim_{h\to 0} \frac{1}{2} \sum_{k=1}^{i_t} (W_{t_k} - W_{t_{k-1}})^2,$$

where all except the first and last terms in the first sum cancel each other. From the definition of the approximate quadratic variation of standard Wiener processes in (5.2.3) we have $[W]_t = t$, see (5.2.5), and $W_0 = 0$, see (3.2.6). Consequently, the value of the Itô integral (5.3.7) is

$$\int_0^t W_s\, dW_s = \frac{1}{2} W_t^2 - \frac{1}{2} [W, W]_t = \frac{1}{2} W_t^2 - \frac{1}{2} [W]_t = \frac{1}{2} W_t^2 - \frac{1}{2} t. \quad (5.3.8)$$

The quantity on the right hand side of this equation is clearly less than $\frac{1}{2} W_t^2$, which would be expected for a differentiable function under classical integration. Note that the difference is equal to half the covariation of integrand and integrator. We shall see below that this property holds more generally.

The above example exhibits striking differences between the Itô integral and the classical integral. Since, in practice, asset price processes with properties similar to those of Wiener processes are typically encountered, these differences turn out to be crucial for the rigorous modeling in finance. For instance, the computation of derivative prices, values of portfolios and other financial quantities may become incorrect, if these differences were ignored. Stochastic calculus which we introduce in this and the following two chapters will allow us to obtain correct quantities.

To illustrate these differences we show in Fig. 5.3.3 the path of a Wiener process together with half of its squared value and the Itô integral

$$I_{W,W}(t) = \int_0^t W_s\, dW_s = \int_0^t \int_0^s dW_z\, dW_s, \quad (5.3.9)$$

for $t \in [0, 1]$. Note in this figure the significant difference between the Itô integral $I_{W,W}(t)$ and the value $\frac{1}{2} W_t^2$ that would be obtained under the classical integration rule.

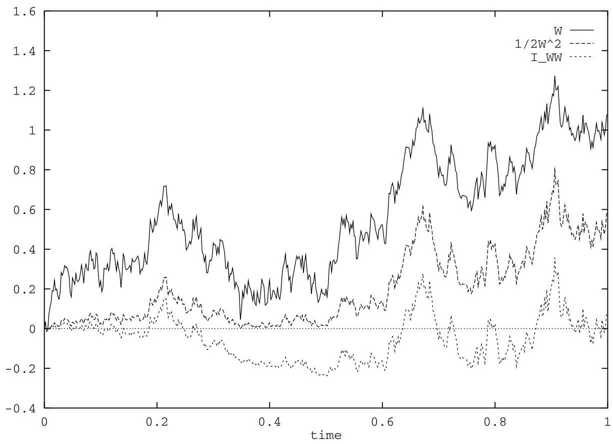



**Fig. 5.3.3.** Paths of $W$, $\frac{1}{2}W^2$ and $I_{W,W}$

In the following analysis it will be shown that the differences between Itô and classical integration relate to the covariation of the processes involved as integrand and integrator. These differences are crucial and impact significantly the area of quantitative finance due to the nature of asset prices.

### General Itô Integrals and Differentials

The definition of an Itô integral as gains from trade, given in (5.3.6), can naturally be extended to include more general integrators. Let us again use, for simplicity, the equidistant time discretization (5.2.1) and denote, as previously, by $Y_{t-}$ the left hand limit of the value of a process $Y = \{Y_t, t \in [0,\infty)\}$ at time $t \in [0,\infty)$. We define for a stochastic process $X = \{X_t, t \in [0,\infty)\}$ as integrator and a predictable process $\xi = \{\xi(t), t \in [0,\infty)\}$ as integrand with

$$\int_0^T \xi(s)^2 \, d[X]_s < \infty \tag{5.3.10}$$

for all $T \in [0,\infty)$ a.s., the Itô integral as the limit in probability

$$\int_0^t \xi(s) \, dX_s \stackrel{P}{=} \lim_{h\to 0} \sum_{k=1}^{i_t} \xi(t_{k-1}) \, (X_{t_k} - X_{t_{k-1}}) \tag{5.3.11}$$

for $t \in [0,\infty)$, provided this limit exists. For details we refer the reader to Protter (2004). Here $i_t$ is the integer index given by (5.3.4) for $t \in [0,\infty)$. We emphasize that in financial applications the Itô integral can be naturally interpreted as gains from trade. Furthermore, one can use almost any adapted process $\xi$, which satisfies (5.3.10), to form an integrand by using its predictable version with left hand limits.

Let $e = \{e_t,\, t \in [0, \infty)\}$ and $f = \{f_t,\, t \in [0, \infty)\}$ be predictable stochastic processes. Consider a stochastic process $Y = \{Y_t,\, t \in [0, \infty)\}$, where

$$Y_t = y_0 + \int_0^t e_s\, ds + \int_0^t f_s\, dW_s \tag{5.3.12}$$

for $t \in [0, \infty)$ and initial value $Y_0 = y_0$. Here $W = \{W_t,\, t \in [0, \infty)\}$ is a standard Wiener process and we assume that appropriate measurability and integrability conditions apply so that the above integrals exist. In particular, the first integral is a random ordinary Riemann-Stieltjes integral for $t \in [0, \infty)$. It exists if

$$\int_0^t |e_s|\, ds < \infty \tag{5.3.13}$$

for all $t \in [0, \infty)$ a.s. The second integral is an Itô integral with respect to the Wiener process $W$, see (5.3.6), where we assume that

$$\int_0^t |f_s|^2\, ds < \infty \tag{5.3.14}$$

for all $t \in [0, \infty)$ a.s. It is common to use the following more compact way of expressing the integral equation (5.3.12): The *Itô differential* $dY_t$ of $Y$ at time $t$ is given by the expression

$$dY_t = e_t\, dt + f_t\, dW_t \tag{5.3.15}$$

for $t \in [0, \infty)$ with $Y_0 = y_0$. This is simply another symbolic way of writing (5.3.12), where one should not forget to add the specification of the initial value $Y_0$. The processes $e$ and $f$ are called *drift* and *diffusion coefficients* of the Itô differential (5.3.15), respectively. The concept of an Itô differential is very powerful. It leads to a compact characterization that can be used to succinctly express the dynamics of rather complicated stochastic processes. Note that no Markovianity is required to characterize a process $Y$ via its stochastic differential. This allows the modeling of very general dynamics and corresponding gains from trade.

For the above process $Y$, given in (5.3.12), consider the Itô integral defined in (5.3.11) with $Y$ replacing $X$. Under rather general conditions it can be shown that

$$\int_0^t \xi(s)\, dY_s = \int_0^t \xi(s)\, e_s\, ds + \int_0^t \xi(s)\, f_s\, dW_s \tag{5.3.16}$$

a.s. for all $t \in [0, \infty)$, see Protter (2004). Therefore, an Itô integral of the above type can be expressed as the sum of a random ordinary Riemann-Stieltjes integral with respect to time and a standard Itô integral with respect to the Wiener process $W$.

These definitions and formulations extend to the case of multi-dimensional integrands $\xi$ and integration with respect to several independent standard Wiener processes. Furthermore, they can be generalized also to hold for more

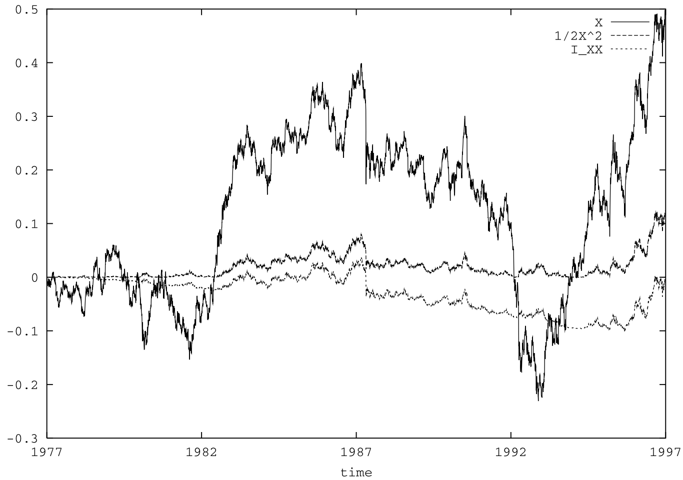



**Fig. 5.3.4.** Log IBM share price $X_t$, $\frac{1}{2}(X_t)^2$ and $I_{X,X}(t)$

general processes as integrators including those with jumps, as we shall see later.

In Fig. 5.3.4 we consider the logarithm $X_t$ of the IBM share price between 1977 and 1997 when normalized to the value one at the beginning. Using $X_t$ we compute also half of its squared value, that is $\frac{1}{2}(X_t)^2$, and plot these values in Fig. 5.3.4 together with the Itô integral $I_{X,X}(t)$ of $X$ with respect to itself. One notes in Fig. 5.3.4 that there is a clear difference between the Itô integral $I_{X,X}(t)$ and what one would expect from a classical integral of a function, which would result in the value $\frac{1}{2}(X_t)^2$ at time $t$. The Itô integral provides here the smaller values, similar as in Fig. 5.3.3.

## 5.4 Itô Integral for Wiener Processes

The Itô integral exhibits a number of important properties and features that are essential in stochastic calculus and thus also for many applications in quantitative finance. The following properties will be repeatedly exploited later, for instance, in the context of pricing and hedging of derivatives.

### Properties of Itô Integrals with Respect to Wiener Processes

Let us consider two $\underline{\mathcal{A}}$-adapted independent Wiener processes $W^1$ and $W^2$. Recall that $(W_t^i - W_s^i)$ is independent of $\mathcal{A}_s$ for $t \in [0, \infty)$, $s \in [0, t]$ and $i \in \{1, 2\}$.

It is useful to specify for $T \in [0, \infty)$ the class $\mathcal{L}_T^2$ of predictable, square integrable stochastic processes $f = \{f_t,\, t \in [0, T]\}$ in the form that

$$\int_0^T E\left((f_t)^2\right) dt < \infty. \qquad (5.4.1)$$

Note that it is convenient to work in a world of square integrable stochastic processes as long as this is possible for the problem at hand. Let us now summarize some fundamental properties of Itô integrals, which are essential and often used in derivations in quantitative finance.

1. *Linearity property*: For $T \in (0,\infty)$, $t \in [0,T]$, $s \in [0,t]$, $Z_1, Z_2 \in \mathcal{L}_T^2$ and $\mathcal{A}_s$-measurable, square integrable random variables $A$ and $B$ it is
$$\int_s^t (A\,Z_1(u) + B\,Z_2(u))\,dW_u^1 = A\int_s^t Z_1(u)\,dW_u^1 + B\int_s^t Z_2(u)\,dW_u^1. \tag{5.4.2}$$
2. *Martingale property*: For $T \in (0,\infty)$, $t \in [0,T]$, $s \in [0,t]$ and $\xi \in \mathcal{L}_T^2$ one has
$$E\left(\int_0^t \xi(u)\,dW_u^1 \,\Big|\, \mathcal{A}_s\right) = \int_0^s \xi(u)\,dW_u^1. \tag{5.4.3}$$
3. *Correlation property*: For $T \in (0,\infty)$, $t \in [0,T]$, independent Wiener processes $W^1$ and $W^2$ and $Z_1, Z_2 \in \mathcal{L}_T^2$ the conditional correlation of two Itô integrals is given by
$$E\left(\int_0^t Z_1(u)\,dW_u^i \int_0^t Z_2(u)\,dW_u^j \,\Big|\, \mathcal{A}_s\right) = \begin{cases} \int_0^t E\left(Z_1(u)\,Z_2(u) \,\Big|\, \mathcal{A}_s\right) du & \text{for } i=j \\ 0 & \text{otherwise} \end{cases} \tag{5.4.4}$$
with $i,j \in \{1,2\}$.
4. *Covariation property*: For $t \in [0,\infty)$, independent Wiener processes $W^1$ and $W^2$ and predictable integrands $Z_1$ and $Z_2$ with $\int_0^t |Z_1(u)\,Z_2(u)|\,du < \infty$ a.s. the covariation of two Itô integrals is
$$\left[\int_0 Z_1(u)\,dW_u^i, \int_0 Z_2(u)\,dW_u^j\right]_t = \begin{cases} \int_0^t Z_1(u)\,Z_2(u)\,du & \text{for } i=j \\ 0 & \text{otherwise} \end{cases} \tag{5.4.5}$$
with $i,j \in \{1,2\}$.
5. *Finite variation property*: For $t \in [0,\infty)$ the covariation between an Itô and a random ordinary Riemann-Stieltjes integral with respect to time vanishes. That is, for predictable $Z_1$ and $Z_2$ one has
$$\left[\int_0 Z_1(u)\,dW_u^1, \int_0 Z_2(u)\,du\right]_t = 0. \tag{5.4.6}$$

In (5.4.5) and (5.4.6) we take the upper end of the integration interval as the time parameter when forming the covariation. Using the martingale property (5.4.3) it can be shown that an Itô integral process is an $(\underline{\mathcal{A}}, P)$-martingale if the integrand is in $\mathcal{L}_T^2$. The above imposed measurability and integrability conditions can be weakened for some of the above stated properties, see Protter (2004).

The following important property of an Itô integral with respect to a Wiener process is very useful in finance. It involves again the notion of a predictable process, as was introduced in Sect. 5.1.

**Lemma 5.4.1.** *If $\xi$ is predictable and it holds for this integrand that*

$$\int_0^T \xi(u)^2 \, du < \infty \tag{5.4.7}$$

*a.s. for all $T \in [0, \infty)$, then the corresponding Itô integral process $I_{\xi,W} = \{I_{\xi,W}(t) = \int_0^t \xi(s) dW_s, \ t \in [0, \infty)\}$ is an $(\underline{\mathcal{A}}, P)$-local martingale.*

The proofs for the above properties and lemma take advantage of the properties of increments of Wiener processes and their relationship to the filtration $\underline{\mathcal{A}}$. Details can be found in Karatzas & Shreve (1991), Kloeden & Platen (1999) or Protter (2004). By application of the Statement (iii) of Lemma 5.2.2 one can derive directly the following result.

**Corollary 5.4.2.** *Assume that $I_{\xi,W}$ is square integrable, then $I_{\xi,W}$ is a square integrable $(\underline{\mathcal{A}}, P)$-martingale if and only if*

$$E\left(\int_0^T \xi(u)^2 \, du\right) < \infty \tag{5.4.8}$$

*for all $T \in [0, \infty)$.*

### Covariation Property

To illustrate the covariation property (5.4.5), Fig. 5.4.1 shows a sample path of a Wiener process together with an Itô integral with respect to this Wiener process using an integrand with value 10 for the first half of the period and value 1 for the rest of the period. The covariation of the Wiener process with this Itô integral is then shown in Fig. 5.4.2. Note that the slope of the covariation is proportional to the integrand of the Itô integral, as is suggested by formula (5.4.5). In Fig. 5.4.3, the quadratic variation of the Itô integral, shown in Fig. 5.4.1, is displayed. Again, as indicated by (5.4.5), the time derivative of the quadratic variation is proportional to the square of the integrand. Consequently, the slope of the quadratic variation in the second period is rather small, about 1% of that of the first period.

### Itô and Deterministic Calculus

The rules that apply to Itô integrals form most of what is called the *Itô* or *stochastic calculus*. This calculus specifies rules for handling stochastic quantities, which involves integration over time. The key relationships in quantitative finance are strongly influenced by these rules.

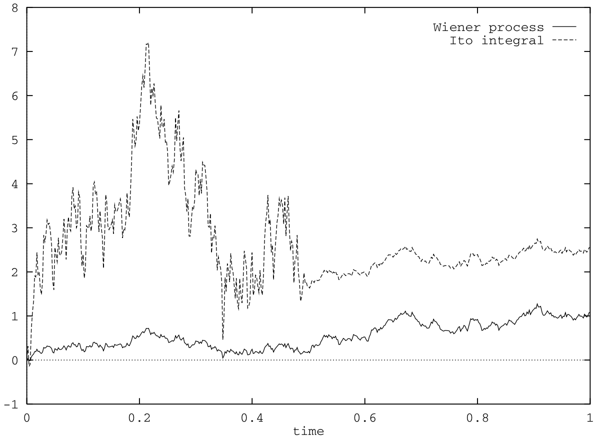



**Fig. 5.4.1.** Wiener process and Itô integral

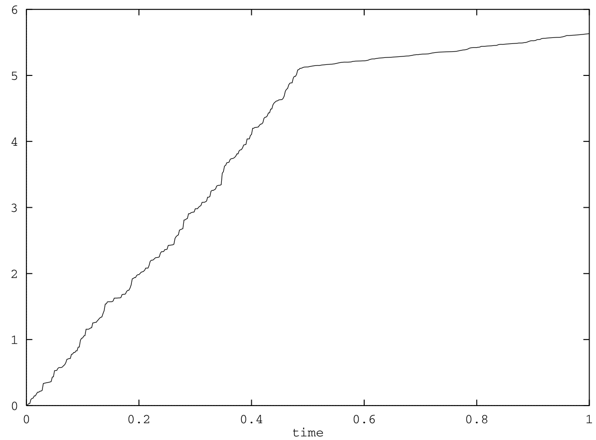



**Fig. 5.4.2.** Covariation of Wiener process and Itô integral

As previously mentioned the rules of Itô calculus are different from those of classical calculus, which is, in general, built on Riemann-Stieltjes integration requiring finite total variation of the integrator. The differences are primarily due to the fact that the Wiener process is of infinite total variation and has trajectories of non-zero, finite quadratic variation, see (5.2.2) and (5.2.25). Thus, the Itô integral has, in general, non-vanishing covariation between its integrand and integrator. The Wiener process and the Itô integral with respect to the Wiener process are continuous processes but not differentiable. Therefore, to ask for the slope or time derivative of a Wiener process, an Itô integral or an asset price when modeled by such process, is a meaningless question.

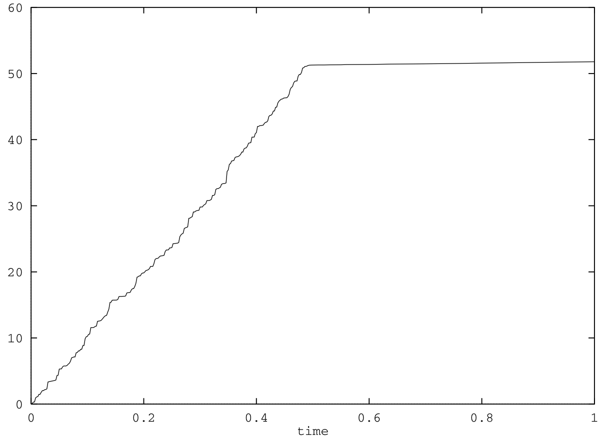



**Fig. 5.4.3.** Quadratic variation of an Itô integral

However as described earlier, the Itô integral is well defined without having to require differentiability of its integrator. As we shall see later, the rules of stochastic calculus provide answers to important problems in quantitative finance, such as how the pricing and hedging of a derivative can be performed or what is the typical dynamics of an asset price or optimal portfolio.

## 5.5 Stochastic Integrals for Semimartingales (*)

In this section we introduce general Itô integrals. These are useful for the formulation of general statements. The most general class of stochastic processes that we mention is that of *semimartingales*. For details on the following results we refer to Protter (2004).

### Semimartingales (*)

From the practical point of view the following class of *semimartingales* is a very rich class of processes. It turns out to be sufficient for the modeling of most finite dimensional problems that appear in finance, insurance, portfolio optimization and other areas of risk management. As we shall see later, staying within this class, is rewarded by rather general and elegant results.

As usual, we assume a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ as introduced in Sect. 5.1. In the following definition of a semimartingale we refer to several notions that we have introduced earlier in this chapter.

**Definition 5.5.1.** *A* semimartingale *is an $\underline{\mathcal{A}}$-adapted, right-continuous stochastic process $X = \{X_t,\, t \in [0, \infty)\}$ with left hand limits, where $X_t$ can be expressed as a sum of the form*

$$X_t = X_0 + A_t + M_t \tag{5.5.1}$$

*for all* $t \in [0, \infty)$*. Here* $A = \{A_t,\, t \in [0, \infty)\}$ *is a process of finite total variation and* $M = \{M_t,\, t \in [0, \infty)\}$ *is an* $(\underline{\mathcal{A}}, P)$*-local martingale.*

When $A$ is predictable, then $X$ is called a *special semimartingale* and the decomposition (5.5.1) is unique. If the discontinuous part

$$A_t^d = \sum_{0 \le s \le t} \Delta A_s$$

of $A$ and the discontinuous part

$$M_t^d = \sum_{0 \le s \le t} \Delta M_s$$

of $M$ are almost surely finite, then each of the processes $A$ and $M$ can be split into a continuous and discontinuous part, that is,

$$A_t = A_t^c + A_t^d \tag{5.5.2}$$

and

$$M_t = M_t^c + M_t^d \tag{5.5.3}$$

for $t \in [0, \infty)$, respectively.

The above defined class of semimartingales includes all stochastic processes that we have introduced so far, in particular, it covers discrete and continuous time Markov chains, diffusion processes, compound Poisson processes and Lévy processes. Note that semimartingales do not need to be Markovian.

For instance, the Wiener process $W = \{W_t,\, t \in [0, \infty)\}$, given in Definition 3.2.2, is a semimartingale. Here the decomposition (5.5.1) is simply so that $X_0 = 0,\ A_t = 0$ and $M_t = M_t^c = W_t$. The Wiener process is a martingale and, thus, by Definition 5.1.2 a local martingale.

A Poisson process $N$ with intensity $\lambda$, as given by Definition 3.5.1, and the compensated Poisson process $q$, defined in (5.1.11), are semimartingales. For the latter, we have $q_0 = X_0 = 0$, where $M_t = M_t^d = N_t - \lambda t = q_t$ is the local martingale, which is here a martingale. Furthermore, $A_t = A_t^c = \lambda t$ characterizes the predictable process $A$ of finite total variation.

In the case of a Lévy process $X$ with the notation given in (3.6.2) and almost surely finite discontinuous martingale part

$$M_t^d = \int_0^t \int_{\mathcal{E}} v(p_\varphi(dv, ds) - \varphi(dv)\, ds), \tag{5.5.4}$$

the initial value is $X_0 = 0$. The continuous local martingale part of $X$ is then of the form

$$M_t^c = \beta\, W_t \tag{5.5.5}$$

and the predictable finite total variation term is continuous and equals

$$A_t = A_t^c = \alpha\, t + \int_0^t \int_{|v|\geq 1} v\, \varphi(dv)\, ds \tag{5.5.6}$$

for $t \in [0, \infty)$, assuming $A_t$ to be finite.

It turns out that the class of semimartingales is stable with respect to important operations and transformations. It is closed with respect to stochastic integration, which when applied to semimartingales as integrands and integrators, generates again semimartingales. Further examples of transformations that map into the class of semimartingales include the application of smooth functions, equivalent changes of measure and time changes. These properties show that the class of semimartingales is a very special class and also highly suitable for financial modeling. The class of semimartingales includes all financial models that we shall cover. However, there are non-semimartingale models being actively studied, such as those based on fractional Brownian motion, see Heyde (1999), Heyde & Liu (2001) and Elliott & van der Hoek (2003).

### Itô Integral for Semimartingales (*)

For an $\underline{\mathcal{A}}$-adapted, right-continuous stochastic process $\xi = \{\xi(t),\, t \in [0, \infty)\}$ let

$$\xi(t-) \stackrel{\text{a.s.}}{=} \lim_{h \to 0+} \xi(t-h)$$

denote again the almost sure left hand limit of $\xi(t)$ at time $t$. Similarly as in (5.3.11), we define for semimartingales $X = \{X_t,\, t \in [0, \infty)\}$ and $\xi = \{\xi(s),\, s \in [0, \infty)\}$ the corresponding *Itô integral* as limit in probability

$$I_{\xi,X}(t) = \int_0^t \xi(s-)\, dX_s \stackrel{P}{=} \lim_{h \to 0} \sum_{k=1}^{i_t} \xi(t_{k-1})\, (X_{t_k} - X_{t_{k-1}}), \tag{5.5.7}$$

using an equidistant time discretization with step size $h$. What is important here is that the integrand is effectively a predictable stochastic process, see Sect. 5.1, since we take always the left hand value in a discretization interval. We could have asked $\xi$ to be a predictable process and could then write in (5.5.7) instead of $\xi(s-)$ simply $\xi(s)$.

The Itô integral enjoys important properties. Most importantly, it is again a semimartingale. If the integrator $X$ is an $(\underline{\mathcal{A}}, P)$-local martingale, then the Itô integral is also an $(\underline{\mathcal{A}}, P)$-local martingale if the integrand is, for example, continuous or locally bounded, see Protter (2004). In the case when $X$ is of finite total variation, then the Itô integral coincides with the random ordinary Riemann-Stieltjes integral.

### Itô Integral for Jump Processes (*)

Let us consider the case when a semimartingale $X$ has jumps, that is, the difference

$$\Delta X_t = X_t - X_{t-} \tag{5.5.8}$$

does not vanish for all $t \in [0, \infty)$. In this case, the following important property of the jumps of the Itô integral $I_{\xi,X}(t)$ applies:

$$\Delta I_{\xi,X}(t) = I_{\xi,X}(t) - I_{\xi,X}(t-) = \xi(t-)\, \Delta X_t \tag{5.5.9}$$

for $t \in [0, \infty)$. This means that at a jump time the value of the Itô integral increases by the value of the integrand before the jump multiplied by the jump size of the integrator. For example, if $N$ is a Poisson process, as given in Definition 3.5.1, then at its $k$th jump time $\tau_k$ we have

$$\Delta N_{\tau_k} = N_{\tau_k} - N_{\tau_{k-1}} = 1$$

for $k \in \{1, 2, \ldots\}$. Consequently, it follows in this case from (5.5.9) that the Itô integral for an integrand $\xi = \{\xi(t),\, t \in [0, \infty)\}$ takes simply the form

$$I_{\xi,N}(t) = \int_0^t \xi(s-)\, dN_s = \sum_{k=1}^{N_t} \xi(\tau_k-)\, \Delta N_{\tau_k} = \sum_{k=1}^{N_t} \xi(\tau_k-) \tag{5.5.10}$$

for $t \in [0, \infty)$. Consider the special case of a finite *pure jump process* $X = \{X_t = \sum_{0 \le s \le t} \Delta X_s,\, t \in [0, \infty)\}$, which jumps at the jump times $\tau_1, \tau_2, \ldots$ of a counting process $p = \{p_t,\, t \in [0, \infty)\}$ with jump size $\Delta X_{\tau_k} = c(k, \tau_k-)$, we obtain for an integrand $\xi = \{\xi(t),\, t \in [0, \infty)\}$ the Itô integral

$$I_{\xi,X}(t) = \int_0^t \xi(s-)\, dX_s = \sum_{k=1}^{p_t} \xi(\tau_k-)\, \Delta X_{\tau_k} = \sum_{k=1}^{p_t} \xi(\tau_k-)\, c(k, \tau_k-) \tag{5.5.11}$$

for $t \in [0, \infty)$. Here it is important to assume that the terms involved are almost surely finite. This means that the sums $\sum_{k=1}^{p_t} \xi(\tau_k-)\, \Delta X_{\tau_k}$ and $\sum_{k=1}^{p_t} \Delta X_{\tau_k}$ almost surely converge to a finite value for all $t \in [0, \infty)$. Note that the integral (5.5.11) covers also the cases of inaccessible, predictable, as well as, deterministic jump times. Thus, discrete time Markov chains and continuous time Markov chains are covered as integrators by the above formula.

### Itô Integral for Poisson Measures (*)

In the case when jump sizes are continuously distributed, as is the case for general Lévy processes, we need to consider the stochastic integration with respect to a Poisson measure. Assume that $p_\varphi(dv, dt)$ is the Poisson measure on $\mathcal{E} \times [0, \infty)$ with intensity measure $q_\varphi(dv, dt) = \varphi(dv)\, dt$, as introduced at the end of Sect. 3.5, satisfying condition (3.5.13). Here $\mathcal{E} = \Re \backslash \{0\}$ is the mark set. We again assume that a Poisson measure is such that for all $h \in [0,\infty)$ and any set $B \in \mathcal{B}(\mathcal{E})$ the $\mathcal{A}_{t+h}$-measurable random variable $p_\varphi(B,[0,t+h]) - p_\varphi(B,[0,t])$ is independent of $\mathcal{A}_t$ for all $t \in [0,\infty)$.

In generalization of relation (5.5.9), we define for a family $(\xi(v))_{v\in\mathcal{E}}$ of a.s. finite adapted processes $\xi(v) = \{\xi(v,t),\, t \in [0,\infty)\}$ with $v \in \mathcal{E}$ the Itô integral

$$I_{\xi,p_\varphi}(t) = \int_0^t \int_{\mathcal{E}} \xi(v,s-)\, p_\varphi(dv,ds) \tag{5.5.12}$$

with respect to $p_\varphi$, such that

$$\begin{aligned}\Delta I_{\xi,p_\varphi}(t) &= \int_0^t \int_{\mathcal{E}} \xi(v,s-)\, p_\varphi(dv,ds) - \int_0^{t-} \int_{\mathcal{E}} \xi(v,s-)\, p_\varphi(dv,ds)\\ &= \int_{\mathcal{E}} \xi(v,t-)\, p_\varphi(dv,\{t\})\end{aligned} \tag{5.5.13}$$

for all $t \in [0,\infty)$. This means that if at a jump time $\tau$ the Poisson measure $p_\varphi$ generates an event with mark $v$, then the change of the value of the corresponding Itô integral is given by the value $\xi(v,\tau-)$ of the integrand $\xi$ for the mark $v$ just before the jump time. Note that we do not have to write always $\xi(v,s-)$ for the integrands in (5.5.9) and (5.5.13) if $\xi(v,\cdot)$ is predictable. However, to emphasize the fact that the integrand has in the case of a jump its value always taken before the jump time, we prefer often the above notation. We refer to Protter (2004) for more details on Itô integrals for Poisson measures.

To illustrate the above definition for the case when $\mathcal{E} = (0,\lambda)$ with $\lambda \in (0,\infty)$, where

$$\varphi(v) = \begin{cases} 1 \text{ for } v \in \mathcal{E} \\ 0 \text{ otherwise,} \end{cases}$$

we obtain for the special case $\xi(v,t) = 1$ the Itô integral

$$I_{1,p_\varphi}(t) = \int_0^t \int_0^\lambda p_\varphi(dv,ds) = N_t \tag{5.5.14}$$

for $t \in [0,\infty)$. Here

$$N = \{N_t = p_\varphi((0,\lambda)\times[0,t]),\, t \in [0,\infty)\} \tag{5.5.15}$$

is a Poisson process with intensity $\lambda$.

As another example let us form the Itô integral for the simple integrand $\xi(v,t) = t$ and use the previous integrator $p_\varphi$. This leads to the Itô integral

$$I_{t,p_\varphi}(t) = \int_0^t \int_0^\lambda s\, p_\varphi(dv,ds) = \sum_{k=1}^{N_t} \tau_k, \tag{5.5.16}$$

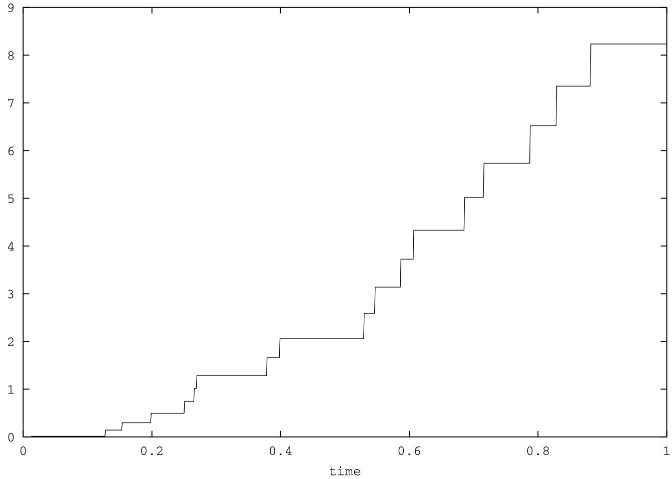



**Fig. 5.5.1.** Itô integral of $\xi(v,t) = t$ with respect to $p_\varphi$

which equals the sum of the jump times of the above Poisson process $N$ given in (5.5.15). Figure 5.5.1 displays for the path of the Poisson process $N$, shown in Fig. 3.5.1, the resulting value of the Itô integral over time. Note that the jump sizes in Fig. 5.5.1 increase as the jump times increase.

Finally, let us discuss an example where the integrand depends on the mark $v$. We choose as integrand the simple function $\xi(v,t) = \frac{v}{\lambda}$. This leads to the Itô integral $I_{\frac{v}{\lambda},p_\varphi}(t)$, which is equivalent to a compound Poisson process, as defined in (3.5.9). Here we have uniformly $U(0,1)$ distributed jump sizes. An example for a path of such an Itô integral can be found in Fig. 3.5.2.

### Itô Integral for a Lévy Process (*)

Similarly as above, we can introduce for a Lévy process $X = \{X_t,\, t \in [0,\infty)\}$ as integrator with decomposition (3.6.2) and for some stochastic process $\xi = \{\xi(t),\, t \in [0,\infty)\}$ the Itô integral

$$\begin{aligned}\int_0^t \xi(s-)\,dX_s = \int_0^t \xi(s)\,\alpha\,ds + \int_0^t \xi(s)\,\beta\,dW_s \\ + \int_0^t \int_{|v|<1} \xi(s-)\,v\,(p_\varphi(dv,ds) - \varphi(dv)\,ds) \\ + \int_0^t \int_{|v|\geq 1} \xi(s-)\,v\,p_\varphi(dv,ds) \qquad (5.5.17)\end{aligned}$$

for $t \in [0,\infty)$, see Protter (2004). Recall that the Poisson measure $p_\varphi(\cdot,\cdot)$ is specified in (3.6.2) under the condition (3.5.13). Here we have split the jump terms according to the representation (3.6.2). In the same manner as for Lévy processes one obtains the Itô integral for general semimartingales by using the decomposition (5.5.1)–(5.5.3) and calculating the different contributing terms.

## 5.6 Exercises for Chapter 5

**5.1.** If we assume that $W^i = \{W^i_t,\, t \in [0,\infty)\}$, $i \in \{1,2\}$, are standard Wiener processes, is the process $Y = \{Y_t = \alpha_1\, W^1_t + \alpha_2\, W^2_t,\, t \in [0,\infty)\}$ for $\alpha_1, \alpha_2 \in \Re$ a martingale?

**5.2.** For a standard Wiener process $W$, is the process $Y = \{Y_t = (W_t)^2,\, t \in [0,\infty)\}$ a martingale, submartingale or supermartingale?

**5.3.** Show that $M = \{M_t = (W_t)^2 - t,\, t \in [0,\infty)\}$ is a martingale, if $W$ is a standard Wiener process.

**5.4.** Let $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ be a filtered probability space with standard Wiener process $W$, geometric Brownian motion $X = \{X_t = \exp\{(r - \frac{1}{2}\,\sigma^2)\,t + \sigma\, W_t\},\, t \in [0,\infty)\}$ and a money account $B = \{B_t = \exp\{r\,t\},\, t \in [0,\infty)\}$ with interest rate $r$. Is the discounted process $\bar{X} = \{\bar{X}_t = \frac{X_t}{B_t},\, t \in [0,\infty)\}$ a martingale, submartingale or supermartingale? Use the fact that an $N(0,1)$ distributed Gaussian random variable $Y$ has Laplace transform

$$E(\exp\{\sigma\, Y\}) = \exp\left\{\frac{1}{2}\,\sigma^2\right\},$$

see (1.3.76).

**5.5.** Compute the quadratic variation $[Y]$ for a transformed Wiener process $Y = \{Y_t = a\,t + b\,W_t,\, t \in [0,\infty)\}$, where $W$ is a standard Wiener process.

**5.6.** Determine the covariation $[Y, W]$ between the transformed Wiener process $Y$ from Exercise 5.5 and the standard Wiener process $W$.

**5.7.** If $X = \{X_t,\, t \in [0,\infty)\}$ is a martingale and $g(\cdot)$ a convex function, is the process

$$g(X) = \{g(X_t),\, t \in [0,\infty)\}$$

a martingale, supermartingale, submartingale or none of these?

**5.8.** (*) Prove the martingale property for Itô integrals with piecewise constant deterministic integrands and the Wiener process as integrator.

**5.9.** (*) Show that the correlation property holds for Itô integrals with piecewise constant deterministic integrands and the Wiener process as integrator.

**5.10.** (*) Derive the linearity property for Itô integrals with piecewise constant deterministic integrands and the Wiener process as integrator.

**5.11.** (*) For a Lévy process $X = \{X_t,\, t \in [0,\infty)\}$ with $E(X_t \,|\, \mathcal{A}_0) = 0$ for all $t \in [0,\infty)$ prove that $X$ is a martingale.

# 6

# The Itô Formula

The price of a security, for instance, a zero coupon bond which generates some future payoff at a maturity date, is often dependent on the value of an underlying process. In many applications, the effect of changes in the underlying process on this price needs to be quantified. In deterministic calculus this type of problem is handled by the chain rule. In stochastic calculus the corresponding generalization of the chain rule is given by the Itô formula. This stochastic chain rule contains terms reflecting the effect due to the stochastic processes involved having non-zero quadratic variation. In this chapter we introduce, apply and derive the Itô formula. It is widely regarded as the main tool in stochastic calculus and is therefore highly important in quantitative finance.

## 6.1 The Stochastic Chain Rule

### The Classical Chain Rule

First consider an example, where the classical deterministic chain rule applies. Suppose we observe in the market the price of a savings account $B_t = \exp\{r\,t\}$, where $r$ denotes a constant continuously compounding interest rate. Then

$$dB_t = r\,B_t\,dt \tag{6.1.1}$$

for $t \in [0, \infty)$ with $B_0 = 1$. Also suppose that we are interested in a financial quantity $u(B_t)$, where $u : \Re \to \Re$ is some differentiable function. For instance, such a quantity could be the square of the value of the savings account, that is, $u(B_t) = (B_t)^2$. Furthermore, suppose that we need to express the evolution of this quantity in terms of properties of $u$ and $B$ with respect to time. In this case, by using the well-known chain rule of deterministic calculus, we can write the equations

$$u(B_t) = u(B_0) + \int_0^t u'(B_s)\,dB_s = u(B_0) + \int_0^t u'(B_s)\,r\,B_s\,ds \tag{6.1.2}$$

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

for $t \in [0, \infty)$. Note from the first line in (6.1.2) that the value of the quantity $u(B_t)$ can be interpreted as the gains from trade with integrand $u'(B_t)$ and integrator $B_t$ for $t \in [0, \infty)$. This means for our simple deterministic example that

$$(B_t)^2 = (B_0)^2 + 2 \int_0^t B_s \, dB_s \tag{6.1.3}$$

for $t \in [0, \infty)$.

### A Stochastic Example

In Sect. 5.3 we considered the Itô integral

$$I_{W,W}(t) = \int_0^t W_s \, dW_s,$$

which is the double Wiener integral for a Wiener process $W = \{W_t, \ t \in [0, \infty)\}$. This stochastic integral was interpreted as the gains from trade, where the number of shares held in the asset whose price was $W$ was equal to its price. By rewriting equation (5.3.8) we obtain

$$(W_t)^2 = 2 \int_0^t W_s \, dW_s + [W]_t = 2 \int_0^t W_s \, dW_s + \int_0^t ds \tag{6.1.4}$$

for $t \in [0, \infty)$. Using the Itô differentials $dW_t$ and $d(W_t)^2$ the equation (6.1.4) can be expressed in the equivalent Itô differential form

$$d(W_t)^2 = 2 \, W_t \, dW_t + dt \tag{6.1.5}$$

for $t \in [0, \infty)$ with initial value $(W_0)^2 = 0$. As previously explained, the equation (6.1.5) is nothing more than an abbreviated form of the stochastic integral equation (6.1.4). This integral equation involves an Itô integral, which is well defined, as discussed in the previous chapter. As a rule in stochastic calculus we shall see later that one can treat $(dW_t)^2$ as $d[W]_t = dt$, see (5.4.5). Note however that $d(W_t)^2$ is different to $(dW_t)^2$. Another rule will suggest setting $(dt)^2 = d[\cdot]_t = 0$ and $dW_t dt = d[W, t]_t = 0$.

### Heuristic Derivation of the Itô Formula

One of the key features of the Itô integral with respect to the Wiener process is its martingale property, described in (5.4.3), which makes it an essential tool for pricing in finance. However, as previously indicated, this fundamental property does not come freely, namely the chain rule of classical calculus does not apply when using Itô integrals. Instead, the stochastic chain rule, the *Itô formula*, has to be applied. We now provide a heuristic derivation of the Itô formula. In Sect. 6.6 a proof of this formula will be presented.

Let $X = \{X_t,\, t \in [0,\infty)\}$ be a stochastic process that is characterized by the Itô differential

$$dX_t = e_t\, dt + f_t\, dW_t \tag{6.1.6}$$

for $t \in [0,\infty)$ with initial value $X_0 = x_0$, see (5.3.15). Here $e = \{e_t,\, t \in [0,\infty)\}$ and $f = \{f_t,\, t \in [0,\infty)\}$ are two stochastic processes with appropriate measurability and integrability properties. Consider a finite difference approximation of the Itô differential (6.1.6) of the form

$$\Delta X_{t_k} = X_{t_{k+1}} - X_{t_k} \approx e_{t_k}\, h + f_{t_k}\, (W_{t_{k+1}} - W_{t_k}) \tag{6.1.7}$$

for $t_k$ from an equidistant time discretization $\{t_\ell = \ell\, h,\, \ell \in \{0,1,\ldots\}\}$ with step size $h > 0$, as introduced in (5.2.1).

We focus our attention on changes in the value $u(t, X_t)$ for a function $u : [0,\infty) \times \Re \to \Re$, resulting from changes in the time $t$ and the value of the underlying $X_t$. Assume that $u$ is differentiable with respect to time $t$ and twice continuously differentiable with respect to the spatial component $x$, that is, the functions $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x}$ and $\frac{\partial^2 u}{\partial x^2}$ exist and are continuous.

To quantify the changes in $u(t, X_t)$ caused by changes in $X_t$ we consider over small time intervals $[t_k, t_{k+1})$ the differences

$$\Delta u(t_k, X_{t_k}) = u(t_{k+1}, X_{t_{k+1}}) - u(t_k, X_{t_k}) \tag{6.1.8}$$

for $k \in \{0,1,\ldots\}$. Since $u$ is assumed to be sufficiently differentiable we can apply a Taylor expansion to obtain the expansion

$$\begin{aligned}\Delta u(t_k, X_{t_k}) = {} & \frac{\partial u(t_k, X_{t_k})}{\partial t}\, h + \frac{\partial u(t_k, X_{t_k})}{\partial x}\, \Delta X_{t_k} \\ & + \frac{1}{2}\, \frac{\partial^2 u(t_k, X_{t_k})}{\partial x^2}\, (\Delta X_{t_k})^2 + R_{t_k},\end{aligned} \tag{6.1.9}$$

where $R_{t_k}$ is the corresponding remainder term.

If the quadratic variation of $X$ were zero, then $h \to 0$ would imply $(\Delta X_{t_k})^2 \to 0$ asymptotically and hence the corresponding term in (6.1.9) would not influence the movements of $u(t, X_t)$. However, in the given stochastic setting this is not the case and, therefore, we need to consider the approximation

$$(\Delta X_{t_k})^2 \approx [X]_{h,t_{k+1}} - [X]_{h,t_k} \approx (f_{t_k})^2\, h, \tag{6.1.10}$$

where $[X]_{h,t}$ denotes the approximate quadratic variation, see (5.2.3). Substituting this expression, together with (6.1.7) into (6.1.9) yields the relation

$$\begin{aligned}\Delta u(t_k, X_{t_k}) = {} & \left(\frac{\partial u(t_k, X_{t_k})}{\partial t} + e_{t_k} \frac{\partial u(t_k, X_{t_k})}{\partial x} + \frac{1}{2} (f_{t_k})^2 \frac{\partial^2 u(t_k, X_{t_k})}{\partial x^2}\right) h \\ & + f_{t_k}\, \frac{\partial u(t_k, X_{t_k})}{\partial x}\, (W_{t_{k+1}} - W_{t_k}) + \bar{R}_{t_k},\end{aligned} \tag{6.1.11}$$

where $\bar{R}_{t_k}$ is the corresponding remainder term.

### Itô Formula

Letting the time discretization become finer and finer in (6.1.11), that is $h \to 0$, results in the one-dimensional version of the *Itô formula*

$$du(t, X_t) = \left( \frac{\partial u(t, X_t)}{\partial t} + e_t \frac{\partial u(t, X_t)}{\partial x} + \frac{1}{2} (f_t)^2 \frac{\partial^2 u(t, X_t)}{\partial x^2} \right) dt + f_t \frac{\partial u(t, X_t)}{\partial x} dW_t \tag{6.1.12}$$

for $t \in [0, \infty)$. This formula will be derived rigorously towards the end of this chapter.

Note again that the Itô differential in (6.1.12) is only a shorthand notation for the *integral representation* of the Itô formula given as

$$u(t, X_t) = u(s, X_s) + \int_s^t \left( \frac{\partial u(z, X_z)}{\partial t} + e_z \frac{\partial u(z, X_z)}{\partial x} + \frac{1}{2} (f_z)^2 \frac{\partial^2 u(z, X_z)}{\partial x^2} \right) dz + \int_s^t f_z \frac{\partial u(z, X_z)}{\partial x} dW_z \tag{6.1.13}$$

for $t \in [0, \infty)$ and $s \in [0, t]$. We remark that by using the notion of quadratic variation, introduced in the previous chapter, we can write the Itô formula (6.1.12) in the compact form

$$du(t, X_t) = \frac{\partial u(t, X_t)}{\partial t} dt + \frac{\partial u(t, X_t)}{\partial x} dX_t + \frac{1}{2} \frac{\partial^2 u(t, X_t)}{\partial x^2} d[X]_t \tag{6.1.14}$$

for $t \in [0, \infty)$. One can read off the following rule

$$(dX_t)^2 = d[X]_t$$

if $X$ is a continuous process. This generalizes the rule that we mentioned after equation (6.1.5). As shown in Föllmer (1981), the Itô formula holds very generally in a pathwise sense requiring almost no technical assumptions.

### Example for a Stochastic Exponential

Let us consider the one dimensional Itô differential

$$dX_t = e_t \, dt + f_t \, dW_t \tag{6.1.15}$$

for $t \in [0, \infty)$ with initial value $X_0 = x_0$. Note by (5.4.5) that

$$d[X]_t = (f_t)^2 \, d[W]_t = (f_t)^2 \, dt.$$

The exponential

$$Y_t = u(X_t) = \exp\{X_t\} \tag{6.1.16}$$

has then by application of the Itô formula (6.1.12) the Itô differential

$$\begin{aligned} dY_t &= d(\exp\{X_t\}) \\ &= \exp\{X_t\} \left(e_t + \frac{1}{2}(f_t)^2\right) dt + \exp\{X_t\}\, f_t\, dW_t \\ &= Y_t \left(e_t + \frac{1}{2}(f_t)^2\right) dt + Y_t\, f_t\, dW_t \end{aligned} \tag{6.1.17}$$

for $t \in [0,\infty)$ with initial value $Y_0 = \exp\{x_0\}$. In the case when $e_t = e$ and $f_t = f$ are constants, the process $X = \{X_t,\, t \in [0,\infty)\}$ is a transformed Wiener process, see (3.2.7), and $Y = \{Y_t,\, t \in [0,\infty)\}$ is a geometric Brownian motion, as is employed under the BS model. One can interpret $Y$ as a solution of a *stochastic differential equation* (SDE) since here some feedback in the drift and diffusion coefficient is built in. In the next chapter we shall study SDEs of more general form.

### Example for Powers of Processes

Let us give another example, where we start again from the Itô differential (6.1.15) for the process $X = \{X_t,\, t \in [0,\infty)\}$. Now we consider for some exponent $k \neq 0$ the power

$$Y_t = u(X_t) = (X_t)^k \tag{6.1.18}$$

for $t \in [0,\infty)$. By application of the Itô formula (6.1.12) we obtain the Itô differential

$$\begin{aligned} dY_t &= k\,(X_t)^{k-1}\,(e_t\,dt + f_t\,dW_t) + \frac{1}{2}\,k\,(k-1)\,(X_t)^{k-2}\,(f_t)^2\,dt \\ &= k\,(Y_t)^{\frac{k-1}{k}}\,(e_t\,dt + f_t\,dW_t) + \frac{1}{2}\,k\,(k-1)\,(Y_t)^{\frac{k-2}{k}}\,(f_t)^2\,dt \end{aligned} \tag{6.1.19}$$

for $t \in [0,\infty)$ with $Y_0 = (x_0)^k$.

## 6.2 Multivariate Itô Formula

In the context of financial modeling, the discussion of functionals of two or more underlying stochastic processes, such as a stock price and a stochastic interest rate, is often required. To enable us to treat such problems properly we consider multi-dimensional stochastic processes or, equivalently, vector valued Itô integrals. For this reason we introduce multi-component Itô differentials with respect to multi-dimensional standard Wiener processes. These then will appear in a multivariate version of the Itô formula to be formulated below.

### Multi-Dimensional Wiener Process

**Definition 6.2.1.** *We call the vector process* $\boldsymbol{W} = \{\boldsymbol{W}_t = (W_t^1, W_t^2, \ldots, W_t^m)^\top,\ t \in [0,\infty)\}$ *an* $m$-dimensional standard Wiener process *if each of its components* $W^j = \{W_t^j,\ t \in [0,\infty)\},\ j \in \{1,2,\ldots,m\}$ *is a scalar* $\underline{\mathcal{A}}$*-adapted standard Wiener process and the Wiener processes* $W^k$ *and* $W^j$ *are independent for* $k \neq j,\ k,j \in \{1,2,\ldots,m\}$.

This means that according to Definition 3.2.2 of a Wiener process, each random variable $W_t^j$ is Gaussian and $\mathcal{A}_t$-measurable with

$$E\left(W_t^j \,\middle|\, \mathcal{A}_0\right) = 0 \tag{6.2.1}$$

and we have independent increments $W_t^j - W_s^j$ such that

$$E\left(W_t^j - W_s^j \,\middle|\, \mathcal{A}_s\right) = 0 \tag{6.2.2}$$

for $t \in [0,\infty),\ s \in [0,t]$ and $j \in \{1,2,\ldots,m\}$. Moreover, we have here additionally the property that

$$E\left(\left(W_t^i - W_s^i\right)\left(W_t^j - W_s^j\right) \,\middle|\, \mathcal{A}_s\right) = \begin{cases} (t-s) & \text{for } i=j \\ 0 & \text{otherwise} \end{cases} \tag{6.2.3}$$

for $t \in [0,\infty),\ s \in [0,t]$ and $i,j \in \{1,2,\ldots,m\}$.

Note that the covariation between different components of the above standard Wiener process is zero, see (5.4.5), that is

$$[W^i, W^j]_t = \begin{cases} t & \text{for} \quad i=j \\ 0 & \text{for} \quad i \neq j \end{cases} \tag{6.2.4}$$

for $t \in [0,\infty)$ and $i,j \in \{1,2,\ldots,m\}$.

To illustrate the above notion in preparation of future examples, Fig. 6.2.1 shows the sample paths of the two components of a two-dimensional standard Wiener process. Each of these two components forms a standard one-dimensional Wiener process and both Wiener processes are independent.

Figure 6.2.2 presents a different visualization of the same pair of trajectories for the two-dimensional Wiener process. Here $W_t^1$ and $W_t^2$ represent the $x$ and $y$ coordinates at time $t$, respectively, that generate a trace similar to the motion of a pollen particle under the microscope. Recall that such a motion was originally observed by Robert Brown, giving rise to it the name Brownian motion. As indicated above, it can be modeled by two independent Wiener processes.

### Vector Itô Differentials

Consider a $d$-dimensional vector function $\boldsymbol{e} :\ [0,\infty)\times\Omega \to \Re^d$ with predictable components $e^k,\ k \in \{1,2,\ldots,d\}$. We have to assume that the components

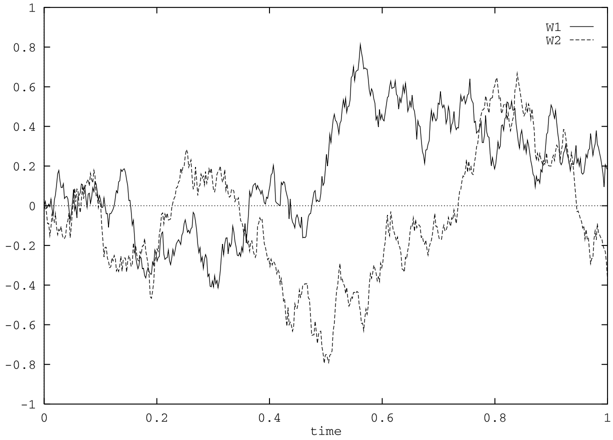



**Fig. 6.2.1.** Components of a two-dimensional standard Wiener process

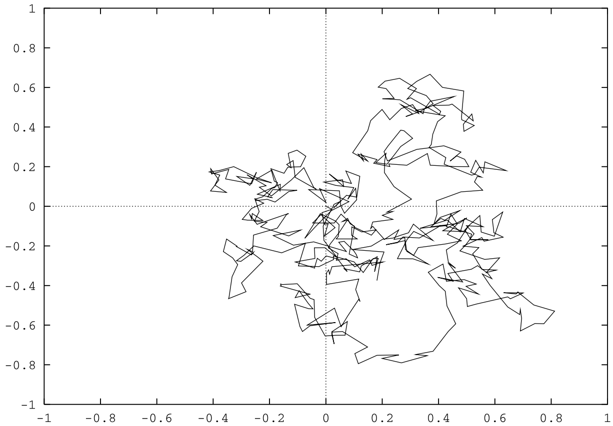



**Fig. 6.2.2.** Trace of a two-dimensional Wiener process

satisfy appropriate integrability and measurability conditions. These are similar to those we introduced for the one-dimensional case. For simplicity, we assume here that

$$\int_0^T |e_z^k| \, dz < \infty \tag{6.2.5}$$

almost surely for $k \in \{1, 2, \ldots, d\}$ and $\boldsymbol{F} : [0, T] \times \Omega \to \Re^{d \times m}$ to be a $d \times m$ matrix valued function with

$$\int_0^T \left(F_z^{i,j}\right)^2 \, dz < \infty \tag{6.2.6}$$

almost surely for $i \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ and all $T \in (0, \infty)$, see Protter (2004). This allows us to introduce a $d$-dimensional stochastic vector process $\boldsymbol{X} = \{\boldsymbol{X}_t = (X_t^1, X_t^2, \ldots, X_t^d)^\top,\, t \in [0,\infty)\}$, where the $k$th component $X^k$ is defined via the Itô integral equation

$$X_t^k - X_0^k = \int_0^t e_z^k\, dz + \sum_{j=1}^m \int_0^t F_z^{k,j}\, dW_z^j \tag{6.2.7}$$

for $t \in [0,\infty)$ and given $\mathcal{A}_0$-measurable initial value $X_0^k \in \Re$, $k \in \{1,2,\ldots,d\}$.

Analogous to the scalar case we denote by $\boldsymbol{e}_t$ and $\boldsymbol{F}_t$ for a given time $t \in [0,\infty)$ the vector and matrix valued random variables, respectively. Then we write the vector valued stochastic integral equation in the form

$$\boldsymbol{X}_t - \boldsymbol{X}_0 = \int_0^t \boldsymbol{e}_z\, dz + \int_0^t \boldsymbol{F}_z\, dW_z \tag{6.2.8}$$

for any $t \in [0,\infty)$ with initial value $\boldsymbol{X}_0 = (X_0^1, \ldots, X_0^d)^\top$. This can be expressed equivalently as the $d$-dimensional *vector Itô differential* given by

$$d\boldsymbol{X}_t = \boldsymbol{e}_t\, dt + \boldsymbol{F}_t\, dW_t, \tag{6.2.9}$$

for $t \in [0,\infty)$ with initial value $\boldsymbol{X}_0 \in \Re^d$. Choosing the dimension $d = 1$, leads to the case of a scalar Itô differential with respect to several independent Wiener processes.

### Multivariate Itô Formula

In the previous section it was noted that for the scalar case with one driving Wiener process the Itô formula involves the quadratic variation of this Wiener process. In the multivariate case, with a multi-dimensional driving Wiener process, it turns out that the covariations between different components of the vector stochastic differential appear in the following *multivariate Itô formula.*

**Theorem 6.2.2.** *Assume that the function* $u : [0,\infty) \times \Re^d \to \Re$ *has continuous partial derivatives* $\frac{\partial u}{\partial t}$, $\frac{\partial u}{\partial x^k}$ *and* $\frac{\partial^2 u}{\partial x^k\, \partial x^i}$ *for* $k, i \in \{1,2,\ldots,d\}$ *and* $\boldsymbol{x} = (x^1, x^2, \ldots, x^d)^\top$. *Define a scalar stochastic process* $Y = \{Y_t,\, t \in [0,\infty)\}$ *by setting*

$$Y_t = u(t, X_t^1, X_t^2, \ldots, X_t^d), \tag{6.2.10}$$

*for* $t \in [0,\infty)$, *where the vector* $\boldsymbol{X}_t = (X_t^1, X_t^2, \ldots, X_t^d)^\top$ *satisfies the vector Itô differential* (6.2.9). *Then the Itô differential for* $Y$ *is of the form*

$$\begin{aligned} dY_t &= du(t, X_t^1, X_t^2, \ldots, X_t^d) \\ &= \left\{ \frac{\partial u}{\partial t} + \sum_{k=1}^d e_t^k \frac{\partial u}{\partial x^k} + \frac{1}{2} \sum_{j=1}^m \sum_{i,k=1}^d F_t^{i,j}\, F_t^{k,j} \frac{\partial^2 u}{\partial x^i\, \partial x^k} \right\} dt \\ &\quad + \sum_{j=1}^m \sum_{i=1}^d F_t^{i,j} \frac{\partial u}{\partial x^i}\, dW_t^j, \end{aligned} \tag{6.2.11}$$

*for* $t \in [0, \infty)$ *with* $Y_0 = u(0, X_0^1, X_0^2, \ldots, X_0^d)$. *Here the partial derivatives in* (6.2.11) *are evaluated at* $(t, X_t^1, X_t^2, \ldots, X_t^d)$.

An informal derivation of this formula, similar to that for the scalar case presented earlier in (6.1.12), provides a quick and insightful way to understand where the various terms appearing in (6.2.11) come from. To see this easily one has simply to apply the corresponding Taylor expansion for the function $u$ and use the rules

$$dW_t^i \, dW_t^j \approx \begin{cases} dt & \text{for} \quad i = j \\ 0 & \text{for} \quad i \neq j \end{cases} \tag{6.2.12}$$

and

$$dW_t^i \, dt \approx 0. \tag{6.2.13}$$

As previously explained, these rules of stochastic calculus yield terms in addition to those usually observed in the deterministic chain rule because of the effects of the covariations between the integrands and integrators involved.

Similar as in formula (6.1.14), by using the notion of covariation, we can write the multivariate Itô formula (6.2.11) in the form

$$du(t, X_t^1, X_t^2, \ldots, X_t^d) = \frac{\partial u}{\partial t} \, dt + \sum_{i=1}^{d} \frac{\partial u}{\partial x^i} \, dX_t^i + \frac{1}{2} \sum_{i,k=1}^{d} \frac{\partial^2 u}{\partial x^i \, \partial x^k} \, d[X^i, X^k]_t \tag{6.2.14}$$

for all $t \in [0, \infty)$. Here the partial derivatives of $u$ on the right hand side of (6.2.14) are taken at $(t, X_t^1, X_t^2, \ldots, X_t^d)$.

## 6.3 Some Applications of the Itô Formula

### Integration-by-Parts Formula

Let us consider two continuous processes $X^1 = \{X_t^1, \, t \in [0, \infty)\}$ and $X^2 = \{X_t^2, \, t \in [0, \infty)\}$ having an Itô differential and finite covariation. Suppose that the Itô differential of the product

$$Y_t = u(t, X_t^1, X_t^2) = X_t^1 \, X_t^2$$

is required. The Itô formula (6.2.14) can then be used to derive for the above product the following *integration-by-parts formula*

$$d(X_t^1 \, X_t^2) = X_t^1 \, dX_t^2 + X_t^2 \, dX_t^1 + d[X^1, X^2]_t. \tag{6.3.1}$$

We consider as an example three cases, where $X^1$ and $X^2$ are standard Wiener processes:

**1.** First assume that the two Wiener processes are the same, that is $X_t^1 = X_t^2 = W_t^1$, where $W_t^1$ is a standard Wiener process. This case was considered in (6.1.5). By rewriting this equation we obtain

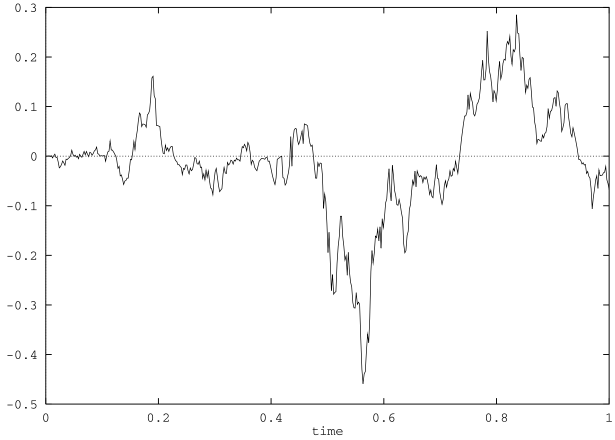



**Fig. 6.3.1.** Product of two independent Wiener processes

$$dY_t = d(W_t^1)^2 = 2\,W_t^1\,dW_t^1 + dt \tag{6.3.2}$$

for $t \in [0, \infty)$. Recall that Fig. 5.3.3 displayed a sample path for $\frac{1}{2}\,Y_t = \frac{1}{2}(W_t^1)^2$.

**2.** In the second case we assume that the two Wiener processes $X^1 = W^1$ and $X^2 = W^2$ are independent, that is $W^1$ and $W^2$ are two independent standard Wiener processes. This then leads by application of the integration-by-parts formula (6.3.1) to the Itô differential

$$dY_t = d(W_t^1\,W_t^2) = W_t^1\,dW_t^2 + W_t^2\,dW_t^1 \tag{6.3.3}$$

for $t \in [0, \infty)$. Note that there is no drift on the right hand side of (6.3.3) since the covariation between the two independent Wiener processes is zero. The formula (6.3.3) coincides with the classic integration by parts formula because there is zero covariation between $W^1$ and $W^2$. In Fig. 6.3.1 we use the same sample path of the two-dimensional standard Wiener process that was shown in Fig. 6.2.1 and Fig. 6.2.2 to generate a corresponding path for the product $Y_t = W_t^1\,W_t^2$.

**3.** The third case assumes that the two standard Wiener processes are correlated, that is we set $X_t^1 = \varrho\,W_t^1 + \sqrt{1-\varrho^2}\,W_t^2$ and $X_t^2 = W_t^1$, where $W^1$ and $W^2$ are independent standard Wiener processes and $\varrho \in [-1, 1]$ is the correlation coefficient, see (1.4.39). We then obtain by the formula (6.3.1) the Itô differential

$$\begin{aligned} dY_t &= d(X_t^1\,X_t^2) \\ &= X_t^1\,dW_t^1 + X_t^2\left(\varrho\,dW_t^1 + \sqrt{1-\varrho^2}\,dW_t^2\right) + \varrho\,dt \\ &= X_t^1\,dX_t^2 + X_t^2\,dX_t^1 + \varrho\,dt \end{aligned} \tag{6.3.4}$$

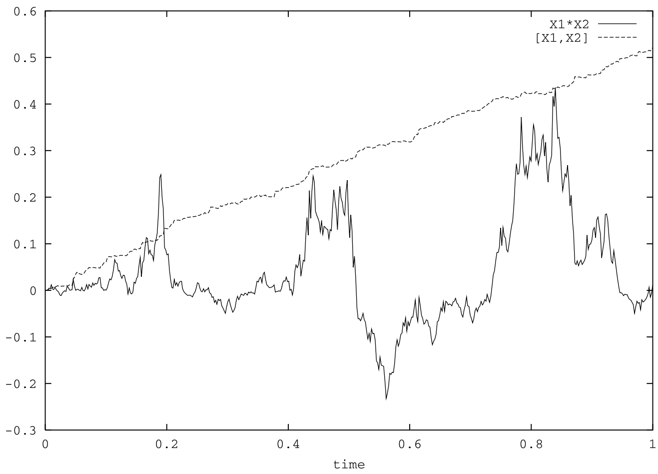



**Fig. 6.3.2.** Product of correlated Wiener processes and their approximate covariation

for $t \in [0, \infty)$. It is easy to see that both (6.3.2) and (6.3.3) are included in formula (6.3.4) for the choices of $\varrho = 1$ and $\varrho = 0$, respectively. Note that for the product of Wiener processes, the correlation coefficient $\varrho$ appears as the drift coefficient in the resulting Itô differential.

Finally, to demonstrate the covariation of correlated Wiener processes, we show in Fig. 6.3.2 the approximate covariation $[X^1, X^2]_{h,t}$, see (5.2.15), together with the product $X_t^1\, X_t^2$ for the correlation coefficient $\varrho = \frac{1}{2}$.

## Example for Geometric Brownian Motion

Let us consider a one-dimensional Itô differential that uses two independent standard Wiener processes and is given by

$$dX_t = e_t^1\, dt + F_t^{1,1}\, dW_t^1 + F_t^{1,2}\, dW_t^2 \tag{6.3.5}$$

with initial value $X_0 = 0$. The functional

$$Y_t = \exp\{X_t\} \tag{6.3.6}$$

has by the Itô formula (6.2.11) the Itô differential

$$\begin{aligned} dY_t &= d\left(\exp\{X_t\}\right) \\ &= Y_t \left( e_t^1 + \frac{1}{2}\left( \left(F_t^{1,1}\right)^2 + \left(F_t^{1,2}\right)^2 \right) \right) dt + Y_t\, F_t^{1,1}\, dW_t^1 + Y_t\, F_t^{1,2}\, dW_t^2 \end{aligned} \tag{6.3.7}$$

for $t \in [0, \infty)$, with initial value $Y_0 = 1$. Note that the process $Y$ is a diffusion process. More precisely, it is a generalized version of geometric Brownian motion, introduced in (4.1.2). Here we have the drift coefficient

$$a(t,x) = x\left(e_t^1 + \frac{1}{2}\left(\left(F_t^{1,1}\right)^2 + \left(F_t^{1,2}\right)^2\right)\right), \tag{6.3.8}$$

that can be compared to (4.3.7). The diffusion coefficients corresponding to $W^1$ and $W^2$ are given by

$$b^1(t,x) = x\,F_t^{1,1} \tag{6.3.9}$$

and

$$b^2(t,x) = x\,F_t^{1,2}, \tag{6.3.10}$$

respectively. These diffusion coefficients generalize what was obtained in (4.3.8), where we had only one driving Wiener process.

For the above Itô differential both the drift and diffusion coefficients appear as products of the asset price with some constants, as was the case in (6.3.7). The constant associated with the drift coefficient is often called the *appreciation rate* or *expected rate of return*. Recall that the constant associated with a diffusion coefficient is the *volatility* component for this diffusion term. If appreciation rate and volatilities are constants, then the corresponding model is called the *Black-Scholes* (BS) *model*.

If one looks at the stochastic differential (6.3.7), then a certain feedback in the drift and diffusion term is modeled. We call an Itô differential that involves some feedback from the state variable, here $Y_t$, a *stochastic differential equation* (SDE). It will be our focus in the next chapter to present results on SDEs. However, within this chapter we continue to study the Itô formula applied to stochastic differentials which cover also SDEs.

### Product of Two Geometric Brownian Motions

Since the Black-Scholes model plays such a central role in asset price modeling, we go in detail through a number of almost elementary applications of the Itô formula. Consider two asset price processes $X^1$ and $X^2$ that are defined as geometric Brownian motions by functionals of the type

$$X_t^i = \exp\left\{\mu^i\,t + \sigma^{i,1}\,W_t^1 + \sigma^{i,2}\,W_t^2\right\}$$

for $i \in \{1,2\}$ and $t \in [0,\infty)$, where $W^1$ and $W^2$ denote two independent standard Wiener processes.

By the Itô formula (6.2.11) we obtain, similarly to (6.3.7), the Itô differentials

$$dX_t^i = X_t^i\left(\mu^i + \frac{1}{2}\left((\sigma^{i,1})^2 + (\sigma^{i,2})^2\right)\right)dt + X_t^i\,\sigma^{i,1}\,dW_t^1 + X_t^i\,\sigma^{i,2}\,dW_t^2 \tag{6.3.11}$$

for $i \in \{1,2\}$ and $t \in [0,\infty)$.

We compute the Itô differential of the product $Y_t = X_t^1\,X_t^2$. Again, by application of the Itô formula (6.2.11) we obtain

$$dY_t = d(X_t^1\, X_t^2)$$
$$= Y_t \left( \mu^1 + \mu^2 + \frac{1}{2} \left(\sigma^{1,1} + \sigma^{2,1}\right)^2 + \frac{1}{2} \left(\sigma^{1,2} + \sigma^{2,2}\right)^2 \right) dt$$
$$+ Y_t \left(\sigma^{1,1} + \sigma^{2,1}\right) dW_t^1 + Y_t \left(\sigma^{1,2} + \sigma^{2,2}\right) dW_t^2 \tag{6.3.12}$$

for $t \in [0, \infty)$. Consequently, the product of two geometric Brownian motions is a geometric Brownian motion, since the drift and diffusion coefficients in (6.3.12) appear as products of $Y_t$ together with some constants. Note also that the appreciation rates and the volatilities of the product of two geometric Brownian motions are obtained by summing the appreciation rates and volatilities of their components.

### Powers of Geometric Brownian Motion

We have seen that products of two geometric Brownian motions are also geometric Brownian motions. We now show that the power of a geometric Brownian motion is also a geometric Brownian motion.

Let $X$ denote a scalar geometric Brownian motion characterized by the Itô differential

$$dX_t = X_t\, a\, dt + X_t\, \sigma\, dW_t, \tag{6.3.13}$$

for $t \in [0, \infty)$ with appreciation rate $a$, volatility $\sigma$ and initial value $X_0 = x$, where $W$ is a standard Wiener process. Then by application of the Itô formula (6.2.11) we obtain for any real valued exponent $k$ and

$$Y_t = (X_t)^k \tag{6.3.14}$$

the Itô differential

$$dY_t = d(X_t)^k = Y_t \left( k\, a + \frac{1}{2}\, k\, (k-1)\, \sigma^2 \right) dt + Y_t\, k\, \sigma\, dW_t \tag{6.3.15}$$

for $t \in [0, \infty)$. This shows that $Y$ is again a geometric Brownian motion, because the drift and diffusion coefficients in (6.3.15) are expressed as products of constants and $Y_t$.

### Inverse of a Geometric Brownian Motion

An interesting phenomenon is observed when considering the dynamics of an inverse of a given geometric Brownian motion, which follows for the exponent $k = -1$ from equation (6.3.15). Taking the stochastic differentials (6.3.12) and (6.3.15) into account, it is clear that not only powers and products but also ratios of geometric Brownian motions are again geometric Brownian motions. These convenient properties certainly had some influence on the historical

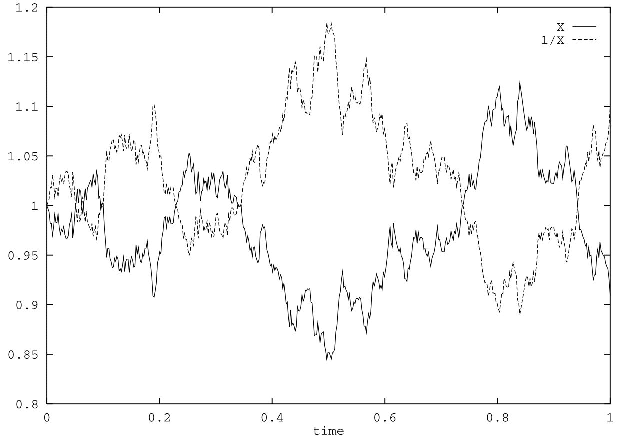



**Fig. 6.3.3.** A geometric Brownian motion and its inverse

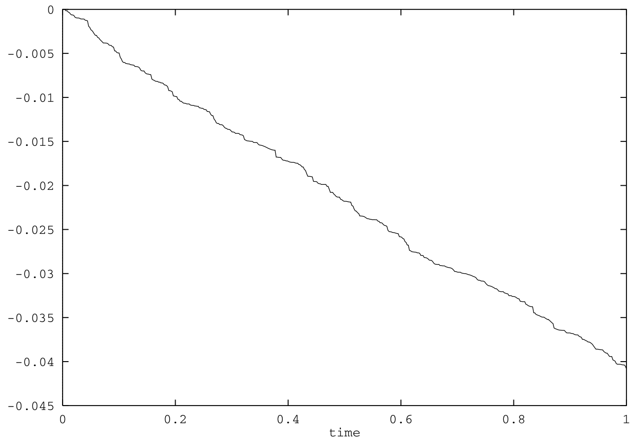



**Fig. 6.3.4.** Approximate covariation between $\frac{1}{X}$ and $X$

development of quantitative finance. In particular, they helped to make the BS model the standard market model.

Figure 6.3.3 shows a sample path of a geometric Brownian motion $X = \{X_t,\ t \in [0, \infty)\}$ with $X_0 = 1,\ a = 0,\ \sigma = 0.2$ together with its inverse $\frac{1}{X_t}$. As is apparent from (6.3.15), in this case the inverse $X_t^{-1}$ has an appreciation rate equal to $\sigma^2$ and is negatively correlated to $X_t$. This negative correlation is visualized in Fig. 6.3.4, which displays the covariation $[X^{-1}, X]_t$ between $X_t^{-1}$ and $X_t$. This covariation, see (5.4.5), is given by

$$[X^{-1}, X]_t = -\sigma^2\, t. \tag{6.3.16}$$

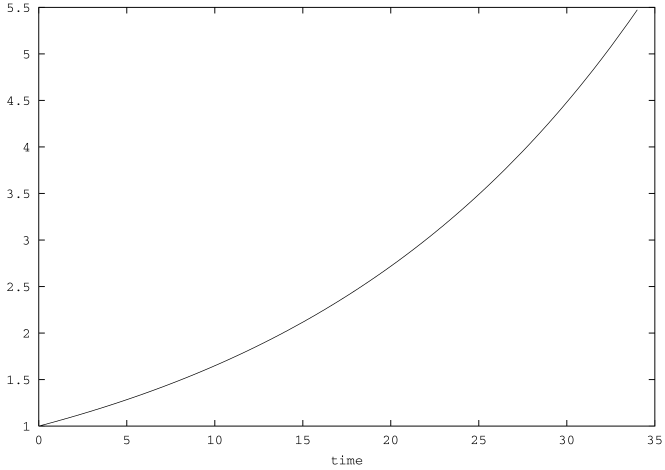



**Fig. 6.3.5.** Savings account

### Black-Scholes Model for a Stock Market

Let us now model a stock market in continuous time with asset prices that follow geometric Brownian motions. The fluctuations of stock prices in the market are driven by continuous trading uncertainty, which is modeled by $d \in \mathcal{N}$ independent standard Wiener processes $W^1, W^2, \ldots, W^d$.

For simplicity, we consider a deterministic, constant short rate $r$. We assume that the interest is continuously accrued. To model the accumulation of interest we form the *savings account* $S_t^0$ at time $t$ as the exponential

$$S_t^0 = \exp\left\{X_t^0\right\} \tag{6.3.17}$$

with

$$X_t^0 = r\,t \tag{6.3.18}$$

for $t \in [0, \infty)$. Obviously,

$$dX_t^0 = r\,dt$$

and, therefore, by the Itô formula when applied to the exponential function (6.3.17), we obtain the differential equation

$$dS_t^0 = S_t^0\,r\,dt \tag{6.3.19}$$

for $t \in [0, \infty)$ with $S_0^0 = 1$. In Fig. 6.3.5 we plot the resulting savings account for a period of $T = 34$ years when choosing a constant interest rate of $r = 0.05$. Note that an initial investment of one dollar in the savings account would have resulted over the given period in a value of about 5.5 dollars. The logarithm $X_t^0 = \ln(S_t^0)$ of the savings account is a linear increasing function, see (6.3.18), with slope equal to the short rate $r$.

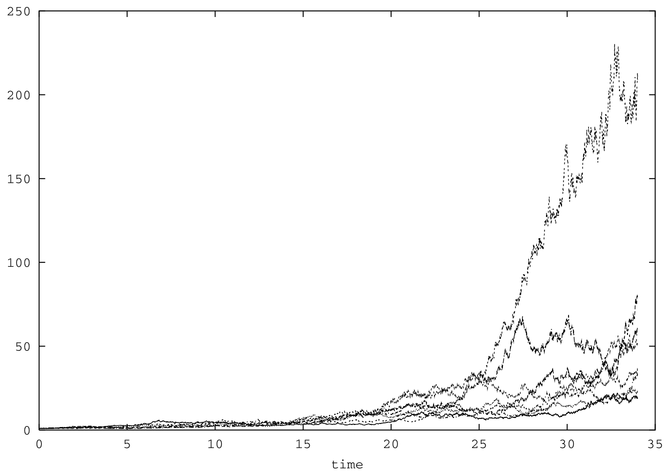



**Fig. 6.3.6.** Stock prices

Now we visualize in Fig. 6.3.6 eight cum dividend stock price processes over the period of 34 years. Here we reinvest all dividends. The $j$th stock price at time $t$ is denoted by $S_t^j$ for $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. For simplicity, we have chosen the volatility matrix to be of the form $\boldsymbol{b} = \sigma \boldsymbol{I}$, where $\sigma = 0.2$ is the volatility parameter and $\boldsymbol{I}$ the unit matrix. In this simple setting each stock evolves independently from all the others. For the simulated scenario we used the volatility parameter $\sigma = 0.2$, the short rate $r = 0.05$ and have set the growth rates to $g^j = 0.1$. As we shall see later, this is a rather poor stock market model since no correlations between the log-returns are modeling the example. Nevertheless such models have been used in practice. It is noticeable in Fig. 6.3.6 that extreme differences in stock prices over the 34 year period can occur. However, it is impossible to predict at any time which of the stocks will outperform the others in the future. They all have in our example the same appreciation rate and volatility.

One notes that the prices in Fig. 6.3.6 evolve quite differently. On average they seem to increase. We constructed these stock prices as exponentials of transformed Wiener processes $X^j = \{X_t^j,\, t \in [0, \infty)\}$, $j \in \{1, 2, \ldots, d\}$, where

$$X_t^j = g^j\, t + \sum_{k=1}^{d} b^{j,k}\, W_t^k \tag{6.3.20}$$

and the $j$th stock price is given as

$$S_t^j = \exp\{X_t^j\} \tag{6.3.21}$$

with $S_0^j > 0$.

The log-price $X_t^j$ of the $j$th stock at time $t$, $j \in \{1, 2, \ldots, d\}$, is therefore modeled by the Itô differential

$$dX_t^j = g^j\, dt + \sum_{k=1}^{d} b^{j,k}\, dW_t^k \tag{6.3.22}$$

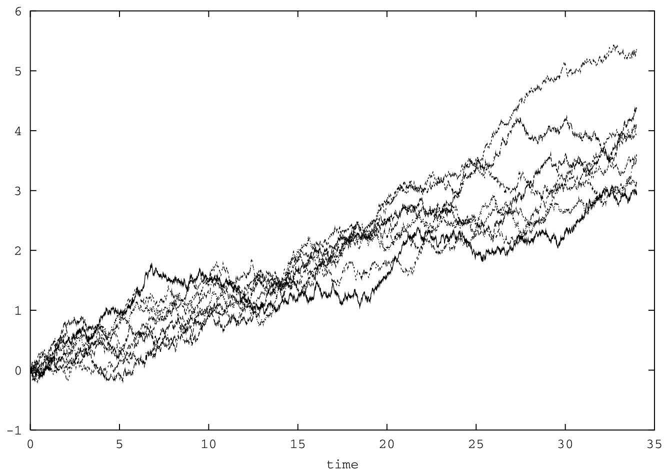



**Fig. 6.3.7.** Logarithms of stock prices

for $t \in [0, \infty)$ with initial value $X_0^j = \ln(S_0^j) \in \Re$, $j \in \{1, 2, \ldots, d\}$. The $j$th *growth rate* $g^j$ and the $j, k$th volatilities $b^{j,k}$ are deterministic constants for $j, k \in \{1, 2, \ldots, d\}$. Here we have the $j$th growth rate

$$g^j = r + p^j - \frac{1}{2} \sum_{k=1}^{d} (b^{j,k})^2. \tag{6.3.23}$$

This leads by application of the Itô formula to the function (6.3.21) for the $j$th stock price to its Itô differential or SDE

$$dS_t^j = S_t^j \left( (r + p^j)\, dt + \sum_{k=1}^{d} b^{j,k}\, dW_t^k \right) \tag{6.3.24}$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. The appreciation rate of the $j$th stock then equals the sum

$$a^j = r + p^j, \tag{6.3.25}$$

where $p^j$ is the $j$th *risk premium* or $j$th *expected excess return*. The matrix $\boldsymbol{b} = [b^{j,k}]_{j,k=1}^d$ denotes the volatility matrix. In Fig.6.3.7 we plot the logarithms $X_t^i$ of the eight stock prices over time.

## Covariation between a Wiener Process and a Functional (*)

Let $g$ denote a twice continuously differentiable function and $W$ a standard Wiener process. Then the covariation, see (5.2.16), between $g(W_t)$ and $W_t$ is given by

$$[g(W), W]_t = \int_0^t g'(W_s)\, ds \tag{6.3.26}$$

for $t \in [0, \infty)$. This can be easily derived by application of both the Itô formula (6.1.12) together with the covariation property (5.4.5) of Itô integrals, which yields

$$d(g(W_t)) = \frac{1}{2}\, g''(W_t)\, dt + g'(W_t)\, dW_t \tag{6.3.27}$$

for $t \in [0, \infty)$. One can also formulate similar statements when the standard Wiener process is substituted by more general processes.

## 6.4 Extensions of the Itô Formula

Let us mention in this section some extensions of the Itô formula that will allow us to derive powerful results for models with jumps covering stochastic processes that are needed for modeling event driven uncertainty in finance and insurance.

### Itô Formula for Jump Processes

The Itô formula can be easily generalized to the case of jump processes. Let us use again our standard notation for the *jump size*

$$\Delta Z_t = Z_t - Z_{t-} \tag{6.4.1}$$

at time $t \in [0, \infty)$ of a given process $Z = \{Z_t,\, t \in [0, \infty)\}$. Here $Z_{t-}$ denotes, as usual, the left hand limit of the process $Z$ at time $t$. Then the value $X_t$ of a pure jump process $X = \{X_t,\, t \in [0, \infty)\}$ can be written at time $t \in [0, \infty)$ as

$$X_t = \sum_{s \in [0,t]} \Delta X_s \tag{6.4.2}$$

if this sum converges almost surely for all $t \in [0, \infty)$. This then allows us to formulate the Itô formula for the given pure jump process in such a simple form that does not need any extra proof.

**Lemma 6.4.1.** *For a pure jump process $X$ and a measurable function $u : \Re \to \Re$ we have the Itô formula*

$$u(X_t) = u(X_0) + \sum_{s \in (0,t]} \Delta u(X_s) \tag{6.4.3}$$

*for $t \in [0, T]$, where $\Delta u(X_t) = u(X_t) - u(X_{t-})$.*

One notes that almost no assumptions are imposed on the function $u(\cdot)$ and the process $X$. What happens in (6.4.3) is that the jumps of $X$ are simply transferred through the function $u$ as they arise.

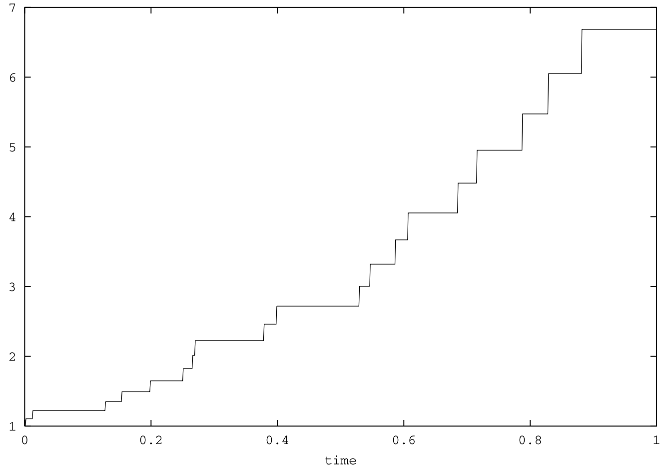



**Fig. 6.4.1.** Path of an exponential of a Poisson process

### Exponential of a Poisson Process

Let us consider an example where a pure jump process plays a role. We denote by $N = \{N_t,\, t \in [0,\infty)\}$ a Poisson process with intensity $\lambda$ as introduced in Sect. 3.5. A path of such a process for $\lambda = 20$ is shown in Fig. 3.5.1. Let us now apply the Itô formula (6.4.2) to obtain for $N$ the differential of the exponential $u(N_t) = \exp\{c\, N_t\}$ with $c > 0$. Since $N$ is a pure jump process that counts the arrival of events we have only to transform its jumps into the jumps of the exponential of $N$. Thus, at the $k$th jump time $\tau_k$ of $N$ we have the identity

$$\exp\{c\, N_{\tau_k}\} = \exp\{c\, N_{\tau_k -}\} + \exp\{c\, N_{\tau_k -}\} \left( \frac{\exp\{c\, N_{\tau_k}\}}{\exp\{c\, N_{\tau_k -}\}} - 1 \right). \tag{6.4.4}$$

In Fig. 3.5.1 we showed a trajectory of a Poisson process $N$. In Fig. 6.4.1 we plot now the corresponding exponential with $c = 0.1$.

By (6.4.4) we obtain the relationship

$$\begin{aligned} \exp\{c\, N_t\} &= \exp\{c\, N_0\} + \int_0^t \left(\exp\{c\, N_s\} - \exp\{c\, N_{s-}\}\right) dN_s \\ &= \exp\{c\, N_0\} + \int_0^t \exp\{c\, N_{s-}\} \left( \frac{\exp\{c\, N_s\}}{\exp\{c\, N_{s-}\}} - 1 \right) dN_s. \end{aligned} \tag{6.4.5}$$

Equivalently, with the notation (6.4.1) we can write the corresponding Itô differential

$$\begin{aligned} d(\exp\{c\, N_t\}) &= \Delta(\exp\{c\, N_t\}) \\ &= \exp\{c\, N_{t-}\}\, (\psi_{\exp}(t-) - 1)\, \Delta N_t \end{aligned} \tag{6.4.6}$$

for $t \in [0,\infty)$ with *jump ratio*

$$\psi_{\exp}(\tau_k-) = \frac{\exp\{c\,N_{\tau_k}\}}{\exp\{c\,N_{\tau_k-}\}} = \exp\{c\,N_{\tau_k} - c\,N_{\tau_k-}\} = \exp\{c\} = e^c \tag{6.4.7}$$

with $\tau_k$ as $k$th jump time. Note that the use of the notion of a jump ratio for the parametrization of the jump size is rather convenient.

### Itô Formula for Semimartingales (*)

After having seen that the inclusion of jumps does not create major problems for an Itô formula, the Itô formula can now be generalized to the case of semimartingales, see Definition 5.5.1. Assume that the vector process $\boldsymbol{X} = \{\boldsymbol{X}_t = (X_t^1, \dots, X_t^\ell)^\top, t \in [0, \infty)\}$ has as its $i$th component the semimartingale $X^i$ with the following decomposition

$$X_t^i = X_0^i + X_t^{i,c} + X_t^{i,d} \tag{6.4.8}$$

for $t \in [0, \infty)$, $i \in \{1, 2, \dots, \ell\}$. Here

$$X_t^{i,c} = A_t^{i,c} + M_t^{i,c} \tag{6.4.9}$$

denotes the $i$th component of the continuous part of $X_t^i$ and $X_t^{i,d}$ that of the pure jump part. This means, we have all jumps absorbed in the term

$$X_t^{i,d} = \sum_{s \in [0,t]} \Delta X_s^i \tag{6.4.10}$$

for $t \in [0, \infty)$ and $i \in \{1, 2, \dots, \ell\}$. Furthermore, $M^{i,c}$ denotes in (6.4.9) a continuous $(\underline{\mathcal{A}}, P)$-local martingale, see (5.2.26), and $A^{i,d}$ a continuous process of finite total variation, see (5.2.25).

**Theorem 6.4.2.** *For a twice continuously differentiable function $u : [0, \infty) \times \Re^\ell \to \Re$, with continuous first derivative with respect to time and second continuous derivatives with respect to the spatial variables, we have the Itô formula*

$$\begin{aligned} u(t, X_t^1, \dots, X_t^\ell) &= u(0, X_0^1, \dots, X_0^\ell) \\ &+ \int_0^t \frac{\partial}{\partial t} u(s, X_s^1, \dots, X_s^\ell)\, ds + \sum_{i=1}^\ell \int_0^t \frac{\partial}{\partial x^i} u(s, X_s^1, \dots, X_s^\ell)\, dX_s^{i,c} \\ &+ \frac{1}{2} \int_0^t \sum_{i,k=1}^\ell \frac{\partial^2}{\partial x^i\, \partial x^k} u(s, X_s^1, \dots, X_s^\ell)\, d[M^{i,c}, M^{k,c}]_s \\ &+ \sum_{s \in (0,t]} \Delta u(s, X_s^1, \dots, X_s^\ell) \end{aligned} \tag{6.4.11}$$

*for $t \in [0, \infty)$. Here the jump size $\Delta u$ of $u$ at time $s$ is defined as in (5.5.8), namely*

$$\Delta u(s, X_s^1, \dots, X_s^\ell) = u(s, X_s^1, \dots, X_s^\ell) - u(s-, X_{s-}^1, \dots, X_{s-}^\ell). \tag{6.4.12}$$

A proof of the general Itô formula (6.4.11) can be found, for instance, in Protter (2004). We remark that in (6.4.11) the jumps are simply transferred through the function $u$ whenever they occur. The Itô formula is almost identical to that for diffusions if there were no jumps. The above general Itô formula can be essential for situations where continuous and event driven uncertainty arises in a model. Similarly to (6.2.14) we can write the Itô formula (6.4.11) in the form

$$du(t, X_t^1, \ldots, X_t^\ell) = \frac{\partial}{\partial t} u(t, X_t^1, \ldots, X_t^\ell)\, dt + \sum_{i=1}^{\ell} \frac{\partial}{\partial x^i} u(t, X_t^1, \ldots, X_t^\ell)\, dX_t^{i,c}$$

$$+\frac{1}{2} \sum_{i,k=1}^{\ell} \frac{\partial^2}{\partial x^i\, \partial x^k} u(t, X_t^1, \ldots, X_t^\ell)\, d[X^{i,c}, X^{k,c}]_t + \Delta u(t, X_t^1, \ldots, X_t^\ell) \quad (6.4.13)$$

for $t \in [0, \infty)$.

**Exponential of Compensated Poisson Process (*)**

Let us continue the example concerning the exponential of a Poisson process by considering the compensated Poisson process $q = \{q_t,\, t \in [0, \infty)\}$, which is a jump martingale, where

$$dq_t = dN_t - \lambda\, dt \quad (6.4.14)$$

for $t \in [0, \infty)$. By the Itô formula (6.4.11) we obtain for $u(q_t) = \exp\{c\, q_t\}$ the stochastic differential

$$\begin{aligned} d\left(\exp\{c\, q_t\}\right) &= -\exp\{c\, q_{t-}\}\, \lambda\, dt + \exp\{c\, q_{t-}\} \left(\psi_{\exp}(t-) - 1\right)\, dN_t \\ &= \exp\{c\, q_{t-}\}\, \lambda\, \left(\psi_{\exp}(t-) - 2\right)\, dt \\ &\quad + \exp\{c\, q_{t-}\} \left(\psi_{\exp}(t-) - 1\right)\, dq_t \\ &= \exp\{c\, q_{t-}\}\, \lambda\, (e^c - 2)\, dt + \exp\{c\, q_{t-}\}\, (e^c - 1)\, dq_t \end{aligned} \quad (6.4.15)$$

for $t \in [0, \infty)$. Here the jump ratio $\psi_{\exp}(t-) = e^c$ remains as in the case of the exponential of a Poisson process. Note that the last part of the sum on the right hand side of (6.4.15) is a martingale differential.

**Exponential for Wiener Process with Jumps (*)**

To provide another example for the above Itô formula (6.4.11) let us add to the dynamics of a Poisson process a Wiener process $W = \{W_t,\, t \in [0, \infty)\}$ and a trend. We consider now the exponential of the process $X = \{X_t,\, t \in [0, \infty)\}$ with Itô differential

$$dX_t = g\, dt + \sigma\, dW_t + c\, (dN_t - \lambda\, dt) \quad (6.4.16)$$

for $t \in [0, \infty)$. Here we use as additional parameters the growth rate $g$, the intensity $\lambda$ and the volatility $\sigma$. The Itô formula (6.4.11) yields for the exponential

$$u(X_t) = \exp\{X_t\}$$

the stochastic differential

$$\begin{aligned} d\left(\exp\{X_t\}\right) &= \exp\{X_t\} \left( g + \frac{1}{2}\,\sigma^2 + \lambda\left(\psi_{\exp}(t) - 2\right)\right) dt + \exp\{X_t\}\,\sigma\,dW_t \\ &\quad + \exp\{X_{t-}\}\left(\psi_{\exp}(t-) - 1\right) dq_t \\ &= \exp\{X_{t-}\}\left(\left(g + \frac{1}{2}\,\sigma^2 + \lambda\left(e^c - 2\right)\right) dt + \sigma\,dW_t + \left(e^c - 1\right) dq_t\right) \end{aligned} \tag{6.4.17}$$

for $t \in [0, \infty)$. We observe that besides the jump terms all other terms are as in the earlier versions of the Itô differential for geometric Brownian motion, see (6.1.17). Therefore, we could call the above exponential a geometric Brownian motion with jumps. The jumps of $X_t$ are transformed by the exponential function, analogous as described already by the identity (6.4.4).

This example indicates that the Itô formula is a powerful tool that allows us to determine the stochastic differential of a function of a given stochastic differential even when jumps are present. We emphasize that the jumps are directly transferred through the given function, which makes the jump part in (6.4.17) very simple to interpret. The above jump diffusion dynamics in (6.4.16) is a special case of the Merton model, see Merton (1976), which we shall study later.

### Itô Formula for Poisson Jump Measure (*)

A particular case of the Itô formula (6.4.11) is obtained when only Wiener processes and Poisson jump measures are involved. Let us assume that $\boldsymbol{W} = \{\boldsymbol{W}_t = (W_t^1, \ldots, W_t^m)^\top, t \in [0, \infty)\}$ is an $m$-dimensional standard Wiener process and $p_{\varphi_r}^r(dv, dt)$ denotes a Poisson measure on $\mathcal{E} \times [0, \infty)$ with intensity measure

$$\nu_{\varphi_r}^r(dv, dt) = \varphi_r(dv)\,dt, \tag{6.4.18}$$

$r \in \{m+1, m+2, \ldots, \bar{\ell}\}$, as introduced in Sect. 3.5 and used in Sect. 5.5. Suppose that the $i$th component $X_t^i$ at time $t$ of the process $\boldsymbol{X}$ has the representation

$$X_t^i = X_0^i + \int_0^t a_s^i\,ds + \sum_{k=1}^m \int_0^t b_s^{i,k}\,dW_s^k + \sum_{r=m+1}^{\bar{\ell}} \int_0^t \int_{\mathcal{E}} c^{i,r}(v, s-)\,p_{\varphi_r}^r(dv, ds) \tag{6.4.19}$$

for $t \in [0, \infty)$ and $i \in \{1, 2, \ldots, \ell\}$, where $a^i$, $b^{i,j}$ and $c^{i,r}$ are appropriately chosen adapted processes and the mark space is given as $\mathcal{E} = \Re\backslash\{0\}$. Then the following version of the Itô formula follows from (6.4.11).

**Corollary 6.4.3.** *For a function $u : [0,\infty) \times \Re^\ell \to \Re$, which is assumed to be differentiable with respect to $t$ and twice differentiable with respect to $x$, for the above process $\boldsymbol{X}$ the Itô formula has the form*

$$
\begin{aligned}
u(t, X_t^1, \ldots, X_t^\ell) = u(0, X_0^1, \ldots, X_0^\ell) + \int_0^t \Bigg( & \frac{\partial u(s, X_s^1, \ldots, X_s^\ell)}{\partial t} \\
& + \sum_{i=1}^{\ell} a_s^i \, \frac{\partial}{\partial x^i} \, u(s, X_s^1, \ldots, X_s^\ell) \\
& + \frac{1}{2} \sum_{i,j=1}^{\ell} \sum_{k=1}^{m} b_s^{i,k} \, b_s^{j,k} \, \frac{\partial^2 u(s, X_s^1, \ldots, X_s^\ell)}{\partial x^i \, \partial x^j} \Bigg) \, ds \\
& + \sum_{k=1}^{m} \sum_{i=1}^{\ell} \int_0^t b_s^{i,k} \, \frac{\partial u(s, X_s^1, \ldots, X_s^\ell)}{\partial x^i} \, dW_s^k \\
& + \sum_{r=m+1}^{\bar{\ell}} \int_0^t \int_{\mathcal{E}} \big( u(s, X_s^1, \ldots, X_s^\ell) \\
& \qquad - u(s, X_{s-}^1, \ldots, X_{s-}^\ell) \big) \, p_{\varphi_r}^r(dv, ds)
\end{aligned}
\tag{6.4.20}
$$

*for* $t \in [0,\infty)$.

By using (6.4.20) it is straightforward to handle problems which include Lévy processes as underlying factors.

## 6.5 Lévy's Theorem (*)

### Identification of Martingales as Wiener Processes (*)

The Wiener process is a basic building block in financial modeling and plays a central role in stochastic calculus. A definition of the Wiener process is given via the properties (3.2.6). By (5.1.5) we saw that the Wiener process is a martingale and from (5.2.5) it followed that its quadratic variation equals time $t$. Note that the converse of this result can be shown, namely that a continuous martingale with a quadratic variation that equals time, is a Wiener process. *Lévy's Theorem* provides this important result, which we formulate below for multi-dimensional continuous martingales. Its derivation relies on an application of the multivariate Itô formula.

**Theorem 6.5.1.** (Lévy) *For $m \in \mathcal{N}$ let $A$ be a given $m$-dimensional vector process $\boldsymbol{A} = \{\boldsymbol{A}_t = (A_t^1, A_t^2, \ldots, A_t^m)^\top, \, t \in [0,\infty)\}$ on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. If each of the processes $A^i = \{A_t^i, \, t \in [0,\infty)\}$ is a* *continuous, square integrable* $(\underline{\mathcal{A}}, P)$*-martingale that starts at* 0 *at time* $t = 0$ *and their covariations are of the form*

$$[A^i, A^k]_t = \begin{cases} t & for \quad i = k \\ 0 & for \quad i \neq k \end{cases} \tag{6.5.1}$$

*for* $i, k \in \{1, 2, \ldots, m\}$, $t \in [0, \infty)$, *then the vector process* $\boldsymbol{A}$ *is an* $m$*-dimensional standard Wiener process on* $[0, \infty)$. *This means that each process* $A^i$ *is a one-dimensional Wiener process that is independent of the other Wiener processes* $A^k$ *for* $k \neq i$.

In particular, one can show that this result implies that a continuous process $X = \{X_t,\, t \in [0, \infty)\}$ is a one-dimensional Wiener process if and only if both the process $X$ and the process $Y = \{Y_t = X_t^2 - t,\, t \in [0, \infty)\}$ are martingales. Furthermore, if one is able to construct for an observed vector process a transformation such that the transformed processes are square integrable continuous martingales with covariations of the form (6.5.1), then one has found the basic building blocks of the given dynamics in the form of a vector of independent Wiener processes. In this case one needs then only to take the inverse of that transformation to arrive at a realistic model. It is a challenge in financial modeling to construct a parsimonious market model with the above property.

### Proof of Lévy's Theorem (*)

To indicate the proof of the above theorem we consider the characteristic function

$$\phi_{\boldsymbol{A}_t - \boldsymbol{A}_s}(\boldsymbol{\theta}) = E\left( \exp\left\{ \imath \sum_{k=1}^{m} \theta^k \left(A_t^k - A_s^k\right) \right\} \Bigg| \, \mathcal{A}_s \right) \tag{6.5.2}$$

for $\boldsymbol{\theta} \in \Re^m$, $t \in [0, \infty)$ and $s \in [0, t]$, with $\imath$ denoting the imaginary unit, see (1.3.77).

By application of a complex valued version of the Itô formula (6.4.11) for semimartingales we obtain

$$\exp\left\{ \imath \sum_{k=1}^{m} \theta^k \, A_t^k \right\} - \exp\left\{ \imath \sum_{k=1}^{m} \theta^k \, A_s^k \right\} = \sum_{k=1}^{m} \int_s^t \imath \, \theta^k \, \exp\left\{ \imath \sum_{l=1}^{m} \theta^l \, A_u^l \right\} dA_u^k$$
$$+ \frac{1}{2} \sum_{k=1}^{m} \int_s^t \left(-(\theta^k)^2\right) \exp\left\{ \imath \sum_{l=1}^{m} \theta^l \, A_u^l \right\} du. \tag{6.5.3}$$

We have introduced the Itô integral with respect to general integrators in (5.3.11). The martingale property for Itô integrals, which follows for integrators that are Wiener processes and integrands that are from $\mathcal{L}_T^2$, when considered on $[0, T]$ with $T \in (0, \infty)$, can be naturally extended to cover the wider class of square integrable martingale integrators with integrands that appear in (6.5.3), see Protter (2004). This means that the terms in the first sum on the right hand side of (6.5.3) are martingales and we have

$$E\left(\int_s^t \exp\left\{\imath \sum_{l=1}^m \theta^l A_u^l\right\} dA_u^k \,\middle|\, \mathcal{A}_s\right) = 0. \tag{6.5.4}$$

Let us now choose any event $\mathcal{F} \in \mathcal{A}_s$ and denote by $\mathbf{1}_{\mathcal{F}}$ the indicator function that equals one if $\mathcal{F}$ occurs. Then multiplying both sides of (6.5.3) by

$$\mathbf{1}_{\mathcal{F}} \exp\left\{-\imath \sum_{k=1}^m \theta^k A_s^k\right\}$$

and taking expectations yields

$$G(t) - P(\mathcal{F}) = -\frac{1}{2} \sum_{k=1}^m (\theta^k)^2 \int_s^t G(u)\, du,$$

where

$$G(u) = E\left(\exp\left\{\imath \sum_{k=1}^m \theta^k \left(A_u^k - A_s^k\right)\right\} \mathbf{1}_{\mathcal{F}}\right)$$

for $u \in [0, t]$. The solution to this ordinary integral equation is given by

$$G(t) = P(\mathcal{F}) \exp\left\{-\frac{1}{2} \sum_{k=1}^m (\theta^k)^2 (t-s)\right\}.$$

Consequently, by the Bayes's formula for conditional means, see (1.1.13) or Karatzas & Shreve (1991), we obtain

$$E\left(\exp\left\{\imath \sum_{k=1}^m \theta^k \left(A_t^k - A_s^k\right)\right\} \,\middle|\, \mathcal{F}\right) = \frac{G(t)}{P(\mathcal{F})} = \exp\left\{-\frac{1}{2} \sum_{k=1}^m (\theta^k)^2 (t-s)\right\}.$$

Clearly, this result holds for any $\mathcal{F} \in \mathcal{A}_s$. Therefore, we have shown that for all $\boldsymbol{\theta} \in \Re^m$, $t \in [0, \infty)$ and $s \in [0, t]$ the characteristic function of the vector increment $\boldsymbol{A}_t - \boldsymbol{A}_s$ is of the form

$$\phi_{\boldsymbol{A}_t - \boldsymbol{A}_s}(\boldsymbol{\theta}) = \exp\left\{-\frac{1}{2} \sum_{k=1}^m (\theta^k)^2 (t-s)\right\}. \tag{6.5.5}$$

It is known, see (1.4.58), that this is the characteristic function of a vector of independent Gaussian distributed random variables, each with mean zero and variance $(t - s)$. Since the characteristic function of a random vector identifies uniquely the joint distribution of this random vector, we see by the Definition 6.2.1 that the process $A$ is an $m$-dimensional standard Wiener process. $\square$

## 6.6 A Proof of the Itô Formula (*)

Since the Itô formula is extremely important in quantitative finance we highlight in the following the main steps of a classical proof of this fundamental tool. For simplicity, we consider the scalar, continuous process $X = \{X_t,\, t \in [0,\infty)\}$, given in (6.1.6), that is

$$X_t = X_0 + \int_0^t e_s\, ds + \int_0^t f_s\, dW_s \tag{6.6.1}$$

for $t \in [0,\infty)$ with initial value $X_0 = x_0$, standard Wiener process $W = \{W_t,\, t \in [0,\infty)\}$ and predictable processes $e$ and $f$, where the second integral is an Itô integral. The proof of the multi-dimensional Itô formula stated in (6.2.11) is a straightforward generalization of what will be given below.

**Theorem 6.6.1.** *If we assume that $u : [0,T] \times \Re \to \Re$ is a function of time $t \in [0,T]$ and state variable $x \in \Re$ such that the partial derivatives $\frac{\partial u(t,x)}{\partial t}$, $\frac{\partial u(t,x)}{\partial x}$ and $\frac{\partial^2 u(t,x)}{\partial x^2}$ exist and are continuous for all $(t,x) \in [0,T] \times \Re$ and $\sqrt{|e|},\, f \in \mathcal{L}_T^2$, see (5.4.1), then the* Itô formula *can be written in the form*

$$du(t,X_t) = \left(\frac{\partial u(t,X_t)}{\partial t} + e_t\, \frac{\partial u(t,X_t)}{\partial x} + \frac{1}{2}\,(f_t)^2\, \frac{\partial^2 u(t,X_t)}{\partial x^2}\right) dt$$
$$+ f_t\, \frac{\partial u(t,X_t)}{\partial x}\, dW_t \tag{6.6.2}$$

*for $t \in [0,T]$.*

### A Lemma (*)

Before we begin with the proof of the Itô formula given in Theorem 6.6.1 let us summarize some application of the Taylor series expansion and the Mean Value Theorem of classical calculus in a simple lemma.

**Lemma 6.6.2.** *Let the function $u : [0,T] \times \Re \to \Re$ be as in Theorem 6.6.1. Then for any $t$, $t + \Delta t \in [0,T]$ and $x$, $x + \Delta x \in \Re$ there exist constants $\alpha, \beta \in [0,1]$ such that*

$$u(t+\Delta t, x+\Delta x) - u(t,x) = \frac{\partial u(t+\alpha\Delta t, x)}{\partial t}\,\Delta t + \frac{\partial u(t,x)}{\partial x}\,\Delta x$$
$$+ \frac{1}{2}\,\frac{\partial^2 u(t, x+\beta\Delta x)}{\partial x^2}\,(\Delta x)^2.$$

**Proof of Theorem 6.6.1 (*)**

**1.** First assume that $e$ and $f$ are deterministic constants, that is, they do not depend on $t$. We choose a continuous sample-path of $X$ and fix a subinterval $[s,t] \subseteq [0,T]$, for which we consider partitions of the form $s = t_1^{(n)} < t_2^{(n)} < \ldots < t_{n+1}^{(n)} = t$ with $\Delta t_j^{(n)} = t_{j+1}^{(n)} - t_j^{(n)}$ and $\delta^{(n)} = \max_{1\le j\le n} \Delta t_j^{(n)}$, where $\lim_{n\to\infty} \delta^{(n)} \overset{\text{a.s.}}{=} 0$. Then

$$u(t, X_t) - u(s, X_s) = \sum_{j=1}^{n} \Delta u_j^{(n)},$$

where

$$\Delta u_j^{(n)} = u\left(t_{j+1}^{(n)}, X_{t_{j+1}^n}\right) - u\left(t_j^{(n)}, X_{t_j^n}\right)$$

for $j \in \{1,2,\ldots,n\}$. Applying Lemma 6.6.2 on each subinterval $[t_j^{(n)}, t_{j+1}^{(n)}]$ for each $\omega \in \Omega$, we have $\alpha_j^{(n)}$, $\beta_j^{(n)} \in [0,1]$ such that

$$\begin{aligned}\Delta u_j^{(n)} &= \frac{\partial u}{\partial t}\left(t_j^{(n)} + \alpha_j^{(n)}\Delta t_j^{(n)}, X_{t_j^n}\right)\Delta t_j^{(n)} \\ &\quad + \frac{\partial u}{\partial x}\left(t_j^{(n)}, X_{t_j^n}\right)\Delta X_j^{(n)} \\ &\quad + \frac{1}{2}\frac{\partial^2 u}{\partial x^2}\left(t_j^{(n)}, X_{t_j^n} + \beta_j^{(n)}\Delta X_j^{(n)}\right)\left(\Delta X_j^{(n)}\right)^2, \qquad (6.6.3)\end{aligned}$$

almost surely, where $\Delta X_j^{(n)} = X_{t_{j+1}^n} - X_{t_j^n}$ for $j \in \{1,2,\ldots,n\}$. By the continuity of $\frac{\partial u}{\partial t}$ and $\frac{\partial^2 u}{\partial x^2}$, and the sample-path continuity of $X$, we have for each $j \in \{1,2,\ldots,n\}$

$$\lim_{n\to\infty} \frac{\partial u}{\partial t}\left(t_j^{(n)} + \alpha_j^{(n)}\Delta t_j^{(n)}, X_{t_j^n}\right) - \frac{\partial u}{\partial t}\left(t_j^{(n)}, X_{t_j^n}\right) \overset{\text{a.s.}}{=} 0, \qquad (6.6.4)$$

and

$$\lim_{n\to\infty} \frac{\partial^2 u}{\partial x^2}\left(t_j^{(n)}, X_{t_j^n} + \beta_j^{(n)}\Delta X_j^{(n)}\right) - \frac{\partial^2 u}{\partial x^2}\left(t_j^{(n)}, X_{t_j^n}\right) \overset{\text{a.s.}}{=} 0. \qquad (6.6.5)$$

Since $e$ and $f$ are independent of $t$, the increments of $X$ are of the form

$$\Delta X_j^{(n)} = e\Delta t_j^{(n)} + f\Delta W_j^{(n)},$$

where $\Delta W_j^{(n)} = W_{t_{j+1}^n} - W_{t_j^n}$ for $j \in \{1,2,\ldots,n\}$. Consequently, it can be shown that the sum

$$\sum_{j=1}^{n}\left\{\left(\Delta X_j^{(n)}\right)^2 - \left(f\Delta W_j^{(n)}\right)^2\right\} = e^2\sum_{j=1}^{n}\left(\Delta t_j^{(n)}\right)^2 + 2ef\sum_{j=1}^{n}\Delta W_j^{(n)}\Delta t_j^{(n)} \qquad (6.6.6)$$

tends to 0 in probability for $\delta^{(n)} \to 0$ as $n \to \infty$. By combining relations (6.6.3)–(6.6.6) we see that under convergence in probability

$$u(t, X_t) - u(s, X_s) \overset{P}{=} \lim_{n\to\infty} \sum_{j=1}^{n} \Delta u_j^{(n)}$$

$$\begin{aligned}
&\overset{P}{=} \lim_{n\to\infty} \sum_{j=1}^{n} \left\{ \frac{\partial u}{\partial t} \left(t_j^{(n)}, X_{t_j^n}\right) + e\, \frac{\partial u}{\partial x} \left(t_j^{(n)}, X_{t_j^n}\right) \right. \\
&\qquad\qquad \left. + \frac{1}{2}\, f^2 \frac{\partial^2 u}{\partial x^2} \left(t_j^{(n)}, X_{t_j^n}\right) \right\} \Delta t_j^{(n)} \\
&\quad + \lim_{n\to\infty} \sum_{j=1}^{n} f\, \frac{\partial u}{\partial x} \left(t_j^{(n)}, X_{t_j^n}\right) \Delta W_j^{(n)} \\
&\quad + \lim_{n\to\infty} \sum_{j=1}^{n} \frac{1}{2} f^2 \frac{\partial^2 u}{\partial x^2} \left(t_j^{(n)}, X_{t_j^n}\right) \left( \left(\Delta W_j^{(n)}\right)^2 - \Delta t_j^{(n)} \right). \qquad (6.6.7)
\end{aligned}$$

The first two terms on the right hand side of (6.6.7) are the terms on the right hand side of (6.6.2). We shall show that the last term in (6.6.7) converges to zero in probability for $n \to \infty$. Let us write $\Gamma_j^{(n)} = (\Delta W_j^{(n)})^2 - \Delta t_j^{(n)}$ with $\mathbf{1}_{n,j}^{(N)}$ denoting the indicator function of the set

$$A_{n,j}^{(N)} = \left\{ \omega \in \Omega : \left| X_{t_i^n} \right| \le N \quad \text{for} \quad i \in \{1, 2, \ldots, j\} \right\}$$

for $j \in \{1, 2, \ldots, n\}$. For fixed $n$ the random variables $\Gamma_j^{(n)}$ are independent with mean $E\left(\Gamma_j^{(n)}\right) = 0$ and variance $E\left(\left(\Gamma_j^{(n)}\right)^2\right) = 2\left(\Delta t_j^{(n)}\right)^2$ for $j \in \{1, 2, \ldots, n\}$. Using this result we obtain the estimate

$$\begin{aligned}
&\lim_{n\to\infty} E\left( \left| \sum_{j=1}^{n} \frac{\partial^2 u}{\partial x^2} \left(t_j^{(n)}, X_{t_j^n}\right) \mathbf{1}_{n,j}^{(N)}\, \Gamma_j^{(n)} \right|^2 \right) \\
&\qquad \overset{P}{=} \lim_{n\to\infty} \sum_{j=1}^{n} E\left( \left| \frac{\partial^2 u}{\partial x^2} \left(t_j^{(n)}, X_{t_j^n}\right) \mathbf{1}_{n,j}^{(N)}\, \Gamma_j^{(n)} \right|^2 \right) \\
&\qquad \overset{P}{\le} \lim_{n\to\infty} C_N \sum_{j=1}^{n} 2 \left(\Delta t_j^{(n)}\right)^2 \\
&\qquad \overset{P}{\le} \lim_{n\to\infty} 2\, C_N |t-s|\, \delta^{(n)} \overset{P}{=} 0.
\end{aligned}$$

Here we have used the upper bound

$$C_N = \max_{\substack{s \le z \le t \\ |x| \le N}} \left| \frac{\partial^2 u}{\partial x^2}(z, x) \right|^2 < \infty.$$

As mentioned in Sect. 2.1, for an event $D$ its *complement* denoted by $D^c$ is given by $D^c = \{\omega \in \Omega : \omega \notin D\}$. Since

$$\bigcup_{j=1}^{n} \left( A_{n,j}^{(N)} \right)^c \subseteq B^{(N)} = \left\{ \omega \in \Omega : \sup_{s \le z \le t} |X_z| > N \right\},$$

so that $\lim_{N \to \infty} P\left(B^{(N)}\right) = 0$, then $\lim_{N \to \infty} P(A_{n,j}^{(N)}) = 1$. Combining these two results it can be shown that the last term in (6.6.7) converges to zero in probability $n \to \infty$. For $e$ and $f$, which do not depend on $t$, the proof is thus complete. We can show that a similar result holds for random step functions $e$ and $f$ since these remain constant within partition subintervals, when conditioned on the sigma-algebra of the last discretization point.

**2.** For general $e$ and $f$ with $\sqrt{|e|}$, $f \in \mathcal{L}_T^2$ we can construct sequences of step functions $(\sqrt{|e^{(n)}|})$, $(f^{(n)})$ in $\mathcal{L}_T^2$ such that the integrals

$$\lim_{n \to \infty} \int_s^t \left| e_z^{(n)} - e_z \right| dz \stackrel{P}{=} 0$$

and

$$\lim_{n \to \infty} \int_s^t \left| f_z^{(n)} - f_z \right|^2 dz \stackrel{P}{=} 0$$

converge in probability to zero. This is because $p$-mean convergence for $p = 1$ implies convergence in probability, see (2.7.6). Then we can show that the sequence defined by

$$X_r^{(n)} = X_s + \int_s^r e_z^{(n)} \, dz + \int_s^r f_z^{(n)} \, dW_z$$

converges in probability to $X_r$ as $n \to \infty$ for each $r \in [0, t]$ and $s \in [0, r]$, that is $\lim_{n \to \infty} X_r^{(n)} \stackrel{P}{=} X_r$. Since the Itô formula has been shown for step functions, then

$$\begin{aligned} u\left(t, X_t^{(n)}\right) - u\left(s, X_s^{(n)}\right) = \int_s^t \bigg( & \frac{\partial u}{\partial t}\left(z, X_z^{(n)}\right) + e_z^{(n)} \frac{\partial u}{\partial x}\left(z, X_z^{(n)}\right) \\ & + \frac{1}{2} \left(f_z^{(n)}\right)^2 \frac{\partial^2 u}{\partial x^2}\left(z, X_z^{(n)}\right) \bigg) dz \\ & + \int_s^t f_z^{(n)} \frac{\partial u}{\partial x}(z, X_z^{(n)}) \, dW_z, \end{aligned} \tag{6.6.8}$$

almost surely for each $n$. Now, from the convergence of $X_z^{(n)}$ to $X_z$ in probability as $n \to \infty$ for $z \in [s, t]$ it follows convergence in probability for the left hand side of (6.6.8), that is

$$\lim_{n\to\infty} (u(t, X_t^{(n)}) - u(s, X_s^{(n)})) \overset{P}{=} u(t, X_t) - u(s, X_s). \tag{6.6.9}$$

Using similar arguments as given in the first part of this proof it can be shown that under convergence in probability

$$\lim_{n\to\infty} \int_s^t \left( \frac{\partial u}{\partial t}\left(z, X_z^{(n)}\right) + e_z^{(n)} \frac{\partial u}{\partial x}\left(z, X_z^{(n)}\right) + \frac{1}{2}\left(f_z^{(n)}\right)^2 \frac{\partial^2 u}{\partial x^2}\left(z, X_z^{(n)}\right)\right) dz$$

$$\overset{P}{=} \int_s^t \left( \frac{\partial u}{\partial t}(z, X_z) + e_z \frac{\partial u}{\partial x}(z, X_z) + \frac{1}{2}(f_z)^2 \frac{\partial^2 u}{\partial x^2}(z, X_z)\right) dz \tag{6.6.10}$$

and

$$\lim_{n\to\infty} \int_s^t f_z^{(n)} \frac{\partial u}{\partial x}\left(z, X_z^{(n)}\right) dW_z \overset{P}{=} \int_s^t f_z \frac{\partial u}{\partial x}(z, X_z)\, dW_z. \tag{6.6.11}$$

As explained at the end of Sect. 2.1, by taking subsequences the above convergences in probability can be considered to hold a.s. Thus, we see by passing to the limit on both sides of equation (6.6.8) for $n \to \infty$ it follows that equation (6.6.2) holds a.s. The processes on the two sides of equation (6.6.2) are continuous and, thus, indistinguishable, see (3.1.6). Note that the integrals appearing on the right hand side of (5.4.1) are well defined as limits in probability. This means that these integrals can be interpreted in a wider sense, namely as limits in probability, when either $\sqrt{\left|\frac{\partial u(\cdot, X(\cdot))}{\partial t}\right|}$, $\sqrt{\left|e(\cdot)\, \frac{\partial u(\cdot, X(\cdot))}{\partial x}\right|}$, $\sqrt{f^2(\cdot)\, \left|\frac{\partial^2 u(\cdot, X(\cdot))}{\partial x^2}\right|}$ or $\left|f(\cdot)\, \frac{\partial u(\cdot, X(\cdot))}{\partial x}\right|$ are not elements of the space $\mathcal{L}_T^2$. In cases where the integrals in (5.4.1) exist in the mean square sense, as described in Sect. 5.3, these limits coincide almost surely with the limits in probability. □

## 6.7 Exercises for Chapter 6

**6.1.** Derive the Itô differential for $(Y_t)^2$ if $Y = \{Y_t = a\,t + b\,W_t,\ t \in [0, \infty)\}$ denotes a transformed Wiener process, where $W$ is a standard Wiener process.

**6.2.** Determine for a geometric Brownian motion $Z_t = Z_0 \exp\{\mu\, t + \sigma\, W_t\}$ the Itô differential for $Z_t$ and $\ln(Z_t)$ by the use of the Itô formula, where $W$ is a standard Wiener process.

**6.3.** What is the Itô differential for the square $(Z_t)^2$ of the geometric Brownian motion in Exercise 6.2?

**6.4.** Derive the Itô differential for the inverse $(Z_t)^{-1}$ of the geometric Brownian motion in Exercise 6.2.

**6.5.** Compute the Itô differential of the product $Y_t\, Z_t$ of the transformed Wiener process $Y$ in Exercise 6.1 and the geometric Brownian motion $Z_t$ in Exercise 6.2.

**6.6.** Consider two transformed Wiener processes with $Y_t^1 = a_1\, t + b_1\, W_t^1$ and $Y_t^2 = a_2\, t + b_2\, W_t^2$, where $W^1$ and $W^2$ are two independent standard Wiener processes. What is the Itô differential for $Y_t^1\, Y_t^2$?

**6.7.** Assume the same transformed Wiener processes as in Exercise 6.6 and compute the Itô differential for the expression $\exp\{Y_t^1\} \exp\{Y_t^2\}$.

**6.8.** Calculate the covariation between a standard Wiener process and its square.

**6.9.** (*) Assume $\xi : [0, \infty) \to \Re$ is a given deterministic function of time and that $X$ is given by an Itô integral, such that

$$X_t = \int_0^t \xi(s)\, dW_s$$

for $t \in [0, \infty)$, where $W$ is a standard Wiener process. Show that $Y = \{Y_t = X_t^2 - [X]_t,\, t \in [0, \infty)\}$ is a martingale.

**6.10.** (*) For a process $X = \{X_t,\, t \in [0, \infty)\}$ with $X_t = \sigma\, W_t + \xi\, N_t$, where $W$ is a standard Wiener process and $N$ a Poisson process with intensity $\lambda > 0$, characterize the stochastic differential of its exponential when $\sigma, \xi > 0$.

**6.11.** (*) For the sum $X_t = a\, N_t^1 + b\, N_t^2$, where $N^1$ and $N^2$ are two independent Poisson processes with intensity $\lambda > 0$, compute the stochastic differential of the exponential.

# 7

# Stochastic Differential Equations

Stochastic differential equations provide a powerful mathematical framework for the continuous time modeling of asset prices and general financial markets. We consider both scalar and vector stochastic differential equations which allow us to model feedback effects in the market. Explicit solutions will be given in certain cases. Furthermore, questions related to the existence and uniqueness of solutions will be discussed. We also mention stochastic differential equations with jumps which allow us to model event driven uncertainty.

## 7.1 Solution of a Stochastic Differential Equation

### Feedback in Asset Price Dynamics

The modeling of changes in financial quantities which depend on their actual values can be achieved by using *stochastic differential equations* (SDEs), see for instance (6.1.17) and (6.3.7). This allows us to discuss situations where the evolution of the asset price depends on some feedback, which arises in any realistic market dynamics. The solution of an SDE coincides with a corresponding stochastic integral equation or Itô differential with integrands that may be functions of the financial quantity itself. The specification of these functions allows the modeling of feedback effects.

As an equation, an SDE contains an unknown, which is its solution process. Therefore, the notion of a solution of an SDE is more complex than that of an Itô differential. It needs additional mathematical clarification. In any case, a solution of an SDE is a stochastic process. To be useful in practical modeling, such a solution needs to *exist* in an appropriate mathematical sense. Furthermore, the *uniqueness* of the solution for an SDE has to be examined to make sure that one achieves the targeted modeling goal.

It is highly efficient, and also rather elegant, to model financial quantities via SDEs rather than discrete time stochastic processes. This does not diminish the important contribution and lasting impact of time series type models in

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

finance, as developed in Engle (1982) and Bollerslev (1986). However, time is evolving continuously. For different observation time step sizes, for instance, daily or hourly, discrete time models provide often rather different calibration results and thus possibly inconsistent answers. A financial market model should be robust when observed for small time step sizes. The framework of SDEs yields such robustness. It is also more compact than the discrete time series approach. In particular, it permits general transformations and other manipulations in continuous time via stochastic calculus without the need to deal with resulting error terms that arise from time discretizations. This has been pointed out by Merton in his important work on continuous time finance summarized in Merton (1992). Continuous dynamics can be efficiently modeled by the drift and diffusion coefficient functions that determine an SDE. Later we shall point out that under the benchmark approach one often needs only to specify the diffusion coefficients in a model. The drift coefficients are automatically determined by the general nature of the market dynamics.

### A Discrete Time Approximation of an SDE

First, as an introduction, let us consider an asset price process $X_h = \{X_h(t), t \in [0,\infty)\}$, which is recursively defined for discrete time points $t_k = k\,h$, $k \in \{0,1,\ldots\}$, with time step size $h > 0$, by the relation

$$\Delta X_h(t_k) = X_h(t_{k+1}) - X_h(t_k) = a\,X_h(t_k)\,h + \sigma\,X_h(t_k)\,(W_{t_{k+1}} - W_{t_k}) \quad (7.1.1)$$

for $k \in \{0,1,\ldots\}$ with initial value $X_h(0) = x_0$. For simplicity, between discretization points one may interpret the solution as being piecewise constant. Here the parameter $a$ is the *appreciation rate* and $\sigma$ is the *volatility* parameter, see Sect. 6.3. The process $W = \{W_t, t \in [0,\infty)\}$ denotes a standard Wiener process.

Inspection of the stochastic difference equation (7.1.1) reveals that the increment $\Delta X_h(t_k)$ is conditionally Gaussian distributed with mean $a\,X_h(t_k)\,h$ and variance $\sigma^2\,(X_h(t_k))^2\,h$. Depending on what the actual value $X_h(t_k)$ of the asset price is at time $t_k$, its increment has a mean that is proportional to this value. Similarly, its deviation $\sigma\,X_h(t_k)\,\sqrt{h}$ is also proportional to its current value. Consequently, current asset price values influence the increments of the asset price. In this manner they are producing a feedback effect. Note that for a given Wiener path a specific trajectory for $X_h$ is obtained recursively for all discretization points $t_k,\ k \in \{0,1,\ldots\}$, by the corresponding solution of the stochastic difference equation (7.1.1).

Stochastic difference equations of the type (7.1.1) can be used for the simulation of asset prices. Many of the figures that we provide within this book are simply generated by an *Euler-Maruyama scheme*, which is a generalization of (7.1.1). For detailed information on the numerical solution of SDEs we refer to Kloeden & Platen (1999) and our introduction to numerical methods in quantitative finance in Chap. 15.

### Limit of a Discrete Time Approximation

In the above example, an important question is: what would be obtained if the time step size $h$ were to converge to zero? It can be shown, see Kloeden & Platen (1999), that $X_h$ converges, for instance, in mean square sense, see (2.7.3), to the process $X = \{X_t,\, t \in [0, \infty)\}$ that satisfies the Itô differential

$$dX_t = a\, X_t\, dt + \sigma\, X_t\, dW_t \tag{7.1.2}$$

for $t \in [0, \infty)$ and $X_0 = x_0$. By comparing (7.1.1) with (7.1.2) this is also what one would naturally expect. As indicated earlier, since there is some feedback captured in the Itô differential (7.1.2) it will be called a *stochastic differential equation* (SDE).

In some sense the SDE (7.1.2) can be interpreted as the asymptotic characterization of the discretely observed stochastic difference equation (7.1.1). Most importantly, for small time step size $h$, it does not significantly depend on the grid parameter $h$.

This is different to the time series approach, which is, see Engle & Bollerslev (1986), usually dependent on the chosen time step size. In some cases it is possible to identify limiting SDEs for popular time series models. By appropriately normalizing the parameters, Nelson (1990) has identified such limiting SDEs for some time series involved models. We shall study such dynamics in Sect. 12.4.

Since we want to take full advantage of the power and elegance of stochastic calculus we shall build financial market models directly in continuous time by using SDEs. This then avoids any discretization effects in the model description and gives full access to manipulations via stochastic calculus when transforming financial quantities.

Let us show that (7.1.2) is the SDE that has geometric Brownian motion as its explicit solution, see (4.1.2). For this purpose we rewrite equation (4.1.2) for geometric Brownian motion in the form

$$X_t = X_0\, \exp\{L_t\} \tag{7.1.3}$$

with some transformed Wiener process $L = \{L_t,\, t \in [0, \infty)\}$, which is characterized by the Itô differential

$$dL_t = g\, dt + b\, dW_t \tag{7.1.4}$$

for $t \in [0, \infty)$ with $L_0 = \ln(x_0)$. By application of the Itô formula (6.1.12) the corresponding Itô differential for $X_t$ is obtained as

$$dX_t = X_t \left( g + \frac{1}{2}\, b^2 \right) dt + X_t\, b\, dW_t \tag{7.1.5}$$

for $t \in [0, \infty)$ with $X_0 = x_0$. Comparing (7.1.2) and (7.1.5) reveals that the SDE (7.1.2) has a solution, which is a geometric Brownian motion, see (4.1.2), with growth rate $g = a - \frac{1}{2}\, \sigma^2$, volatility $b = \sigma$ and initial value $X_0 = x_0$.

When combining (7.1.3) and (7.1.4), one can identify at time $t$ an explicit solution of the SDE (7.1.5) as a function of the standard Wiener process value $W_t$ that is given in the form

$$X_t = x_0 \exp\left\{\left(a - \frac{1}{2}\,\sigma^2\right) t + \sigma\, W_t\right\} \tag{7.1.6}$$

for $t \in [0, \infty)$. Since this function is a solution for the SDE (7.1.2), it also means that there exists a stochastic process $X = \{X_t,\, t \in [0, \infty)\}$ such that (7.1.2) is satisfied for all $t \in [0, \infty)$. Note that the appreciation rate $a = g + \frac{1}{2}\,\sigma^2$ is larger than the growth rate $g$. This is an effect that results from stochastic calculus and would not appear if $W_t$ were differentiable.

The existence of a solution of an SDE does not automatically mean that this solution is also unique. There may also be other solutions that satisfy the given SDE. However, as we shall see later, the above solution is indeed unique. Recall that we have previously displayed sample paths of geometric Brownian motions, for instance, in Fig. 4.1.2. These paths can be interpreted as trajectories of the solution of an SDE of the form (7.1.5).

### Solution of an SDE

More generally, we say that a stochastic process $Y = \{Y_t,\, t \in [t_0, \infty)\}$ is a *solution* of a given SDE

$$dY_t = \mu(t, Y_t)\, dt + b(t, Y_t)\, dW_t \tag{7.1.7}$$

for $t \in [t_0, \infty)$, with initial value $Y_{t_0} = y_0$, $0 \leq t_0 < \infty$, and driving standard Wiener process $W$, if the process $Y$ has for all $t \in [t_0, \infty)$ an Itô differential of the form (7.1.7). More precisely, a *solution* of the SDE (7.1.7) is a pair $(Y, W)$ of adapted stochastic processes that are defined on the given filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. This pair needs to be such that $W$ is a standard Wiener process and the continuous process $Y$ satisfies a.s. for each $t \in [t_0, \infty)$ the Itô integral equation

$$Y_t = y_0 + \int_{t_0}^{t} \mu(s, Y_s)\, ds + \int_{t_0}^{t} b(s, Y_s)\, dW_s. \tag{7.1.8}$$

Here we need to assume that both integrals on the right hand side of (7.1.8) exist. For instance, it is sufficient when the processes $\sqrt{|\mu|}$ and $b$ are chosen from a set $\mathcal{L}_T^2$, see (5.4.1), for some $T \in [0, \infty)$. However, weaker assumptions are also sufficient, see Chap. 5.

As was the case for Itô differentials, the SDE (7.1.7) is only a shorthand notation for the integral equation (7.1.8). Conditions that ensure the existence and uniqueness of a solution of an SDE will be discussed later.

## 7.2 Linear SDE with Additive Noise

Linear SDEs are those that have linear drift and diffusion coefficients. The BS-SDE (7.1.2) is an example of a linear SDE. We call a solution of an SDE an explicit solution, if it has an analytic representation which does not use the solution itself. Linear SDEs form a class of SDEs that have explicit solutions. When an SDE has an explicit solution, then not only the question of the existence of a solution is resolved, one also can efficiently study most quantitative problems related to this SDE. In this section we shall study some properties of linear SDEs.

### Linear SDE with Additive Noise

Let us first consider the *linear SDE with additive noise*

$$dX_t = (a_1(t)\, X_t + a_2(t))\, dt + b_2(t)\, dW_t \tag{7.2.1}$$

for $t \in [t_0, \infty)$, where the coefficients $a_1$, $a_2$ and $b_2$ are deterministic functions of time $t$. In the drift coefficient we have a feedback effect introduced, which is linear in $X_t$. Here the initial value $X_{t_0}$ is a given constant and $W$ is a standard Wiener process. The SDE (7.2.1) covers, for instance, the Ornstein-Uhlenbeck process, see (4.2.3), and short rate models proposed in Merton (1973a), Ho & Lee (1986) and Vasicek (1977).

We form a homogeneous equation from (7.2.1) by setting $a_2(t) = b_2(t) = 0$. For this *ordinary differential equation* (ODE) of the form

$$d\Phi_{t,t_0} = a_1(t)\, \Phi_{t,t_0}\, dt \tag{7.2.2}$$

with initial condition $\Phi_{t_0,t_0} = 1$ we obtain its *fundamental solution*

$$\Phi_{t,t_0} = \exp\left\{ \int_{t_0}^{t} a_1(s)\, ds \right\}$$

for $t \in [t_0, \infty)$, which is unique due to the linearity of (7.2.2). Assume that a solution of the SDE (7.2.1) exists. Applying the Itô formula (6.1.13) to the transformation $U(t,x) = \Phi_{t,t_0}^{-1}\, x$ for the solution $X_t$ of (7.2.1), we obtain

$$\begin{aligned} d\left(\Phi_{t,t_0}^{-1} X_t\right) &= \left( \frac{d\Phi_{t,t_0}^{-1}}{dt} X_t + (a_1(t)\, X_t + a_2(t))\, \Phi_{t,t_0}^{-1} \right) dt + b_2(t)\, \Phi_{t,t_0}^{-1}\, dW_t \\ &= a_2(t)\, \Phi_{t,t_0}^{-1}\, dt + b_2(t)\, \Phi_{t,t_0}^{-1}\, dW_t, \end{aligned} \tag{7.2.3}$$

since

$$\frac{d\Phi_{t,t_0}^{-1}}{dt} = -\Phi_{t,t_0}^{-1}\, a_1(t)$$

for $t \in [t_0, \infty)$.

**Explicit Solution**

The right hand side of the SDE (7.2.3) only involves known functions of time $t$ as integrands. Consequently, these can be integrated to give

$$\Phi_{t,t_0}^{-1} X_t = \Phi_{t_0,t_0}^{-1} X_{t_0} + \int_{t_0}^{t} a_2(s)\, \Phi_{s,t_0}^{-1}\, ds + \int_{t_0}^{t} b_2(s)\, \Phi_{s,t_0}^{-1}\, dW_s.$$

Since $\Phi_{t_0,t_0} = 1$ this leads to the *explicit solution*

$$X_t = \Phi_{t,t_0} \left( X_{t_0} + \int_{t_0}^{t} a_2(s)\, \Phi_{s,t_0}^{-1}\, ds + \int_{t_0}^{t} b_2(s)\, \Phi_{s,t_0}^{-1}\, dW_s \right) \tag{7.2.4}$$

for $t \in [t_0, \infty)$ for the linear SDE (7.2.1) with additive noise.

**Ornstein-Uhlenbeck Examples**

As an example we consider a standard Ornstein-Uhlenbeck process, see (4.2.3) and (4.5.6). This is a diffusion process with drift function $a(s,x) = -x$ and diffusion coefficient function $b(s,x) = \sqrt{2}$. Taking into account our discussion on diffusion processes in the context of relations (4.3.6) and (7.1.7), one notes that the standard OU process solves a linear SDE with additive noise given by

$$dX_t = -X_t\, dt + \sqrt{2}\, dW_t \tag{7.2.5}$$

for $t \in [t_0, \infty)$ with initial value $X_{t_0}$ at time $t_0$. From the relation (7.2.4) we obtain the explicit solution of (7.2.5) in the form

$$X_t = \exp\{-(t - t_0)\}\, X_{t_0} + \int_{t_0}^{t} \sqrt{2}\, \exp\{-(t-s)\}\, dW_s \tag{7.2.6}$$

for $t \in [t_0, \infty)$. Recall that a sample path of a standard OU process was shown in Fig. 4.2.4 and that the transition density of the standard OU process is Gaussian, see (4.2.3).

More generally, one can show that the SDE

$$dX_t = \gamma_t\, (\bar{X}_t - X_t)\, dt + \beta_t\, dW_t \tag{7.2.7}$$

of an *Ornstein-Uhlenbeck* (OU) *process* for $t \in [t_0, \infty)$ with initial value $X_{t_0}$ at time $t_0 \in [0, \infty)$ has the explicit solution

$$\begin{aligned} X_t = X_{t_0} \exp\left\{ -\int_{t_0}^{t} \gamma_s\, ds \right\} &+ \int_{t_0}^{t} \exp\left\{ -\int_{s}^{t} \gamma_u\, du \right\} \gamma_s\, \bar{X}_s\, ds \\ &+ \int_{t_0}^{t} \exp\left\{ -\int_{s}^{t} \gamma_u\, du \right\} \beta_s\, dW_s \end{aligned} \tag{7.2.8}$$

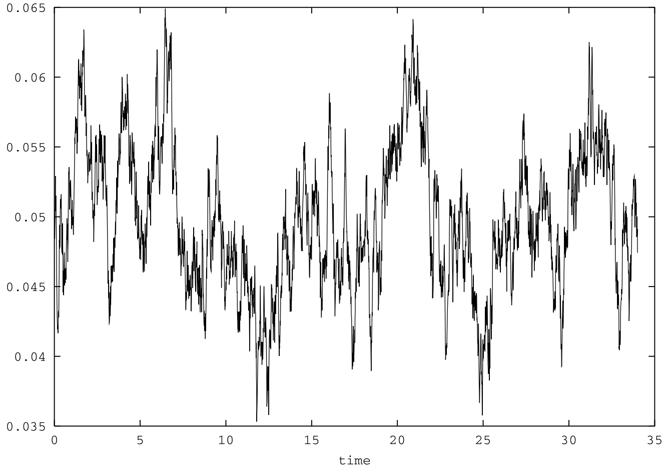



**Fig. 7.2.1.** Vasicek interest rate dynamics

for $t \in [t_0, \infty)$. Here $\gamma_t$ is the *speed of adjustment* at time $t$, $\bar{X}_t$ the *reference level* or *mean reversion level* and $\beta_t$ the diffusion coefficient at time $t$. For $\gamma_t > 0$ the solution of the OU SDE reverts always back to the reference level $\bar{X}_t$. The transition density for the solution of the OU SDE, see also (4.2.3), is Gaussian with mean and variance that can be obtained from (7.2.8), as will be discussed below. The SDE (7.2.7) with solution (7.2.8) can be interpreted as that of the Vasicek (1977) interest rate model when setting the parameters constant. The reference level $\bar{X}_t = \bar{X}$ is then the long-term average value of the interest rate. The parameter $\beta_t = \beta$ characterizes the magnitude of the fluctuations of the short rate. The speed of adjustment parameter $\gamma_t = \gamma$ determines how long it takes for a shock to the short rate loses its impact. In Fig. 7.2.1 we show typical short rate dynamics under the Vasicek model with $t_0 = 0$, $\bar{X} = X_0 = 0.05$, $\beta = 0.01$ and $\gamma = 2$. Note the mean reverting feature of the path of the OU process.

## 7.3 Linear SDE with Multiplicative Noise

### General Linear SDE

We now consider a *general linear SDE* which also covers the case of multiplicative noise. It is given by

$$dX_t = (a_1(t)\, X_t + a_2(t))\, dt + (b_1(t)\, X_t + b_2(t))\, dW_t \tag{7.3.1}$$

for $t \in [t_0, \infty)$, where $a_1$, $a_2$, $b_1$ and $b_2$ are appropriate deterministic functions of time and $W$ is a standard Wiener process. We assume that for the possibly random initial value $X_{t_0}$ its mean and variance are given and $X_{t_0}$ is independent of the Wiener process $W$.

### Explicit Solution

Using similar arguments to those above, it can be shown by the Itô formula that the linear SDE (7.3.1) has an explicit solution of the form

$$X_t = \Psi_{t,t_0} \left( X_{t_0} + \int_{t_0}^{t} (a_2(s) - b_1(s)\, b_2(s))\, \Psi_{s,t_0}^{-1}\, ds + \int_{t_0}^{t} b_2(s)\, \Psi_{s,t_0}^{-1}\, dW_s \right), \tag{7.3.2}$$

where

$$\Psi_{t,t_0} = \exp\left\{ \int_{t_0}^{t} \left( a_1(s) - \frac{1}{2}\, b_1^2(s) \right) ds + \int_{t_0}^{t} b_1(s)\, dW_s \right\} \tag{7.3.3}$$

for $t \in [t_0, \infty)$. It is easy to see for $b_1(s) = 0$ that for $s \in [0, \infty)$ that the expression (7.3.2) reduces to equation (7.2.4).

### Moment Equations

The explicit solution (7.2.4) turns out to be a Gaussian process whenever the initial value $X_{t_0}$ is either a constant or a Gaussian random variable. Its mean and second moment both satisfy ODEs. These are stated below for the general linear SDE (7.3.1). We remark that the solution of the general linear SDE given by (7.3.2) and (7.3.3) is not always Gaussian.

If we take the expectation of the integral form of equation (7.3.1) and use the fact that $X$ can be shown to be square integrable, then the martingale property (5.4.3) of an Itô integral can be exploited and we obtain for the mean

$$\mu(t) = E(X_t) \tag{7.3.4}$$

the ODE

$$d\mu(t) = (a_1(t)\, \mu(t) + a_2(t))\, dt \tag{7.3.5}$$

for $t \in [t_0, \infty)$ with $\mu(t_0) = E(X_{t_0})$.

Figure 7.3.1 shows the mean $\mu(t)$ for an OU process with initial value $t_0 = 0$, $X_{t_0} = 1$, where $a_1(t) = -1$, $a_2(t) = 0$, $b_1(t) = 0$ and $b_2(t) = \sqrt{2}$ for $t \in [t_0, 5]$. Note that in this example the mean converges exponentially over time, according to the formula $\mu(t) = \exp\{-t\}$, towards the reference level or mean reversion level, which is here zero. More generally, the solution of the ODE (7.3.5) for the mean has the form

$$\mu(t) = E(X_{t_0})\, \exp\left\{ \int_{t_0}^{t} a_1(s)\, ds \right\} + \int_{t_0}^{t} a_2(s)\, \exp\left\{ \int_{s}^{t} a_1(u)\, du \right\} ds \tag{7.3.6}$$

for $t \in [t_0, \infty)$.

By similar arguments it can also be shown that the second moment

$$P(t) = E(X_t^2) \tag{7.3.7}$$

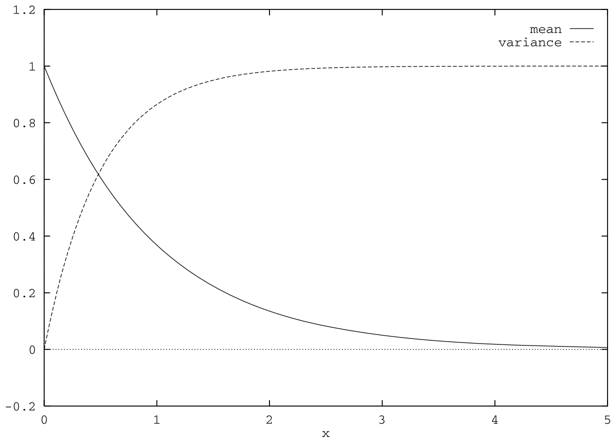



**Fig. 7.3.1.** Mean and variance of an Ornstein-Uhlenbeck process

satisfies the ODE

$$dP(t) = \Big(\left(2a_1(t) + b_1^2(t)\right) P(t) + 2\mu(t)\left(a_2(t) + b_1(t)\, b_2(t)\right) + b_2^2(t)\Big)\, dt \quad (7.3.8)$$

for $t \in [t_0, \infty)$ and

$$P(t_0) = E\left(X_{t_0}^2\right).$$

To derive (7.3.8) we apply the Itô formula to obtain an SDE for $(X_t)^2$ and then take the expectation of the integral form of this equation. Here we use the fact that $X_t$ can be shown to be square integrable. Both (7.3.5) and (7.3.8) are linear ODEs and can be solved explicitly as a special case of the linear SDE (7.2.1).

Figure 7.3.1 displays also the variance $v(t)$ for an OU process. It is easy to see that the variance

$$v(t) = P(t) - (\mu(t))^2 = 1 - e^{-2(t-t_0)} \quad (7.3.9)$$

starts at zero and converges to its long term average value of one.

More generally, the solution of the ODE (7.3.8) is

$$\begin{aligned} P(t) = P(t_0)\, \exp\left\{\int_{t_0}^t \left(2\, a_1(s) + b_1^2(s)\right) ds\right\} \\ + \int_{t_0}^t \left(2\,\mu(s)\left(a_2(s) + b_1(s)\, b_2(s)\right) + b_2^2(s)\right) \\ \times \exp\left\{\int_s^t \left(2\, a_1(u) + b_1^2(u)\right) du\right\} ds \end{aligned} \quad (7.3.10)$$

for $t \in [t_0, \infty)$. Therefore, we obtain the variance as the difference

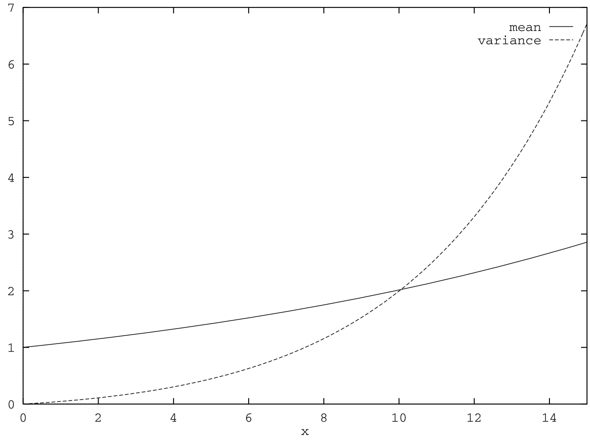



**Fig. 7.3.2.** Mean and variance in Black-Scholes model

$$v(t) = P(t) - (\mu(t))^2 \tag{7.3.11}$$

for $t \in [t_0, \infty)$, which uses (7.3.10) and the square of (7.3.6).

### Moments of Black-Scholes Model

As an example of a particular linear SDE that is important in finance consider the SDE of the BS model given by

$$dX_t = a\, X_t\, dt + \sigma\, X_t\, dW_t, \tag{7.3.12}$$

for $t \in [t_0, \infty)$. Using the moment equations (7.3.6), (7.3.9) and (7.3.10), it is easy to see that we have the mean

$$\mu(t) = X_{t_0}\, \exp\{a\,(t - t_0)\} \tag{7.3.13}$$

and variance

$$v(t) = P(t) - (\mu(t))^2 = (X_{t_0})^2\, \exp\{2\,a\,(t - t_0)\}\, \left(\exp\{\sigma^2\,(t - t_0)\} - 1\right) \tag{7.3.14}$$

for $t \in [t_0, \infty)$. These formulas show for $a > 0$ that the mean and the variance both diverge as $t \to \infty$. These two quantities are shown in Fig. 7.3.2 for $t_0 = 0$, $X_{t_0} = 1$, $a = 0.07$ and $\sigma = 0.2$.

## 7.4 Vector Stochastic Differential Equations

### Vector SDE

Obviously, multi-dimensional versions of stochastic processes as solutions of SDEs are needed for the modeling of financial markets. The relationship between vector and scalar SDEs is analogous to that between vector and scalar Itô differentials.

We recall that $\boldsymbol{W} = \{\boldsymbol{W}_t = (W_t^1, W_t^2, \ldots, W_t^m)^\top,\, t \in [t_0, \infty)\}$ is an $m$-dimensional standard Wiener process with components $W^1$, $W^2$, …, $W^m$. These are independent scalar Wiener processes, as described in Sect. 5.1.

We use a $d$-dimensional vector function $\boldsymbol{a} : [t_0, \infty) \times \Re^d \to \Re^d$ and a $d{\times}m$-matrix function $\boldsymbol{b} : [t_0, \infty) \times \Re^d \to \Re^{d \times m}$ to form a $d$-dimensional *vector stochastic differential equation*

$$d\boldsymbol{X}_t = \boldsymbol{a}(t, X_t)\, dt + \boldsymbol{b}(t, X_t)\, d\boldsymbol{W}_t \tag{7.4.1}$$

for $t \in [t_0, \infty)$ with initial value $\boldsymbol{X}_{t_0} \in \Re^d$. As was the case for Itô differentials we have to interpret the differential (7.4.1) as an Itô integral equation of the form

$$\boldsymbol{X}_t = \boldsymbol{X}_{t_0} + \int_{t_0}^t \boldsymbol{a}(s, X_s)\, ds + \int_{t_0}^t \boldsymbol{b}(s, X_s)\, d\boldsymbol{W}_s, \tag{7.4.2}$$

for $t \in [t_0, \infty)$. Here the random ordinary Riemann-Stieltjes integral and the Itô integral are determined componentwise. The $i$th component of (7.4.2) is then given by the SDE

$$X_t^i = X_{t_0}^i + \int_{t_0}^t a^i(s, X_s)\, ds + \sum_{\ell=1}^m \int_{t_0}^t b^{i,\ell}(s, X_s)\, dW_s^\ell, \tag{7.4.3}$$

for $t \in [t_0, \infty)$ and $i \in \{1, 2, \ldots, d\}$. This equation shows how the different components of the vector solution of the SDE feed into the elements of the drift and diffusion coefficients and how the different Wiener processes drive the components of the vector solution. Note that the drift and diffusion coefficients of each component can also depend on all other components. The form of the vector SDE given in (7.4.1) provides a compact description of the set of components expressed by (7.4.3).

Questions concerning the existence and uniqueness of solutions of vector SDEs naturally arise. It can be shown under appropriate assumptions that a unique solution of the vector SDE (7.4.1) exists. For details we refer to the existence and uniqueness theorem at the end of this chapter, see also Krylov (1980).

### Explicit Solutions of Multi-Dimensional Linear SDEs

Consider a $d$-dimensional *linear SDE* of the form

$$d\boldsymbol{X}_t = (\boldsymbol{A}_t\, \boldsymbol{X}_t + \boldsymbol{\alpha}_t)\, dt + \sum_{\ell=1}^m (\boldsymbol{B}_t^{\boldsymbol{\ell}}\, \boldsymbol{X}_t + \boldsymbol{\beta}_t^{\boldsymbol{\ell}})\, dW_t^\ell, \tag{7.4.4}$$

where $\boldsymbol{A}$, $\boldsymbol{B^1}$, $\boldsymbol{B^2}$, …, $\boldsymbol{B^m}$ are $d \times d$-matrix and $\boldsymbol{\alpha}$, $\boldsymbol{\beta^1}$, $\boldsymbol{\beta^2}$, …, $\boldsymbol{\beta^m}$ $d$-dimensional vector valued deterministic functions of time. Similar as was shown in Sect. 7.3 for the scalar case, we obtain an explicit solution of (7.4.4) in the form

$$\boldsymbol{X}_t = \boldsymbol{\Psi}_{t,t_0}\left(\boldsymbol{X}_{t_0} + \int_{t_0}^t \boldsymbol{\Psi}_{s,t_0}^{-1}\left(\boldsymbol{\alpha}_s - \sum_{\ell=1}^m \boldsymbol{B}_s^{\boldsymbol{\ell}}\boldsymbol{\beta}_s^{\boldsymbol{\ell}}\right)ds + \sum_{\ell=1}^m \int_{t_0}^t \boldsymbol{\Psi}_{s,t_0}^{-1}\,\boldsymbol{\beta}_s^{\boldsymbol{\ell}}\,dW_s^{\ell}\right) \tag{7.4.5}$$

for $t \in [t_0, \infty)$. Here $\boldsymbol{\Psi}_{t,t_0}$ is the $d{\times}d$ *fundamental matrix* at time $t$ with $\boldsymbol{\Psi}_{t_0,t_0} = \boldsymbol{I}$, where $\boldsymbol{I}$ is denoting the unit matrix. The fundamental matrix satisfies the *matrix SDE*

$$d\boldsymbol{\Psi}_{t,t_0} = \boldsymbol{A}_t\,\boldsymbol{\Psi}_{t,t_0}\,dt + \sum_{\ell=1}^m \boldsymbol{B}_t^{\boldsymbol{\ell}}\,\boldsymbol{\Psi}_{t,t_0}\,dW_t^{\ell}, \tag{7.4.6}$$

which can be interpreted elementwise similar to a vector SDE, see (7.4.1) and (7.4.3). One can check the above explicit solution by applying the Itô formula (6.2.11).

### Moments of Multi-Dimensional Linear SDEs

For the vector SDE (7.4.4), we can derive vector and matrix ODEs for the vector mean

$$\boldsymbol{\mu}(t) = E(\boldsymbol{X}_t)$$

and the $d{\times}d$ matrix of second moments

$$\boldsymbol{P}(t) = E(\boldsymbol{X}_t\,\boldsymbol{X}_t^\top),$$

respectively.

Recall that for $d$-dimensional vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ the product $\boldsymbol{x}\boldsymbol{y}^\top$ is a $d{\times}d$ matrix with $(i,j)$th component $x^i y^j$. One can then show that the *vector mean* satisfies the vector ODE

$$d\boldsymbol{\mu}(t) = (\boldsymbol{A}_t\boldsymbol{\mu}(t) + \boldsymbol{\alpha}_t)\,dt. \tag{7.4.7}$$

Furthermore, the *matrix of second moments* satisfies the matrix ODE

$$d\boldsymbol{P}(t) = \Bigg(\boldsymbol{A}_t\,\boldsymbol{P}(t) + \boldsymbol{P}(t)\,\boldsymbol{A}_t^\top + \sum_{\ell=1}^m \boldsymbol{B}_t^{\boldsymbol{\ell}}\,\boldsymbol{P}(t)\,\left(\boldsymbol{B}_t^{\boldsymbol{\ell}}\right)^\top + \boldsymbol{\alpha}_t\,\boldsymbol{\mu}(t)^\top + \boldsymbol{\mu}(t)\,\boldsymbol{\alpha}_t^\top$$
$$+ \sum_{\ell=1}^m \left(\boldsymbol{B}_t^{\boldsymbol{\ell}}\,\boldsymbol{\mu}(t)\left(\boldsymbol{\beta}_t^{\boldsymbol{\ell}}\right)^\top + \boldsymbol{\beta}_t^{\boldsymbol{\ell}}\,\boldsymbol{\mu}(t)^\top\,\boldsymbol{B}_t^{\boldsymbol{\ell}} + \boldsymbol{\beta}_t^{\boldsymbol{\ell}}\left(\boldsymbol{\beta}_t^{\boldsymbol{\ell}}\right)^\top\right)\Bigg)\,dt, \tag{7.4.8}$$

with initial conditions $\boldsymbol{\mu}(t_0) = E\left(\boldsymbol{X}_{t_0}\right)$ and $\boldsymbol{P}(t_0) = E\left(\boldsymbol{X}_{t_0}\,\boldsymbol{X}_{t_0}^\top\right)$.

## 7.5 Constructing Explicit Solutions of SDEs

In Kloeden & Platen (1999) a collection of explicit solutions of SDEs is given. It appears that many of these are, in principle, transformations of the solutions of systems of linear SDEs obtained via the Itô formula. The following property of certain linear SDEs has importance in modeling financial markets.

### Commutativity

If the matrices $\boldsymbol{A}$, $\boldsymbol{B^1}$,$\boldsymbol{B^2}$, ..., $\boldsymbol{B^m}$ are constant and *commute*, that is, if

$$\boldsymbol{A}_t\,\boldsymbol{B}^{\boldsymbol{\ell}}_t = \boldsymbol{B}^{\boldsymbol{\ell}}_t\,\boldsymbol{A}_t \quad \text{and} \quad \boldsymbol{B}^{\boldsymbol{\ell}}_t\,\boldsymbol{B}^{\boldsymbol{k}}_t = \boldsymbol{B}^{\boldsymbol{k}}_t\,\boldsymbol{B}^{\boldsymbol{\ell}}_t \tag{7.5.1}$$

for all $k, \ell \in \{1, 2, \ldots, m\}$ and $t \in [0, \infty)$, then an explicit solution of the fundamental matrix SDE (7.4.6) can be obtained. It is at time $t$ given by the fundamental matrix

$$\boldsymbol{\Psi}_{t,t_0} = \exp\left\{\int_{t_0}^{t}\left(\boldsymbol{A}_s - \frac{1}{2}\sum_{\ell=1}^{m}\left(\boldsymbol{B}^{\boldsymbol{\ell}}_s\right)^2\right)ds + \sum_{\ell=1}^{m}\int_{t_0}^{t}\boldsymbol{B}^{\boldsymbol{\ell}}_s\,dW^{\ell}_s\right\}, \tag{7.5.2}$$

where the exponential is taken elementwise. In the special case, where the matrices $\boldsymbol{B}^{\boldsymbol{\ell}}_t$, $\ell \in \{1, 2, \ldots, m\}$, are all identically zero this formula reduces to the matrix expression

$$\boldsymbol{\Psi}_{t,t_0} = \exp\left\{\int_{t_0}^{t}\boldsymbol{A}_s\,ds\right\}, \tag{7.5.3}$$

which is the fundamental solution of the deterministic linear vector ODE

$$d\boldsymbol{\Psi}_{t,t_0} = \boldsymbol{A}_t\,\boldsymbol{\Psi}_{t,t_0}\,dt$$

for $t \in [t_0, \infty)$.

The relations (7.4.5)–(7.5.2) have demonstrated that an explicit solution can be obtained for a multi-dimensional linear SDE with matrices that commute in the sense of (7.5.1). However, even if the matrices in (7.4.6) do not commute, then the first and second moment relations (7.4.7) and (7.4.8) remain valid and can be conveniently exploited.

### Multi-Asset Black-Scholes Model

The risky assets in a BS model that corresponds to a setup similar to that mentioned in Sect. 6.3, can be written in the form (7.4.4), where $\boldsymbol{A}_t$, $\boldsymbol{B}^{\boldsymbol{1}}_t$, ..., $\boldsymbol{B}^{\boldsymbol{m}}_t$ are diagonal matrices and the vectors $\boldsymbol{\alpha}_t = \boldsymbol{\beta}^{\boldsymbol{1}}_t = \ldots = \boldsymbol{\beta}^{\boldsymbol{m}}_t$ are zero vectors.

To be precise, we denote by $\boldsymbol{S}_t$ a diagonal matrix with $j$th diagonal element $S^j_t$, $j \in \{1, 2, \ldots, d\}$, representing the $j$th stock price at time $t \in [0, \infty)$. The vector of appreciation rates at time $t$ is given as $\boldsymbol{a}_t = (a^1_t, a^2_t, \ldots, a^d_t)^\top$. The vector of volatilities with respect to the $k$th Wiener process is described as $\boldsymbol{b}^k_t = (b^{1,k}_t, b^{2,k}_t, \ldots, b^{d,k}_t)^\top$. The SDE for the $j$th Black-Scholes asset price $S^j_t$ is then defined in the form

$$dS^j_t = S^j_t\left(a^j_t\,dt + \sum_{k=1}^{d} b^{j,k}_t\,dW^k_t\right) \tag{7.5.4}$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. To fit this type of SDE into the framework given by the fundamental matrix SDE (7.4.6) we use now the diagonal matrices $\boldsymbol{A}_t = [A_t^{i,j}]_{i,j=1}^d$ with

$$A_t^{i,j} = \begin{cases} a_t^j & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{7.5.5}$$

and $\boldsymbol{B}_t^{\boldsymbol{k}} = [B_t^{k,i,j}]_{i,j=1}^d$ with

$$B_t^{k,i,j} = \begin{cases} b_t^{j,k} & \text{for } i = j \\ 0 & \text{otherwise} \end{cases} \tag{7.5.6}$$

for $k, i, j \in \{1, 2, \ldots, d\}$ and $t \in [0, \infty)$. Note that these matrices commute due to their diagonal structure. This allows us to write the SDE (7.5.4) as matrix SDE

$$d\boldsymbol{S}_t = \boldsymbol{A}_t \, \boldsymbol{S}_t \, dt + \sum_{k=1}^d \boldsymbol{B}_t^{\boldsymbol{k}} \, \boldsymbol{S}_t \, dW_t^k \tag{7.5.7}$$

for $t \in [0, \infty)$. We note that this corresponds to the matrix SDE (7.4.6) for the fundamental matrix solution. Consequently, by (7.5.2) we obtain the explicit solution

$$S_t^j = S_0^j \exp\left\{ \int_0^t \left( a_s^j - \frac{1}{2} \sum_{k=1}^d \left( b_s^{j,k} \right)^2 \right) ds + \sum_{k=1}^d b_s^{j,k} \, dW_s^k \right\} \tag{7.5.8}$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. Thus, an explicit solution is available for the BS model when formulated for an entire market. This is important from the practical point of view and supports the fact that the BS model is the standard market model.

### A Representation of the Square Root Process

Let us now study another process, which is highly important in finance. It appears, for instance, in the Cox, Ingersoll, Ross (CIR) interest rate model, see also Sect. 4.5. We shall need this process below to describe the dynamics of a particular financial market model, which we shall later derive under the benchmark approach.

We have seen in Chap. 6 that the Itô formula can be used to create new diffusion processes as functionals of known ones. In addition, it can also be used to understand links between certain diffusion processes and their functionals, as will be shown below. We shall relate the dynamics of the *square root* (SR) *process*, see Sect. 4.4, to that of OU processes, see Sect. 4.2.

Consider $n$ OU processes of the form

$$dX_t^i = -c \, X_t^i \, dt + b \, dW_t^i \tag{7.5.9}$$

for $t \in [0, \infty)$, with $X_0^i = x_0$, $c > 0$, $b > 0$, and independent standard Wiener processes $W^i$ for $i \in \{1, 2, \ldots, n\}$. We show in Fig. 7.5.1 the sample paths of

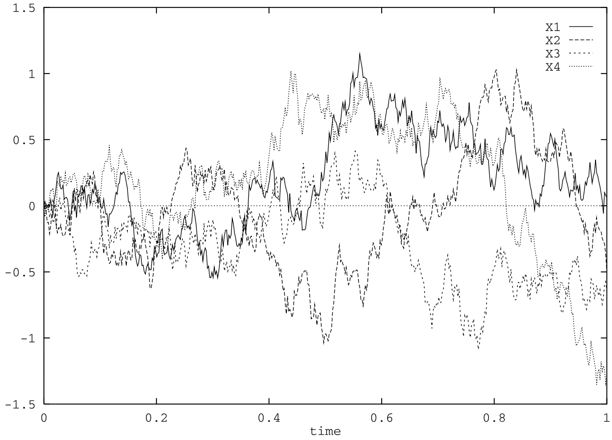



**Fig. 7.5.1.** Four independent Ornstein-Uhlenbeck processes

$n = 4$ independent standard OU processes, see (4.2.3). These will be used to construct an SR process.

Consider now the square of each of the above OU processes, which by (6.2.11) satisfy the Itô differential

$$d(X_t^i)^2 = \left(-2c\,(X_t^i)^2 + b^2\right)dt + 2b\,X_t^i\,dW_t^i, \tag{7.5.10}$$

for $t \in [0, \infty)$ and $i \in \{1, 2, \ldots, n\}$. We shall now form the sum

$$Y_t = \sum_{i=1}^{n} (X_t^i)^2 \tag{7.5.11}$$

of the $n$ squared OU processes. Figure 7.5.2 displays the sample path of this sum $Y_t$ for the paths shown in Fig. 7.5.1. The value $Y_t$ satisfies by (7.5.10) and (7.5.11) the Itô differential

$$\begin{aligned} dY_t &= d\left(\sum_{i=1}^{n}(X_t^i)^2\right) \\ &= \left(\sum_{i=1}^{n}\left(-2c\,(X_t^i)^2\right) + n\,b^2\right)dt + 2b\sum_{i=1}^{n} X_t^i\,dW_t^i \end{aligned} \tag{7.5.12}$$

for $t \in [0, \infty)$. To simplify this Itô differential let $\bar{W} = \{\bar{W}_t,\, t \in [0, \infty)\}$ denote the process that is given by the expression

$$\bar{W}_t = \int_0^t d\bar{W}_s = \sum_{i=1}^{n} \int_0^t \frac{X_s^i}{\sqrt{Y_s}}\,dW_s^i \tag{7.5.13}$$

for $t \in [0, \infty)$

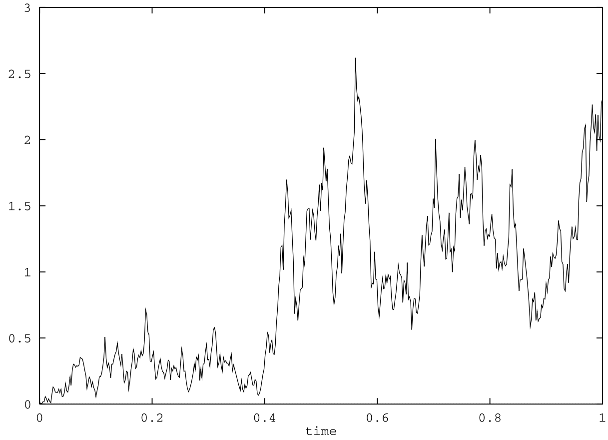



**Fig. 7.5.2.** Sum of four squared Ornstein-Uhlenbeck-processes

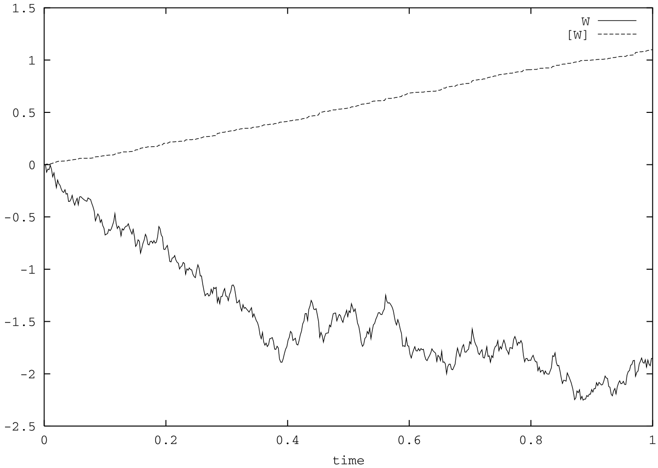



**Fig. 7.5.3.** The process $\bar{W}$ and its quadratic variation $[\bar{W}]$

Figure 7.5.3 displays the approximate quadratic variation of $\bar{W}_t$. Note that it seems almost exactly to equal the quadratic variation of a standard Wiener process with $[\bar{W}]_t = t$. Indeed it turns out that $\bar{W}$ is a continuous martingale that starts at zero and has quadratic variation

$$[\bar{W}]_t = \int_0^t \sum_{i=1}^n \frac{(X_s^i)^2}{Y_s}\, ds = t. \tag{7.5.14}$$

By applying Lévy's Theorem, see (6.2.1), we see that $\bar{W}$ is a standard Wiener process.

From (7.5.13) and (7.5.14) it is apparent that the Itô differential (7.5.12) has the form

$$dY_t = (n\,b^2 - 2\,c\,Y_t)\,dt + 2\,b\,\sqrt{Y_t}\,d\bar{W}_t \tag{7.5.15}$$

for $t \in [0,\infty)$ with $Y_0 = n\,(x_0)^2$. One observes that the process $Y$ is an SR process similar to that used in the CIR short rate model, as can be seen from its drift and diffusion coefficient, see Sect. 4.3 and Sect. 4.4. It is important to note that the SR process is linear mean reverting in its drift with *reference level* $\frac{n\,b^2}{2\,c}$ and *speed of adjustment* $2c$. This result shows that the sum of squared OU processes yields an SR process. Since the independent OU processes, given by (7.5.12), have Gaussian transition densities the resulting SR process must have a chi-square transition density with $n$ degrees of freedom, see Sect. 1.2. Indeed, it can be shown that the transition density given in formula (4.4.6) is exactly of this type for $n \in \{2,3,\ldots\}$. One calls $n$ the *dimension* of the SR process. Later we shall consider SR processes of general dimension $n \in (0,\infty)$.

### Minimal Market Model

By the use of a square root (SR) process one can formulate a stylized version of the *minimal market model* (MMM), as suggested in Platen (2001, 2002). We shall show later in Chap. 13 that this model arises naturally from economic considerations. It is a model that expresses the dynamics of a normalized market index

To describe the model we introduce a Wiener process $W$. This Wiener process drives an SR process $Y = \{Y_t,\, t \in [0,\infty)\}$, which expresses the normalized index. Here we set

$$dY_t = (1 - \eta\,Y_t)\,dt + \sqrt{Y_t}\,dW_t \tag{7.5.16}$$

for $t \in [0,\infty)$ with $Y_0 > 0$, where the *net growth rate* $\eta > 0$ is the key parameter. One notes that $\frac{1}{\eta}$ is the reference level for the SR process $Y$ and $\eta$ is the speed of adjustment of $Y$. Let us denote by $B = \{B_t,\, t \in [0,\infty)\}$ the savings account process with constant short rate $r_t = r \geq 0$, such that

$$B_t = \exp\{r\,t\} \tag{7.5.17}$$

for $t \in [0,\infty)$. Under the stylized MMM the market index $S = \{S_t,\, t \in [0,\infty)\}$ is modeled as the product

$$S_t = B_t\,Y_t\,\alpha(t) \tag{7.5.18}$$

with an exponential function

$$\alpha(t) = \alpha_0\,\exp\{\eta\,t\} \tag{7.5.19}$$

for $t \in [0,\infty)$ with $\alpha_0 > 0$. One notes that the function $B_t\alpha(t)$, when used for normalization, yields the normalized market index $Y_t = \frac{S_t}{B_t\alpha(t)}$.

Using the Itô formula for the discounted market index it follows from (7.5.18) and (7.5.16) that

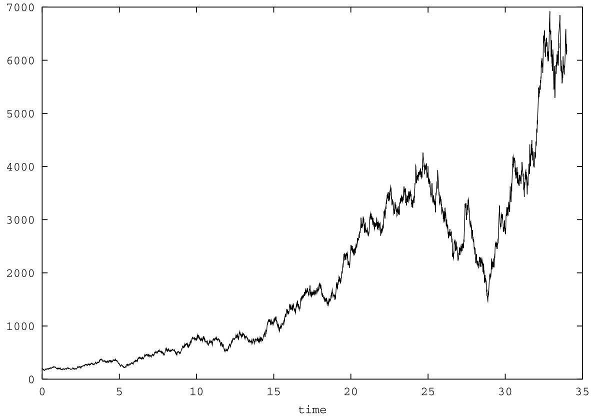



**Fig. 7.5.4.** Market index under the MMM

$$\bar{S}_t = \frac{S_t}{B_t} = Y_t\,\alpha(t), \tag{7.5.20}$$

where

$$d\bar{S}_t = d(Y_t\,\alpha(t)) = Y_t\,\alpha(t)\left(\left(\frac{1}{Y_t} - \eta + \eta\right)dt + \frac{1}{\sqrt{Y_t}}\,dW_t\right)$$

$$= \bar{S}_t\left(\frac{1}{Y_t}\,dt + \frac{1}{\sqrt{Y_t}}\,dW_t\right) = \alpha(t)\,dt + \sqrt{\alpha(t)\,\bar{S}_t}\,dW_t \tag{7.5.21}$$

for $t \in [0,\infty)$. In (7.5.21) we see that $\frac{1}{\sqrt{Y_t}}$ is the volatility of $\bar{S}$. In this sense the MMM has a stochastic volatility. The market index $S_t = \bar{S}_t B_t$ satisfies then, by (7.5.21) and application of the Itô formula, the SDE

$$d(Y_t\,\alpha(t)\,B_t) = dS_t = S_t\left(r\,dt + \frac{1}{\sqrt{Y_t}}\left(\frac{1}{\sqrt{Y_t}}\,dt + dW_t\right)\right) \tag{7.5.22}$$

for $t \in [0,\infty)$. Note that under the MMM the volatility $\frac{1}{\sqrt{Y_t}}$ of the market index equals the inverse of the square root of the normalized index. Consequently, the market index and its volatility are negatively correlated. In Fig. 7.5.4 we show a trajectory of a simulated market index for $\eta = 0.05$, $\alpha_0 = 10$, $r = 0.05$ and $Y_0 = 20$. To visualize the resulting negative correlation of the market index with its volatility, we plot the corresponding volatility $\frac{1}{\sqrt{Y_t}}$ in Fig. 7.5.5.

## 7.6 Jump Diffusions (*)

### Merton's Jump Diffusion Model (*)

One of the pioneers in the use of continuous time models in finance, in particular, for asset prices with jumps, has been Merton, see Merton (1976). He

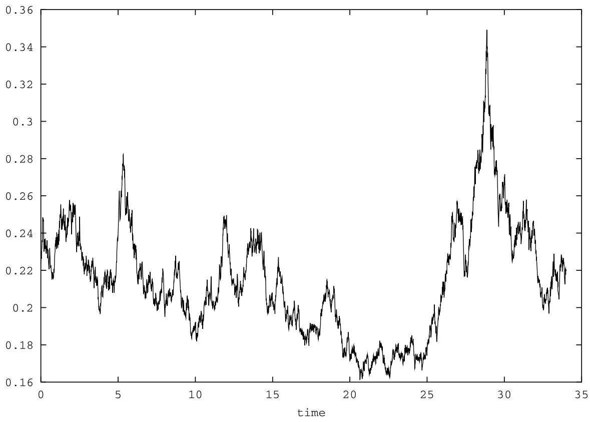



**Fig. 7.5.5.** Volatility of market index under the MMM

modeled the dynamics of a stock price $S_t$ by an SDE of the type

$$dS_t = S_{t-}\,(a\,dt + \sigma\,dW_t + dY_t) \tag{7.6.1}$$

for $t \in [0,\infty)$ and $S_0 > 0$. In this *Merton model* (MM) $Y = \{Y_t,\, t \in [0,\infty)\}$ denotes a compound Poisson process, see (3.5.9), where

$$Y_t = \sum_{k=1}^{N_t} \xi_k \tag{7.6.2}$$

and thus

$$dY_t = \xi_{N_{t-}+1}\,dN_t \tag{7.6.3}$$

for $t \in [0,\infty)$. The process $N = \{N_t,\, t \in [0,\infty)\}$ is a Poisson process, see (3.5.1), with intensity $\lambda > 0$ and $\xi_1, \xi_2, \ldots$ are i.i.d random variables with mean

$$\hat{\xi} = E(\xi_i) < \infty,$$

which are independent of $W$ and $N$. The constant $a$ is the instantaneous expected rate of return of the stock if there were no jumps, whereas $\sigma$ is the constant volatility parameter. The compound Poisson process $Y$ has finite total variation, see (5.2.25). Therefore, the Itô differential $dY_t$, see Sect. 5.5, can be interpreted in the ordinary Riemann-Stieltjes sense. As usual, we take $S$, $N$ and $Y$ to be right continuous. Recall that the left hand limit $S_{t-}$ of $S_t$ at time $t$ denotes the value just before a potential jump at time $t$. Let us denote the $k$th jump time of the Poisson process $N$ by $\tau_k$. Then we have for $Y$ the jump size

$$\Delta Y_{\tau_k} = Y_{\tau_k} - Y_{\tau_k-} = \xi_k \tag{7.6.4}$$

at time $\tau_k$ for $k \in \{1,2,\ldots\}$. The dynamics specified in (7.6.1) yields, therefore, for $S$ the $k$th jump increment

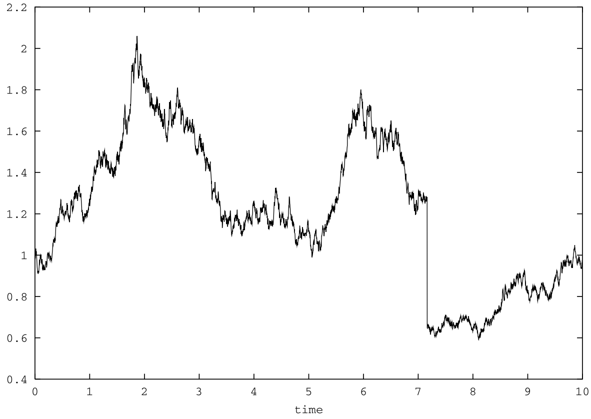



**Fig. 7.6.1.** Path of Merton's jump diffusion model

$$\Delta S_{\tau_k} = S_{\tau_k} - S_{\tau_k -} = S_{\tau_k -}\, \Delta Y_{\tau_k} = S_{\tau_k -}\, \xi_k$$

at time $\tau_k$ and thus

$$S_{\tau_k} = S_{\tau_k -}\, (\xi_k + 1). \tag{7.6.5}$$

One can interpret

$$\Psi(\tau_k -) = \frac{S_{\tau_k}}{S_{\tau_k -}} = \xi_k + 1 \tag{7.6.6}$$

as the *jump ratio* of $S$ at $\tau_k$. To ensure that $S_t$ does not become negative one needs to assume that

$$\xi_k \geq -1, \tag{7.6.7}$$

or equivalently that the jump ratio is nonnegative, that is, by (7.6.6) and (7.6.5)

$$\Psi(\tau_k -) \geq 0 \tag{7.6.8}$$

for all $k \in \{1, 2, \ldots\}$. The value $\lambda \hat{\xi} t$ compensates the accumulated jumps of $Y$ until time $t$ in the sense that

$$\tilde{Y}_t = Y_t - \lambda\, \hat{\xi}\, t \tag{7.6.9}$$

forms an $(\underline{\mathcal{A}}, P)$-martingale. This means that we can rewrite (7.6.1) in the form

$$dS_t = S_t\, (a + \lambda\, \hat{\xi})\, dt + S_t\, \sigma\, dW_t + S_{t-}\, d\tilde{Y}_t \tag{7.6.10}$$

for $t \in [0, \infty)$. The last two terms on the right hand side of the SDE (7.6.10) each form a martingale and the expected return of $S$ over a period of length $\Delta$ equals $(a + \lambda\, \hat{\xi})\Delta$.

In Fig. 7.6.1 we show a path of a solution of the SDE (7.6.10) with $S_0 = 1$, $a = 0.1$, $\sigma = 0.2$, $\lambda = 0.1$ and $\xi_k = -0.5$. It could model the stock of a company that at default recovers half of its previous value. In the above example there is on average one default in 10 years and we observe in Fig. 7.6.1 one such default or credit event. Of course, for another scenario a different number of jumps may arise.

### Explicit Solution of Merton's Jump Diffusion Model (*)

In the given case we have the *explicit solution* of the SDE (7.6.1) in the form

$$S_t = S_0 \exp\left\{\left(a - \frac{1}{2}\,\sigma^2\right) t + \sigma\, W_t\right\} \prod_{k=1}^{N_t} (\xi_k + 1) \tag{7.6.11}$$

for $t \in [0, \infty)$. The explicit solution in (7.6.11) generalizes that of the geometric Brownian motion in (7.1.6). Additionally, the product of the jump ratios, see (7.6.6), appears. Solutions of this type were used by Merton to derive a formula for the value of an option on $S$. Important is the question of how to specify realistically the distribution of the random variables $\xi_k$. In Merton (1976) it was assumed that the $k$th jump ratio $\Psi(\tau_k-) = \xi_k + 1$ is lognormal, whereas, for instance, in Kou (2002) a log-Laplace distribution is assumed. The above model can also be interpreted as an extension of the Cramér-Lundberg model for the surplus process of an insurance company, see (3.5.10), where the surplus has some continuous uncertainty.

### SDE with Jumps (*)

Let us consider a stochastic process $Z = \{Z_t,\, t \in [0, \infty)\}$ with jumps with stochastic differential

$$dZ_t = g_t\, dt + \sigma_t\, dW_t + c_{t-}\, dN_t \tag{7.6.12}$$

for $t \in [0, \infty)$ with $Z_0 = z_0 \in \Re$. Here $W$ is a standard Wiener process and $N$ a Poisson process with intensity $\lambda > 0$, see Definition 3.5.1. The functions $g$, $\sigma$ and $c$ are assumed to be given deterministic functions of time. It then follows for the exponential

$$X_t = \exp\{Z_t\} \tag{7.6.13}$$

by application of the Itô formula (6.4.20) the SDE

$$dX_t = X_{t-}\left[\left(g_t + \frac{1}{2}\,\sigma_t^2\right) dt + \sigma_t\, dW_t + (\exp\{c_{t-}\} - 1)\, dN_t\right] \tag{7.6.14}$$

for $t \in [0, \infty)$, see (6.4.6). Note that since $W$ is a martingale and the compensated Poisson process $q = \{q_t,\, t \in [0, \infty)\}$ with

$$q_t = N_t - \lambda\, t$$

is also a martingale. The equation for the mean $\mu(t) = E(X_t)$ follows from (7.6.14). It is the ODE

$$d\mu(t) = \mu(t)\left(g_t + \frac{1}{2}\,\sigma_t^2 + \lambda(\exp\{c_t\} - 1)\right) dt \tag{7.6.15}$$

for $t \in [0, \infty)$ with $\mu(0) = \exp\{z_0\}$. Obviously, by (7.6.13),

$$X_t^2 = \exp\{2\, Z_t\} \tag{7.6.16}$$

and by similar arguments as above we obtain for the second moment

$$P(t) = E(X_t^2)$$

of $X_t$ the ODE

$$dP(t) = P(t)\Big(2\left(g_t + \sigma_t^2\right) + \lambda\left(\exp\{2\, c_t\} - 1\right)\Big)dt \tag{7.6.17}$$

for $t \in [0, \infty)$ with $P(0) = \exp\{2\, z_0\}$. From the above ODEs one can obtain the expression for the variance of $X_t$ and other functionals.

### SDE for Jump Diffusions (*)

It is important to incorporate event driven uncertainty not only for credit risk but also for insurance risk and operational risk. Such uncertainty generate not only defaults or rating changes but also insurance claims, operational failures and catastrophes. Additionally, some continuous uncertainty from trading noise is often present in an asset price dynamics. As we have seen previously, SDEs for realistic asset prices should be such that their coefficients do not only depend on time but generate also a feedback effect in dependence on the level of the asset price itself.

To give an example, let $W = \{W_t,\, t \in [0, \infty)\}$ denote a standard Wiener process and $N = \{N_t,\, t \in [0, \infty)\}$ a Poisson process with intensity $\lambda > 0$. A typical scalar SDE for an asset price $X_t$ could then take the form

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t + c(t-, X_{t-})\, dN_t \tag{7.6.18}$$

for $t \in [0, \infty)$, with initial value $X_0 \in (0, \infty)$. Here $a(\cdot, \cdot)$ is called the drift coefficient and $b(\cdot, \cdot)$ the diffusion coefficient, which controls the magnitude of continuous fluctuations. Furthermore, $c(\cdot, \cdot)$ denotes the jump coefficient, which determines the jump size at an event. An example for such an SDE was given by (7.6.14) with the solution shown in Fig. 7.6.1. We call the resulting process $X$ a *jump diffusion* or an *Itô process with jumps*.

### Multi-Dimensional SDEs for Jump Diffusions (*)

To model a financial market, many sources of continuous and event driven uncertainty have to be taken into account. Furthermore, there are several interacting factors that need to play a role in a reasonably realistic financial market model. Let us denote by $W^1, \ldots, W^m$ independent standard Wiener processes. Furthermore, $N^1, \ldots, N^n$ denote $n$ Poisson processes with corresponding intensities $\lambda^1(t, \boldsymbol{X}_t), \ldots, \lambda^n(t, \boldsymbol{X}_t)$ at time $t$. These intensities depend on time and also on the vector $\boldsymbol{X}_t = (X_t^1, X_t^2, \ldots, X_t^d)^\top$ of factors. The model for the factors is conveniently described by an SDE. For the $i$th factor the SDE has the form

$$dX_t^i = a^i(t, \boldsymbol{X}_t)\, dt + \sum_{k=1}^{m} b^{i,k}(t, \boldsymbol{X}_t)\, dW_t^k + \sum_{\ell=1}^{n} c^{i,\ell}(t-, \boldsymbol{X}_{t-})\, dN_t^\ell \tag{7.6.19}$$

for $t \in [0, \infty)$ with $X_0^i \in \Re$ and $i \in \{1, 2, \ldots, d\}$. Obviously, the drift coefficients $a^i(\cdot, \cdot)$, diffusion coefficients $b^{i,k}(\cdot, \cdot)$ and jump coefficients $c^{i,\ell}(\cdot, \cdot)$ need to satisfy appropriate conditions. These must ensure the existence and uniqueness of the solution of the system of SDEs (7.6.19). In Sect. 7.7 we shall mention appropriate conditions that allow us to work with an SDE of the above type, see also Ikeda & Watanabe (1989).

### SDE of Exponential Lévy Models (*)

One direction for the generalization of Merton's jump diffusion model (7.6.11) is to use an exponential, as in (3.6.1), of the form

$$S_t = S_0 \exp\{X_t\}, \tag{7.6.20}$$

where $X = \{X_t,\, t \in [0, \infty)\}$ is a Lévy process, as defined in (3.6.2). This means, one considers the exponential of the expression

$$X_t = \alpha\, t + \beta\, W_t + \int_0^t \int_{|v|<1} v(p_\varphi(dv, ds) - \varphi(dv)\, ds) + \int_0^t \int_{|v|\geq 1} v\, p_\varphi(dv, ds) \tag{7.6.21}$$

for $t \in [0, \infty)$. Recall that $W$ is a standard Wiener process, $p_\varphi$ a Poisson measure with Lévy measure $\varphi(dv)$ so that (3.5.13) is satisfied. By application of the Itô formula (6.4.11) we obtain for the resulting *exponential Lévy model* the SDE

$$\begin{aligned} dS_t = S_{t-} \Bigg[ \left( \alpha + \frac{1}{2}\beta^2 - \int_{|v|<1} v\, \varphi(dv) \right) dt + \beta\, dW_t \\ + \int_{-\infty}^{\infty} (\exp\{v\} - 1)\, p_\varphi(dv, dt) \Bigg] \end{aligned} \tag{7.6.22}$$

for $t \in [0, \infty)$. We assume that all expressions on the right hand sides of (7.6.21) and (7.6.22) exist. Note that formula (7.6.20) together with (7.6.21) provide an explicit solution for the SDE (7.6.22) and, thus, for the asset price under the exponential Lévy model.

Asset price models of the type (7.6.21) have been proposed and studied by many authors, for instance, by Madan & Seneta (1990), Madan & Milne (1991), Eberlein & Keller (1995) and Barndorff-Nielsen & Shephard (2001) and Miyahara & Novikov (2002).

### General SDEs Driven by Jump Measures (*)

In the case when there are events with state dependent intensities that are related, which may impact several factors at the same time, it is appropriate to use SDEs driven by Poisson measures, as described in Sect. 3.5. To give an example, we extend the SDE (7.6.19) for jump diffusions and use instead of Poisson processes Poisson jump measures $p^{\ell}_{\varphi_\ell}(\cdot,\cdot)$, $\ell \in \{1,2,\ldots,n\}$. Leaving the other terms in the SDE (7.6.19) unchanged we arrive at a general SDE for the $i$th factor in the form

$$dX^i_t = a^i(t,\boldsymbol{X}_t)\,dt + \sum_{k=1}^{m} b^{i,k}(t,\boldsymbol{X}_t)\,dW^k_t + \sum_{\ell=1}^{n}\int_{\mathcal{E}} c^{i,\ell}(v,t-,\boldsymbol{X}_{t-})\,p^{\ell}_{\varphi_\ell}(dv,dt) \tag{7.6.23}$$

for $t \in [0,\infty)$ with $X^i_t \in \Re$, $i \in \{1,2,\ldots,d\}$. Here the $(i,\ell)$th jump coefficient is not only a function of time and factors but also depends on the mark $v \in \mathcal{E}$ as an element of the mark set $\mathcal{E} = \Re\backslash\{0\}$, see (3.5.11). It is important to satisfy condition (3.5.13), that is,

$$\int_{\mathcal{E}} \min(1,v^2)\,\varphi_\ell(dv) < \infty \tag{7.6.24}$$

for $\ell \in \{1,2,\ldots,n\}$. This condition guarantees the existence of the integrals under consideration. Similarly, as with the SDE for Lévy processes, one can model by the SDE (7.6.23) certain dynamics that involve infinitely many small jumps. Furthermore, the intensity for those jumps that impact the factors is controllable via the jump coefficients since these depend on the mark $v$. Even though we model the Poisson jump measure $p^{\ell}_{\varphi_\ell}$ in a standard way, its impact on the factors can depend in a very flexible manner on the actual values of all factors. It is clear that if we choose

$$\varphi_\ell(dv) = \mathbf{1}_{\{v\in[0,\lambda]\}}\,dv,$$

and the jump coefficients independent of the marks, then we recover the SDE (7.6.19) for jump diffusions. Such SDEs that are driven by jump measures have been used in financial modeling, for instance, in Björk, Kabanov & Runggaldier (1997) or Christensen & Platen (2005).

### Stochastic Exponentials of Semimartingales (*)

As we have already seen in Sect. 5.5 and Sect. 6.4, semimartingales form an extremely rich class of stochastic processes. We shall now mention a few results on solutions of SDEs for semimartingales. These SDEs generalize, in principle, all previously mentioned types of SDEs and permit the formulation of very general models and statements.

For simplicity, we consider the one-dimensional case. The corresponding multi-dimensional generalization is rather obvious, but requires cumbersome notation. Let $Z = \{Z_t,\, t \in [0,\infty)\}$ denote a one-dimensional semimartingale, see (5.5.1), where $Z^c$ denotes its continuous part, which can be expressed in the form

$$Z_t^c = Z_t - \sum_{0<s\le t} \Delta Z_s - Z_0 \tag{7.6.25}$$

if its discontinuous part $\sum_{0\le s\le t} \Delta Z_s$ is almost surely finite for all $t \in [0,\infty)$. Here we assume that $Z_0 \in \Re$ and use again our notation for the jump size

$$\Delta Z_t = Z_t - Z_{t-}$$

for a potential jump of $Z$ at time $t$. Consider now the stochastic process $X = \{X_t,\, t \in [0,\infty)\}$ with

$$X_t = \exp\{Z_t\} = X_0 \, \exp\{Z_t^c\} \prod_{0<s\le t} \exp\{\Delta Z_s\} \tag{7.6.26}$$

for $t \in [0,\infty)$, where we assume the product $\prod_{0<s\le t} \exp\{\Delta Z_s\}$ to be almost surely finite for all $t \in [0,\infty)$. The SDE for $X_t$ then follows by the Itô formula for semimartingales (6.4.11) in the form

$$dX_t = X_t\, dZ_t^c + \frac{1}{2}\, X_t\, d[Z_t^c] + X_{t-}\, (\exp\{\Delta Z_t\} - 1) \tag{7.6.27}$$

for $t \in [0,\infty)$ with initial value $X_0 = \exp\{Z_0\}$. Note that the last expression on the right hand side of (7.6.27) for the jumps changes the value of $X_{t-}$ by the jump ratio $\exp\{\Delta Z_t\}$ if there is a jump at time $t$.

In the above sense the expression (7.6.26), which we call the *stochastic exponential* of $Z$, provides an explicit solution for the semimartingale SDE (7.6.27). Note that the process $X = \{X_t,\, t \in [0,\infty)\}$ in (7.6.26) remains always strictly positive. In the stochastic exponential (7.6.26) the terms are similar to those usually observed for a geometric Brownian motion. Only the product of the exponentials of the jumps $\Delta Z_t$ appears as an extra factor. The exponentials of the jumps $\Delta Z_t$ represent the jump ratios for $X_t$, see (7.6.6). The expression (7.6.26) provides an explicit solution for the SDE (7.6.27). Note that for a strictly positive process $X$, satisfying an SDE of the form (7.6.27), we can identify the process $Z = \{Z_t,\, t \in [0,\infty)\}$ with the logarithm of $X$, that is

$$Z_t = \ln(X_t) = Z_0 + Z_t^c + \sum_{0<s\le t} \Delta Z_s \tag{7.6.28}$$

for $t \in [0,\infty)$ if $\sum_{0<s\le t} \Delta Z_s$ is almost surely finite.

## 7.7 Existence and Uniqueness (*)

For any model that uses an SDE it is essential that it has a solution. Furthermore, it is important that it has a unique solution according to some appropriate criterion. One such criterion is described below in detail, which is based on a notion of strong uniqueness. Usually one can only formulate sufficient conditions to establish uniqueness. The techniques presented in the literature for proving existence and uniqueness of a solution of an SDE are rather similar. They typically assume Lipschitz continuity of the drift and diffusion coefficients. We aim to provide here some insight into typical issues that arise when ensuring the existence and uniqueness of a solution of an SDE. We shall outline the typical arguments that are used when proving the existence and uniqueness of a solution of an SDE.

### On the Existence of Solutions of SDEs (*)

As before we work in a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. For simplicity, we focus on a scalar SDE of the form

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, dW_t \tag{7.7.1}$$

for $t \in [t_0, T]$ with fixed $T \in (t_0, \infty)$, $t_0 \in [0, \infty)$. Note that analogous results hold in the case of vector SDEs, as given in (7.4.1) and for SDEs with jumps as mentioned in Sect. 7.6. For more general SDEs the techniques are rather similar, see Protter (2004). The above Wiener process $W$ is assumed to be $\underline{\mathcal{A}}$-adapted and the increments $(W_t - W_s)$ are supposed to be independent of $\mathcal{A}_s$ for $t \in [t_0, T],\ s \in [t_0, t]$. Recall that the SDE (7.7.1) is a short hand notation for the Itô integral equation

$$X_t = X_{t_0} + \int_{t_0}^{t} a(s, X_s)\, ds + \int_{t_0}^{t} b(s, X_s)\, dW_s, \tag{7.7.2}$$

where the first integral is a random ordinary Riemann-Stieltjes integral and the second integral is an Itô integral. For (7.7.2) to make sense $X$ needs to be $\underline{\mathcal{A}}$-adapted. This leads us to the following definition.

**Definition 7.7.1.** *We call a pair $(X, W)$, consisting of a stochastic process $X = \{X_t,\ t \in [t_0, T]\}$ and an $\underline{\mathcal{A}}$-adapted standard Wiener process $W$, a* strong solution *of the Itô integral equation (7.7.2) if $X$ is $\underline{\mathcal{A}}$-adapted, the integrals on the right hand side are well-defined and the equality in (7.7.2) holds almost surely.*

As discussed previously, the mentioned integrals are, for instance, well-defined if $\sqrt{|a(\cdot, X_\cdot)|}$ and $b(\cdot, X_\cdot)$ belong to the set $\mathcal{L}_T^2$, see (5.4.1), for all $T \in (0, \infty)$.

### On the Uniqueness of Solutions of SDEs (*)

For fixed coefficient functions $a$ and $b$, any solution $X$ will usually depend on the particular initial value $X_{t_0}$ and the sample path of the Wiener process $W$

under consideration. For a specified initial value $X_{t_0}$ the *uniqueness* of strong solutions of the SDE (7.7.1) refers to the *indistinguishability*, see (3.1.6), of the solution processes.

**Definition 7.7.2.** *If any two strong solutions $X$ and $\tilde{X}$ are indistinguishable on $[t_0, T]$, that is if*

$$X_t = \tilde{X}_t \tag{7.7.3}$$

*a.s. for all $t \in [t_0, T]$, then we say that the solutions of (7.7.1) are* pathwise unique *on $[t_0, T]$. In general, we call a pathwise unique strong solution a* unique strong solution.

In some papers one considers the *uniqueness in law* for a given SDE. This kind of uniqueness arises if whenever $(X, W)$ and $(\tilde{X}, \tilde{W})$ are two solutions, which may be defined on different probability spaces with $X_{t_0} = \tilde{X}_{t_0}$ and the laws of $X$ and $\tilde{X}$ are equal. Such a solution is referred to as a *unique weak solution*.

The following result is due to Yamada & Watanabe (1971). It illustrates the advantage of having strong uniqueness for the solutions of an SDE.

**Theorem 7.7.3.** *Suppose that strong uniqueness holds for the solutions of the SDE (7.7.1). Then their uniqueness in law follows.*

### Example for a Unique Weak Solution (*)

Let $W = \{W_t,\, t \in [0, \infty)\}$ denote a Wiener process. The following well-known example, which is due to Tanaka (1963), is quite illustrative. For *Tanaka's SDE*

$$X_t = \int_0^t \operatorname{sgn}(X_s)\, dW_s \tag{7.7.4}$$

with

$$\operatorname{sgn}(x) = \begin{cases} 1 & \text{for} \quad x \geq 0 \\ -1 & \text{for} \quad x < 0 \end{cases}$$

strong uniqueness does not hold. Indeed, if $(X, W)$ is a solution of (7.7.4), then $(-X, W)$ is also a solution. Note however that one can show that (7.7.4) possesses a unique weak solution, see Revuz & Yor (1999). We emphasize that $X$ is here a Wiener process.

### Standard Assumptions (*)

Let us now formulate the typical hypotheses of an existence and uniqueness theorem. The coefficient functions $a$, $b : [t_0, T] \times \Re \to \Re$ are assumed to be given.

**(I)** *Measurability*: The coefficient functions $a$ and $b$ are assumed to be jointly $\mathcal{L}_T^2$-measurable in $(t, x) \in [t_0, T] \times \Re$.

**(II)** *Lipschitz condition*: There exists a finite constant $K > 0$ such that

$$|a(t,x) - a(t,y)| \le K\,|x - y|$$

and

$$|b(t,x) - b(t,y)| \le K\,|x - y|$$

for all $t \in [t_0, T]$ and $x,\ y \in \Re$.

**(III)** *Linear growth bound*: There exists a constant $K > 0$ such that

$$|a(t,x)|^2 \le K^2(1 + |x|^2)$$

and

$$|b(t,x)|^2 \le K^2(1 + |x|^2)$$

for all $t \in [t_0, T]$ and $x \in \Re$.

**(IV)** *Initial value*: $X_{t_0}$ is $\mathcal{A}_{t_0}$-measurable with $E(|X_{t_0}|^2) < \infty$.

Note that in the case of vector SDEs the above norms appear as Euclidean norms. If the drift function $a(\cdot,\cdot)$ has a bounded first derivative with respect to $x$, then it is obviously Lipschitz continuous. In particular, assumptions (II) and (III) ensure that a solution of an SDE does not explode.

## Gronwall Inequality (*)

To prove an existence and uniqueness result for the SDE (7.7.1) under the above conditions, typically the *Gronwall inequality* is exploited, which we state in the following lemma.

**Lemma 7.7.4.** *Let $\alpha, \beta : [t_0, T] \to \Re$ be integrable functions with the property*

$$0 \le \alpha(t) \le \beta(t) + L \int_{t_0}^{t} \alpha(s)\, ds$$

*for $t \in [t_0, T]$ and some fixed $L > 0$. Then*

$$\alpha(t) \le \beta(t) + L \int_{t_0}^{t} \exp\{L(t-s)\}\, \beta(s)\, ds$$

*for $t \in [t_0, T]$.*

## Strong Uniqueness (*)

Using the measurability assumption (I) and the Lipschitz condition (II) and assuming that solutions of (7.7.2) exist, one can show their strong uniqueness. This result is formulated in the following lemma.

**Lemma 7.7.5.** *If conditions (I), (II) and (IV) hold, then the solutions of* (7.7.2) *with sample paths corresponding to the same initial value and the same Wiener process are strongly unique.*

**Proof**: Let $X$ and $\tilde{X}$ be two solutions of (7.7.2) on $[t_0, T]$, which have continuous sample paths. Since they may not have finite second moments, we shall use the following localization method:
For $N > 0$ and $t \in [t_0, T]$ define

$$\mathbf{1}_t^{(N)} = \begin{cases} 1 & \text{for} \quad |X_u|, |\tilde{X}_u| \leq N \quad \text{and} \quad u \in [t_0, t] \\ 0 & \text{otherwise.} \end{cases}$$

Obviously, $\mathbf{1}_t^{(N)}$ is $\mathcal{A}_t$-measurable and $\mathbf{1}_t^{(N)} = \mathbf{1}_t^{(N)} \mathbf{1}_s^{(N)}$ for $s \in [t_0, t]$. Consequently, using the Lipschitz condition (II) the integrals in the equation

$$\begin{aligned} Z_t^{(N)} &= \mathbf{1}_t^{(N)} (X_t - \tilde{X}_t) \\ &= \mathbf{1}_t^{(N)} \int_{t_0}^t \mathbf{1}_s^{(N)} \left( a(s, X_s) - a(s, \tilde{X}_s) \right) ds \\ &\quad + \mathbf{1}_t^{(N)} \int_{t_0}^t \mathbf{1}_s^{(N)} \left( b(s, X_s) - b(s, \tilde{X}_s) \right) dW_s, \end{aligned} \tag{7.7.5}$$

are well-defined for $t \in [t_0, T]$. Applying again the Lipschitz condition (II), we obtain

$$\begin{aligned} \max &\left\{ \left| \mathbf{1}_s^{(N)} \left( a(s, X_s) - a(s, \tilde{X}_s) \right) \right|, \left| \mathbf{1}_s^{(N)} \left( b(s, X_s) - b(s, \tilde{X}_s) \right) \right| \right\} \\ &\leq K \, \mathbf{1}_s^{(N)} \left| X_s - \tilde{X}_s \right| \leq 2 \, K \, N \end{aligned} \tag{7.7.6}$$

for $s \in [t_0, t]$. Therefore, the second moment exists for $Z_t^{(N)}$ and thus also the two integrals in (7.7.5). Using (7.7.5), the inequality $(a + b)^2 \leq 2(a^2 + b^2)$, the inequality (1.4.65) with $r = 2$ and the correlation property (5.4.4) of an Itô integral we can write by applying Fubini's Theorem

$$\begin{aligned} E\left( \left| Z_t^{(N)} \right|^2 \right) &\leq 2E\left( \left| \int_{t_0}^t \mathbf{1}_s^{(N)} \left( a(s, X_s) - a(s, \tilde{X}_s) \right) ds \right|^2 \right) \\ &\quad + 2E\left( \left| \int_{t_0}^t \mathbf{1}_s^{(N)} (b(s, X_s) - b(s, \tilde{X}_s)) \, dW_s \right|^2 \right) \\ &\leq 2(T - t_0) \int_{t_0}^t E\left( \left| \mathbf{1}_s^{(N)} \left( a(s, X_s) - a(s, \tilde{X}_s) \right) \right|^2 \right) ds \\ &\quad + 2 \int_{t_0}^t E\left( \left| \mathbf{1}_s^{(N)} \left( b(s, X_s) - b(s, \tilde{X}_s) \right) \right|^2 \right) ds \end{aligned}$$

for $t \in [t_0, T]$. This estimate can be combined with (7.7.6) to obtain

$$E\left(\left|Z_t^{(N)}\right|^2\right) \leq L \int_{t_0}^{t} E\left(\left|Z_s^{(N)}\right|^2\right) ds \tag{7.7.7}$$

for $t \in [t_0, T]$, where $L = 2(T - t_0 + 1)K^2$.

Now, applying the Gronwall inequality given in Lemma 7.7.4 with $\alpha(t) = E(|Z_t^{(N)}|^2)$ and $\beta(t) \equiv 0$ we can conclude that

$$E\left(\left|Z_t^{(N)}\right|^2\right) = E\left(\left|\mathbf{1}_t^{(N)}\left(X_t - \tilde{X}_t\right)\right|^2\right) = 0,$$

for $t \in [t_0, T]$, and hence $\mathbf{1}_t^{(N)} X_t = \mathbf{1}_t^{(N)} \tilde{X}_t$ a.s. for each $t \in [t_0, T]$.

Since the sample paths are continuous they are also bounded for each $\omega \in \Omega$. This allows us to make the probability

$$P\left(\mathbf{1}_t^{(N)} \neq 1 \text{ for all } t \in [t_0, T]\right) \leq P\left(\sup_{t \in [t_0, T]} |X_t| > N\right) + P\left(\sup_{t \in [t_0, T]} \left|\tilde{X}_t\right| > N\right)$$

arbitrarily small by taking the truncation level $N$ sufficiently large. Therefore, since

$$\begin{aligned} P(X_t \neq \tilde{X}_t) &= P\left(\mathbf{1}_t^{(N)}(X_t - \tilde{X}_t) \neq 0\right) + P\left(\left(1 - \mathbf{1}_t^{(N)}\right)(X_t - \tilde{X}_t) \neq 0\right) \\ &\leq P\left(\mathbf{1}_t^{(N)}(X_t - \tilde{X}_t) \neq 0\right) + P(\mathbf{1}_t^{(N)} \neq 1), \end{aligned}$$

we obtain $P(X_t \neq \tilde{X}_t) = 0$ for each $t \in [t_0, T]$ and hence $P(X_t \neq \tilde{X}_t \text{ for all } t \in D) = 0$ for any countable dense subset $D$ of $[t_0, T]$.

Since the solutions are continuous and coincide on a dense subset of $[t_0, T]$, they must coincide a.s. on the entire interval $[t_0, T]$ that means they are indistinguishable, see (3.1.6). Thus (7.7.3) holds, that is the two solutions $X$ and $\tilde{X}$ of (7.7.2) are strong unique. □

### Existence and Uniqueness Theorem (*)

So far it has not been clarified whether the SDE (7.7.1) has a strong solution. The following result establishes this property and uses also the uniqueness, which we had established already.

**Theorem 7.7.6.** *Under conditions (I) - (IV), the SDE (7.7.1) has a unique strong solution $X = \{X_t, t \in [t_0, T]\}$ on $[t_0, T]$ with*

$$\sup_{t \in [t_0, T]} E\left(|X_t|^2\right) < \infty.$$

**Proof**:
**1.** Because of Lemma 7.7.5 we only have to establish the existence of a strong solution $X$ on $[t_0, T]$ for a given Wiener process $W = \{W_t,\, t \in [t_0, T]\}$. We shall do this by the *method of successive approximations*.

**2.** Define $X_t^{(0)} = X_{t_0}$ and recursively

$$X_t^{(n+1)} = X_{t_0} + \int_{t_0}^t a\left(s, X_s^{(n)}\right) ds + \int_{t_0}^t b\left(s, X_s^{(n)}\right) dW_s \tag{7.7.8}$$

for $n \in \{0, 1, \ldots\}$. If for a fixed $n \geq 0$ the approximation $X_t^{(n)}$ is $\mathcal{A}_t$-measurable and continuous on $[t_0, T]$, then it follows from assumptions (I), (II) and (III) that the integrals in (7.7.8) are well-defined and that the resulting process $X^{(n+1)}$ is $\underline{\mathcal{A}}$-adapted and can be chosen to be continuous on $[t_0, T]$. As $X_t^{(0)}$ is $\mathcal{A}_t$-measurable and continuous on $[t_0, T]$, it follows by induction that each $X_t^{(n)}$ for $n \in \mathcal{N}$ is also $\mathcal{A}_t$-measurable and continuous.

From assumption (IV) and the definition of $X_t^{(0)}$ it is clear that

$$\sup_{t \in [t_0, T]} E\left(\left|X_t^{(0)}\right|^2\right) < \infty.$$

Applying the inequality $(a+b+c)^2 \leq 3(a^2+b^2+c^2)$, the inequality (1.4.65) with $r = 2$, the identity (5.4.4) and the linear growth bound (III) to (7.7.8) we obtain

$$\begin{aligned} E\left(\left|X_t^{(n+1)}\right|^2\right) &\leq 3E\left(\left|X_{t_0}\right|^2\right) + 3E\left(\left|\int_{t_0}^t a\left(s, X_s^{(n)}\right) ds\right|^2\right) \\ &\quad + 3E\left(\left|\int_{t_0}^t b\left(s, X_s^{(n)}\right) dW_s\right|^2\right) \\ &\leq 3E\left(\left|X_{t_0}\right|^2\right) + 3(T-t_0)\, E\left(\int_{t_0}^t \left|a\left(s, X_s^{(n)}\right)\right|^2 ds\right) \\ &\quad + 3E\left(\int_{t_0}^t \left|b\left(s, X_s^{(n)}\right)\right|^2 ds\right) \\ &\leq 3E\left(\left|X_{t_0}\right|^2\right) + 3(T-t_0+1)K^2 E\left(\int_{t_0}^t \left(1 + \left|X_s^{(n)}\right|^2\right) ds\right) \end{aligned}$$

for $n \in \{0, 1, \ldots\}$. By induction this means that

$$\sup_{t \in [t_0, T]} E\left(\left|X_t^{(n)}\right|^2\right) \leq C_0 < \infty \tag{7.7.9}$$

for $n \in \mathcal{N}$, where $C_0$ is a constant that does not depend on $n$.

Using similar arguments to those that were applied to prove inequality (7.7.7) we can show that

$$E\left(\left|X_t^{(n+1)} - X_t^{(n)}\right|^2\right) \leq L \int_{t_0}^{t} E\left(\left|X_s^{(n)} - X_s^{(n-1)}\right|^2\right) ds \tag{7.7.10}$$

for $t \in [t_0, T]$ and $n \in \mathcal{N}$, where $L = 2(T - t_0 + 1)K^2$. Application of the Cauchy formula

$$\int_{t_0}^{t} \int_{t_0}^{t_{n-1}} \cdots \int_{t_0}^{t_1} f(s)\, ds\, dt_1 \ldots dt_{n-1} = \frac{1}{(n-1)!} \int_{t_0}^{t} (t-s)^{n-1} f(s)\, ds$$

in repeated iterations of the inequality (7.7.10), shows that

$$E\left(\left|X_t^{(n+1)} - X_t^{(n)}\right|^2\right) \leq \frac{L^n}{(n-1)!} \int_{t_0}^{t} (t-s)^{n-1} E\left(\left|X_s^{(1)} - X_s^{(0)}\right|^2\right) ds \tag{7.7.11}$$

for $t \in [t_0, T]$ and $n \in \mathcal{N}$.

Using the growth condition (III) instead of the Lipschitz condition (II) in the derivation of (7.7.10) for $n = 0$, we note that

$$\begin{aligned} E\left(\left|X_t^{(1)} - X_t^{(0)}\right|^2\right) &\leq L \int_{t_0}^{t} \left(1 + E\left(\left|X_s^{(0)}\right|^2\right)\right) ds \\ &\leq L\,(T - t_0)\left(1 + E\left(|X_{t_0}|^2\right)\right) = C_1. \end{aligned}$$

Substituting this result into (7.7.11) provides the estimate

$$E\left(\left|X_t^{(n+1)} - X_t^{(n)}\right|^2\right) \leq \frac{C_1 L^n (t - t_0)^n}{n!}$$

for $t \in [t_0, T]$ and $n \in \{0, 1, \ldots\}$ and therefore

$$\sup_{t \in [t_0, T]} E\left(\left|X_t^{(n+1)} - X_t^{(n)}\right|^2\right) \leq \frac{C_1 L^n (T - t_0)^n}{n!} \tag{7.7.12}$$

for $n \in \{0, 1, \ldots\}$. This result establishes in a mean square sense the convergence of the successive approximations on $[t_0, T]$.

**3.** To prove almost sure convergence of the sample paths of the successive approximations uniformly on $[t_0, T]$ we define

$$Z_n = \sup_{t \in [t_0, T]} \left|X_t^{(n+1)} - X_t^{(n)}\right|$$

for $n \in \{0, 1, \ldots\}$, such that, from (7.7.8)

$$Z_n \le \int_{t_0}^{T} \left| a(s, X_s^{(n)}) - a(s, X_s^{(n-1)}) \right| ds$$
$$+ \sup_{t\in[t_0,T]} \left| \int_{t_0}^{t} \left( b(s, X_s^{(n)}) - b(s, X_s^{(n-1)}) \right) dW_s \right|.$$

By application of the inequality (1.4.65) with $r = 2$, (5.4.4), (5.2.12) and the Lipschitz condition (II) we obtain the estimate

$$E\left(|Z_n|^2\right) \le 2\,(T - t_0)\,K^2 \int_{t_0}^{T} E\left(\left|X_s^{(n)} - X_s^{(n-1)}\right|^2\right) ds$$
$$+ 2\,K^2 \int_{t_0}^{T} E\left(\left|X_s^{(n)}) - X_s^{(n-1)}\right|^2\right) ds$$
$$\le 2\,(T - t_0 + 1)\,K^2 \int_{t_0}^{T} E\left(\left|X_s^{(n)} - X_s^{(n-1)}\right|^2\right) ds.$$

This result can be combined with (7.7.12) so that

$$E\left(|Z_n|^2\right) \le \frac{C_2\,L^{n-1}(T-t_0)^{n-1}}{(n-1)!} \tag{7.7.13}$$

for $n \in \mathcal{N}$, where $C_2 = 2C_1K^2(T-t_0+4)(T-t_0)$. Now, applying the Markov inequality (1.3.57) to each term and summing up, we have the inequality

$$\sum_{n=1}^{\infty} P\left(Z_n > \frac{1}{n^2}\right) \le C_2 \sum_{n=1}^{\infty} \frac{n^4}{(n-1)!} L^{n-1}(T-t_0)^{n-1},$$

where the series on the right hand side converges. Therefore the series on the left hand side also converges. Hence by the Borel-Cantelli Lemma, stated in Sect. 2.7, we can conclude that the $Z_n$ converge a.s. to zero. This means, the successive approximations $X_t^{(n)}$ converge almost surely uniformly on $[t_0, T]$ to the limit $\tilde{X}_t$ defined by

$$\tilde{X}_t = X_{t_0} + \sum_{n=0}^{\infty} \left\{ X_t^{(n+1)} - X_t^{(n)} \right\}. \tag{7.7.14}$$

**4.** We obtain from (7.7.9) that $\tilde{X}$ is mean square bounded on $[0,\infty)$. As the almost sure limit of $\underline{\mathcal{A}}$-adapted processes, $\tilde{X}$ is $\underline{\mathcal{A}}$-adapted. The uniform limit of continuous processes $\tilde{X}$ is also continuous. Taking this into account and also the growth condition (III), the right-hand side of the integral equation (7.7.2) is well defined if we replace the process $X$ with $\tilde{X}$. It remains to prove that it then equals the left hand side of (7.7.2). However, taking the almost sure limit as $n \to \infty$ on both sides of (7.7.8) it can be shown that $\tilde{X}$ is indeed a solution of (7.7.2).

To see this, note that the left hand side of (7.7.8) converges a.s. to $\tilde{X}_t$ uniformly on $[t_0, T]$. Comparing the right hand sides of (7.7.2) and (7.7.8), we obtain by the Lipschitz condition (II) the estimates

$$\left| \int_{t_0}^{t} a(s, X_s^{(n)})\, ds - \int_{t_0}^{t} a(s, \tilde{X}_s)\, ds \right| \leq K \int_{t_0}^{t} \left| X_s^{(n)} - \tilde{X}_s \right| \, ds \tag{7.7.15}$$

and

$$\int_{t_0}^{t} \left| b(s, X_s^{(n)}) - b(s, \tilde{X}_s) \right|^2 \, ds \leq K^2 \int_{t_0}^{t} \left| X_s^{(n)} - \tilde{X}_s \right|^2 \, ds, \tag{7.7.16}$$

which both converge a.s. to zero as $n \to \infty$ for each $t \in [t_0, T]$. This implies that we have the limits

$$\lim_{n\to\infty} \int_{t_0}^{t} a(s, X_s^{(n)})\, ds \stackrel{\text{a.s.}}{=} \int_{t_0}^{t} a(s, \tilde{X}_s)\, ds$$

and

$$\lim_{n\to\infty} \int_{t_0}^{t} b(s, X_s^{(n)})\, dW_s \stackrel{\text{a.s.}}{=} \int_{t_0}^{t} b(s, \tilde{X}_s)\, dW_s$$

for each $t \in [t_0, T]$. By choosing an appropriate subsequence we have the last limit also converging a.s. for this subsequence, see Sect. 2.7. Therefore, the right-hand side of (7.7.8) converges a.s. to the right-hand side of (7.7.2). Thus, the limit process $\tilde{X}$ satisfies the stochastic integral equation (7.7.2) a.s. □

Note that with the exception of the square integrability of $X$, the main result of Theorem 7.7.6 may remain valid if we impose certain weaker assumption instead of the square integrability of the initial value $X_{t_0}$.

### Yamada Condition (*)

Several financial models that are of practical interest have factors that satisfy SDEs, which do not have Lipschitz continuous coefficient functions. An example of such an SDE is given by the square root process, see Sect. 4.4, which appears, for instance, in the CIR model and the MMM, see Sect. 4.3 and Sect. 7.5. Its diffusion coefficient function, which is the square root of the process value itself, has infinite slope at zero. Consequently, the Lipschitz condition (II) is not satisfied. Another example, where the Lipschitz condition may fail is the CEV model, see (4.3.12).

It is possible to replace the Lipschitz condition (II) on the diffusion coefficient $b$ by the weaker *Yamada condition*, which assumes that there exists an increasing function $\varrho : [0, \infty) \to \Re$ with

$$\varrho(0) = 0 \quad \text{and} \quad \int_{0}^{\varepsilon} \varrho^{-2}(u)\, du = +\infty$$

for any $\varepsilon > 0$ such that

$$|b(t,x) - b(t,y)| \leq \varrho\left(|x-y|\right) \tag{7.7.17}$$

for all $x$, $y \in \Re$ and $t \in [t_0, T]$. This condition is sufficient to establish an existence and uniqueness theorem for solutions of corresponding SDEs. For example, $\varrho$ may be given by $\varrho(u) = u^\alpha$ for $\alpha \in [\frac{1}{2}, 1]$ or $\varrho(u) = \sqrt{|u \ln(u)|}$. With the first of these choices, for instance, it can be concluded that the SDE

$$dX_t = |X_t|^\alpha \, dW_t \tag{7.7.18}$$

has a pathwise unique solution for $X_{t_0} > 0$ when $\alpha \in [\frac{1}{2}, 1)$. Counterexamples exist which show that this is not the case for $\alpha < \frac{1}{2}$. Making zero an absorbing state for $X$ leads to a unique solution for $\alpha \in (0, \frac{1}{2})$.

Note that the CEV model, see (4.3.11) and (4.3.12), has an SDE similar to the above SDE (7.7.18). Therefore, one has to be careful in constructing an asset price model which relies on a diffusion coefficient that resembles a power of the asset price itself. In particular, it appears to be critical when the exponent $\alpha$ is below one half. Using the Yamada condition, the existence and uniqueness of the solution of the SDE (7.5.15) for the square root process can be shown. Corresponding proofs can be found in Karatzas & Shreve (1991), Ikeda & Watanabe (1989) or Cherny (2000). We shall see in Sect. 8.7 that SR processes are transformations of *squared Bessel processes*.

It is not trivial to establish the existence and uniqueness of the solution of an SDE for a *Bessel process* $Z = \{Z_t,\, t \in [0, \infty)\}$ of dimension $\delta > 1$. A Bessel process is the square root of a squared Bessel process, see Sect. 8.7. The SDE of $Z_t$ is of the form

$$dZ_t = \frac{\delta - 1}{2\, Z_t} \, \mathbf{1}_{\{Z_t \neq 0\}} \, dt + dW_t \tag{7.7.19}$$

for $t \in [0, \infty)$ with $Z_0 > 0$. In Cherny (2000) the following results are shown.

**Lemma 7.7.7.** (Cherny)

*(i) For $\delta \geq 2$ the SDE (7.7.19) has a unique strong solution.*
*(ii) If $\delta \in (1, 2)$ or $Z_0 = 0$, then there exist other strong solutions of the SDE (7.7.19) with the same $Z_0$ and Wiener process $W = \{W_t,\, t \in [0, \infty)\}$.*

This result indicates that one has to be careful when describing a stochastic process via some SDE without properly checking the existence and uniqueness of its solutions. For instance, one can show that the squared Bessel process, which corresponds to the square of the solution of (7.7.19), has an SDE with a unique strong solution. However, the square root of this process, that is the Bessel process itself, satisfies an SDE where the uniqueness of its strong solutions breaks down for a range of dimensions and some initial value. This is not a purely technical issue. It is highly relevant in practice because Bessel processes arise naturally as factors in financial modeling, as we shall see later.

The above discussion also indicates that it is preferable to work, when possible, in the modeling of financial quantities with squared Bessel processes as underlying factors instead of Bessel processes.

## 7.8 Markovian Solutions of SDEs (*)

For many index, equity, foreign exchange rate and interest rate models the solutions of the corresponding SDEs are *Markov processes*. This is an important property since it allows the application of powerful analytical tools to solutions and functionals of these types of SDEs.

### Underlying SDE (*)

To illustrate the Markov property of solutions of SDEs consider the scalar SDE

$$dX_t = a(t, X_t)\,dt + b(t, X_t)\,dW_t \tag{7.8.1}$$

for $t \in [t_0, T]$, with initial value $X_{t_0} = x_0$, where $W$ denotes a standard Wiener process.

Under the standard assumptions (I) - (IV) in Sect. 7.7 one can prove that the solution of the SDE (7.8.1) is a diffusion process, which is a Markov process. That is, its transition density satisfies the conditions (4.3.1)–(4.3.3) for the drift and diffusion coefficients. For the corresponding proof and for applications it is helpful to have some basic estimates for the higher order moments of solutions of the above SDE, which we summarize below.

### Bounds on Higher Order Moments (*)

The following lemma provides an example of some useful upper estimates for the higher even moments of the solution of the SDE (7.8.1).

**Lemma 7.8.1.** *Suppose that the Standard Assumptions (I)–(IV) of Sect. 7.7 hold and that the 2nth initial moment*

$$E\left(|X_{t_0}|^{2n}\right) < \infty$$

*is finite for some integer $n \geq 1$. Then the 2nth moment of the solution $X$ of (7.8.1) at time $t \in [t_0, T]$ satisfies the inequality*

$$E\left(|X_t|^{2n}\right) \leq \left(1 + E\left(|X_{t_0}|^{2n}\right)\right) \exp\{C\,(t - t_0)\} \tag{7.8.2}$$

*and also the estimate*

$$E\left(|X_t - X_{t_0}|^{2n}\right) \leq D\left(1 + E\left(|X_{t_0}|^{2n}\right)\right)(t - t_0)^n \exp\{C\,(t - t_0)\}, \tag{7.8.3}$$

*where $T < \infty$, $C = 2n(2n+1)K^2$ and $D$ is a positive constant depending only on $n$, $K$ and $T - t_0$.*

Here the constant $K$ appears in the standard conditions (II) and (III). Consequently, if the initial value $X_{t_0}$ is a constant, then this lemma implies the existence of all higher order moments. The proof of the above lemma uses techniques similar to those applied in the proof of Theorem 7.7.6, see Kloeden & Platen (1999).

### Solutions of SDEs as Diffusion Processes (*)

We now formulate a theorem that characterizes the solution of the SDE (7.8.1) as a diffusion process, see Sect. 4.3. The conditions (II) and (III) and the additional assumption that the drift and diffusion coefficient functions are continuous in time imply condition (I), which already follows for time homogeneous drift and diffusion coefficient functions from the Lipschitz condition (II). Provided a unique strong solution of the SDE (7.8.1) exists, the assumptions needed to prove the result below can be considerably weakened.

**Theorem 7.8.2.** *Assume that the coefficient functions $a$ and $b$ are continuous and conditions (II) and (III) of the Standard Assumption in Sect. 7.7 hold. Then for any fixed initial value $X_{t_0}$ the solution $X$ of (7.8.1) is a diffusion process on $[t_0, T]$ with drift coefficient $a(t,x)$ and diffusion coefficient $b(t,x)$.*

### Proof of Theorem 7.8.2 (*)

Let us denote by

$$X^{s,x} = \{X_t^{s,x},\, t \in [s,T]\}$$

a stochastic process that starts at time $s \in [t_0, T]$ with initial value

$$X_s^{s,x} = x.$$

Under the assumption of Theorem 7.7.6 one can show that for $t \in [t_0, T]$ and $s \in [t_0, t]$ we have that

$$X_t^{t_0,x} = X_t^{s, X_s^{s,x}} \tag{7.8.4}$$

a.s., see Rogers & Williams (2000), which encapsulates the Markov property.

**1.** From the inequality (7.8.3) with $n = 2$ we have

$$E\left(|X_t^{s,x} - x|^4\right) \leq C\,|t-s|^2$$

for $t \in [t_0, T]$, $s \in [t_0, t]$ and some constant $C$, which depends on $t_0$, $T$ and $x$. Hence, for any $\varepsilon > 0$ the transition density $p(s,x;t,y)$ satisfies the inequalities

$$\int_{|y-x|>\varepsilon} p(s,x;t,y)\,dy \leq \varepsilon^{-4} \int_{\Re} |y-x|^4\, p(s,x;t,y)\,dy \leq C\,\varepsilon^{-4}\,|t-s|^2,$$

so that

$$\lim_{t\downarrow s}\frac{1}{t-s}\int_{|y-x|>\varepsilon} p(s,x;t,y)\,dy=0,$$

which is condition (4.3.1) in Sect. 4.3.

**2.** To verify the other two limits it suffices to show that

$$\lim_{t\downarrow s}\frac{1}{t-s}\,E\left(X_t^{s,x}-x\right)=a(s,x) \tag{7.8.5}$$

and

$$\lim_{t\downarrow s}\frac{1}{t-s}\,E\left(\left|X_t^{s,x}-x\right|^2\right)=b^2(s,x). \tag{7.8.6}$$

Taking the expectation on both sides of the integral version of equation (7.8.1), and using the martingale property (5.4.3) of Itô integrals, we obtain

$$\begin{aligned} E\left(X_t^{s,x}-x\right) &= E\left(\int_s^t a(u,X_u^{s,x})\,du\right)\\ &= (t-s)\,E\left(\int_0^1 a\Big(s+v(t-s),X_{s+v(t-s)}^{s,x}\Big)\,dv\right). \end{aligned} \tag{7.8.7}$$

Since the sample paths of $X^{s,x}$ are a.s. continuous and the function $a$ is continuous, then

$$\lim_{t\downarrow s} a\Big(s+v(t-s),X_{s+v(t-s)}^{s,x}\Big)\overset{\text{a.s.}}{=}a(s,x).$$

**3.** Furthermore, from the growth condition (III) it can be concluded that

$$\left|a\Big(s+v(t-s),X_{s+v(t-s)}^{s,x}\Big)\right|^2\le K^2\left(1+\left|X_{s+v(t-s)}^{s,x}\right|^2\right)$$

a.s. Combining this with the mean-square boundedness of the solution $X^{s,x}$, see (7.8.2), we have

$$E\left(\int_0^1\left|a\left(s+v(t-s),X_{s+v(t-s)}^{s,x}\right)\right|^2\,dv\right)<\infty.$$

Hence by Fubini's theorem and using the growth condition (II) and (7.8.2) we can interchange in (7.8.7) the order of integration to conclude with the Dominated Convergence Theorem that

$$\lim_{t\downarrow s}E\left(\int_0^1 a\Big(s+v(t-s),X_{s+v(t-s)}^{s,x}\Big)\,dv\right)$$

$$= \lim_{t \downarrow s} \int_0^1 E\left(a\Big(s + v(t-s), X^{s,x}_{s+v(t-s)}\Big)\right) dv$$

$$= \int_0^1 \lim_{t \downarrow s} E\left(a\Big(s + v(t-s), X^{s,x}_{s+v(t-s)}\Big)\right) dv$$

$$= \int_0^1 a(s,x)\, dv \;=\; a(s,x)$$

from which (7.8.5) then follows.

**4.** The remaining limit (7.8.6) can be established in a similar way. However, here one should first use the Itô formula to obtain a stochastic integral equation for $(X^{s,x}_t - x)^2$. $\square$

## 7.9 Exercises for Chapter 7

**7.1.** Compute the mean and variance as functions of time of the standard Ornstein-Uhlenbeck process with initial value $X_{t_0} = 1$.

**7.2.** Determine the mean and the variance as functions of time for a geometric Brownian motion with appreciation rate $a = 0.05$, volatility $\sigma = 0.2$ and initial value $X_0 = 1$.

**7.3.** Solve explicitly the scalar SDE

$$dX_t = -\frac{1}{2}\, X_t\, dt + X_t\, dW^1_t + X_t\, dW^2_t,$$

where $W^1$ and $W^2$ are independent standard Wiener processes.

**7.4.** (*) Let $N = \{N_t,\, t \in [0,\infty)\}$ be a Poisson process with intensity $\lambda > 0$ and $W$ a standard Wiener process. For the process $Z = \{Z_t,\, t \in [0,\infty)\}$ with

$$Z_t = a\, t + b\, W_t + c\, N_t$$

compute the SDE for $X = \{X_t = \exp\{k\, Z_t\},\, t \in [0,\infty)\}$ for $a, b, c, k > 0$.

**7.5.** (*) For the process $X$ in Exercise 7.4 compute the expectation

$$u(t) = E(X_t).$$

**7.6.** (*) Show the explicit solution of the SDE

$$dX_t = (a_1\, X_t + a_2)\, dt + (b_1\, X_t + b_2)\, dW_t$$

for $t \in [0,\infty)$ with $X_0 > 0$ and prove that this solution solves the above SDE.

# 8

# Introduction to Option Pricing

In the previous chapters we have prepared mathematical tools that allow us to model in continuous time the dynamics of financial securities, for instance, stocks. Now, we shall study prices of derived financial securities. A *derivative security*, for instance an option, is a financial instrument whose value is dependent upon the values of an underlying more fundamental security. In this chapter we give an introduction into derivatives, in particular, European options. For simplicity, we focus our discussion on options under the BS model. Furthermore, we introduce at the end of the chapter important results on squared Bessel processes because these will be crucial for the understanding of the following chapters.

## 8.1 Options

Options have been introduced to provide some optionality to the buyer or seller of a security. In the simplest case the holder of an *option* has the right but not the obligation to buy or sell an underlying security for an agreed price at a preset date. We discuss now options as a particular type of derivative to highlight important general features of derivative securities.

### European Call Option

Let us denote by $S_t$, the price of a security at time $t \in [0, \infty)$, measured in units of the domestic currency. This can be, for instance, a stock index. We call $S = \{S_t,\, t \in [0, \infty)\}$ the price process of the underlying security. A *European call option* on an underlying security $S$ gives the owner the *right to buy* the security at a preset *strike price* $K$ at the *expiration date* $T \in (0, \infty)$. The price at time $t$ for this right is the *European call option price* $c_{T,K}(t, S_t)$, which is paid when the option contract is entered at time $t$. Note that there is an initial payment at the time when the contract is signed. An *American option* has the same payoff function as a European option. However, the holder has the

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

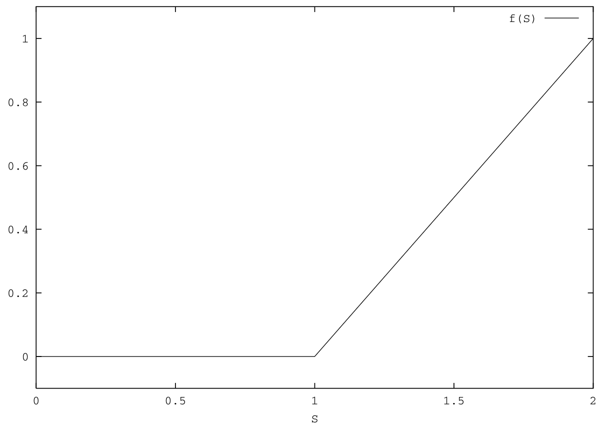



**Fig. 8.1.1.** Payoff function of a European call option for $K = 1$

right to exercise it at any time before the maturity date. Figure 8.1.1 shows the payoff function of a *call option*

$$H(S) = (S - K)^+ \tag{8.1.1}$$

with strike price $K = 1$, where we use the notation $a^+ = \max(0, a)$.

A European call option with expiration date $T \in (0, \infty)$ is at time $t \in [0, T]$ said to be *in-the-money*, *at-the-money* or *out-of-the-money*, if $S_t > K$, $S_t = K$ or $S_t < K$, respectively. The function

$$H(S_t) = (S_t - K)^+ \tag{8.1.2}$$

is called the *intrinsic value* of the call option at time $t \in [0, T]$.

As an example, consider a European call option at the beginning of 1995 on the S&P500 index, displayed in Fig. 3.1.1, with a strike price of $K = 400$ and expiration date at the end of 1995. Figure 3.1.1 shows that the S&P500 was at the end of 1995 approximately at \$500. This means that the value of the option was at the end of 1995 at a level of about \$100. We shall see from the theoretical pricing formulas presented in this chapter that the realized payoff of about \$100 would have considerably exceeded the original price of the option at the beginning of 1995. Of course, if the S&P500 stayed below \$400 during 1995, then the owner of the call option would have received nothing and would have lost the original option price that he or she paid when the option contract was written. This shows that there is substantial leverage involved when using options.

### European Put Option

For market participants who aim to sell an underlying security at a future date, the purchase of a *European put option* might be of advantage. This

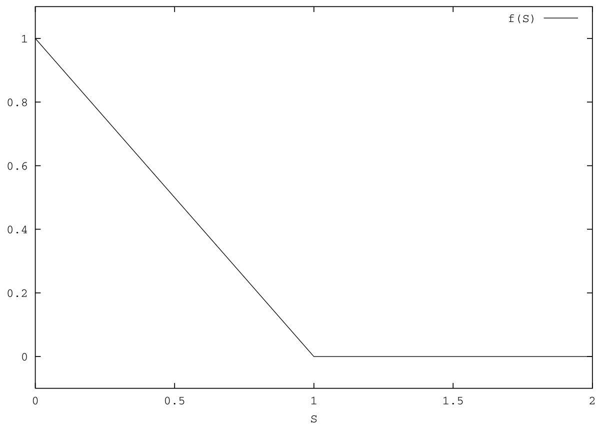



**Fig. 8.1.2.** Payoff function of a European put option, $K = 1$

financial contract is similar to the European call option but gives the holder the *right to sell* a security for a specified strike price $K$ at an expiration date $T$. We denote the *European put option price* by $p_{T,K}(t, S_t)$. Figure 8.1.2 displays the payoff function of a European put option

$$H(S_T) = (K - S_T)^+ \tag{8.1.3}$$

with strike price $K = 1$. A European put option is at time $t \in [0, T]$ *in-the-money*, *at-the-money* or *out-of-the money* if $S_t < K$, $S_t = K$ or $S_t > K$, respectively. The quantity

$$H(S_t) = (K - S_t)^+ \tag{8.1.4}$$

is called the *intrinsic value* of a put option at time $t \in [0, T]$.

It is important to specify whether someone is the owner of an underlying security or derivative. A market participant is *long* in a security, if he or she is the owner of that security. On the other hand, one is *short* in a security if one borrows it, sells it and has the obligation of giving it back at a later date. Owning a negative unit of a security is therefore possible through the practice of *short-selling*.

### Combinations of European Put and Call Options

To implement special hedging or speculative trading strategies it is common to form portfolios that consist of combinations of European call and put options. As an example, a *butterfly spread* is constructed by buying a call with strike price $K_1$, selling two calls with strike price $K_2 > K_1$ and buying another call with strike price $K_3 > K_2$. Figure 8.1.3 shows the resulting payoff function $H(S)$ of a butterfly spread with $K_1 = 0.6$, $K_2 = 1$, $K_3 = 1.4$. The butterfly spread has zero payoff outside the interval $[K_1, K_3]$. It allows to create at maturity a cash flow when the underlying security is near the strike price.

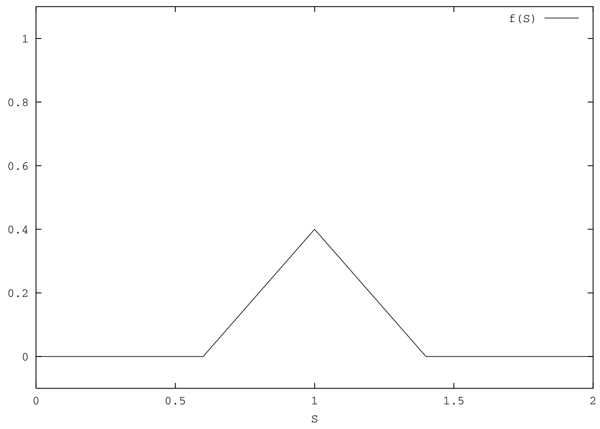



**Fig. 8.1.3.** Payoff of a butterfly spread

Theoretically one can approximate almost any reasonable payoff function at a given expiration date by portfolios of European calls and puts because a corresponding portfolio of butterfly spreads can concentrate a desired payoff close to each possible value of the underlying security.

## Options

More generally, we call a derivative a *European option* if it gives the right to realize a given payoff according to a given function $H : [0, \infty) \to \Re$ of the underlying $S_T$ at a specified expiration date $T \in [0, \infty)$. If the payoff can be exercised on or before the expiration date, then the contract is called an *American option*. The call and put options, introduced previously, are examples of European options. An American option is, in general, more expensive than a corresponding European option because it provides additionally the right to exercise early. One can show that the price of an American call option on an underlying security that pays no dividend is the same as its European counterpart.

In the following we denote by $V(t, S_t)$ the *value* at time $t \in [0, T]$ of a European option with payoff function $H$ and maturity date $T \in [0, \infty)$. Here $H$ has to fulfill some integrability condition which we do not specify at this stage. The *pricing function* $V : [0, T) \times [0, \infty) \to \Re$ for a European option can, in general, be shown to be differentiable with respect to time and twice differentiable with respect to the underlying security. This smoothness property will be exploited later for its computation. The efficient evaluation of this function is of importance both for the pricing and the hedging of these contracts. We shall show later that in certain cases explicit pricing formulas are available. However, in general, one needs to apply numerical methods.

## 8.2 Options under the Black-Scholes Model

We now consider options for the particular dynamics of the Black-Scholes (BS) model for the underlying security.

### Black-Scholes Model

For simplicity, let us use the BS model, see Sect. 7.5, as a description for the dynamics of the underlying security. It has been established historically as the *standard market model* for option pricing, see Black & Scholes (1973). This model supposes that the underlying security price $S = \{S_t,\, t \in [0,T]\}$ follows a geometric Brownian motion, see (6.3.6), with time dependent, deterministic appreciation rate $a = \{a_t,\, t \in [0,T]\}$ and strictly positive, deterministic volatility $\sigma = \{\sigma_t,\, t \in [0,T]\}$, that is

$$dS_t = a_t\, S_t\, dt + \sigma_t\, S_t\, dW_t \tag{8.2.1}$$

for $t \in [0,T]$ with given initial value $S_0 > 0$. Here $W$ denotes a standard Wiener process $W = \{W_t,\, t \in [0,T]\}$. Furthermore, there is a domestic savings account $B = \{B_t,\, t \in [0,T]\}$, which accrues the deterministic interest $r = \{r_t,\, t \in [0,T]\}$. We assume

$$dB_t = r_t\, B_t\, dt \tag{8.2.2}$$

for $t \in [0,T]$ with initial value

$$B_0 = 1. \tag{8.2.3}$$

The domestic savings account is also called the *locally riskless* asset since there is no noise term in its differential equation (8.2.2). Typically, in the standard BS model one sets the volatility $\sigma$, the appreciation rate $a$ and the short rate $r$ to be constant, which yields the basic model for option pricing. In the following analysis we typically allow these parameters to be time dependent. We shall later show in Sect. 10.6 that the savings account can be defined more precisely as a limit of a roll-over short term bond account.

### Hedge Portfolio

For the following let us fix the maturity date at $T$. From the practical point of view it is most important to realize that the writer of a European option can *replicate* the payoff $H(S_T)$ at the expiration date $T$. To achieve this, a *hedge portfolio* has to be established, which consists at time $t$ of $\delta_t^1$ units of the underlying security $S_t$ and $\delta_t^0$ units of the domestic savings account $B_t$. At time $t \in [0,T]$ the value of this portfolio is then set to the value $V(t, S_t)$ of the option. That is, the hedge portfolio has the value

$$V(t, S_t) = \delta_t^0\, B_t + \delta_t^1\, S_t \tag{8.2.4}$$

at time $t \in [0,T]$. By the Itô formula (6.4.11) we obtain

$$dV(t,S_t) = \delta_t^0\, dB_t + \delta_t^1\, dS_t + B_t\, d\delta_t^0 + S_t\, d\delta_t^1 + d[\delta^1, S]_t \tag{8.2.5}$$

at time $t \in [0,T]$.

### Self-Financing Portfolios

We assume that the hedge portfolio is *self-financing*. This means that all changes in the value of the portfolio are caused by gains from trade, that is, by changes in the savings account $B$ and the underlying security $S$. We can express the self-financing property of the portfolio $V(t,S_t)$ in differential form by assuming the SDE

$$dV(t,S_t) = \delta_t^0\, dB_t + \delta_t^1\, dS_t \tag{8.2.6}$$

for $t \in [0,T]$. Note that by (8.2.6) and (8.2.5) for the above hedge portfolio to be self-financing we have to satisfy the condition

$$B_t\, d\delta_t^0 + S_t\, d\delta_t^1 + d[\delta^1, S]_t = 0 \tag{8.2.7}$$

for all time $t \in [0,T]$.

We call the process $\delta = \{\delta_t = (\delta_t^0, \delta_t^1)^\top,\ t \in [0,T]\}$ a *self-financing strategy* if $\delta^0 = \{\delta_t^0,\ t \in [0,T]\}$ and $\delta^1 = \{\delta_t^1,\ t \in [0,T]\}$ are predictable processes and both are such that the hedge portfolio, whose value is given in (8.2.4), satisfies (8.2.6). We say that the hedge portfolio *replicates* the payoff $H(S_T)$ at the expiration date $T$, if

$$V(T,S_T) = H(S_T). \tag{8.2.8}$$

Furthermore, we need to assume the existence of the involved gains from trade or, equivalently, the corresponding Itô integrals. For our setup it is sufficient to assume that $\delta^1(\cdot)\sigma(\cdot)S(\cdot)$, $\sqrt{\delta^1(\cdot)\, a(\cdot)\, S(\cdot)}$ and $\sqrt{|\delta^0(\cdot)\, r(\cdot)\, B(\cdot)|}$ are in $\mathcal{L}_T^2$, see (5.4.1). Note however, for other models one may require weaker integrability conditions. Without further mentioning, we consider in the following only self-financing portfolios and strategies and omit the phrase self-financing.

We allow the hedge portfolio to be rebalanced continuously. Furthermore, we assume, for simplicity, that there are no additional costs, such as transaction costs, involved in hedging. One typically characterizes this setup as *continuous hedging* in a *frictionless market*.

### Discounted Value Function

To identify in a simple way an appropriate hedging strategy it is convenient to consider the corresponding *discounted value function* $\bar{V} : [0,T] \times [0,\infty) \to [0,\infty)$ given by

$$\bar{V}(t, \bar{S}_t) = \frac{V(t, S_t)}{B_t} \tag{8.2.9}$$

and the *discounted underlying security*

$$\bar{S}_t = \frac{S_t}{B_t} \tag{8.2.10}$$

for $t \in [0, T]$. By the Itô formula (6.2.11) we obtain from (8.2.1) and (8.2.2) the SDE

$$d\bar{S}_t = (a_t - r_t)\, \bar{S}_t\, dt + \sigma_t\, \bar{S}_t\, dW_t \tag{8.2.11}$$

for $t \in [0, T]$ with $\bar{S}_0 = S_0$. By discounting with the savings account one is taking the time value of money into account. This is extremely important for an investor who always can invest into the locally riskless asset, the savings account $B$. In this sense it is understandable when investors prefer to denominate a security in units of the savings account instead of denominating it in units of the currency.

### Profit and Loss Process

A hedger who has an option in her or his trading book faces at time $t$ a *profit and loss* (P&L) that is denoted by $C_t$ for $t \in [0, T]$. The ultimate goal of the hedger is to achieve zero P&L throughout the hedge. Then the selling of options and hedging these becomes ideally a riskless business.

To take for the P&L the time value of money into account, we consider the *discounted profit and loss*

$$\bar{C}_t = \frac{C_t}{B_t} \tag{8.2.12}$$

at time $t$. For a given strategy $\delta$ the discounted P&L $\bar{C}_t$ at time $t \in [0, T]$ is obtained as the corresponding discounted value of the hedge portfolio minus the discounted gains from trade and minus the initial value of the discounted portfolio. It can be written in the form

$$\bar{C}_t = \bar{V}(t, \bar{S}_t) - I_{\delta^1, \bar{S}}(t) - \bar{V}(0, \bar{S}_0) \tag{8.2.13}$$

for $t \in [0, T]$. Here we use the gains from trade $I_{\delta^1, \bar{S}}$, see (5.3.11), with respect to the discounted security $\bar{S}$, which according to (8.2.11) is of the form

$$I_{\delta^1, \bar{S}}(t) = \int_0^t \delta_u^1\, d\bar{S}_u = \int_0^t \delta_u^1\, (a_u - r_u)\, \bar{S}_u\, du + \int_0^t \delta_u^1\, \sigma_u\, \bar{S}_u\, dW_u \tag{8.2.14}$$

for $t \in [0, T]$. Obviously, with respect to the constant discounted domestic savings account

$$\bar{B}_t = 1 \tag{8.2.15}$$

there is zero gains from trade $I_{\delta^0, \bar{B}}(t) = 0$ for $t \in [0, T]$.

When the option contract is established at time $t = 0$, then the hedger receives from the buyer of the option the payment $V(0, S_0)$. This is equivalent to the discounted value

$$\bar{V}(0, \bar{S}_0) = \frac{V(0, S_0)}{B_0},$$

see (8.2.9) and (8.2.3). Thus, we have according to (8.2.13) and (8.2.14) zero initial discounted P&L

$$\bar{C}_0 = 0. \tag{8.2.16}$$

The discounted P&L $\bar{C}_t$ is then the actual discounted portfolio value that a hedger holds at time $t$.

### No-Arbitrage for P&L Process

Now, let us discuss some notion of arbitrage, which is fundamental for the modeling of financial markets. If a market participant is able to generate by her or his nonnegative total portfolio of investable securities some strictly positive wealth out of nothing, then this is interpreted as *arbitrage*. Any reasonable financial market model should avoid the modeling of arbitrage. We shall introduce a precise definition of arbitrage later in Sect. 10.2. At the present introductory level we call it an arbitrage if the market model allows to form a nonnegative portfolio that starts at zero and attains with strictly positive probability a strictly positive value at some later date. The nonnegativity of the portfolio reflects the *limited liability* of each investor for her or his total portfolio of investable wealth.

By excluding arbitrage a hedger can run a nonnegative hedge book with zero total initial value only such that its value remains always zero. This means that the P&L process of this business starts at zero and remains at zero all the time. Therefore, we aim to identify under no arbitrage a hedging strategy $\delta$ for which the discounted P&L remains zero, that is,

$$\bar{C}_t = 0 \tag{8.2.17}$$

for all $t \in [0, T]$. We call this a *perfect hedge* and the corresponding hedge portfolio $V = \{V(t, S_t),\ t \in [0, T]\}$ that returns the payoff at maturity $T$ is then a replicating portfolio.

### Discounted P&L Increments

We shall now demonstrate how an appropriate hedging strategy $\delta$ can be constructed. For this purpose we examine the increments of the discounted P&L process $\bar{C}$. With the definition of the discounted P&L given in (8.2.13) its increments can be expressed in the form

$$\bar{C}_t - \bar{C}_s = \bar{V}(t, \bar{S}_t) - \bar{V}(s, \bar{S}_s) - \int_s^t \delta_u^1 \, d\bar{S}_u \tag{8.2.18}$$

for $t \in [0,T]$ and $s \in [0,t]$. Assuming that the discounted pricing function $\bar{V}(\cdot,\cdot)$ is differentiable with respect to time and twice differentiable with respect to the discounted underlying security value, the Itô formula (6.2.11) can be applied and we obtain from (8.2.18) the relation

$$\bar{C}_t - \bar{C}_s = \int_s^t \left[ \frac{\partial \bar{V}(u,\bar{S}_u)}{\partial u} + \frac{1}{2}\,\sigma_u^2\,\bar{S}_u^2\,\frac{\partial^2 \bar{V}(u,\bar{S}_u)}{\partial \bar{S}^2} + (a_u - r_u)\,\bar{S}_u \left( \frac{\partial \bar{V}(u,\bar{S}_u)}{\partial \bar{S}} - \delta_u^1 \right) \right] du + \int_s^t \sigma_u\,\bar{S}_u \left( \frac{\partial \bar{V}(u,\bar{S}_u)}{\partial \bar{S}} - \delta_u^1 \right) dW_u \tag{8.2.19}$$

for $t \in [0,T)$, $s \in [0,t]$. The formula (8.2.19) provides an explicit representation for the increments of the discounted P&L.

## Discounted Black-Scholes PDE

Note that a strategy $\delta$ that minimizes the fluctuations of the discounted P&L process $\bar{C}$ is obtained if the second integral on the right hand side of (8.2.19) vanishes for the choice of the *hedge ratio* $\delta_t^1$ given by

$$\delta_t^1 = \frac{\partial \bar{V}(t,\bar{S}_t)}{\partial \bar{S}} \tag{8.2.20}$$

for $t \in [0,T)$. It can be seen that when taking (8.2.20) into account, then the first term in (8.2.19) disappears if the discounted value function $\bar{V}$ satisfies the PDE

$$\frac{\partial \bar{V}(t,\bar{S})}{\partial t} + \frac{1}{2}\,\sigma_t^2\,\bar{S}^2\,\frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} = 0 \tag{8.2.21}$$

for $t \in [0,T)$ and $\bar{S} \in (0,\infty)$. The resulting PDE (8.2.21) is not sufficient to determine fully the function $\bar{V}(\cdot,\cdot)$. However, it would keep by (8.2.19) and (8.2.20) any discounted P&L constant. Additionally, some condition at the terminal time $T$ needs to be specified to make sure that we start from a zero initial discounted P&L. To ensure this and, thus, the replication of the payoff at the expiration date $T$, see (8.2.8) and (8.2.9), we have to satisfy the *terminal condition*

$$\bar{V}(T,\bar{S}) = \frac{H(\bar{S}\,B_T)}{B_T} = \frac{H(S)}{B_T} \tag{8.2.22}$$

for $\bar{S} \in (0,\infty)$. We call the PDE (8.2.21) together with its terminal condition (8.2.22) the *discounted Black-Scholes partial differential equation* (discounted BS-PDE). This PDE determines a discounted pricing function $\bar{V}(\cdot,\cdot)$ that allows a perfect hedge for the corresponding European payoff.

For instance, for European call and put options it can be shown that the discounted BS-PDE has a unique solution and, thus, determines uniquely the option price. The uniqueness of the solution of a PDE in the above form is, in general, not trivially established, as we shall see in Chap. 12.

### Black-Scholes PDE

By a transformation of variables, see (8.2.9) and (8.2.10), the above discounted BS-PDE can be rewritten for the undiscounted option pricing function $V(\cdot,\cdot)$ in the form

$$\frac{\partial V(t,S)}{\partial t} + r_t\, S\, \frac{\partial V(t,S)}{\partial S} + \frac{1}{2}\,\sigma_t^2\, S^2\, \frac{\partial^2 V(t,S)}{\partial S^2} - r_t\, V(t,S) = 0 \qquad (8.2.23)$$

for $t \in [0,T)$ and $S \in (0,\infty)$ with terminal condition, see (8.2.8),

$$V(T,S) = H(S) \qquad (8.2.24)$$

for $S \in (0,\infty)$. We call (8.2.23) together with (8.2.24) the BS-PDE. Note that the BS-PDE and the discounted BS-PDE do not depend on the values of the appreciation rate $a_t$ of the underlying security. This is a remarkable fact, which results from the choice of $\delta_t^1$ in (8.2.20) that eliminated in (8.2.19) any potential impact of $a_t$.

### Option Price

In the formula (8.2.20) for the hedge ratio we describe the number $\delta_t^1$ of units to be held in the underlying security. By dividing equation (8.2.4) on both sides by the savings account and using equations (8.2.10) and (8.2.9), we can now determine the number of units that needs to be held in the domestic savings account. It is given by the relation

$$\delta_t^0 = \delta_t^0\,\bar{B}_t = \bar{V}(t,\bar{S}_t) - \delta_t^1\,\bar{S}_t \qquad (8.2.25)$$

for $t \in [0,T]$.

The option price obtained at time $t$ is, of course, just $V(t,S_t)$. The appropriate value of the hedge portfolio at time $t$ in units of the domestic currency can, therefore, be calculated, see (8.2.9), via the formula

$$V(t,S_t) = \bar{V}(t,\bar{S}_t)\,B_t \qquad (8.2.26)$$

for $t \in [0,T]$.

### Numeraire Invariance

Let us now check whether the above construction of a hedge portfolio identifies a self-financing strategy $\delta$. As mentioned previously, this is a strategy that changes the portfolio value only through changes in gains from trade, see (8.2.6) and (8.2.7). For our discounted securities we have from (8.2.18) because of zero discounted P&L $\bar{C}_t = 0$ for all $t \in [0,T]$ that

$$d\bar{V}(t,\bar{S}_t) = \delta_t^1\, d\bar{S}_t \qquad (8.2.27)$$

for all $t \in [0, T]$. This means that the portfolio $\bar{V}(t, \bar{S}_t)$ is self-financing when denominated in units of the savings account, because all changes in $\bar{V}(t, \bar{S}_t)$ are due to changes in $\bar{S}_t$. It is now of interest that the portfolio is also shown to be self-financing when using other numeraires, for instance, if denominated in units of the domestic currency. For this case we multiply $\bar{V}(t, S_t)$ by the savings account $B_t$ and obtain from (8.2.26) and (8.2.27) by the integration-by-parts formula (6.3.1) the SDE

$$\begin{aligned} dV(t, S_t) &= d(\bar{V}(t, \bar{S}_t)\, B_t) \\ &= B_t\, d\bar{V}(t, \bar{S}_t) + \bar{V}(t, \bar{S}_t)\, dB_t + d[B_\cdot, \bar{V}(\cdot, \bar{S}_\cdot)]_t \\ &= B_t\, \delta_t^1\, d\bar{S}_t + (\delta_t^0 + \delta_t^1\, \bar{S}_t)\, dB_t + \delta_t^1\, d[B, \bar{S}]_t \\ &= \delta_t^0\, dB_t + \delta_t^1\, \left(B_t\, d\bar{S}_t + \bar{S}_t\, dB_t + d[B, \bar{S}]_t\right) \\ &= \delta_t^0\, dB_t + \delta_t^1\, d(B_t\, \bar{S}_t) \\ &= \delta_t^0\, dB_t + \delta_t^1\, dS_t \end{aligned} \tag{8.2.28}$$

for $t \in [0, T]$. This proves the condition (8.2.7), which ensures that the resulting portfolio is self-financing when expressed in units of the domestic currency.

Consequently, the changes in the portfolio value are only a result of gains from trade in the underlying security $S$ and the savings account $B$. The above result in (8.2.28) is important, since it shows that a portfolio that is self-financing in one denomination is also self-financing in another denomination. Note that such a result holds more generally, as will be shown in (9.6.18) and towards the end of Chap. 14. This means that a change in numeraire does not impact on the self-financing property. We could select any strictly positive portfolio as numeraire and would see, similarly as above, that a portfolio, which is self-financing in one denomination is also self-financing under this numeraire.

The discounted P&L process $\bar{C}_t$ starts at zero, see (8.2.16), and has zero increments, see (8.2.19)–(8.2.21). Therefore, it is zero for the above identified hedging strategy. The undiscounted P&L process $C = \{C_t,\, t \in [0, T]\}$ with

$$C_t = \bar{C}_t\, B_t = 0 \tag{8.2.29}$$

for $t \in [0, T]$, see (8.2.12) and (8.2.17), equals then also zero. Consequently, the resulting nonnegative P&L process does not permit arbitrage, as was required.

The above hedging approach for determining the value of an option is essentially based on the Itô formula. This fundamental tool allows us to obtain in continuous time a perfect hedging strategy together with the corresponding option price. Note that no expectation has been taken to determine the option price.

We shall see later in Chap. 10 that the above approach for finding a perfect hedge and a corresponding price for a derivative security can be generalized to more complex payoff structures and more general asset price models. Certain PDEs, similar to those given in (8.2.21) and (8.2.22), arise also for other payoffs and security dynamics. What differs are the volatility specification and the boundary conditions.

## 8.3 The Black-Scholes Formula

In this section we study the solution of the BS-PDE (8.2.23) with its terminal condition (8.2.24) in the case of a European call option.

### Black-Scholes Formula

Let us describe the price of a European call option for an underlying security $S = \{S_t,\, t \in [0,T]\}$ that follows the SDE (8.2.1). The payoff is according to (8.1.1) of the form

$$H(S) = (S - K)^+ \tag{8.3.1}$$

for $S \in (0,\infty)$ with strike price $K > 0$ and matures at the terminal date $T$.

In their Nobel prize winning work Black, Scholes and Merton provided the explicit description of the price $c_{T,K}(t, S_t)$ at time $t$ for the European call option with expiry date $T$ and strike price $K$, see Black & Scholes (1973) and Merton (1973b). This result is widely known as the *Black-Scholes formula* (BS formula). It takes the form

$$c_{T,K}(t, S_t) = S_t\, N(d_1(t)) - K\, \frac{B_t}{B_T}\, N(d_2(t)) \tag{8.3.2}$$

with

$$d_1(t) = \frac{\ln\left(\frac{S_t}{K}\right) + \int_t^T \left(r_s + \frac{1}{2}\,\sigma_s^2\right)\, ds}{\sqrt{\int_t^T \sigma_s^2\, ds}} \tag{8.3.3}$$

and

$$d_2(t) = d_1(t) - \sqrt{\int_t^T \sigma_s^2\, ds} \tag{8.3.4}$$

for $t \in [0,T)$. Here $B_t$ is again the domestic savings account at time $t$, see (8.2.2). Furthermore, $N(\cdot)$ denotes the standard Gaussian distribution function, see (1.2.7), with density

$$N'(x) = \frac{1}{\sqrt{2\pi}}\, \exp\left\{-\frac{x^2}{2}\right\} \tag{8.3.5}$$

for all $x \in \Re$, see (1.2.8). It can be shown by direct calculation that the above European call option pricing function $c_{T,K}(\cdot,\cdot)$ solves the BS-PDE given in (8.2.23) for the payoff function (8.3.1), see Exercise 8.1. One observes in the

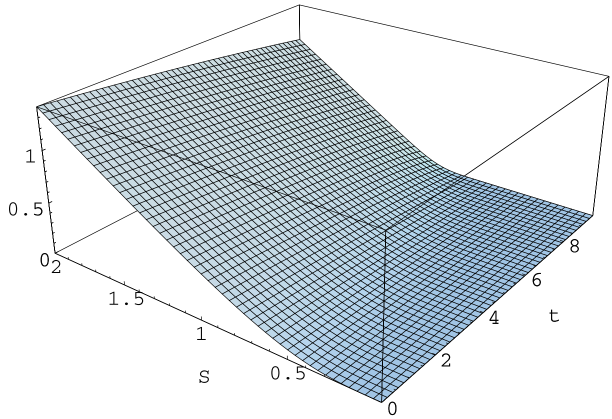



**Fig. 8.3.1.** Black-Scholes European call option price

BS formula that the option price does not depend on the specific choice of the appreciation rate $a_t$ of the underlying security, which we explained earlier.

Noting the form of the BS formula (8.3.2), a heuristic guess for the number $\delta_t^1$ of units of the risky asset to be held in the hedge portfolio would be $N(d_1(t))$. We show below that this is correct. However, this result is not as obvious as it may seem because $d_1(t)$ and $d_2(t)$ depend on $S_t$.

For small values of $S_t$, the expressions $d_1(t)$ and $d_2(t)$ and also $N(d_1(t))$ and $N(d_2(t))$ are small, see Fig. 1.2.4. Thus, for small $S_t$ the European call option has almost no value. However, for large underlying security price $S_t$ the quantities $d_1(t)$ and $d_2(t)$ are both large so that $N(d_1(t))$ and $N(d_2(t))$ are approximately one, as can be seen in Fig. 1.2.4. Consequently, by (8.3.2) the option value equals approximately $S_t - K \frac{B_t}{B_T}$ in this case.

The BS formula (8.3.2) can be interpreted as being an analytical formula. However, the Gaussian distribution function $N(\cdot)$ needs still to be approximated by other more basic functions or obtained by numerical evaluation of the integral of the Gaussian density $N'(\cdot)$ given in (8.3.5). In (1.2.7) a reasonably accurate and efficient approximation for the standard Gaussian distribution function has been provided.

### European Call Option Price

To give an idea about the shape of the pricing function $c_{T,K}$ we show in Fig. 8.3.1 the European call option price as a function of time $t$ and the underlying security price $S$ with volatility $\sigma = 0.2$, strike price $K = 1$, expiration date $T = 10$ years and short rate $r = 0.05$. Figure 8.3.1 depicts prices for up to ten years to display some long term features of the typical Black-Scholes option price. Note that close to the expiration date $T = 10$ the option price has approximately the value of the hockey stick like payoff function (8.3.1). As previously mentioned, for small values of the underlying security the option price remains close to zero and for large security prices $S$ the option has a price close to $S - K\,\frac{B_t}{B_T}$.

## 8.4 Sensitivities for European Call Option

The pricing function $c_{T,K}$, see (8.3.2), for the European call option depends on several variables. These are the underlying security price $S_t$, the time to maturity $T-t$, the volatility $\sigma$ of the underlying security, the interest rate $r$ and the strike price $K$. Changes in any of these variables influence the option price. Therefore, it is of practical importance to know how sensitive the pricing function $c_{T,K}$ is with respect to these variables.

It is informative to use the classical Taylor formula to expand the increments of the value of the derivative security over a small time interval $[t, t+h]$ in dependence on the above mentioned variables. By omitting higher order terms one obtains

$$
\begin{aligned}
& V(t+h, S_{t+h}) - V(t, S_t) \\
&\quad \approx \frac{\partial V(t,S_t)}{\partial S}\,(S_{t+h}-S_t) + \frac{1}{2}\,\frac{\partial^2 V(t,S_t)}{\partial S^2}\,(S_{t+h}-S_t)^2 \\
&\qquad + \frac{\partial V(t,S_t)}{\partial t}\,h + \frac{\partial V(t,S_t)}{\partial \sigma}\,(\sigma_{t+h}-\sigma_t) + \frac{\partial V(t,S_t)}{\partial r}\,(r_{t+h}-r_t) \\
&\quad = \Delta(S_{t+h}-S_t) + \frac{1}{2}\,\Gamma\,(S_{t+h}-S_t)^2 - \Theta\,h + \mathcal{V}\,(\sigma_{t+h}-\sigma_t) + \varrho\,(r_{t+h}-r_t),
\end{aligned}
\tag{8.4.1}
$$

where

$$
\Delta = \frac{\partial V(t,S_t)}{\partial S}, \quad \Gamma = \frac{\partial^2 V(t,S_t)}{\partial S^2}, \quad \Theta = \frac{\partial V(t,S_t)}{\partial t},
$$
$$
\mathcal{V} = \frac{\partial V(t,S_t)}{\partial \sigma} \quad \text{and} \quad \varrho = \frac{\partial V(t,S_t)}{\partial r}
$$

for $t \in [0,T]$. Here the letters $\Delta$, $\Gamma$, $\Theta$, $\mathcal{V}$ and $\varrho$ denote the corresponding partial derivatives which are called *sensitivities* or *greeks*. The expansion (8.4.1) shows how the above greeks influence the increments of the Black-Scholes option price. Note that for obtaining a first order approximation one needs to include the second order derivative $\Gamma$ since the conditional expectation

$$
E\left((S_{t+h}-S_t)^2 \,\middle|\, \mathcal{A}_t\right) \approx \sigma_t^2\, S_t^2\, h.
$$

is of order $h$.

In the following we discuss some of the above greeks for the Black-Scholes European call option price. In the figures displayed below, we choose as default parameter the strike price $K = 1$ and maturity date $T = 10$. We consider the

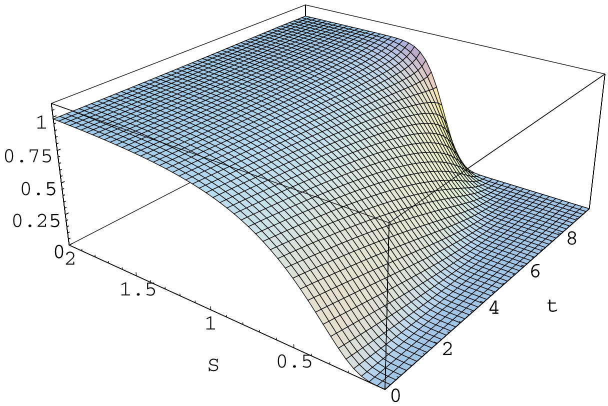



**Fig. 8.4.1.** Delta as a function of $t$ and $S_t$

parameters $a$, $\sigma$ and $r$ to be constant and fix $T$ and $K$, unless we study a sensitivity with respect to such a parameter. This means, we study sensitivities for the standard BS model.

## Delta

The *delta* has been previously mentioned as hedge ratio, see (8.2.20). It measures the sensitivity of the option price with respect to changes in the price of the underlying security $S_t$. We set

$$\Delta = \frac{\partial V(t, S_t)}{\partial S} = \delta_t^1 \tag{8.4.2}$$

and obtain from (8.3.2) and (8.3.3) the expression

$$\Delta = N(d_1(t)), \tag{8.4.3}$$

which can be shown to equal the partial derivative appearing in (8.2.20), see Exercise 8.2.

Figure 8.4.1 shows for constant $\sigma = 0.2$ and $r = 0.05$ the delta for the European call option as a function of time $t$ and asset price $S_t$.

Note that the delta for a European call option is always positive and bounded by one. Close to expiration and strike price $K = 1$, delta behaves almost like a step function moving from level zero to one. This makes hedging quite difficult in this situation.

## Gamma

The sensitivity of the hedge ratio delta, with respect to the security price $S_t$ is called *gamma*. This greek is important for the length of re-balancing intervals in practical hedging under transaction costs. A large gamma reflects

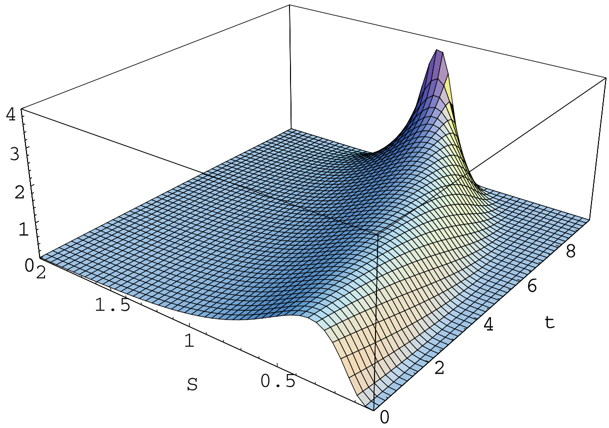



**Fig. 8.4.2.** Gamma as a function of $t$ and $S_t$

large changes in the hedge ratio and thus typically large transaction costs. The gamma is set to

$$\Gamma = \frac{\partial \Delta}{\partial S} = \frac{\partial^2 V(t, S_t)}{\partial S^2}. \tag{8.4.4}$$

Using (8.4.3) and (8.3.3), it can be shown that

$$\Gamma = N'(d_1(t)) \frac{1}{S_t \, \sigma \sqrt{T - t}} \tag{8.4.5}$$

for $t \in [0, T)$. Note that gamma is always positive. Figure 8.4.2 displays gamma for the European call option as a function of time $t$ and the underlying security price $S_t$, using the same parameter values as in Fig. 8.4.1.

Close to maturity the gamma has a profile in spatial direction similar to that of the bell shaped curve of the Gaussian density, see Fig. 1.2.3. It becomes extremely large close to expiration for security prices that are near the strike price, which is here set to K=1.

### Theta

The *theta* of a hedge portfolio measures the dependence of the option price on the remaining time to expiration $(T - t)$. The parameter theta, often called the time decay of the portfolio, provides an estimate of the time sensitivity of the option price and is given by the expression

$$\Theta = -\frac{\partial V(t, S_t)}{\partial (T - t)}. \tag{8.4.6}$$

From (8.3.2) we obtain

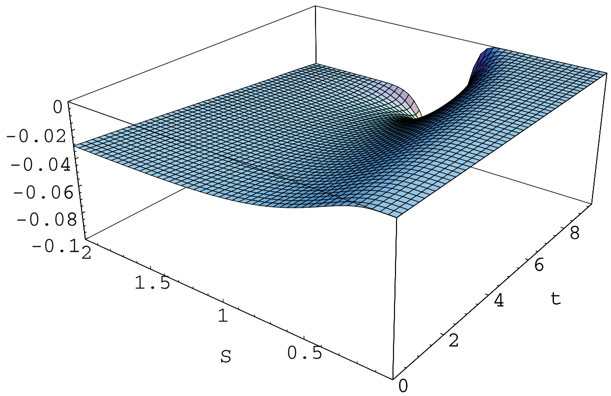



**Fig. 8.4.3.** Theta as a function of $t$ and $S_t$

$$\Theta = -N'(d_1(t))\,\frac{S_t\,\sigma}{2\sqrt{T-t}} - r\,K\,\exp\{-r(T-t)\}\,N(d_2(t)) \tag{8.4.7}$$

for $t \in [0, T)$. Figure 8.4.3 displays theta for the European call option as a function of time $t$ and security price $S_t$ for the same parameter values as used in Fig. 8.4.1.

### Vega

In the standard BS model a constant volatility $\sigma$ is assumed. However, in practice volatility is difficult to estimate and changes over time. It is important to see how differences in volatilities influence derivative prices. The sensitivity of the option price with respect to volatility is called *vega*, which is given by

$$\mathcal{V} = \frac{\partial V(t, S_t)}{\partial \sigma}. \tag{8.4.8}$$

Using (8.3.2), it can be shown that

$$\mathcal{V} = N'(d_1(t))\,S_t\,\sqrt{T-t} \tag{8.4.9}$$

for $t \in [0, T)$. Figure 8.4.4 shows vega as a function of the volatility $\sigma$ and the time $t$, where we have set $r = 0.05$, $K = 1$ and $S_t = 1$. Vega is positive and decreases substantially close to expiration. Its maximum value can be found for a volatility value that is close to $\sqrt{2\,r}$.

### Rho

In the standard BS model interest rates are assumed to be constant. However, in practice interest rates vary. The sensitivity of the option price with respect to the interest rate $r$ can be analyzed through *rho*, which is defined as

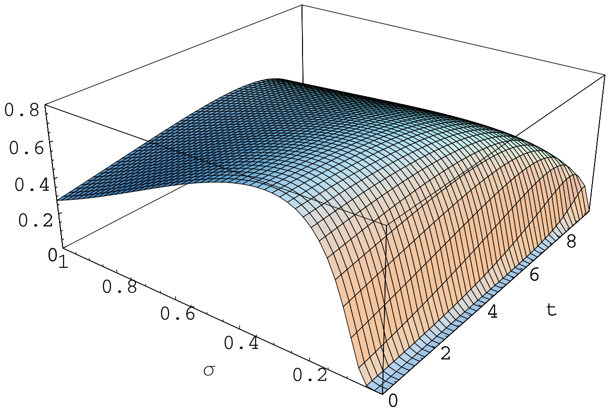



**Fig. 8.4.4.** Vega as a function of $t$ and $\sigma$

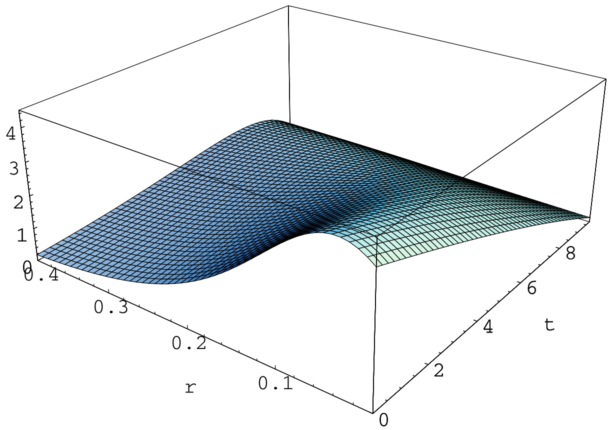



**Fig. 8.4.5.** Rho as a function of $t$ and $r$

$$\varrho = \frac{\partial V(t, S_t)}{\partial r}.$$

From (8.3.2), the rho can be derived as the expression

$$\varrho = N(d_2(t))\,(T-t)\,K\,\exp\{-r\,(T-t)\} \tag{8.4.10}$$

for $t \in [0, T)$. Note that rho is always positive for a European call option. Figure 8.4.5 shows rho as a function of time $t$ and interest rate $r$ for $\sigma = 0.2$, $K = 1$ and $S_t = 1$. Rho appears to be larger for large time to maturity and largest for an interest rate close to $\frac{1}{2}\sigma^2$.

## 8.5 European Put Option

In this section we present a key relationship between European put and call options under the BS model. Additionally, the greeks of put options and their properties will be discussed.

### Put-Call Parity

*Put-call parity* provides a simple way to determine the price of a European put option if the corresponding call option price for the same strike and maturity has been already computed. For the BS model the put-call parity relation can be expressed in the form

$$c_{T,K}(t, S_t) = p_{T,K}(t, S_t) + S_t - K\,\frac{B_t}{B_T} \tag{8.5.1}$$

for $t \in [0, T]$. This relation can be derived from the fact that the payoff function for the terminal value of the quantity on the left hand side of equation (8.5.1) equals the payoff function of that on the right hand side, which is

$$(S_T - K)^+ = (K - S_T)^+ + S_T - K. \tag{8.5.2}$$

For a wide range of models a similar put-call parity holds. This property of put and call prices is not restricted to the BS model because it reflects the general relationship (8.5.2) between their payoffs.

### European Put Option Price

Using put-call parity, the pricing function $p_{T,K}$ for a *European put option* for the BS model with constant volatility $\sigma$ and constant interest rate $r$ is given by the formula

$$\begin{aligned} p_{T,K}(t, S_t) &= S_t\,(N(d_1(t)) - 1) - K\,\frac{B_t}{B_T}\,(N(d_2(t)) - 1) \\ &= -S_t\,N(-d_1(t)) + K\,\frac{B_t}{B_T}\,N(-d_2(t)), \end{aligned} \tag{8.5.3}$$

where $d_1(t)$ and $d_2(t)$ are given in (8.3.3) and (8.3.4). Figure 8.5.1 shows the European put option price as a function of time $t$ and security price $S_t$ for volatility $\sigma = 0.2$, strike price $K = 1$, expiration date $T = 10$ and interest rate $r = 0.05$.

It is interesting to compare the European call option price in Fig. 8.3.1 with the corresponding put option price displayed in Fig. 8.5.1. Inspection of both figures shows that, at the expiration date, the corresponding ramp like payoff functions are matched by the pricing functions.

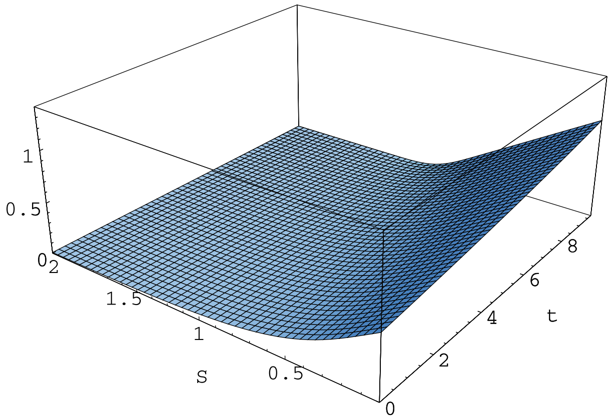



**Fig. 8.5.1.** European put option price

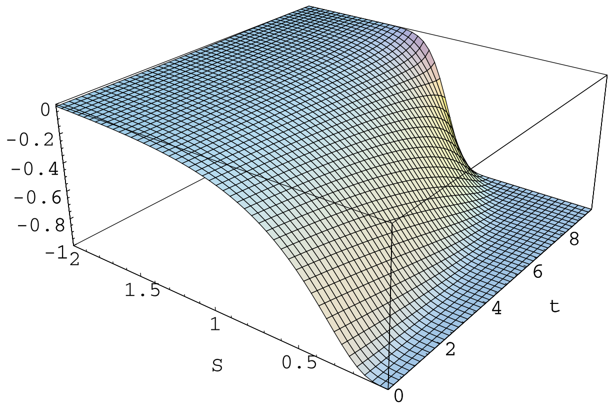



**Fig. 8.5.2.** Delta for the European put as a function of $t$ and $S_t$

### Greeks for the European Put Option

As with European calls, the sensitivities of the Black-Scholes European put option price (8.5.3) can be examined with respect to changes in various variables using the same notation. By using the put-call parity in (8.5.1) or the European put price (8.5.3) one obtains easily the corresponding sensitivities. The delta for the European put option is, according to (8.4.2) and (8.5.1), given by

$$\Delta = \frac{\partial p_{T,K}}{\partial S} = N(d_1(t)) - 1. \tag{8.5.4}$$

Figure 8.5.2 shows the delta for the European put option as a function of time $t$ and security price $S_t$. Note for the European put option that delta is always negative and bounded between $-1$ and 0. Comparing formulas (8.4.3) with (8.5.4) reveals that the delta of the put equals that of the call minus one.

The sensitivity of the delta with respect to the underlying security price $S_t$ is again called gamma, which for European puts, see (8.4.4) and (8.5.4), is given by the expression

$$\Gamma = \frac{\partial^2 p_{T,K}}{\partial S^2} = N'(d_1(t)) \frac{1}{S_t \, \sigma \sqrt{T-t}}. \tag{8.5.5}$$

This is the same formula as that for the European call given in (8.4.5). Thus, Fig. 8.4.2 also shows the shape of the gamma for the European put.

The other greeks for European puts, similar to those mentioned earlier, are given by the relations:

$$\Theta = \frac{\partial p_{T,K}}{\partial (T-t)} = N'(d_1(t)) \frac{S_t \, \sigma}{2\sqrt{T-t}} + r\, K \, \exp\{-r\,(T-t)\}\,(N(d_2(t)) - 1), \tag{8.5.6}$$

$$\mathcal{V} = \frac{\partial p_{T,K}}{\partial \sigma} = N'(d_1(t))\, S_t \sqrt{T-t} \tag{8.5.7}$$

and

$$\varrho = \frac{\partial p_{T,K}}{\partial r} = (T-t)\, K \, \exp\{r\,(T-t)\}\,(N(d_2(t)) - 1). \tag{8.5.8}$$

**Bounds for European Calls and Puts**

There exist some simple bounds for European call and put option prices on a stock that pays no dividends. From the BS formula (8.3.2) it follows for the European call

$$c_{T,K}(t, S_t) \le S_t \tag{8.5.9}$$

for $t \in [0, T]$. By forming at time $t$ a portfolio that consists of a European call together with $K \frac{B_t}{B_T}$ units of the savings account, the payoff at maturity $T$ will be

$$\max(S_T, K) \ge S_T.$$

Therefore, it follows that

$$c_{T,K}(t, S_t) \ge \left( S_t - K \frac{B_t}{B_T} \right)^+ \tag{8.5.10}$$

for $t \in [0, T]$. This holds for all European call option prices. Any derivative that gives the holder more rights is more expensive. Therefore, an American call price $C_{T,K}(t, S_t)$ at time $t$ with maturity $T$ and strike $K$ is larger than the corresponding European call and by (8.5.10) we obtain

$$C_{T,K}(t, S_t) \ge c_{T,K}(t, S_t) \ge \left( S_t - K \frac{B_t}{B_T} \right)^+ \ge (S_t - K)^+ \tag{8.5.11}$$

for $t \in [0, T]$. This means, the American call price $C_{T,K}(t, S_t)$, see Sect. 8.1, is always larger than the intrinsic value $(S_t - K)^+$ and will therefore never be early exercised. Thus, we have

$$C_{T,K}(t, S_t) = c_{T,K}(t, S_t) \tag{8.5.12}$$

for $t \in [0, T]$. This interesting feature is not model dependent.

On the other hand, from (8.5.3) the upper bound

$$p_{T,K}(t, S_t) \leq K \, \frac{B_t}{B_T} \tag{8.5.13}$$

for $t \in [0, T]$ can be obtained. By put-call parity and the positivity of call prices it also follows that

$$p_{T,K}(t, S_t) \geq K \, \frac{B_t}{B_T} - S_t. \tag{8.5.14}$$

for $t \in [0, T]$.

Note that the above bounds do, in principle, not depend on the choice of the model for the underlying security dynamics if one substitutes $\frac{B_t}{B_T}$ by the corresponding zero coupon bond of the respective model. They hold generally because they are a consequence of the shape of the put and call payoff functions.

## 8.6 Hedge Simulation

In Sect. 8.1 we identified by hedging arguments the discounted BS-PDE for discounted option prices. This led in Sect. 8.3 to the BS formula, which provides the solution for the BS-PDE. By using a *hedge simulation* we show now how a hedge portfolio works in detail. This type of continuous trading is called *delta hedging*. In the following, we construct a hedge portfolio for a European call option under the standard BS model with constant appreciation rate $a$, volatility $\sigma$ and short rate $r$. We examine the evolution of the hedge portfolio for two different scenarios along an equidistant time discretization with $t_k = k\,h$, $k \in \{0, 1, \ldots\}$, for some small time step size $h > 0$. In this sense we shall perform an approximate hedge, which can be interpreted as a continuous hedge in a frictionless market. For each of the two scenarios we shall check whether the payoff is replicated by the hedge portfolio and the P&L remains approximately zero as predicted by the theoretical results presented in Sect. 8.1.

### Hedging Strategy

The hedge ratio, that is the delta $\delta_t^1$, has according to (8.2.20), (8.4.3) and (8.3.3) the value

$$\begin{aligned} \delta_t^1 &= N(d_1(t)) \\ &= N\left(\frac{\ln\left(\frac{S_t}{K}\right) + \left(r + \frac{1}{2}\,\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned} \tag{8.6.1}$$

for $t \in [0, T)$. For the number of units held in the domestic savings account we obtain from (8.2.25), (8.2.26), (8.2.10), (8.3.2) and (8.6.1) the relation

$$\begin{aligned} \delta_t^0 &= -\frac{K}{B_T} N(d_2(t)) \\ &= -\frac{K}{B_T} N\left(\frac{\ln\left(\frac{S_t}{K}\right) + \left(r - \frac{1}{2}\sigma^2\right)(T-t)}{\sigma\sqrt{T-t}}\right) \end{aligned} \tag{8.6.2}$$

for $t \in [0, T]$. Recall that the price of the call option at time $t \in [0, T)$, see (8.2.4), is given by

$$c_{T,K}(t, S_t) = \delta_t^1 \, S_t + \delta_t^0 \, B_t. \tag{8.6.3}$$

Furthermore, from (8.2.13) the discounted P&L takes the form

$$\bar{C}_t = \bar{V}(t, \bar{S}_t) - \int_0^t \delta_s^1 \, d\bar{S}_s - \bar{V}(0, \bar{S}_0) \tag{8.6.4}$$

for $t \in [0, T)$. Let us now recall that the discounted P&L remains zero. This follows, for instance, by a straightforward application of the Itô formula (6.2.11) for $\bar{V}$, where we obtain

$$\bar{V}(t, \bar{S}_t) = \bar{V}(0, \bar{S}_0) + \int_0^t \delta_s^1 \, d\bar{S}_s \tag{8.6.5}$$

and, thus, with (8.6.4) it must hold

$$\bar{C}_t = 0 \tag{8.6.6}$$

for $t \in [0, T]$. As previously explained, this is a consequence of the fact that the terms on the right hand side of (8.2.19) vanish by the choice of the hedging strategy.

## In-the-Money Scenario

For illustration, let us generate linearly interpolated values of the underlying security price $S_t$, say a stock index, from a sample path of a geometric Brownian motion starting at $S_0 = 1$ with appreciation rate $a = 0.05$ and volatility $\sigma = 0.2$, using the time points $t_i = i\,h \in [0, 10]$ for $i \in \{0, 1, \ldots, 500\}$ with time step size $h = 0.02$. This path is shown in Fig. 8.6.1 together with the corresponding hedge ratio $\delta_t^1$, see (8.6.1), for the European call option with expiration date $T = 10$, strike price $K = 1$ and interest rate $r = 0.05$. Note that for the given sample path the security price ends up in-the-money, that is we have $S_T > K$. For this scenario we observe that the hedge ratio converges to the value $\delta_T^1 = 1$ as $t$ tends to $T$. This is the correct value for the hedge ratio since the security price ends up in-the-money and the option will be exercised.

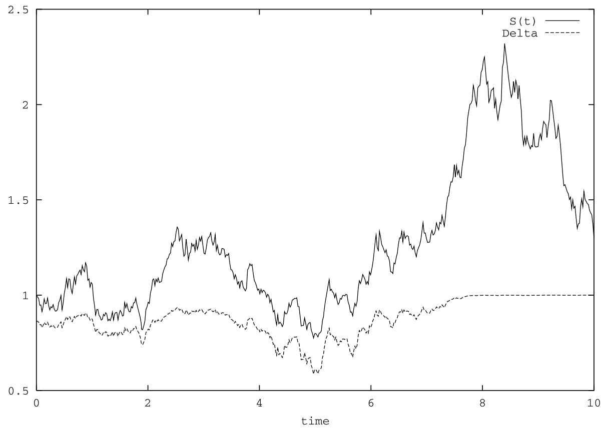



**Fig. 8.6.1.** Underlying security and hedge ratio for in-the-money call

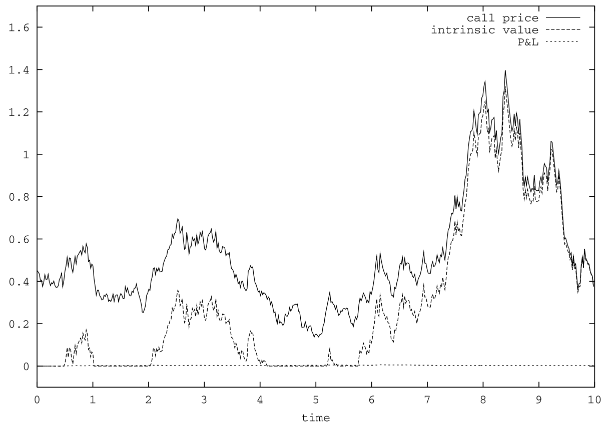



**Fig. 8.6.2.** Price, intrinsic value and P&L for in-the-money call

The evolution of the value of the corresponding hedge portfolio, which equals the call option price $c_{T,K}(t, S_t)$, is shown in Fig. 8.6.2 in dependence on time $t$. For comparison, Fig. 8.6.2 also displays the intrinsic value of the call option, that is the value

$$H(S_t) = (S_t - K)^+$$

for $t \in [0, T]$, see (8.1.2). Figure 8.6.2 shows for this sample path that the hedge portfolio replicates the payoff of the option at the expiration date $T = 10$ since the option price converges to its intrinsic value for $t$ approaching $T$.

For illustration, Fig. 8.6.2 also displays for the obtained self-financing strategy $\delta$ the P&L $C_t$ of the hedge portfolio. According to (8.2.13), the discounted P&L equals the value of the discounted portfolio minus the gains from trade in the discounted security $\bar{S}$ using the strategy $\delta$ minus the initial price. Note in Fig. 8.6.2 that the P&L remains almost perfectly zero over time as is expected from equation (8.6.6).

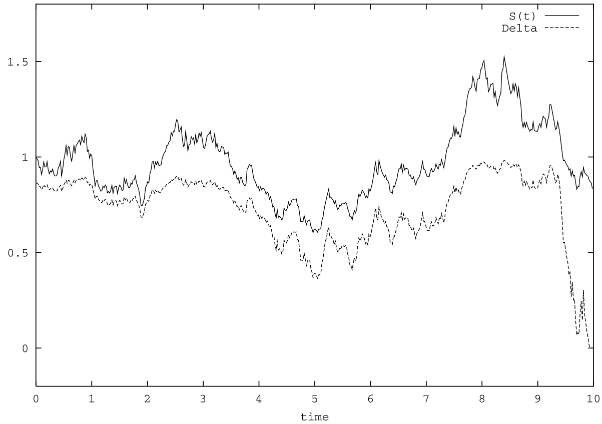



**Fig. 8.6.3.** Underlying security and hedge ratio for out-of-the-money call

### Out-of-the-Money Scenario

The replication of the payoff through delta hedging does not depend on the sample path of the underlying security, as can be seen from (8.2.19). To illustrate this we change the sample path by assuming a zero appreciation rate $a = 0$ in the above example. This brings the previous sample path of the underlying security down, as is evident from Fig. 8.6.3 and Fig. 8.6.1. It shows that the call option expires now out-of-the-money, that is $S_T < K = 1$. Consequently, the delta, that is the hedge ratio $\delta_t^1$, converges to zero for $t$ tending towards $T$. Figure 8.6.4 shows the corresponding sample path of the call option price $c_{T,K}(t, S_t)$ and its intrinsic value

$$H(S_t) = (S_t - K)^+$$

together with the P&L for the hedge portfolio. It is apparent that also in this case the payoff is replicated at the expiration date $T = 10$ and the P&L remains approximately zero over time, see (8.6.6).

We have used the same sample path of the driving Wiener process to generate both the in- and out-of-the-money scenarios for $a = 0.05$ and $a = 0$, respectively. The hedge simulations can be compared with each other via the corresponding graphs in Figs. 8.6.1–8.6.4. Note that the initial option prices $c_{T,K}(0, S_0)$ at time $t = 0$ for both scenarios are the same. Changing the appreciation rate $a$ in the BS model has not altered any part of our formulas and final hedging results. This striking phenomenon is a key feature of hedging. Independently of the realized scenario and the underlying appreciation rate the previously identified perfect hedge replicates the given payoff.

### Hedging a European Call Option on the S&P500

Let us now apply the above Black-Scholes delta hedging to an S&P500 index option. Figure 3.1.1 shows this index for the years from 1993 up until 1998

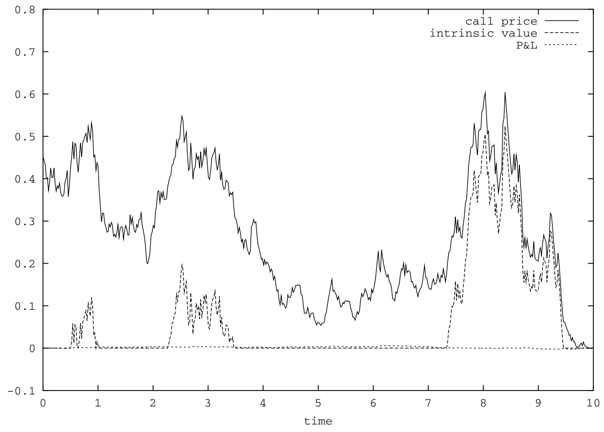



**Fig. 8.6.4.** Price, intrinsic value and P&L for out-of-the-money call

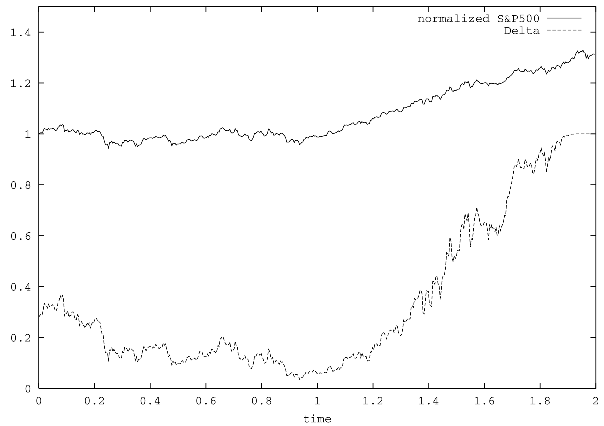



**Fig. 8.6.5.** Normalized S&P500 and hedge ratio for $K = 1.2$

and Fig. 5.2.6 its logarithm and quadratic variation. To make the following study similar to the above hedge simulation we divide the S&P500 data by its value at January, 3, 1994 and use the 520 observations of the normalized index for the years 1994 and 1995 as scenario of the underlying security. Figure 8.6.5 depicts the normalized S&P500 values for these two years and Fig. 8.6.6 the approximate quadratic variation, see (5.2.3), of the logarithm of the normalized S&P500. The quadratic variation seems to be reasonably linear for this period. According to formula (5.2.14), which provides some definition of volatility, we can read off from the plotted graph of the quadratic variation in Fig. 8.6.6 an average volatility of approximately $\sigma \approx \sqrt{\frac{0.016}{2}} \approx 0.09$. Furthermore, we set the USD short rate to the constant value $r \approx 0.05$, which is reasonable for the period under consideration.

Now, let us consider a European call option on the normalized S&P500 sample path as underlying security, which expires at the end of the period,

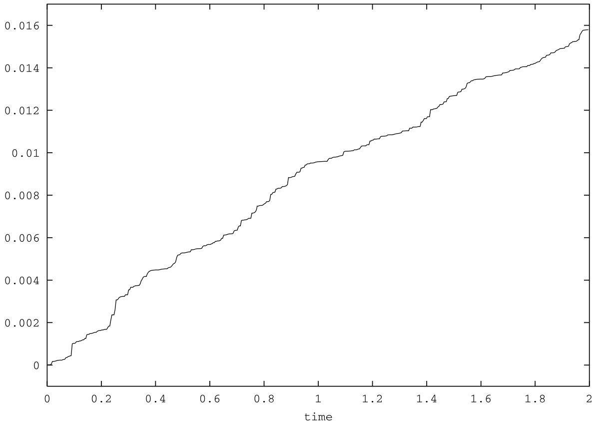



**Fig. 8.6.6.** Quadratic variation of log-S&P500

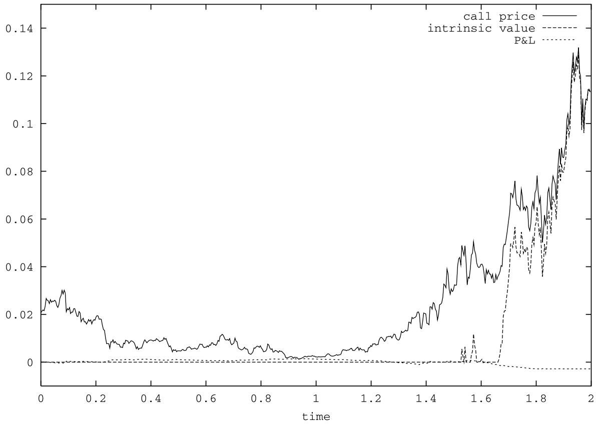



**Fig. 8.6.7.** Call price on S&P500, intrinsic value and P&L

that is in December 1995. We perform delta hedging according to what would be obtained for the BS model with the above parameters and use the same procedure that was employed for the previous hedge simulation. To study an in-the-money call option scenario we consider first a strike price of $K = 1.2$. Figure 8.6.5 shows the hedge ratio $\delta^1_t$ for this European call option. Figure 8.6.7 displays the corresponding evolution of the call option price, intrinsic value and P&L similar to Fig. 8.6.2 and Fig. 8.6.4. Note that the payoff is reasonably well replicated at the expiration date. However, the P&L is not as close to zero as was the case for the simulated BS model, previously examined. Note from Fig. 8.6.6 that the volatility of the underlying security was not fluctuating greatly during the chosen period. For longer dated options over longer time periods changes in the P&L can be shown to be more dramatic and the BS model is then not sufficient to provide an acceptable hedge. The nonzero P&L is clearly a consequence of the fact that the S&P500 does not exactly follow the BS model. The result can only be improved by using alternative asset price models which allow a volatility that is stochastic and reflects better reality. A paper by Bakshi, Cao & Chen (1997) shows that for the hedging of short dated options the BS model performs reasonably well. However, for the prices of these options the authors pointed out that the BS model seems to be not sufficiently accurate.

We shall later study various models that generate volatility which is stochastic. Some of these models involve squared Bessel processes, which we introduce in the following section.

## 8.7 Squared Bessel Processes (*)

As we shall see, many quantities that involve Bessel processes can be expressed in terms of Bessel functions. This gives this class of processes its name. We summarize in this section important results on squared Bessel processes because some of these will be crucial for the understanding of the following chapters presenting the benchmark approach.

To facilitate the explicit computation of derivative prices and other quantities under various models, including the CIR model, the CEV model and the MMM, we list in this section properties of square root and squared Bessel processes. Most of these properties are scattered in the literature. Some of them can be found, for instance, in Karatzas & Shreve (1991), Revuz & Yor (1999) or Jeanblanc, Yor & Chesney (2009).

The following results on time transformed squared Bessel processes will also be important for the understanding of the typical dynamics of financial markets. We shall give an example for a local martingale that is not a martingale. This example will turn out to be potentially closely linked to the real market dynamics.

### Squared Bessel Process (*)

Let us introduce the *squared Bessel process* ($\mathrm{BESQ}_x^\delta$) $X = \{X_\varphi,\, \varphi \in [0,\infty)\}$ of *dimension* $\delta \geq 0$ given by the SDE

$$dX_\varphi = \delta\, d\varphi + 2\sqrt{|X_\varphi|}\, dW_\varphi \tag{8.7.1}$$

for $\varphi \in [0,\infty)$ with $X_0 = x \geq 0$, where $W = \{W_\varphi,\, \varphi \in [0,\infty)\}$ is a standard Wiener process on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ starting at the initial $\varphi$-time, $\varphi = 0$, at zero. This means for $\varphi \in [0,\infty)$ that

$$[W]_\varphi = \varphi$$

for all $\varphi \in [0,\infty)$. Here we assume that $X$ is reflected at zero if it reaches the level zero. It turns out that the absolute sign under the square root in (8.7.1) can be removed. $X_\varphi$ remains nonnegative in this case and (8.7.1) has a unique strong solution, see Revuz & Yor (1999).

We have the following *scaling property*:
If $X = \{X_\varphi, \varphi \in [0,\infty)\}$ is a $\text{BESQ}^\delta_x$, then $Z = \{Z_\varphi, \varphi \in [0,\infty)\}$ with $Z_\varphi = \frac{1}{a} X_{a\varphi}$ is a $\text{BESQ}^\delta_{\frac{x}{a}}$ for all $a > 0$.

For $\delta \in \mathcal{N}$ and $x \geq 0$ the dynamics of a $\text{BESQ}^\delta_x$ $X$ can be expressed as the sum of the squares of $\delta$ independent standard Wiener processes $W^1$, $W^2$, ..., $W^\delta$, where

$$x = \sum_{k=1}^{\delta} (w^k)^2. \tag{8.7.2}$$

Here one sets

$$X_\varphi = \sum_{k=1}^{\delta} (w^k + W^k_\varphi)^2 \tag{8.7.3}$$

for $\varphi \in [0,\infty)$. Note that this construction is invariant with respect to the particular choice of $w^k$, $k \in \{1,2,\ldots,\delta\}$, when (8.7.2) is satisfied. Clearly, the function (8.7.3) is a function of components of the solution of a simple linear system of SDEs, where each component represents a Wiener process. By an application of the Itô formula we obtain

$$dX_\varphi = \delta\, d\varphi + 2 \sum_{k=1}^{\delta} \left(w^k + W^k_\varphi\right) dW^k_\varphi \tag{8.7.4}$$

for $\varphi \in [0,\infty)$ with

$$X_0 = \sum_{k=1}^{\delta} (w^k)^2 = x. \tag{8.7.5}$$

By setting

$$dW_\varphi = |X_\varphi|^{-\frac{1}{2}} \sum_{k=1}^{\delta} \left(w^k + W^k_\varphi\right) dW^k_\varphi \tag{8.7.6}$$

we obtain the SDE (8.7.1). Note that we have for $W_\varphi$ the quadratic variation

$$[W]_\varphi = \int_0^\varphi \frac{1}{X_s} \sum_{k=1}^{\delta} \left(w^k + W^k_s\right)^2 ds = \varphi.$$

Thus, $W_\varphi$ in (8.7.6) forms by Lévy's theorem, see Theorem 6.5.1, a Wiener process in the $\varphi$ time scale.

In Fig. 8.7.1 we show the path of a squared Bessel process of dimension $\delta = 4$ in the $\varphi$ time scale, which starts at $X_0 = 100$, where we set $w^k = 5$ for $k \in \{1,2,3,4\}$. Note the tendency of the process to increase over time, which is typical.

Squared Bessel processes have the following important *additivity property*, see Shiga & Watanabe (1973):
Let $X = \{X_\varphi, \varphi \in [0,\infty)\}$ be a $\text{BESQ}^\delta_x$ and $Y = \{Y_\varphi, \varphi \in [0,\infty)\}$ an

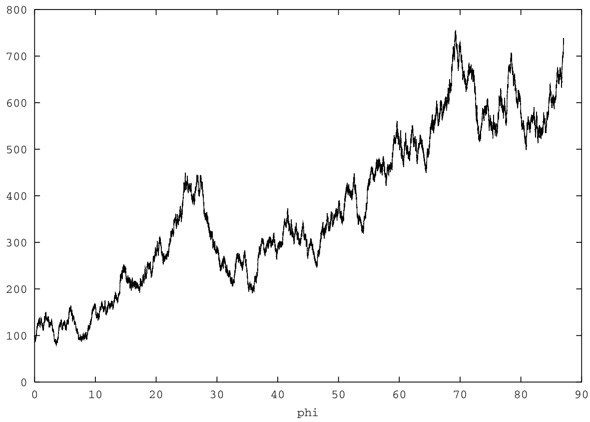



**Fig. 8.7.1.** Squared Bessel process of dimension $\delta = 4$ in $\varphi$-time

independent $\mathrm{BESQ}_y^{\delta'}$ with $x, y, \delta, \delta' \geq 0$. Then the process $Z = \{Z_\varphi,\, \varphi \in [0,\infty)\}$ where $Z_\varphi = X_\varphi + Y_\varphi$ is a $\mathrm{BESQ}_{x+y}^{\delta+\delta'}$.

It can be shown that a squared Bessel process $\mathrm{BESQ}_x^\delta$ of dimension $\delta > 2$ with $X_0 = x > 0$ stays always strictly positive, that is

$$P\left(\inf_{0\leq\varphi<\infty} X_\varphi > 0\right) = 1, \tag{8.7.7}$$

see Karatzas & Shreve (1998). In this case $X_\varphi$ tends to infinity as $\varphi$ goes to infinity. For the case $\delta = 2$ one has

$$P\left(\inf_{0\leq\varphi<\infty} X_\varphi > 0\right) = 0.$$

Furthermore, for a $\mathrm{BESQ}_x^\delta$ $X$ process with $\delta \in [0, 2)$ and $X_0 = x > 0$, there is a strictly positive probability that $X$ will hit zero before any fixed $\varphi$-time $\varphi' \in (0,\infty)$, that is

$$P\left(\inf_{0\leq\varphi\leq\varphi'} X_\varphi = 0\right) > 0. \tag{8.7.8}$$

This means $X_\varphi$ reaches zero in finite time with strictly positive probability.

For $\delta > 0$ and $x > 0$ the transition density for a $\mathrm{BESQ}_x^\delta$ process $X$ starting at the $\varphi$-time $\varrho \in [0,\infty)$ in $x$ being at time $\varphi \in (\varrho,\infty)$ in $y$ is given as

$$p_\delta(\varrho, x; \varphi, y) = \frac{1}{2(\varphi-\varrho)} \left(\frac{y}{x}\right)^{\frac{\nu}{2}} \exp\left\{-\frac{x+y}{2(\varphi-\varrho)}\right\} I_\nu\left(\frac{\sqrt{x\,y}}{\varphi-\varrho}\right), \tag{8.7.9}$$

see Revuz & Yor (1999), where $I_\nu$ is the modified Bessel function of the first kind, see (1.2.15), with index $\nu$. Here the index is defined as

$$\nu = \frac{\delta}{2} - 1. \tag{8.7.10}$$

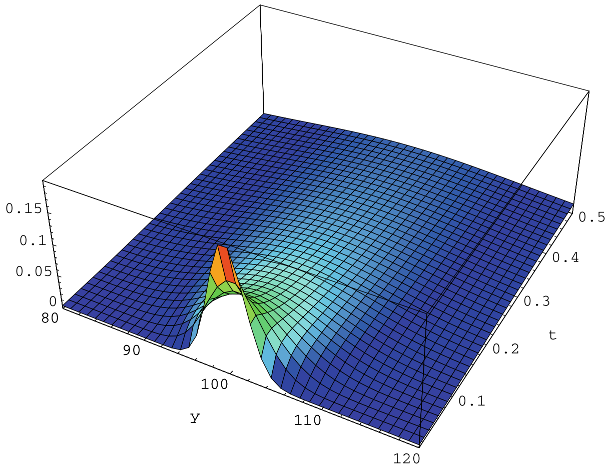



**Fig. 8.7.2.** Transition density of squared Bessel process for $\delta = 4$

In Fig. 8.7.2 we show the transition density of a squared Bessel process of dimension four, $\delta = 4$, which means index $\nu = 1$, starting at $x = 100$.

For small values of $z$ one has

$$I_\nu(z) \approx \frac{1}{\nu\,\Gamma(\nu)} \left(\frac{z}{2}\right)^\nu \tag{8.7.11}$$

for $\nu > 0$. Therefore, the transition density of a $\text{BESQ}_0^\delta$ process $X$, which starts at time zero at $x = 0$, is

$$p_\delta(0,0;\varphi,y) = (2\,\varphi)^{-\frac{\delta}{2}} \frac{y^{\frac{\delta}{2}-1}}{\Gamma(\frac{\delta}{2})} \exp\left\{-\frac{y}{2\,\varphi}\right\}. \tag{8.7.12}$$

From (8.7.9) and (1.2.14) one notices that for fixed $\delta > 2$, $x, y \geq 0$ and $\varphi > 0$ the transition density $p_\delta(0,x;\varphi,y)$ is the density of a non-central chi-square distributed random variable $Y = \frac{X_\varphi}{\varphi}$, see (1.2.13) with dimension $\delta$, and non-centrality parameter $\ell = \frac{x}{\varphi}$. Consequently, by (1.2.13) we obtain

$$P\left(\frac{X_\varphi}{\varphi} < u\right) = \sum_{k=0}^{\infty} \frac{\exp\left\{-\frac{\ell}{2}\right\}\left(\frac{\ell}{2}\right)^k}{k\,!} \left(1 - \frac{\Gamma\left(\frac{u}{2};\frac{\delta+2k}{2}\right)}{\Gamma\left(\frac{\delta+2k}{2}\right)}\right), \tag{8.7.13}$$

where $\Gamma(\cdot;\cdot)$ is the incomplete gamma function, see (1.2.12).

Furthermore, for $\alpha > -\frac{\delta}{2}$, $\varphi \in (0,\infty)$ and $\delta > 2$ one can show that

$$E\left(X_\varphi^\alpha \,\middle|\, \mathcal{A}_0\right) = \begin{cases} (2\,\varphi)^\alpha \exp\left\{\frac{-X_0}{2\,\varphi}\right\} \sum_{k=0}^\infty \left(\frac{X_0}{2\,\varphi}\right)^k \frac{\Gamma\left(\alpha+k+\frac{\delta}{2}\right)}{k\,!\,\Gamma\left(k+\frac{\delta}{2}\right)} & \text{for} \quad \alpha > -\frac{\delta}{2} \\ \infty & \text{for} \quad \alpha \leq -\frac{\delta}{2}, \end{cases} \tag{8.7.14}$$

see Exercise 8.8. By (8.7.1) it follows that

$$E\left(X_\varphi \,\middle|\, \mathcal{A}_0\right) = X_0 + \delta\,\varphi \tag{8.7.15}$$

for $\varphi \in [0,\infty)$. Thus, for $\alpha \in (-\frac{\delta}{2}, 0]$, $\varphi \in (0,\infty)$ and $\delta > 2$ it follows by the monotonicity of the gamma function

$$E\left(X_\varphi^\alpha \,\middle|\, \mathcal{A}_0\right) \le (2\varphi)^\alpha \exp\left\{\frac{-X_0}{2\varphi}\right\} \left(\frac{\Gamma\left(\alpha + \frac{\delta}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)} + \exp\left\{\frac{X_0}{2\varphi}\right\}\right)$$
$$< \infty, \tag{8.7.16}$$

see Exercise 8.8. Let us remark, by using the property $\frac{\Gamma(k+1)}{\Gamma(k+2)} = \frac{1}{k+1}$ of the gamma function and an expansion of the exponential function, that one obtains from (8.7.14) for $\delta = 4$ the explicit expression

$$E\left(X_\varphi^{-1} \,\middle|\, \mathcal{A}_0\right) = X_0^{-1}\left(1 - \exp\left\{\frac{-X_0}{2\,\varphi}\right\}\right) \tag{8.7.17}$$

for $\varphi \in (0,\infty)$.

If one absorbs a squared Bessel process with dimension $\delta \in [0,2)$ at zero, the transition density (8.7.9) changes, see Borodin & Salminen (2002), such that $I_{|\nu|}$ appears in the formula instead of $I_\nu$. That is, one has for $x > 0$ and $\varphi \in [0,\infty)$

$$p_\delta(0,x;\varphi,y) = \frac{1}{2\,\varphi}\left(\frac{y}{x}\right)^{\frac{\nu}{2}} \exp\left\{-\frac{x+y}{2\,\varphi}\right\} I_{|\nu|}\left(\frac{\sqrt{x\,y}}{\varphi}\right). \tag{8.7.18}$$

Let $P_x^\delta$ denote the law of a $\text{BESQ}_x^\delta$ process $X = \{X_\varphi,\ \varphi \in [0,\infty)\}$ of dimension $\delta$ with initial value $X_0 = x > 0$ at time $\varphi = 0$. In Revuz & Yor (1999) one can find the following important result. If we introduce the stopping time $\tau = \inf\{\varphi \in [0,\infty) :\ X_\varphi = 0\}$, then for $\delta > 2$ the relation holds:

$$P_x^{4-\delta}\Big|_{\mathcal{A}_\varphi \cap \{\varphi < \tau\}} = \left(\frac{x}{X_\varphi}\right)^{\frac{\delta}{2}-1} P_x^\delta\Big|_{\mathcal{A}_\varphi} \tag{8.7.19}$$

for all $\varphi \in (0,\infty)$. In principle, on the left hand side of the above relationship we consider squared Bessel processes with absorption at zero and on the right hand side squared Bessel processes that never reach zero. The same relationship (8.7.19) yields for $\delta < 2$ the equation

$$P_x^{\delta}\Big|_{\mathcal{A}_\varphi \cap \{\varphi < \tau\}} = \left(\frac{x}{X_\varphi}\right)^{1-\frac{\delta}{2}} P_x^{4-\delta}\Big|_{\mathcal{A}_\varphi}, \tag{8.7.20}$$

see also Exercises 8.9 and 8.10.

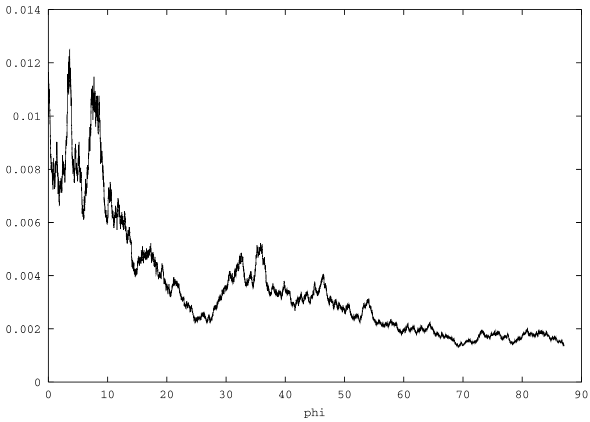



**Fig. 8.7.3.** Inverse of a squared Bessel process of dimension $\delta = 4$ in $\varphi$-time

### Examples of Strict Local Martingales (*)

We now present an example of a local martingale that is not a martingale, see Definition 5.1.1 and Definition 5.2.1. We consider the inverse $Z = \{Z_\varphi = X_\varphi^{-1}, \varphi \in [0,\infty)\}$ of a squared Bessel process $X = \{X_\varphi, \varphi \in [0,\infty)\}$ of dimension four, as given in (8.7.1), with $X_0 > 0$. Then it follows by the Itô formula that

$$dZ_\varphi = -2\, Z_\varphi^{\frac{3}{2}}\, dW_\varphi \tag{8.7.21}$$

for $\varphi \in [0,\infty)$, where

$$Z_0 = X_0^{-1}. \tag{8.7.22}$$

By the *driftless* SDE (8.7.21) the process $Z$ turns out to be a local martingale in $\varphi$ time, see Sect. 5.2 and Sect. 5.5 or Protter (2004). From (8.7.17) it follows that

$$\begin{aligned} E\left(Z_\varphi \,\middle|\, \mathcal{A}_0\right) &= E\left(X_\varphi^{-1} \,\middle|\, \mathcal{A}_0\right) \\ &= Z_0 \left(1 - \exp\left\{\frac{-1}{2\, Z_0\, \varphi}\right\}\right) < Z_0 \end{aligned} \tag{8.7.23}$$

for $\varphi \in (0,\infty)$. This relation is *not* consistent with $Z$ being an $(\underline{\mathcal{A}}, P)$-martingale. It actually proves that $Z$ cannot be a martingale according to equation (5.1.2). Thus, the inverse $Z$ of a squared Bessel process of dimension four is a continuous local martingale that is not a martingale. We say that such a local martingale is a *strict local martingale*. This observation will be very important for realistic financial modeling and derivative pricing, as we shall see later. Similarly, from relations (5.1.7) and (8.7.23) we can conclude that $Z$ is a strict supermartingale.

In Fig. 8.7.3 we exhibit the inverse of the path of a squared Bessel process of dimension four in $\varphi$-time, which refers to the example with the path in

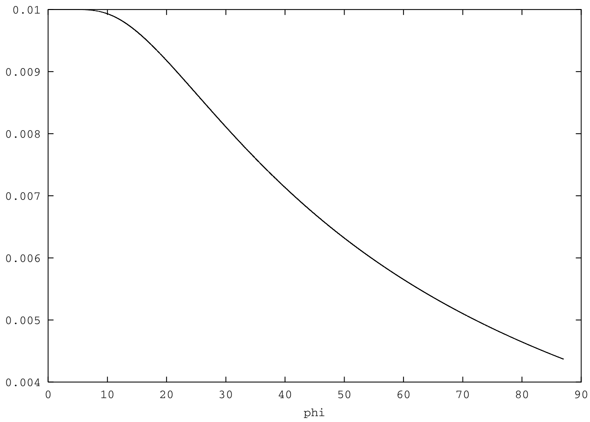



**Fig. 8.7.4.** Expectation of the inverse of the squared Bessel process for $\delta = 4$ in $\varphi$-time

Fig. 8.7.1. Note that this path is typical of that of a strict supermartingale. Here its current observation is larger than the best forecast of its future values.

In Fig. 8.7.4 we plot for the above example, by using formula (8.7.23), the expectation at time 0 of the inverse of the four dimensional squared Bessel process for varying $\varphi$-time. We clearly see the decline in this expectation over $\varphi$-time as was already indicated by the sample path in Fig. 8.7.3.

More generally, see Göing-Jaeschke & Yor (2003), for real valued dimension $\delta > 2$ one can show that the process

$$Z = \{Z_\varphi = X_\varphi^{1-\frac{\delta}{2}},\ \varphi \in [0, \infty)\} \tag{8.7.24}$$

is a *strict local martingale* if $X = \{X_\varphi,\ \varphi \in [0, \infty)\}$ is a $BESQ_x^\delta$ process of dimension $\delta > 2$ with $X_0 = x > 0$. One can see this from the relationship (8.7.19) since the expectation of $Z_\varphi$ is strictly less than one, because of the possible absorption of a squared Bessel process of dimension $4 - \delta < 2$, see (8.7.8). Alternatively, by application of the transition density (8.7.9) it follows that

$$\begin{aligned} E\left(Z_\varphi \,\middle|\, \mathcal{A}_0\right) &= E\left(X_\varphi^{1-\frac{\delta}{2}} \,\middle|\, \mathcal{A}_0\right) = \int_0^\infty y^{1-\frac{\delta}{2}}\, p_\delta(0, x; \varphi, y)\, dy \\ &= x^{1-\frac{\delta}{2}} \int_0^\infty p_{4-\delta}(0, y; \varphi, x)\, dy \\ &= x^{1-\frac{\delta}{2}} \left(1 - \frac{\Gamma(\frac{\delta}{2} - 1; \frac{x}{2\,\varphi})}{\Gamma(\frac{\delta}{2} - 1)}\right) < x^{1-\frac{\delta}{2}} \end{aligned} \tag{8.7.25}$$

for $\varphi \in (0, \infty)$. Here $\Gamma(\cdot)$ is again the gamma function, see (1.2.10), and $\Gamma(\cdot;\cdot)$ is the incomplete gamma function, see (1.2.12). Note that for the special case $\delta = 4$ we obtain from (8.7.25) and (1.2.12) the relation (8.7.23). Furthermore, the inequality in (8.7.25) is strict for $\varphi > 0$, which shows that $Z$ is a strict supermartingale.

### Time Transformation (*)

Using a squared Bessel process we can derive by transformations more general processes. These include, for instance, the *square root* (SR) *process* that was mentioned previously in (4.4.6).

Let $b : [0, \infty) \to \Re$ and $c : [0, \infty) \to (0, \infty)$ be given deterministic functions of time. We introduce the exponential

$$s_t = s_0 \exp\left\{\int_0^t b_u \, du\right\} \tag{8.7.26}$$

and the $\varphi$-time

$$\varphi(t) = \varphi(0) + \frac{1}{4} \int_0^t \frac{c_u^2}{s_u} \, du \tag{8.7.27}$$

for $t \in [0, \infty)$ and $s_0 > 0$ in dependence on time. Note that by (8.7.27) and (8.7.26) we have for constant $b < 0$ and $c \neq 0$ that

$$\varphi(t) = \varphi(0) + \frac{c^2}{4\, b\, s_0} \left(1 - \exp\{-b\, t\}\right) \tag{8.7.28}$$

for $t \in [0, \infty)$ and the time

$$t(\varphi) = -\frac{1}{b} \ln\left(1 - \frac{4\, b\, s_0}{c^2} \left(\varphi - \varphi(0)\right)\right) \tag{8.7.29}$$

for $\varphi \in [\varphi(0), \infty)$. For illustration we plot in Fig. 8.7.5 the time in units of $\varphi$-time, when we set $\varphi(0) = 0$, $c = 1$, $b = -0.05$ and $s_0 = 0.2$.

In Fig. 8.7.6 we show the path $X_{\varphi(t)}$ of the squared Bessel process $X$ in dependence on time $t$. It will be suggested in Sect. 13.2 under the MMM that such a time transformed squared Bessel process of dimension $\delta = 4$ is closely matching the dynamics of the discounted market portfolio. For comparison we plot for the previous example of a squared Bessel process in Fig. 8.7.7 the expected value $E(X_{\varphi(t)} | \mathcal{A}_0)$ of $X_{\varphi(t)}$, see (8.7.15), in dependence on time $t$ when using the above default parameters.

Let us visualize for the above example in Fig. 8.7.8 also the expected value of the squared Bessel process of dimension $\delta = 4$ with respect to time $t$. If one compares the Figs. 8.7.4 and 8.7.8, then one notes that after about five years, that is $t = 5$, the expected value of the inverse of the squared Bessel process starts to decline noticeably in our example.

We then show in the case of our example for comparison the inverse of the squared Bessel process of dimension $\delta = 4$, which we plotted in Fig. 8.7.3 in $\varphi$-time, in Fig. 8.7.9 in dependence on time $t$. One observes the typical systematic decline of a strict supermartingale.

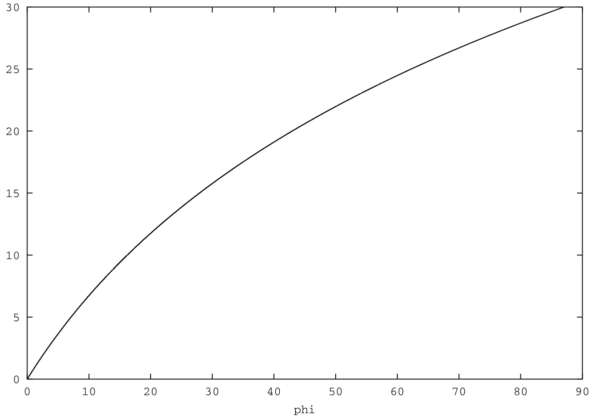



**Fig. 8.7.5.** Time $t(\varphi)$ against $\varphi$ time

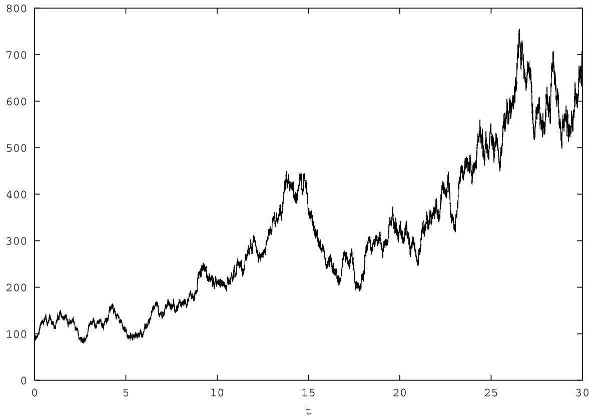



**Fig. 8.7.6.** Squared Bessel process in dependence on time $t$

## Square Root Process (*)

We shall now demonstrate the close relationship of a *square root* (SR) *process* with a squared Bessel process $X$. Given a squared Bessel process $X$ of dimension $\delta > 0$ and using our previous notation we introduce then the SR process

$$Y = \left\{Y_t = s_t \, X_{\varphi(t)}, \, t \in [0, \infty)\right\}$$

of dimension $\delta > 0$ in dependence on time $t$, by the transformation

$$Y_t = s_t \, X_{\varphi(t)} \tag{8.7.30}$$

for $t \in [0, \infty)$, see also Delbaen & Shirakawa (1997). Using (8.7.1), (8.7.26), (8.7.27), (8.7.30) and applying the Itô formula (6.2.11), the SDE for the SR process $Y$ follows as

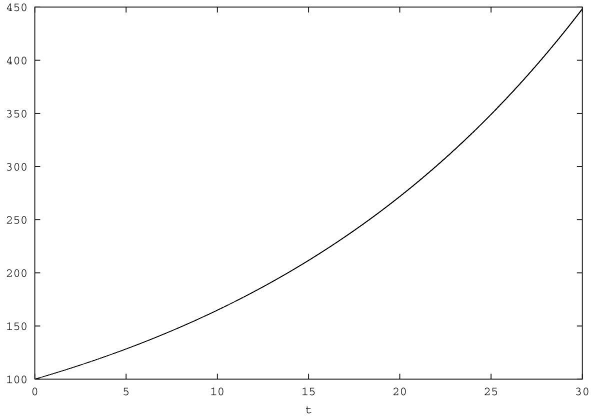



**Fig. 8.7.7.** Expectation of a squared Bessel process in dependence on time $t$

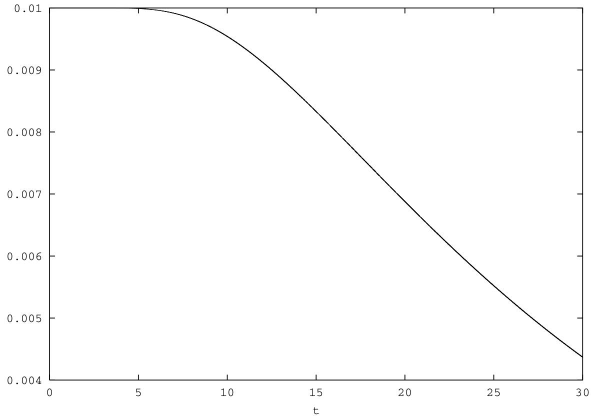



**Fig. 8.7.8.** Expectation of the inverse of a squared Bessel process in dependence on time $t$

$$\begin{aligned} dY_t &= s_t\, dX_{\varphi(t)} + X_{\varphi(t)}\, ds_t \\ &= s_t\, \delta\, d\varphi(t) + s_t\, 2\sqrt{X_{\varphi(t)}}\, dW_{\varphi(t)} + X_{\varphi(t)}\, s_t\, b_t\, dt \\ &= \left( \frac{\delta}{4}\, c_t^2 + b_t\, Y_t \right) dt + c_t \sqrt{Y_t} \sqrt{\frac{4\, s_t}{c_t^2}}\, dW_{\varphi(t)} \qquad (8.7.31) \end{aligned}$$

for $t \in [0, \infty)$ and $Y_0 = s_0\, X_{\varphi(0)} > 0$. Note that $W = \{W_\varphi,\, \varphi \in [\varphi(0), \infty)\}$ is a Wiener process in the transformed $\varphi$-time $\varphi(t) \in [\varphi(0), \infty)$, which is linked to the time $t$ by (8.7.27). The martingale $U = \{U_t,\, t \in [0, \infty)\}$ with the stochastic differential

$$dU_t = \sqrt{\frac{4\, s_t}{c_t^2}}\, dW_{\varphi(t)} \qquad (8.7.32)$$

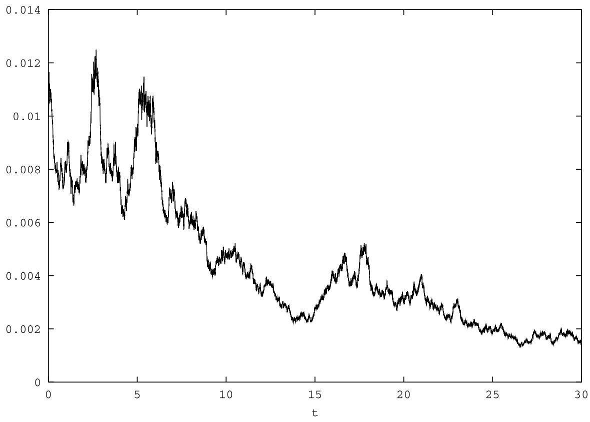



**Fig. 8.7.9.** Inverse of squared Bessel process in dependence on time $t$

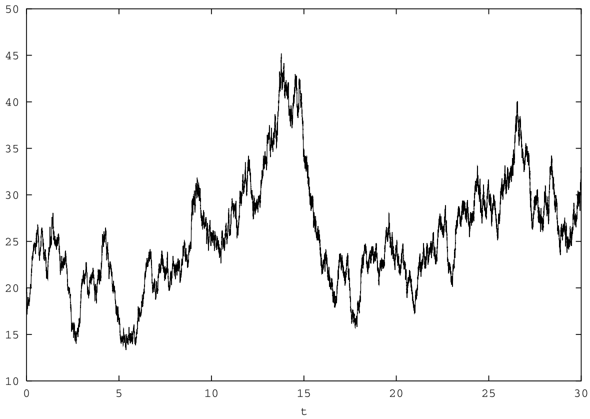



**Fig. 8.7.10.** Sample path of a square root process in dependence on time $t$

has the quadratic variation

$$[U]_t = \int_0^t \frac{4\, s_z}{c_z^2}\, d\varphi(z) = t. \tag{8.7.33}$$

By Lévy's theorem, see Theorem 6.5.1, the process $U = \{U_t,\, t \in [0,\infty)\}$ is then a Wiener process with respect to $t \in [0,\infty)$ on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. Thus, we have from (8.7.31) and (8.7.32) the SDE

$$dY_t = \left(\frac{\delta}{4}\, c_t^2 + b_t\, Y_t\right) dt + c_t\, \sqrt{Y_t}\, dU_t \tag{8.7.34}$$

for $t \in [0,\infty)$ for the SR process $Y$ with $Y_0 = s_0\, X_{\varphi(0)}$. For an appropriate choice of $b$, $c$ and $\delta$ the process $Y$ expresses the SR process mentioned in (4.4.6) and (7.5.15). Figure 8.7.10 displays for our example the path of the corresponding SR process of dimension $\delta = 4$ in dependence on time $t$. For the visualization of the transition density of an SR process we refer to Fig. 4.4.1.

For $\delta > 2$ the transformation (8.7.30) allows us to reduce the characterization of the probability density for $Y_t$, see (8.7.43), to that of determining $p_\delta(\varphi(0), \frac{Y_0}{s_0}; \varphi(t), \frac{Y_t}{s_t})$, which is given in (8.7.9). It follows from (8.7.27), (8.7.14) and (8.7.16) for $\alpha > -\frac{\delta}{2}$, $t \in (0, \infty)$ and $\delta > 2$ the $\alpha$th moment

$$E\left(Y_t^\alpha \,\middle|\, \mathcal{A}_0\right) = (2\,\bar{\varphi}_t\,\bar{s}_t)^\alpha \exp\left\{-\frac{Y_0}{2\,\bar{\varphi}_t}\right\} \sum_{k=0}^\infty \left(\frac{Y_0}{2\,\bar{\varphi}_t}\right)^k \frac{\Gamma(\alpha + k + \frac{\delta}{2})}{k!\,\Gamma(k + \frac{\delta}{2})} \tag{8.7.35}$$

and if additionally $\alpha \in (-\frac{\delta}{2}, 0)$ the estimate

$$E\left(Y_t^\alpha \,\middle|\, \mathcal{A}_0\right) \leq (2\,\bar{\varphi}_t\,\bar{s}_t)^\alpha \exp\left\{-\frac{Y_0}{2\,\bar{\varphi}_t}\right\} \left(\frac{\Gamma(\alpha + \frac{\delta}{2})}{\Gamma(\frac{\delta}{2})} + \exp\left\{\frac{Y_0}{2\,\bar{\varphi}_t}\right\}\right)$$
$$< \infty, \tag{8.7.36}$$

where

$$\bar{s}_t = \frac{s_t}{s_0} = \exp\left\{\int_0^t b_u\,du\right\} \tag{8.7.37}$$

and

$$\bar{\varphi}_t = s_0\,(\varphi(t) - \varphi(0)) = \frac{1}{4}\int_0^t \frac{c_u^2}{\bar{s}_u}\,du \tag{8.7.38}$$

for $t \in [0, \infty)$. Note that $\bar{\varphi}_t$ and the above moments do not depend on the choice of the parameter $s_0$, which cancels due to the structure of the functions $\varphi(t)$ and $s_t$.

By using the SDE (8.7.34) the first moment of the SR process value $Y_t$ can be shown to have the form

$$E\left(Y_t \,\middle|\, \mathcal{A}_0\right) = E\left(Y_0 \,\middle|\, \mathcal{A}_0\right) \exp\left\{\int_0^t b_s\,ds\right\} + \int_0^t \frac{\delta}{4}\,c_s^2 \exp\left\{\int_s^t b_z\,dz\right\} ds \tag{8.7.39}$$

for $t \in [0, \infty)$.

For the special case $\delta = 4$ and $\alpha = -1$ we obtain from (8.7.17) and (8.7.27)–(8.7.30) the first order negative moment of $Y_t$ in the form

$$E\left(Y_t^{-1} \,\middle|\, \mathcal{A}_0\right) = \frac{1 - \exp\left\{-\frac{Y_0}{2\,\bar{\varphi}_t}\right\}}{Y_0\,\bar{s}_t}. \tag{8.7.40}$$

For $\delta > 2$, $c_t^2 = c^2 > 0$ and $b_t = b < 0$ the resulting SR process $Y = \{Y_t,\, t \in [0, \infty)\}$ with SDE (8.7.34) is ergodic, see Sect. 4.5. For the case $\delta = 4$ it has linear mean reversion with speed of adjustment parameter $-b$ and reference level $-\frac{c^2}{b}$. Thus, we obtain by (8.7.39) for an ergodic SR process $Y$ the long term mean

$$\lim_{t \to \infty} E\left(Y_t \,\middle|\, \mathcal{A}_0\right) = -\frac{c^2}{b} \tag{8.7.41}$$

and the first order negative moment

$$\lim_{t\to\infty} E\left(Y_t^{-1}\,\middle|\,\mathcal{A}_0\right) = -2\,\frac{b}{c^2}. \tag{8.7.42}$$

We have for the SR process $Y = \{Y_t,\, t \in [0,\infty)\}$ an analytical transition density $p(s, Y_s; t, Y_t)$ that follows from (8.7.9) and (8.7.30) in the form

$$p(s, Y_s; t, Y_t) = \frac{p_\delta\left(\varphi(s), \frac{Y_s}{s_s}; \varphi(t), \frac{Y_t}{s_t}\right)}{s_t} \tag{8.7.43}$$

for $0 \le s < t < \infty$. In the case when $\delta > 2$, $b_t = b < 0$ and $c_t = c \neq 0$ the ergodic SR process $Y$ has the transition density

$$p(0, x; t, y) = \frac{1}{2\bar{s}_t\,\bar{\varphi}_t}\left(\frac{y}{x\,\bar{s}_t}\right)^{\frac{\nu}{2}} \exp\left\{-\frac{x + \frac{y}{\bar{s}_t}}{2\,\bar{\varphi}_t}\right\} I_\nu\left(\frac{\sqrt{x\,\frac{y}{\bar{s}_t}}}{\bar{\varphi}_t}\right) \tag{8.7.44}$$

for $0 < t < \infty$ and $x, y \in (0,\infty)$, where $\nu = \frac{\delta}{2} - 1$, $\bar{s}_t = \exp\{b\,t\}$ and $\bar{\varphi}_t = \frac{c^2}{4\,b}(1 - \frac{1}{\bar{s}_t})$. It has then as stationary density a gamma density, which can be obtained via (4.5.20) in the form

$$p_{Y_\infty}(y) = \frac{\left(\frac{-2\,b}{c^2}\right)^{\frac{\delta}{2}}\, y^{\frac{\delta}{2}-1}}{\Gamma(\frac{\delta}{2})}\, \exp\left\{\frac{2\,b}{c^2}\,y\right\}. \tag{8.7.45}$$

The variance equals in this case

$$E\left((Y_t - E(Y_t))^2\,\middle|\,\mathcal{A}_0\right) = Y_0\,\frac{c^2}{b}\,(\exp\{2\,b\,t\} - \exp\{b\,t\}) + \frac{\delta\,c^4}{8\,b^2}\,(1 - \exp\{b\,t\})^2 \tag{8.7.46}$$

for $t \in [0,\infty)$.

### Affine Process (*)

Let us now further transform the above SR process given by (8.7.30) to cover the class of *affine processes*, see Duffie & Kan (1994) and Sect. 4.5. These processes have affine, that is linear, drift and linear squared diffusion coefficient functions. Here, we simply shift the SR process by a nonnegative, differentiable, deterministic function of time $a : [0,\infty) \to [0,\infty)$ defined through its derivative

$$a'_t = \frac{da_t}{dt} \tag{8.7.47}$$

for $t \in [0,\infty)$ with $a_0 \in [0,\infty)$. More precisely, we define the process $R = \{R_t,\, t \in [0,\infty)\}$ with

$$R_t = Y_t + a_t \tag{8.7.48}$$

for $t \in [0, \infty)$. Since $Y$ is nonnegative also $R$ remains nonnegative. By the Itô formula we obtain from (8.7.48) and (8.7.47) the SDE

$$dR_t = \left( \frac{\delta}{4} \, c_t^2 + a_t' - b_t \, a_t + b_t \, R_t \right) dt + c_t \sqrt{R_t - a_t} \, dU_t \tag{8.7.49}$$

for $t \in [0, \infty)$ with $R_0 = Y_0 + a_0$. This means that the transform

$$R_t = s_t \, X_{\varphi(t)} + a_t \tag{8.7.50}$$

of a squared Bessel process $X$ of dimension $\delta$ yields an *affine diffusion process*, see (4.5.14) and (4.5.15), which satisfies the SDE (8.7.49).

## 8.8 Exercises for Chapter 8

**8.1.** Show for the BS model that the discounted European call option price of the discounted Black-Scholes formula satisfies the discounted Black-Scholes partial differential equation with corresponding terminal condition.

**8.2.** Derive the expression for the hedge ratio of the European call option in the BS model using the discounted BS-PDE.

**8.3.** Derive the gamma of a European put option for the BS model.

**8.4.** Compute, for a European put option under the BS model, the number of units $\delta_t^0$ to be held at a given time $t$ in the domestic savings account.

**8.5.** Transform the discounted BS-PDE for a discounted European option price into a corresponding BS-PDE for the corresponding undiscounted option price as a function of time and undiscounted underlying security.

**8.6.** Show for the European put option under the BS model that the corresponding P&L process is zero.

**8.7.** Derive the first moment of a square root process with constant parameters $c > 0$, $b < 0$ and dimension $\delta > 2$ satisfying the SDE

$$dY_t = \left( \frac{\delta}{4} \, c^2 + b \, Y_t \right) dt + c \sqrt{Y_t} \, dW_t$$

for $t \in [0, \infty)$ and $Y_0 > 0$, where $W$ is a Wiener process.

**8.8.** (*) Derive the moments for the squared Bessel process with dimension $\delta > 2$ including moments of negative order, as long as they exist, and show estimates of the type (8.7.16).

**8.9.** (*) Show by using the transition density $p_\delta$ of a squared Bessel process of dimension $\delta > 2$ that

$$\int_0^\infty y^{1-\frac{\delta}{2}}\, p_\delta(0,x;\varphi,y)\,dy = x^{1-\frac{\delta}{2}} \int_0^\infty p_\delta(0,y;\varphi,x)\,dy.$$

**8.10.** (*) Show with the transition density $p_\delta$ of a squared Bessel process of dimension $\delta > 2$ that

$$\int_0^\infty p_\delta(0,y;\varphi,x)\,dy = \left(1 - \frac{\Gamma\left(\frac{\delta}{2}-1, \frac{x}{2\varphi}\right)}{\Gamma(\frac{\delta}{2}-1)}\right).$$

# 9

# Various Approaches to Asset Pricing

A fundamental result of this chapter is that prices can be generally obtained under the benchmark approach in situations where other approaches are not available. This chapter also clarifies relationships between real world pricing under the benchmark approach and the pricing by other means in the areas of finance and insurance. Furthermore, it presents the Girsanov transformation, the change of numeraire technique and the Feynman-Kac formula, which are all highly relevant to derivative pricing.

## 9.1 Real World Pricing

### Various Pricing Approaches

In the literature, pricing concepts for risky securities have been developed in several seemingly different approaches. Often one determines the price of an asset by reference to its underlying economic value. General equilibrium based models, such as the *intertemporal capital asset pricing model* (ICAPM), see Merton (1973a), provide examples of this approach. The *actuarial pricing approach*, see Bühlmann (1970) and Gerber (1997), which is common in insurance and accounting, provides another important example in this direction. The above mentioned approaches aim to provide an economic explanation for the value of prices and why asset prices move if changes in economic variables occur.

A much less ambitious question is asked in pricing approaches which arise when one is marking to market. Given the prices of some assets that securitize uncertainty in the market, one analyzes under such approach what consequences this has for the values of other securities in this market. The securities to be priced are typically derivatives. The previously described option pricing methodology of Chap. 8 provides an example for such a pricing approach, which is based on the assumption that there is *no arbitrage*, see Ross (1976), Harrison & Kreps (1979) and Harrison & Pliska (1981). As we

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

shall see in Sect. 9.4, within the *arbitrage pricing theory* (APT) the *risk neutral* pricing approach has been developed that allows convenient *numeraire changes* and corresponding changes of pricing measures.

It is a challenge to reconcile presently used different pricing approaches and to highlight their specific features in a consistent framework. The *benchmark approach* uses the *growth optimal portfolio* (GOP) as reference unit or numeraire. As we shall see, the GOP is the portfolio that maximizes expected logarithmic utility from terminal wealth, see Kelly (1956) and Long (1990). The GOP exists in all financial market models that we shall consider. In the next chapter it will be made clear for a continuous market what is the composition of the GOP. Chapter 14 will generalize this result to jump diffusion markets. For the purpose of this chapter we keep the market model as BS model and, therefore, the GOP very simple. We shall unify in a natural way some of the mentioned pricing methods under the benchmark approach by using the concept of *real world pricing*. To illustrate the different asset pricing methodologies we explore in this chapter various alternative ways to price a future payoff. We discuss the different approaches typically in the context of the BS model, which considerable simplifies our presentation. However, most of the conclusions apply also for other models, as we shall see later in Chaps. 10–14.

First, we introduce in the following the real world pricing concept that allows prices for payoffs to be obtained as conditional expectations under the real world probability measure. We then show in later sections how the benchmark approach relates to other pricing concepts. The advantage of the benchmark approach is that as soon as the GOP exists one can always perform real world pricing. Other approaches may have extra conditions to satisfy which may not allow to form derivative prices for certain models of interest.

### Portfolios under the BS Model

In the previous chapter, we have identified via no-arbitrage and hedging arguments a price for a European option under the BS model. If one wants to exclude arbitrage, then there is no alternative to this price. We now translate this result into a pricing concept that is based on some conditional expectation. To achieve this we express this price as a conditional expectation of the option payoff. The expectation will be taken under the real world probability measure $P$. This is the probability measure that models the market as it evolves and as we can observe it by exploiting empirical evidence. Only under this measure one can estimate model parameters historically. We shall show later in Sect. 10.6 how to obtain the GOP without reference to a specific model and the estimation of particular parameters. The key question that has to be resolved is: In the denomination of which *numeraire* should one express the payoff to apply an expectation under the real world probability measure?

With this goal in mind, we ask whether there exists a strictly positive process, for instance, a market index, which when used as numeraire or *bench-*

*mark*, generates realistic benchmarked derivative price processes that are martingales with respect to the real world probability measure. This means that benchmarked derivative prices then represent the best forecast of their future benchmarked values. In this way a natural pricing method could be established via conditional expectation under the real world probability measure. The described use of the GOP as numeraire portfolio follows the line of arguments in Long (1990), Bajeux-Besnainou & Portait (1997), Becherer (2001) and Bühlmann & Platen (2003). Since in the case of a European option under the BS model we have already identified the corresponding no-arbitrage price, we now aim to identify the corresponding benchmark that, when used as numeraire, yields this price that allows to replicate the given payoff.

As already indicated, for simplicity, we consider here a simple Black-Scholes (BS) market. It contains an underlying security with price process $S = \{S_t,\ t \in [0,T]\}$, as given by (8.2.1), which satisfies the SDE

$$dS_t = a_t\, S_t\, dt + \sigma_t\, S_t\, dW_t \tag{9.1.1}$$

for $t \in [0,T]$ with $S_0 > 0$, where $T \in [0,\infty)$. Furthermore, our BS model has a domestic savings account with value process $B = \{B_t,\ t \in [0,T]\}$, see (8.2.2), where

$$dB_t = r_t\, B_t\, dt \tag{9.1.2}$$

for $t \in [0,T]$ and $B_0 = 1$.

A self-financing strategy $\delta = \{\delta_t = (\delta_t^0, \delta_t^1)^\top,\ t \in [0,T]\}$, see (8.2.4)–(8.2.7), with $\delta_t^0$ units held at time $t$ in the domestic savings account and $\delta_t^1$ units invested in the underlying security, has the corresponding portfolio value

$$S_t^\delta = \delta_t^0\, B_t + \delta_t^1\, S_t \tag{9.1.3}$$

with

$$\begin{aligned} dS_t^\delta &= \delta_t^0\, dB_t + \delta_t^1\, dS_t \\ &= (\delta_t^0\, r_t\, B_t + \delta_t^1\, a_t\, S_t)\, dt + \delta_t^1\, \sigma_t\, S_t\, dW_t \\ &= S_t^\delta \left( (\pi_\delta^0(t)\, r_t + \pi_\delta^1(t)\, a_t)\, dt + \pi_\delta^1(t)\, \sigma_t\, dW_t \right) \end{aligned} \tag{9.1.4}$$

for $t \in [0,T]$. Note that the SDE (9.1.4) is such that it guarantees the self-financing property of the portfolio, where all changes of its value are due to changes in the securities. Here we use the corresponding *fractions*

$$\pi_\delta^0(t) = \delta_t^0\, \frac{B_t}{S_t^\delta} \tag{9.1.5}$$

and

$$\pi_\delta^1(t) = \delta_t^1\, \frac{S_t}{S_t^\delta} \tag{9.1.6}$$

that are held in the respective securities. Obviously, these fractions add up to one, that is,

$$\pi_\delta^0(t) + \pi_\delta^1(t) = 1 \tag{9.1.7}$$

for $t \in [0, T]$. Note that the notion of a fraction makes only sense as long as the portfolio value is not zero.

### Growth Optimal Portfolio

Let us derive for the given BS market the *growth optimal portfolio* (GOP) which will be shown in Chap. 10 to be the portfolio that maximizes the drift of its logarithm, see Long (1990), Karatzas & Shreve (1998) or Platen (2002). By the Itô formula we obtain from (9.1.4) and (9.1.7) for the logarithm $\ln(S_t^\delta)$ of a strictly positive portfolio the SDE

$$d\ln(S_t^\delta) = g_t^\delta\, dt + \pi_\delta^1(t)\, \sigma_t\, dW_t \tag{9.1.8}$$

with *growth rate*

$$g_t^\delta = r_t + \pi_\delta^1(t)\,(a_t - r_t) - \frac{1}{2}\,(\pi_\delta^1(t))^2\, \sigma_t^2 \tag{9.1.9}$$

for $t \in [0, T]$.

**Definition 9.1.1.** *Under the BS model the GOP is the portfolio process* $S^{\delta_*} = \{S_t^{\delta_*},\ t \in [0, T]\}$ *with optimal growth rate* $g_t^{\delta_*}$ *at time t such that*

$$g_t^\delta \leq g_t^{\delta_*} \tag{9.1.10}$$

*almost surely for all* $t \in [0, T]$ *and strictly positive portfolio processes* $S^\delta$*.*

Let us now choose the fraction $\pi_\delta^1(t)$ such that the growth rate $g_t^\delta$ is maximized for each $t \in [0, T]$, which will give us the GOP. Note that the choice of the reference unit is *not* relevant for the corresponding optimization problem. By application of the first order condition to maximize the growth rate $g_t^\delta$ in (9.1.9) with respect to the fraction $\pi_\delta^1(t)$ we obtain the condition

$$\frac{\partial g_t^\delta}{\partial \pi_\delta^1(t)} = a_t - r_t - \pi_{\delta_*}^1(t)\, \sigma_t^2 = 0 \tag{9.1.11}$$

for $t \in [0, T]$. Therefore, we obtain the *optimal fraction* in the underlying security

$$\pi_{\delta_*}^1(t) = \frac{a_t - r_t}{\sigma_t^2} \tag{9.1.12}$$

and, thus, by (9.1.7) the optimal fraction in the savings account

$$\pi_{\delta_*}^0(t) = 1 - \pi_{\delta_*}^1(t) \tag{9.1.13}$$

for $t \in [0, T]$. Because of (9.1.9) and (9.1.12) the *optimal growth rate* is then of the form

$$g_t^{\delta_*} = r_t + \frac{1}{2}\left(\frac{a_t - r_t}{\sigma_t}\right)^2 \tag{9.1.14}$$

for $t \in [0,T]$. Now, we obtain from (9.1.4), (9.1.12) and (9.1.13) the GOP as the wealth process $S^{\delta_*} = \{S_t^{\delta_*},\, t \in [0,T]\}$, which satisfies the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*}\, \big((r_t + \theta_t^2)\, dt + \theta_t\, dW_t\big) \tag{9.1.15}$$

with initial value $S_0^{\delta_*} > 0$ and GOP volatility

$$\theta_t = \pi_{\delta_*}^1(t)\, \sigma_t = \frac{a_t - r_t}{\sigma_t} \tag{9.1.16}$$

for $t \in [0,T]$. The quantity $\theta_t$ in (9.1.16) is the, so-called, *market price of risk* at time $t$.

According to (9.1.14) and (9.1.16) the optimal growth rate for the given BS model equals

$$g_t^{\delta_*} = r_t + \frac{1}{2}\,\theta_t^2 \tag{9.1.17}$$

for $t \in [0,T]$. This reveals a close link between the squared volatility and the optimal growth rate of the GOP. For the *discounted* GOP

$$\bar{S}_t^{\delta_*} = \frac{S_t^{\delta_*}}{B_t} \tag{9.1.18}$$

we derive by the Itô formula with (9.1.15) and (9.1.2) the SDE

$$d\bar{S}_t^{\delta_*} = \bar{S}_t^{\delta_*}\, \theta_t\, (\theta_t\, dt + dW_t) \tag{9.1.19}$$

for $t \in [0,T]$, see (10.2.8). Note that the drift of the discounted GOP is determined as the square of its diffusion coefficient. This observation is crucial and holds also more generally for continuous financial markets, as we shall see in Chap. 10. Within this chapter we keep our BS market very simple. Therefore, the GOP is here only a composition of two securities.

### Benchmarked Savings Account

Let us now introduce the notion of *benchmarking*. Any security when expressed in units of the GOP we call a *benchmarked security*. For instance, the savings account $B$, when denominated in units of the GOP, is called the *benchmarked savings account* $\hat{S}^0 = \{\hat{S}_t^0,\, t \in [0,T]\}$, where

$$\hat{S}_t^0 = \frac{B_t}{S_t^{\delta_*}} \tag{9.1.20}$$

for $t \in [0,T]$. By application of the Itô formula (6.2.11) to the inverse of $\bar{S}_t^{\delta_*}$ in (9.1.19) or the relation (9.1.20), the differential equation (9.1.2) and the SDE (9.1.15), it follows for the benchmarked savings account $\hat{S}_t^0$ the SDE

$$d\hat{S}^0_t = -\theta_t\,\hat{S}^0_t\,dW_t \tag{9.1.21}$$

for $t \in [0,T]$. This means that the benchmarked savings account is driftless.

Since the process $\hat{S}^0 = \{\hat{S}^0_t,\ t \in [0,T]\}$ is a geometric Brownian motion it follows by (5.4.1) and (7.3.8) that $\hat{S}^0 \in \mathcal{L}^2_T$. Thus, the Itô integral on the right hand side of the integral version of (9.1.21) is, by the martingale property (5.4.3) of Itô integrals, a martingale. This means that the benchmarked savings account process $\hat{S}^0$ is under the given BS model an $(\underline{\mathcal{A}}, P)$-martingale. We shall see later that this is a particular property of the BS model and may not hold for other models.

### Benchmarked Underlying Security

Let us now benchmark in our BS market the underlying security $S$. That is, we consider the *benchmarked* security price

$$\hat{S}^1_t = \frac{S_t}{S^{\delta_*}_t} \tag{9.1.22}$$

for $t \in [0,T]$. Then by the Itô formula (6.2.11) together with (9.1.1), (9.1.15) and equation (9.1.16) the SDE for $\hat{S}^1_t$ becomes

$$\begin{aligned} d\hat{S}^1_t &= \hat{S}^1_t\,\big((a_t - r_t - \sigma_t\,\theta_t)\,dt + (\sigma_t - \theta_t)\,dW_t\big) \\ &= \hat{S}^1_t\,(\sigma_t - \theta_t)\,dW_t \end{aligned} \tag{9.1.23}$$

for $t \in [0,T]$. Consequently, according to (9.1.23), the GOP when used as benchmark, has the property that the resulting SDE for the benchmarked security $\hat{S}^1$ is driftless. By similar arguments as above one can show that the geometric Brownian motion $\hat{S}^1$ is in $\mathcal{L}^2_T$. The process $\hat{S}^1$ is, therefore, by (5.4.3) an $(\underline{\mathcal{A}}, P)$-martingale.

### Benchmarked Option Price

Consider a European option on the underlying security $S$ under the BS model with value for its hedge portfolio

$$V(t) = V(t, S_t), \tag{9.1.24}$$

as determined in Sect. 8.2 by equation (8.2.4). Using (8.2.9) and (9.1.20), we obtain for the benchmarked European option price the expression

$$\hat{V}(t) = \frac{V(t)}{S^{\delta_*}_t} = \frac{V(t,S_t)}{S^{\delta_*}_t} = \bar{V}(t,\bar{S}_t)\,\hat{S}^0_t \tag{9.1.25}$$

for $t \in [0,T]$. Here we have the benchmarked payoff

$$\hat{V}(T) = \frac{H(S_T)}{S_T^{\delta_*}} \tag{9.1.26}$$

at maturity $T$. By application of the Itô formula (6.2.11) we obtain from (8.2.11) and (8.2.21) for the discounted value $\bar{V}(t, \bar{S}_t) = \frac{V(t,S_t)}{B_t}$ of the option with $\bar{S}_t = \frac{S_t}{B_t}$ the SDE

$$\begin{aligned} d\bar{V}(t, \bar{S}_t) &= \left( \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial t} + (a_t - r_t)\, \bar{S}_t\, \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}} + \frac{1}{2}\, \sigma_t^2\, \bar{S}_t^2\, \frac{\partial^2 \bar{V}(t, \bar{S}_t)}{\partial \bar{S}^2} \right) dt \\ &\quad + \sigma_t\, \bar{S}_t\, \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}}\, dW_t \\ &= \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}}\, \bar{S}_t\, ((a_t - r_t)\, dt + \sigma_t\, dW_t) \end{aligned} \tag{9.1.27}$$

for $t \in [0, T]$. On the other hand, by using the Itô formula (6.2.11) and also the relations (9.1.25), (9.1.27), (8.2.9), (9.1.16) and (9.1.21) we obtain the SDE

$$\begin{aligned} d\hat{V}(t) &= d(\bar{V}\, \hat{S}_t^0) \\ &= \hat{S}_t^0\, d\bar{V} + \bar{V}\, d\hat{S}_t^0 + d[\bar{V}, \hat{S}^0]_t \\ &= \hat{S}_t^0\, (a_t - r_t)\, \bar{S}_t\, \frac{\partial \bar{V}}{\partial \bar{S}}\, dt + \hat{S}_t^0\, \sigma_t\, \bar{S}_t\, \frac{\partial \bar{V}}{\partial \bar{S}}\, dW_t \\ &\quad - \theta_t\, \bar{V}\, \hat{S}_t^0\, dW_t - \sigma_t\, \bar{S}_t\, \frac{\partial \bar{V}}{\partial \bar{S}}\, \hat{S}_t^0\, \theta_t\, dt \end{aligned} \tag{9.1.28}$$

for $t \in [0, T)$, where, for simplicity, we have suppressed in our notation the dependence of $\bar{V}$ on $(t, \bar{S}_t)$. By using (9.1.16) the SDE (9.1.28) can be rewritten in the form

$$d\hat{V}(t) = \hat{S}_t^0 \left( \sigma_t\, \bar{S}_t\, \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}} - \theta_t\, \bar{V}(t, \bar{S}_t) \right) dW_t \tag{9.1.29}$$

for $t \in [0, T)$. Note that the SDE for the benchmarked option price is driftless. Also here one can show that the diffusion coefficient in (9.1.29) is in $\mathcal{L}_T^2$. Therefore, by (5.4.3), the benchmarked option price process $\hat{V} = \{\hat{V}(t),\, t \in [0, T]\}$, is an $(\underline{\mathcal{A}}, P)$-martingale.

We have seen that the property of the GOP when used as numeraire or benchmark, to convert benchmarked prices into martingales, seems to apply quite generally under the BS model. In the literature the GOP is therefore also known as the *numeraire portfolio*, see Long (1990).

### Real World Pricing

Summarizing the above analysis, we conclude under the BS model that the GOP is the numeraire portfolio for the domestic savings account $B$, the underlying security price $S$ and the option price $V$. When used as denominator it makes the corresponding benchmarked price processes $\hat{S}^0$, $\hat{S}^1$ and $\hat{V}$ into $(\underline{\mathcal{A}}, P)$-martingales. This implies, by the martingale property (5.1.2), that these prices, when expressed in units of the GOP, are the best forecast of their future benchmarked values. We obtain this remarkable fact as a consequence of outstanding properties of the GOP, which we shall discuss in the next chapter.

Intuitively, the martingale property of benchmarked prices relates to the common notion of what constitutes a fair price. The following definition will be applied generally throughout the book for all models and not only for the BS model.

**Definition 9.1.2.** *A security price process $V = \{V_t,\, t \in [0,\infty)\}$ is called* fair *if its benchmarked value $\hat{V}_t = \frac{V_t}{S_t^{\delta_*}}$ forms an $(\underline{\mathcal{A}}, P)$-martingale.*

This leads by application of the martingale property of $\hat{V}$ directly to the following pricing formula.

**Corollary 9.1.3.** *For any fair security price process $V = \{V_t,\, t \in [0,\infty)\}$ one has for any time $t \in [0,\infty)$ and $T \in (t,\infty)$ the* real world pricing formula

$$V_t = S_t^{\delta_*}\, E\left(\frac{V_T}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t\right). \tag{9.1.30}$$

It is most important to emphasize that the expectation in (9.1.30) is taken under the real world probability measure $P$. The numeraire is here the GOP. Note that by application of the optional sampling theorem, see (5.1.19), it follows that $T$ can also be a bounded stopping time in the real world pricing formula (9.1.30).

Under the BS model the savings account, the underlying security and European option price form fair price processes since their benchmarked price processes are $(\underline{\mathcal{A}}, P)$-martingales. Note that a real world option price forms a fair price process and is under the given BS model consistent with the hedging arguments previously applied in Chap. 8. In this sense the real world or fair option price is a no-arbitrage price.

As we shall see later in this chapter, *real world pricing* can be generally applied and will turn out to be the natural pricing concept under the benchmark approach. It only requires the existence of a GOP, as can be seen from the real world pricing formula (9.1.30).

### A Martingale Representation

The SDE (9.1.29) for the benchmarked option price process $\hat{V}$ can be rewritten by using (9.1.20), (8.2.9) and (8.2.10) in the integral form

$$\hat{V}(T) = \hat{V}(t) + \int_t^T \left(\sigma_z\, \frac{S_z}{S_z^{\delta_*}}\, \frac{\partial \bar{V}(z, \bar{S}_z)}{\partial \bar{S}} - \theta_z\, \hat{V}(z)\right) dW_z \tag{9.1.31}$$

for $t \in [0, T]$. This provides a representation of the benchmarked option payoff

$$\hat{V}(T) = \frac{H(S_T)}{S_T^{\delta_*}}.$$

Since $\hat{V}$ is a martingale under the real world probability $P$, we call (9.1.31) the *real world martingale representation* of $\frac{H(S_T)}{S_T^{\delta_*}}$. By taking the conditional expectation $E(\cdot \,|\, \mathcal{A}_t)$ on both sides of equation (9.1.31), it follows by the martingale property of $\hat{V}$ that

$$E\left(\hat{V}(T) \,\big|\, \mathcal{A}_t\right) = \hat{V}(t) \tag{9.1.32}$$

for $t \in [0, T]$. Now, when we multiply both sides of equation (9.1.32) by $S_t^{\delta_*}$, then we obtain by (9.1.25) the fair option price $V(t)$ at time $t$ in the form

$$V(t) = S_t^{\delta_*}\, \hat{V}(t) = S_t^{\delta_*}\, E\left(\hat{V}(T) \,\big|\, \mathcal{A}_t\right) \tag{9.1.33}$$

for $t \in [0, T]$. Therefore, due to (9.1.25) and (8.2.24), we can express the European option price with payoff $H(S_T)$ at maturity $T$ by

$$V(t) = S_t^{\delta_*}\, E\left(\frac{H(S_T)}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) \tag{9.1.34}$$

for all $t \in [0, T]$. This recovers the real world pricing formula (9.1.30). It is most important to emphasize that this pricing formula uses the conditional expectation under the real world probability measure $P$ and not under any transformed measure.

The fair price $V(t)$, when expressed in units of the domestic currency at time $t$, is simply obtained by multiplying the fair benchmarked price $\hat{V}(t)$ by the GOP value $S_t^{\delta_*}$, that is

$$V(t) = S_t^{\delta_*}\, \hat{V}(t) \tag{9.1.35}$$

for $t \in [0, T]$, as is described by the real world pricing formula (9.1.34), see also (9.1.25) and (9.1.33).

We shall apply the real world pricing formula later quite generally when determining the fair price of derivatives. Once the GOP is identified in a model one can determine the fair value of any integrable benchmarked payoff by the real world pricing formula. As we shall see, it is possible to derive from this pricing formula several other common derivative pricing and asset pricing rules.

### Benchmarked Portfolios

For a given general portfolio $S^\delta$ we can also compute the SDE for its benchmarked value

$$\hat{S}^\delta_t = \frac{S^\delta_t}{S^{\delta_*}_t} \tag{9.1.36}$$

for $t \in [0,T]$. It follows from (9.1.4), (9.1.15), (9.1.16), (9.1.6) and by application of the Itô formula that

$$d\hat{S}^\delta_t = \hat{S}^\delta_t \left(\pi^1_\delta(t)\,\sigma_t - \theta_t\right) dW_t = \left(\delta^1_t\,\hat{S}^1_t\,\sigma_t - \hat{S}^\delta_t\,\theta_t\right) dW_t \tag{9.1.37}$$

for $t \in [0,T]$ with $\hat{S}^\delta_0 = \frac{S^\delta_0}{S^{\delta_*}_0}$. Obviously, $\hat{S}^\delta$ is an $(\underline{\mathcal{A}}, P)$-local martingale, see Lemma 5.4.1. It follows by Lemma 5.2.3 that under the given BS model any nonnegative benchmarked portfolio is an $(\underline{\mathcal{A}}, P)$-supermartingale.

In the case when $\hat{S}^\delta$ is such that the conditions (ii) or (iii) of Lemma 5.2.2 are satisfied, then $\hat{S}^\delta$ is also a true $(\underline{\mathcal{A}}, P)$-martingale and not just a supermartingale. This indicates that there may exist portfolio processes that when benchmarked are not martingales.

## An Unfair Portfolio

The following simple example demonstrates that even in a simple BS market there exist perfectly reasonable nonnegative portfolio processes that are *unfair*, which means that they are not fair. Since we have above observed that nonnegative benchmarked portfolios are always supermartingales an unfair portfolio is a supermartingale that is not a martingale.

To provide an example, let us introduce the inverse $Z = \{Z_t,\, t \in [0,T]\}$ of a squared Bessel process of dimension four, which we have shown in Sect. 8.7 to be a strict local martingale, see also Revuz & Yor (1999). It satisfies the SDE

$$dZ_t = -2\,(Z_t)^{\frac{3}{2}}\,dW_t \tag{9.1.38}$$

for $t \in [0,T]$, where we set $Z_0 = 1$. The process $Z$ is an $(\underline{\mathcal{A}}, P)$-local martingale but not an $(\underline{\mathcal{A}}, P)$-martingale. By Lemma 5.2.3 it is a strict $(\underline{\mathcal{A}}, P)$-supermartingale.

We can now identify a strategy $\delta$ with initial benchmarked portfolio value

$$\hat{S}^\delta_0 = Z_0 = 1 \tag{9.1.39}$$

that matches in the SDE (9.1.37) the diffusion coefficient such that

$$\hat{S}^\delta_t \left(\pi^1_\delta(t)\,\sigma_t - \theta_t\right) = -2\,(Z_t)^{\frac{3}{2}} \tag{9.1.40}$$

for all $t \in [0,T]$. Then it follows for the fraction

$$\pi^1_\delta(t) = \left(\theta_t - \frac{2\,(Z_t)^{\frac{3}{2}}}{\hat{S}^\delta_t}\right)\frac{1}{\sigma_t} \tag{9.1.41}$$

of wealth that is invested in the underlying security that the resulting self-financing portfolio $S^\delta$ has at time $t$ the benchmarked value

$$\hat{S}_t^\delta = Z_t \tag{9.1.42}$$

for all $t \in [0,T]$. This means, $\hat{S}^\delta$ equals the strict supermartingale $Z$, see (8.7.21). Note by (9.1.42) and (9.1.41) that the fraction simplifies in this case to the expression

$$\pi_\delta^1(t) = \frac{1}{\sigma_t}\left(\theta_t - 2\sqrt{Z_t}\right), \tag{9.1.43}$$

for $t \in [0,T]$. We emphasize that this yields a perfectly reasonable self-financing portfolio. As we have pointed out in Sect. 8.7, the formula (8.7.17) for the first negative moment of a squared Bessel processes of dimension four yields

$$E\left(\hat{S}_t^\delta \,\middle|\, \mathcal{A}_0\right) = E\left(\hat{S}_t^\delta\right) = \hat{S}_0^\delta\left(1 - \exp\left\{\frac{-1}{2\,\hat{S}_0^\delta\, t}\right\}\right) < \hat{S}_0^\delta \tag{9.1.44}$$

for $t \in (0,T]$.

By the strict inequality (9.1.44) we see that the $(\underline{\mathcal{A}}, P)$-supermartingale $\hat{S}^\delta$ is here not a martingale. This example demonstrates that even under a BS model not all integrable, nonnegative, benchmarked portfolios are $(\underline{\mathcal{A}}, P)$-martingales.

We shall see later that generally under the benchmark approach all nonnegative benchmarked portfolios are supermartingales. This is a fundamental property for the wide class of financial market models that we consider in this book. There exist several popular pricing concepts that we shall discuss below. For some of these we can show that they correspond to real world pricing in the sense that their benchmarked price processes are martingales under the real world probability measure.

## 9.2 Actuarial Pricing

In this section we show that the common actuarial pricing or net present value pricing methodology, which is widely used in insurance and accounting, can be derived from real world pricing.

### Setup for the GOP

To illustrate the actuarial pricing method in a simple, familiar setting we consider again a BS market. The underlying security $S_t$ is not of relevance for the following analysis since the payoff $H$ that shall be priced, will be assumed to be independent of the GOP. However, it will be essential for our arguments that the financial market model has a GOP $S^{\delta_*} = \{S_t^{\delta_*},\, t \in [0,T]\}$ which, for simplicity, we assume to be of the same form as in the SDE (9.1.15). Since the GOP satisfies then a Black-Scholes dynamics of the type (7.3.12) we obtain from (7.3.3) an explicit expression for the GOP value at time $t$ in the form

$$S_t^{\delta_*} = S_0^{\delta_*} \exp\left\{\int_0^t \left(r_s + \frac{\theta_s^2}{2}\right) ds + \int_0^t \theta_s \, dW_s\right\} \tag{9.2.1}$$

for $t \in [0, T]$.

## Fair Zero Coupon Bond

To illustrate the actuarial pricing methodology let us at first determine at time $t$ the fair value of a *zero coupon bond*. This is the value at time $t$ for the payment of one monetary unit at time $T$, obtained under the real world pricing formula (9.1.34). Obviously, this corresponds to a European payoff $H = 1$. If we denote the fair value of this payoff at time $t \in [0, T]$ by $P(t, T)$, then we obtain by (9.1.34) the *fair zero coupon bond* price in the form

$$P(t, T) = S_t^{\delta_*} \, E\left(\frac{1}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t\right). \tag{9.2.2}$$

In the given case we can explicitly compute this value. Since $r$ and $\theta$ are deterministic, it follows from (9.2.2) and (9.2.1) that

$$\begin{aligned} P(t, T) &= E\left(\exp\left\{-\int_t^T r_s \, ds - \frac{1}{2}\int_t^T \theta_s^2 \, ds - \int_t^T \theta_s \, dW_s\right\} \,\bigg|\, \mathcal{A}_t\right) \\ &= \exp\left\{-\int_t^T r_s \, ds\right\} E\left(\exp\left\{-\int_t^T \frac{\theta_s^2}{2} \, ds - \int_t^T \theta_s \, dW_s\right\} \,\bigg|\, \mathcal{A}_t\right) \end{aligned} \tag{9.2.3}$$

for $t \in [0, T]$. Using the Laplace transform (1.3.76) of a Gaussian random variable it follows that the conditional expectation on the right hand side of (9.2.3) equals the real value one. Alternatively, we can use the fact that the exponential under the conditional expectation forms an $(\underline{\mathcal{A}}, P)$-martingale, see Sect. 5.1. Therefore, we obtain as fair zero coupon bond price at time $t$ the value

$$P(t, T) = \exp\left\{-\int_t^T r_s \, ds\right\} = \frac{B_t}{B_T}. \tag{9.2.4}$$

Note that the value $\exp\{-\int_t^T r_s ds\}$, if invested at time $t = 0$ in a savings account, has the value of one monetary unit at time $T$. Under the benchmark approach it will be always possible to establish the fair price of a zero coupon bond. However, if the exponential under the conditional expectation in (9.2.3) is a strict supermartingale, then the conditional expectation is less than one and $P(t, T)$ is less than the right hand side of (9.2.4). This is similar to the effect that yielded inequality (9.1.44). We shall study such cases later in more detail.

### Fair Price of an Independent Payoff

Now, let us consider at the fixed maturity date $T$ a random $\mathcal{A}_T$-measurable payoff $H > 0$, which is *independent* of the GOP value $S_T^{\delta_*}$. For instance, this could be a life insurance claim or a payoff based on a weather index. Such a claim may be by its nature independent of the GOP. The payoff $H$ at time $T$ could also model operational failures in a company during a period that finishes at maturity $T$. Alternatively, it could, for instance, model the total sum of insurance claims from a particular group of cars in the year prior to $T$. The key assumption is here that the above random payoff $H$ is independent of the random value $S_T^{\delta_*}$ of the GOP at the maturity date $T$. To be precise, we assume that $H$ is independent of $S_T^{\delta_*}$, see (1.1.13) and (1.4.22), and that the expectation of the benchmarked payoff

$$E\left(\left|\frac{H}{S_T^{\delta_*}}\right|\right) < \infty \tag{9.2.5}$$

is finite.

Then we can compute the fair price $U_H(t)$ at time $t \in [0, T]$ for the payoff $H$ according to the real world pricing formula (9.1.30). We obtain its fair price in the form

$$U_H(t) = S_t^{\delta_*}\, E\left(\frac{H}{S_T^{\delta_*}} \,\middle|\, \mathcal{A}_t\right).$$

Recall that the expectation of a product of independent random variables is the product of their expectations, see (1.4.25). Since we have assumed that $H$ is independent of $S_T^{\delta_*}$ we obtain by this property the expression

$$U_H(t) = S_t^{\delta_*}\, E\left(\frac{1}{S_T^{\delta_*}} \,\middle|\, \mathcal{A}_t\right) E\left(H \,\middle|\, \mathcal{A}_t\right).$$

By using now the fair zero coupon bond price $P(t, T)$ in (9.2.2), it follows the widely used *actuarial pricing formula*

$$U_H(t) = P(t, T)\, E\left(H \,\middle|\, \mathcal{A}_t\right). \tag{9.2.6}$$

Under this formula one computes the conditional expectation of a future cash flow at time $T$ and discounts it back to the present time $t$ by using the corresponding fair zero coupon bond price. This takes into account the evolution of the time value of money. The procedure is also known as *net present value* calculation. It is widely used in practice. Thus, we recover from real world pricing in the case of independence of payoff and GOP, the well-known formula of actuarial and net present value pricing.

Note that in the actuarial pricing formula (9.2.6) we do not require the knowledge of the dynamics of the GOP. We even do not need to observe the GOP in this case. One only needs to know the expectation of the payoff under the real world probability measure and the fair price of a zero coupon bond, which is given in the market.

In our simple BS model we obtain from (9.2.6) the following version of the actuarial pricing formula

$$U_H(t) = \frac{B_t}{B_T} E\left(H \,\big|\, \mathcal{A}_t\right) \tag{9.2.7}$$

for $t \in [0, T]$. We see in formula (9.2.7) the simple discounting rule for the expected future payoff, as is most common in actuarial and accounting practice. We emphasize that the conditional expectations in (9.2.6) and (9.2.7) are taken with respect to the real world probability measure $P$ and that these formulas are derived for the case when the payoff $H$ is independent of the GOP value $S_T^{\delta_*}$. The actuarial pricing formula (9.2.6) can be shown to hold generally for payoffs independent of the GOP for the models that we consider in this book. In this sense actuarial pricing turns out to be a particular case of real world pricing. On the other hand, when starting from a benchmarked actuarial price process $\hat{U}_H = \{\hat{U}_H(t) = \frac{U_H(t)}{S_t^{\delta_*}},\, t \in [0, \infty)\}$ with $H$ independent of $S_T^{\delta_*}$, it follows from the actuarial pricing formula (9.2.6) that the benchmarked actuarial price

$$\hat{U}_H(t) = \frac{P(t,T)}{S_t^{\delta_*}} E\left(H \,\big|\, \mathcal{A}_t\right) \tag{9.2.8}$$

is, as the product of independent martingales, an $(\underline{\mathcal{A}}, P)$-martingale.

## 9.3 Capital Asset Pricing Model

### Risk Premium for the GOP

Later we shall derive for a general continuous financial market the influential intertemporal capital asset pricing model (ICAPM), see Merton (1973a). It is the continuous time generalization of the *capital asset pricing model* (CAPM), due to Sharpe (1964), Lintner (1965) and Mossin (1966). In practice, the ICAPM has been widely used for pricing securities in an approximate sense. We illustrate in the context of the BS model how the ICAPM can be used for pricing.

First, let us define what we mean by a risk premium. The *risk premium* $p_V(t)$ at time $t$ for a security price process $V = \{V(t),\, t \in [0, T]\}$ is defined as the *expected excess return* above the short rate $r_t$, which is given as the almost sure limit

$$p_V(t) \stackrel{\text{a.s.}}{=} \lim_{h \downarrow 0} \frac{1}{h} E\left(\frac{V(t+h) - V(t)}{V(t)} \,\bigg|\, \mathcal{A}_t\right) - r_t \tag{9.3.1}$$

for $t \in [0, T]$.

The GOP value $S_t^{\delta_*}$ at time $t$ satisfies according to (9.1.15) the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*} \left( (r_t + p_{S^{\delta_*}}(t))\, dt + \sqrt{p_{S^{\delta_*}}(t)}\, dW_t \right) \tag{9.3.2}$$

for $t \in [0, T]$ with *risk premium*

$$p_{S^{\delta_*}}(t) = \theta_t^2 = \left( \frac{a_t - r_t}{\sigma_t} \right)^2, \tag{9.3.3}$$

see (9.1.16). Note that the risk premium $p_{S^{\delta_*}}(t)$ in the SDE (9.3.2) of the GOP equals the square of its volatility.

**Risk Premium of the Underlying Security**

By using in the SDE (9.1.1) of the underlying security $S_t$ the formula (9.1.16) for the GOP volatility $\theta_t$, we obtain the SDE

$$dS_t = S_t \left( r_t\, dt + \sigma_t \left( \theta_t\, dt + dW_t \right) \right) \tag{9.3.4}$$

for $t \in [0, T]$. It follows that the risk premium $p_S(t)$ of the underlying security $S$ equals according to (9.3.1) and (9.3.4) the product

$$p_S(t) \overset{\text{a.s.}}{=} \lim_{h \downarrow 0} \frac{1}{h} E \left( \frac{S_{t+h} - S_t}{S_t} \,\middle|\, \mathcal{A}_t \right) - r_t \overset{\text{a.s.}}{=} \sigma_t\, \theta_t \tag{9.3.5}$$

almost surely for all $t \in [0, T]$. Note that the risk premium of the underlying security equals, as $h \to 0$, the normalized covariance of the returns of the underlying security and the GOP, that is,

$$\begin{aligned} p_S(t) &\overset{\text{a.s.}}{=} \lim_{h \downarrow 0} \frac{1}{h} E \left( \left( \frac{S_{t+h} - S_t}{S_t} \right) \left( \frac{S_{t+h}^{\delta_*} - S_t^{\delta_*}}{S_t^{\delta_*}} \right) \,\middle|\, \mathcal{A}_t \right) \\ &\overset{\text{a.s.}}{=} \lim_{h \downarrow 0} \frac{1}{h} E \left( \int_t^{t+h} \sigma_s\, dW_s \int_t^{t+h} \theta_s\, dW_s \,\middle|\, \mathcal{A}_t \right) \overset{\text{a.s.}}{=} \sigma_t\, \theta_t \end{aligned} \tag{9.3.6}$$

almost surely for $t \in [0, T]$.

There is also an alternative way of characterizing the risk premium (9.3.5). The risk premium can be obtained by forming the time derivative of the covariation between the logarithm $\ln(S_t)$ of the underlying security and the logarithm $\ln(S_t^{\delta_*})$ of the GOP. More precisely, by the Itô formula and the covariation property (5.4.5) of Itô integrals we can express the risk premium of $S$ in the form

$$p_S(t) = \frac{d}{dt} \left[ \ln(S), \ln(S^{\delta_*}) \right]_t = \sigma_t\, \theta_t \tag{9.3.7}$$

for $t \in [0, T]$. We shall see later that such a result holds generally in a continuous financial market.

### Risk Premium of a Portfolio

Now, let us calculate risk premia for portfolios. It follows for a portfolio value $S_t^\delta$ at time $t$ from the SDE (9.1.4) and equation (9.1.7) and (9.1.16) the SDE

$$dS_t^\delta = S_t^\delta \left( r_t \, dt + \pi_\delta^1(t) \, \sigma_t \, (\theta_t \, dt + dW_t) \right) \tag{9.3.8}$$

for $t \in [0, T]$. As defined above in (9.3.1), its risk premium $p_{S^\delta}(t)$ at time $t$ equals the expected excess return

$$p_{S^\delta}(t) \overset{\text{a.s.}}{=} \lim_{h \downarrow 0} \frac{1}{h} \, E\left( \frac{S_{t+h}^\delta - S_t^\delta}{S_t^\delta} \,\middle|\, \mathcal{A}_t \right) - r_t \tag{9.3.9}$$

for $t \in [0, T)$. It follows from (9.3.8) and (9.3.9) that we obtain for the fraction $\pi_\delta^1(t)$ the risk premium

$$p_{S^\delta}(t) = \pi_\delta^1(t) \, \sigma_t \, \theta_t \tag{9.3.10}$$

at time $t \in [0, T]$. The risk premium of a portfolio equals the product of market price of risk and portfolio volatility. As in (9.3.7), it follows from the form of the portfolio SDE (9.3.8) that this risk premium equals the normalized covariance between the return of the portfolio and that of the GOP. We have then

$$p_{S^\delta}(t) \overset{\text{a.s.}}{=} \lim_{h \downarrow 0} \frac{1}{h} \, E\left( \left( \frac{S_{t+h}^\delta - S_t^\delta}{S_t^\delta} \right) \left( \frac{S_{t+h}^{\delta_*} - S_t^{\delta_*}}{S_t^{\delta_*}} \right) \,\middle|\, \mathcal{A}_t \right) \overset{\text{a.s.}}{=} \pi_\delta^1(t) \, \sigma_t \, \theta_t \tag{9.3.11}$$

for $t \in [0, T)$. Alternatively, by the Itô formula and the covariation property (5.4.5) of Itô integrals it also follows

$$p_{S^\delta}(t) = \frac{d}{dt} \left[ \ln(S^\delta), \ln(S^{\delta_*}) \right]_t = \pi_\delta^1(t) \, \sigma_t \, \theta_t \tag{9.3.12}$$

for $t \in [0, T]$. As we shall see later, this type of formula holds in a general continuous financial market and not only under the BS model.

Note that it follows from the above formula (9.3.12) that the risk premium $p_B(t)$ of the savings account is zero, as should be expected.

### Portfolio Beta

The ICAPM uses the *market portfolio* (MP) as reference portfolio. One can choose, for instance, the MP as the portfolio of all tradable securities. In practice, this is convenient but difficult to specify explicitly. One can always argue about the exact composition of the MP. In any case, in reality the MP is a reasonably broadly diversified portfolio. The Morgan Stanley capital weighted world stock accumulation index (MSCI) arises as a possible proxy for the MP. We shall show later, see also Platen (2005b), that diversified portfolios can be expected in reality to be close to each other and also close to the GOP. This means, under general assumptions we shall see that diversified portfolios approximate the GOP. This fundamental fact is model independent. In the following, we use the GOP as proxy for the MP. Its movements can be interpreted to model the movements of the market as a whole, thus, modeling *general market risk* or *systematic risk*. We shall later consider *specific market risk*, which describes the movements of a portfolio that are not in line with those of the market index, see Platen & Stahl (2003).

When using the ICAPM one aims to measure for a given portfolio $S_t^\delta$ its *systematic risk parameter* $\beta_{S^\delta}(t)$, which is the, so-called, *beta*. The beta equals the ratio of covariations

$$\beta_{S^\delta}(t) = \frac{\frac{d[\ln(S^\delta), \ln(S^{\delta_*})]_t}{dt}}{d[\ln(S^{\delta_*})]_t} \tag{9.3.13}$$

for $t \in [0,T]$. Obviously, the beta equals one if the portfolio $S^\delta$ moves similarly to the market as a whole. If $S^\delta$ moves totally independent of the GOP, then its beta is zero. By using (9.3.12) it follows that

$$\beta_{S^\delta}(t) = \frac{\pi_\delta^1(t)\,\sigma_t\,\theta_t}{\theta_t^2} = \frac{p_{S^\delta}(t)}{p_{S^{\delta_*}}(t)} \tag{9.3.14}$$

for $t \in [0,T]$. This means that the portfolio beta is the normalized risk premium, where the normalizing quantity is the risk premium of the MP.

Obviously, the beta for the savings account is zero, that is

$$\beta_B(t) = 0 \tag{9.3.15}$$

for $t \in [0,T]$. This expresses the fact that there is no systematic risk in the savings account. Under the given BS model the beta of the underlying security $S$ is by (9.3.14) and (9.3.5) obtained as

$$\beta_S(t) = \frac{\sigma_t}{\theta_t} \tag{9.3.16}$$

for $t \in [0,T]$. This beta is close to one if the underlying security fluctuates similarly to the GOP and, thus, the MP.

A portfolio has a small absolute value of beta if its fluctuations are almost independent of those of the GOP. This means that there is then little systematic or general market risk in this portfolio.

## ICAPM Pricing Rule

By using relation (9.3.14) we obtain for the risk premium $p_{S^\delta}(t)$ of a portfolio $S^\delta$ the *ICAPM formula*

$$p_{S^\delta}(t) = \beta_{S^\delta}(t)\,p_{S^{\delta_*}}(t) \tag{9.3.17}$$

for $t \in [0, T]$. The portfolio beta $\beta_{S^\delta}(t)$, as defined in (9.3.14), has in the given case for a portfolio $S^\delta$ with fraction $\pi^1_\delta(t)$ the value

$$\beta_{S^\delta}(t) = \pi^1_\delta(t)\, \frac{\sigma_t}{\theta_t} \tag{9.3.18}$$

for $t \in [0, T]$.

Under the given BS model a portfolio beta is all that needs to be known about the portfolio's risk characteristics when using the ICAPM formula. The formula (9.3.17) does not contain prices explicitly. It only refers to risk premia. However, the ICAPM can be used in practice for approximate asset pricing. To explain this, we go back to the definition of a return. By the ICAPM formula (9.3.17) we have approximately for a portfolio $S^\delta$ with fraction $\pi^1_\delta(t)$ over a small period $[t, t+h]$ the expected return

$$E\left( \frac{S^\delta_{t+h} - S^\delta_t}{S^\delta_t} \,\middle|\, \mathcal{A}_t \right) = \frac{E\left(S^\delta_{t+h} \,\middle|\, \mathcal{A}_t\right) - S^\delta_t}{S^\delta_t}$$
$$\approx (r_t + p_{S^\delta}(t))\, h = (r_t + \beta_{S^\delta}(t)\, p_{S^{\delta_*}}(t))\, h.$$

Therefore, it follows for small $h > 0$ by (9.3.3) and (9.3.18) that approximately

$$\frac{E\left(S^\delta_{t+h} \,\middle|\, \mathcal{A}_t\right)}{S^\delta_t} \approx 1 + (r_t + \beta_{S^\delta}(t)\, p_{S^{\delta_*}}(t))\, h = 1 + \left( r_t + \frac{\delta^1_t\, S_t\, \sigma_t\, \theta_t}{S^\delta_t} \right) h.$$

This yields the *ICAPM pricing rule*

$$S^\delta_t \approx \frac{E\left(S^\delta_{t+h} \,\middle|\, \mathcal{A}_t\right)}{1 + (r_t + \beta_{S^\delta}(t)\, \theta^2_t)\, h} \tag{9.3.19}$$

or, similarly, by using the above relations and (9.1.16), the self-interpreting pricing rule

$$S^\delta_t \approx \frac{E\left(S^\delta_{t+h} \,\middle|\, \mathcal{A}_t\right) - \delta^1_t\, S_t\, (a_t - r_t)\, h}{1 + r_t\, h} \tag{9.3.20}$$

for $t \in [0, T]$. We emphasize that (9.3.19) and (9.3.20) are approximate formulas. It is interesting to note that the ICAPM pricing rule (9.3.19) uses the portfolio beta and the expected future value of the portfolio as main inputs. Notice that the conditional expectation of the future portfolio value is taken under the real world probability measure, as is the case under real world pricing.

Via the benchmark approach we derive in Sect. 11.2 under general assumptions the ICAPM for continuous financial markets. This means we shall provide the basis for the ICAPM pricing rule (9.3.19). This approximate pricing formula is, of course, not fully consistent with real world pricing. However, it is a reasonable description of the fair price when $h$ is small. The ICAPM pricing rule (9.3.19) is widely applied in practice. It provides another example where commonly accepted relationships in finance, insurance or accounting can be naturally derived under the benchmark approach by using the GOP as central building block.

## 9.4 Risk Neutral Pricing

By referring to the results from Chap. 8 on option pricing under the BS model, we now illustrate the widely used standard *risk neutral* pricing methodology, which one could interpret as the core of the *arbitrage pricing theory* (APT) and its generalizations, see for instance, Black & Scholes (1973), Ross (1976), Harrison & Kreps (1979), Harrison & Pliska (1981), Föllmer & Sondermann (1986), Föllmer & Schweizer (1991) and Delbaen & Schachermayer (1994, 1998, 2006).

### Drifted Wiener Process

In the classical literature on derivative pricing it has been standard to use the domestic savings account $B = \{B_t,\, t \in [0,T]\}$ as reference unit or numeraire. For obtaining an option price one introduces an appropriate probability measure, the *risk neutral probability measure* $P_\theta$. It allows to interpret the Black-Scholes pricing formula as a conditional expectation under this measure. As we shall see, this method provides an elegant and compact description of option prices in the case of the BS model. We shall show that the change to the risk neutral probability measure $P_\theta$ is equivalent to a *change of variables* with a corresponding probabilistic interpretation. Most importantly, we shall emphasize the fact that a number of assumptions have to be made to perform this change of variables, which may not be satisfied for realistic models.

Let us reformulate the SDE (9.1.1) for the underlying security $S$ under the BS model, where we assume now, for simplicity, constant short rate $r$, constant volatility $\sigma$, constant appreciation rate $a$ and, therefore, also constant market price of risk $\theta$. We perform this change of variable in such a way that the SDE (9.1.1) shows formally the same appreciation rate $r$ as the domestic savings account $B$, see (9.1.2). To achieve this it is necessary to introduce a corresponding driving process $W_\theta$ that no longer equals the Wiener process $W$. This is the, so-called, *drifted Wiener process* $W_\theta = \{W_\theta(t),\, t \in [0,T]\}$ with

$$W_\theta(t) = W_t + \theta\, t \tag{9.4.1}$$

for $t \in [0,T]$. Recall that the market price of risk is for our BS model of the form

$$\theta = \frac{a-r}{\sigma}, \tag{9.4.2}$$

see (9.1.16). This allows us to rewrite the SDE (9.1.1) for the underlying security in the form

$$dS_t = (a - \sigma\,\theta)\, S_t\, dt + \sigma\, S_t\, dW_\theta(t) = r\, S_t\, dt + \sigma\, S_t\, dW_\theta(t) \tag{9.4.3}$$

for $t \in [0,T]$. According to (6.3.7), (6.3.6) and (9.4.1), the geometric Brownian motion $S = \{S_t,\, t \in [0,T]\}$ has then the explicit representation

$$S_t = S_0 \exp\left\{\left(r - \frac{1}{2}\sigma^2\right) t + \sigma W_\theta(t)\right\} \tag{9.4.4}$$

for $t \in [0, T]$. Note that for $\theta \neq 0$ the process $W_\theta$ is *not* a Wiener process under the real world probability measure $P$.

### Radon-Nikodym Derivative

We shall show that the process $W_\theta$ is a standard Wiener process under the *risk neutral measure* $P_\theta$. This measure is characterized by its *Radon-Nikodym derivative*

$$\Lambda_\theta(T) = \frac{dP_\theta}{dP}\bigg|_{\mathcal{A}_T} = \frac{\hat{S}^0_T}{\hat{S}^0_0}. \tag{9.4.5}$$

Recall that

$$\hat{S}^0_T = \frac{B_T}{S^{\delta_*}_T}$$

is the benchmarked domestic savings account at time $T$, see (9.1.22). The Radon-Nikodym derivative $\Lambda_\theta(T)$ defines the risk neutral measure $P_\theta$, which is given in the form

$$P_\theta(A) = \int_A \Lambda_\theta(T)\, dP(\omega) = \int_A \frac{dP_\theta(\omega)}{dP(\omega)}\, dP(\omega) \tag{9.4.6}$$

for all subsets $A \in \Omega$.

Note that the measure $P_\theta$ is not automatically a probability measure. For risk neutral pricing to be useful in practice, we need the property that the risk neutral measure $P_\theta$ is a probability measure. This is equivalent to the request that a corresponding change of variables in an integration can be performed.

The following definition will be used generally throughout the book.

**Definition 9.4.1.** *Two measures are* equivalent *if they have the same sets of events of measure zero.*

The equivalence of the risk neutral and the real world probability measure is a fundamental requirement of the risk neutral approach. In the case of the above BS model one is able to apply the risk neutral approach since the measure $P_\theta$ is a probability measure and equivalent to $P$. The model generates geometric Brownian motions on $[0, T]$ under $P$ and under $P_\theta$ with the same sets of events that have probability zero under both measures. However, there is already a problem even under the BS model if one wants to extend the time horizon $T$ to infinity and aims to consider asymptotics for $T \to \infty$. Details on a construction allowing some risk neutral pricing in such a case can be found, for instance, in Karatzas & Shreve (1998).

As discussed in Sect. 9.1, under the BS model the benchmarked savings account $\hat{S}^0 = \{\hat{S}^0_t,\ t \in [0, T]\}$ is an $(\underline{\mathcal{A}}, P)$-martingale with initial value

$$\hat{S}_0^0 = \frac{1}{S_0^{\delta_*}}. \tag{9.4.7}$$

Thus, for the BS model due to (9.4.5) the *Radon-Nikodym derivative process* $\Lambda_\theta = \{\Lambda_\theta(t),\ t \in [0,T]\}$ with

$$\Lambda_\theta(t) = \frac{\hat{S}_t^0}{\hat{S}_0^0} \tag{9.4.8}$$

is an $(\underline{\mathcal{A}}, P)$-martingale that starts at $\Lambda_\theta(0) = 1$.

We shall see later that the martingale property of the Radon-Nikodym derivative process is crucial for the risk neutral approach. It makes sure that the measure $P_\theta$ is having a total mass of one, allowing it to be a probability measure.

We remark that the Radon-Nikodym derivative process is referred to in the literature also as *state price density*, *pricing kernel*, *deflator* or *stochastic discount factor*, see Hansen & Jagannathan (1991), Constatinides (1992), Rogers (1997), Cochrane (2001) and Duffie (2001).

Later it will become clear that the just mentioned Radon-Nikodym process simply expresses the benchmarked savings account when normalized to one. The corresponding risk neutral pricing method can, thus, be derived from real world pricing.

### Risk Neutral Measure Transformation

To illustrate the measure transformation that is performed under the risk neutral approach, let us demonstrate under the given BS model that $W_\theta$ is a Wiener process under the risk neutral probability measure $P_\theta$. The above Radon-Nikodym derivative process $\Lambda_\theta$, see (9.4.8), has the representation

$$\Lambda_\theta(t) = \exp\left\{-\frac{1}{2}\,\theta^2\, t - \theta\, W_t\right\} \tag{9.4.9}$$

for $t \in [0,T]$. By using the Laplace transform (1.3.76) of a Gaussian random variable we have by the martingale property of $\Lambda_\theta$ the total risk neutral probability

$$P_\theta(\Omega) = E(\Lambda_\theta(T)) = E(\Lambda_\theta(T)\,\big|\,\mathcal{A}_0) = \Lambda_\theta(0) = 1. \tag{9.4.10}$$

This shows that $P_\theta$ is a probability measure.

For fixed $\tilde{y} \in \Re,\ t \in [0,T]$ and $s \in [0,t)$ let $A$ be the event

$$A = \{\omega \in \Omega :\ W_\theta(t,\omega) - W_\theta(s,\omega) < \tilde{y}\}.$$

Here we indicate in the notation $W_\theta(t,\omega)$ its dependence on the outcome $\omega \in \Omega$. Using relation (9.4.1), this event can equivalently be written in the form

$$A = \{\omega \in \Omega : \, W(t,\omega) - W(s,\omega) < \tilde{y} - \theta\,(t-s)\}\,.$$

Since $W$ is a Wiener process on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, then $A \in \mathcal{A}_t$ where $A$ is independent of $\mathcal{A}_s$. Combining these facts, it follows by using the indicator function $\mathbf{1}_A$, with $E_\theta$ denoting expectation under $P_\theta$, that

$$\begin{aligned} P_\theta(A) &= E_\theta(\mathbf{1}_A) = E(\Lambda_\theta(T)\,\mathbf{1}_A) \\ &= E\left(\Lambda_\theta(t)\,\mathbf{1}_A\,\frac{\Lambda_\theta(T)}{\Lambda_\theta(t)}\right) = E(\Lambda_\theta(t)\,\mathbf{1}_A) \\ &= E\left(\Lambda_\theta(s)\,\frac{\Lambda_\theta(t)}{\Lambda_\theta(s)}\,\mathbf{1}_A\right) = E\left(\Lambda_\theta(s)\right)\,E\left(\frac{\Lambda_\theta(t)}{\Lambda_\theta(s)}\,\mathbf{1}_A\right) \\ &= E\left(\frac{\Lambda_\theta(t)}{\Lambda_\theta(s)}\,\mathbf{1}_A\right). \end{aligned} \tag{9.4.11}$$

We know that $W_t - W_s$ is Gaussian distributed with mean zero and variance $(t-s)$. Therefore, we obtain for the event $A$ with (9.4.9) the $P_\theta$-probability

$$\begin{aligned} P_\theta(A) &= \int_{-\infty}^{\tilde{y}-\theta(t-s)} \exp\left\{-\frac{\theta^2}{2}\,(t-s) - \theta\,y\right\} \frac{1}{\sqrt{2\,\pi\,(t-s)}}\,\exp\left\{-\frac{y^2}{2\,(t-s)}\right\} dy \\ &= \int_{-\infty}^{\tilde{y}-\theta(t-s)} \frac{1}{\sqrt{2\,\pi\,(t-s)}}\,\exp\left\{-\frac{(y+\theta\,(t-s))^2}{2\,(t-s)}\right\} dy \\ &= \int_{-\infty}^{\tilde{y}} \frac{1}{\sqrt{2\,\pi\,(t-s)}}\,\exp\left\{-\frac{z^2}{2\,(t-s)}\right\} dz. \end{aligned} \tag{9.4.12}$$

This equation shows that $W_\theta(t) - W_\theta(s)$ is Gaussian distributed under $P_\theta$ with mean zero and variance $(t-s)$. Note that we have only changed variables for the integration in (9.4.12), which is permitted due to the properties of the Gaussian density. From the properties (3.2.6) of the Wiener process $W$ and relation (9.4.1) we conclude under the given BS model that $W_\theta(0) = 0$. Using arguments similar to those applied in (9.4.11), it follows that $W_\theta$ has independent increments. Therefore, by using (3.2.6), we can formulate the following simple version of the following *Cameron-Martin Girsanov Theorem* when applied to the BS model.

**Theorem 9.4.2.** (Cameron-Martin Girsanov) *Under the BS model the process $W_\theta$ is a standard Wiener process in the filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P_\theta)$, which is defined under the risk neutral probability measure $P_\theta$.*

In the next section we shall provide a more general version of this theorem.

### Risk Neutral Pricing Formula

As previously mentioned, the probability measure $P_\theta$, which is called the risk neutral probability measure, can be used for option pricing. This is the probability measure under which the process $W_\theta$, see (9.4.1), becomes an $(\underline{\mathcal{A}}, P_\theta)$-Wiener process. Let us now obtain from the real world pricing formula (9.1.34) the price of a European option with payoff $H(S_T)$ under the given BS model.

Using (9.4.5), (9.4.9) and the explicit expression (9.4.4) for the geometric Brownian motion $S$, see (9.4.3), we can rewrite the real world pricing formula (9.1.34) for $t = 0$ in the form

$$
\begin{aligned}
V(0, S_0) &= E\left(\frac{S_0^{\delta_*}}{S_T^{\delta_*}}\, H(S_T) \,\Big|\, \mathcal{A}_0\right) = E\left(\frac{\frac{B_T}{S_T^{\delta_*}}}{\frac{B_0}{S_0^{\delta_*}}} \left(\frac{H(S_T)}{B_T}\right) \,\Big|\, \mathcal{A}_0\right) \\
&= E\left(\frac{\hat{S}_T^0}{\hat{S}_0^0} \left(\frac{H(S_T)}{B_T}\right) \,\Big|\, \mathcal{A}_0\right) = E\left(\frac{\Lambda_\theta(T)}{\Lambda_\theta(0)} \left(\frac{H(S_T)}{B_T}\right) \Big|\, \mathcal{A}_0\right) \\
&= E\left(\exp\left\{-\frac{\theta^2}{2}\, T - \theta\, W_T\right\} \left(\exp\{-r\, T\}\, H(S_T)\right) \Big|\, \mathcal{A}_0\right) \\
&= \int_{-\infty}^{\infty} \left[\exp\{-r\, T\}\, H\left(S_0 \exp\left\{\left(a - \frac{1}{2}\,\sigma^2\right) T + \sigma\, y\right\}\right)\right] \\
&\qquad \times \exp\left\{-\frac{\theta^2}{2}\, T - \theta\, y\right\} \frac{1}{\sqrt{T}}\, N'\left(\frac{y}{\sqrt{T}}\right) dy.
\end{aligned}
$$

With the change of variables $\tilde{y} = y + \theta\, T = y + (\frac{a-r}{\sigma})T$ we then obtain

$$
\begin{aligned}
V(0, S_0) = \exp\{-r\, T\} \int_{-\infty}^{\infty} &\left[H\left(S_0 \exp\left\{\left(r - \frac{1}{2}\,\sigma^2\right) T + \sigma\, \tilde{y}\right\}\right)\right] \\
&\times \frac{1}{\sqrt{T}}\, N'\left(\frac{\tilde{y}}{\sqrt{T}}\right) d\tilde{y}.
\end{aligned}
$$

When written in the following form, the above result provides the *risk neutral pricing formula*

$$
V(0, S_0) = \exp\{-r\, T\}\, E_\theta\left(H(S_T) \,\big|\, \mathcal{A}_0\right) = E_\theta\left(\frac{H(S_T)}{B_T} \,\Big|\, \mathcal{A}_0\right). \tag{9.4.13}
$$

Here $E_\theta$ denotes the expectation with respect to the risk neutral probability measure $P_\theta$ and $N'(\cdot)$ is the standard Gaussian density, see (1.2.8).

The above derivation shows that the fair price at time $t = 0$ of an option under the BS model can be rewritten as a conditional expectation $E_\theta$ under the risk neutral probability measure $P_\theta$ of a savings account discounted payoff $\frac{H(S_T)}{B_T}$. The above risk neutral pricing formula has been widely used in derivative pricing. In the current literature the risk neutral pricing formula appears to be the standard pricing tool. However, note that certain assumptions need to be satisfied to apply this pricing formula.

We exploited a number of mathematical properties that are automatically guaranteed under the BS model. As we shall see later, for certain more realistic asset price models, for instance the MMM, the martingale property of the Radon-Nikodym derivative $\Lambda_\theta$, does not hold and an equivalent risk neutral probability measure does not exist. Since the real world pricing concept does not require the existence of an equivalent risk neutral probability measure one can always apply the real world pricing formula as long as the GOP exists and the expectation of the benchmarked payoff is finite.

### Risk Neutral SDEs

Note that under the risk neutral probability measure $P_\theta$ the discounted underlying security price $\bar{S}$, see (8.2.10), satisfies under the BS model according to (9.4.3) and by application of the Itô formula the SDE

$$d\bar{S}_t = \sigma\, \bar{S}_t\, dW_\theta(t) \tag{9.4.14}$$

for $t \in [0, T]$. Thus, $\bar{S}$ is driftless under $P_\theta$ and can be shown for the BS model to be an $(\underline{\mathcal{A}}, P_\theta)$-martingale, see Exercise 9.1. Furthermore, it follows by the Itô formula, (8.2.21) and (9.4.1) that the SDE for the discounted option price $\bar{V}$, see (8.2.9), is given by

$$d\bar{V}(t, \bar{S}_t) = \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}}\, \sigma\, \bar{S}_t\, dW_\theta(t) \tag{9.4.15}$$

for $t \in [0, T]$. This means that also the SDE for $\bar{V}$ is driftless under $P_\theta$. One can show for the given BS model that $\bar{V}$ is an $(\underline{\mathcal{A}}, P_\theta)$-martingale, see Exercise 9.2. Obviously, the discounted savings account $\bar{B}$, see (8.2.15), is a constant and, thus, trivially an $(\underline{\mathcal{A}}, P_\theta)$-martingale. For both $(\underline{\mathcal{A}}, P_\theta)$-martingales $\bar{V}$ and $\bar{B}$ it is easy to see from Sect. 9.1 that their benchmarked values $\hat{V}(t) = \frac{\bar{V}(t,\bar{S}_t)}{\bar{S}_t^{\delta_*}}$ and $\hat{S}_t^0 = \frac{\bar{B}_t}{\bar{S}_t^{\delta_*}}$ form $(\underline{\mathcal{A}}, P)$-martingales.

### Risk Neutral SDE for Portfolios

Generally, for the above BS model all discounted portfolio prices can be shown to form $(\underline{\mathcal{A}}, P_\theta)$-local martingales. This property follows from the SDE (9.1.37) and Lemma 5.4.1 since by application of Itô's formula

$$\begin{aligned} d\bar{S}_t^\delta &= d(\bar{S}_t^{\delta_*}\, \hat{S}_t^\delta) \\ &= \bar{S}_t^\delta\, \pi_\delta^1(t)\, \sigma\, (\theta\, dt + dW_t) \\ &= \bar{S}_t^\delta\, \pi_\delta^1(t)\, \sigma\, dW_\theta(t) \end{aligned} \tag{9.4.16}$$

for $t \in [0, T]$. By Lemma 5.2.3 any nonnegative discounted portfolio is, therefore, an $(\underline{\mathcal{A}}, P)$-supermartingale. If $\bar{S}^\delta$ is an $(\underline{\mathcal{A}}, P_\theta)$-martingale, then the risk

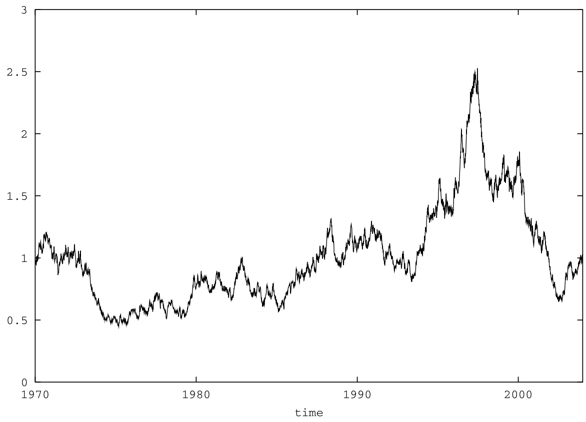



**Fig. 9.4.1.** A Radon-Nikodym derivative process for a BS model

neutral pricing formula (9.4.13) holds for $\bar{S}^\delta$. Furthermore, since the Radon-Nikodym derivative process $\Lambda_\theta$ is here an $(\underline{\mathcal{A}}, P)$-martingale, we shall see later that in this case the benchmarked portfolio value $\hat{S}_t^\delta = \frac{\bar{S}^\delta}{\bar{S}_t^{\delta_*}}$ forms an $(\underline{\mathcal{A}}, P)$-martingale.

We have seen that the real world pricing formula (9.1.34) does not hold for an unfair portfolio as constructed in (9.1.38)–(9.1.43). Similarly, for such a portfolio also the risk neutral pricing formula fails. Thus, one should *not* expect all discounted portfolios to be automatically $(\underline{\mathcal{A}}, P_\theta)$-martingales under the risk neutral probability measure $P_\theta$, even under a simple BS model. Unfortunately, some literature gives the impression that this is the case.

Observe in the derivation of (9.4.13) that we have performed a change of variables from $W_t$ to $W_\theta(t)$ with the interpretation that $W$ and $W_\theta$ are Wiener processes under $P$ and $P_\theta$, respectively. The only variable that is random in the risk neutral pricing formula (9.4.13) is $S_T$, as compared to the real world pricing formula (9.1.34), where also the random GOP value $S_T^{\delta_*}$ is involved. Thus, the computation of option prices by using the risk neutral approach is simplified for the case of the BS model. This simplification relies on the existence of the equivalent risk neutral probability measure $P_\theta$ under the BS model.

We shall see in the next chapter that the benchmark approach, with its real world pricing concept, handles more general models than those permitted under the risk neutral approach. An equivalent risk neutral probability measure need not exist under the benchmark approach. This freedom in modeling will become important when we are going to model realistically the typical market dynamics.

In Fig. 9.4.1 we show a path of an exponential martingale from a geometric Brownian motion with volatility $\theta = 0.2$. We know that the path in Fig. 9.4.1 is that of a martingale. Here the actual value is the best forecast of future values. Similar to equation (9.4.8) one can show that the candidate Radon-

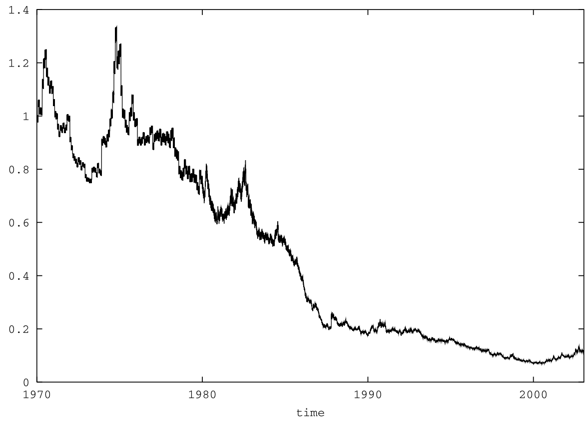



**Fig. 9.4.2.** Candidate Radon-Nikodym derivative of hypothetical risk neutral measure

Nikodym derivative process $\Lambda = \{\Lambda_t,\, t \in [0,T]\}$ for a hypothetical equivalent risk neutral probability measure for a range of continuous financial markets is given by the benchmarked savings account, see (9.4.8) and Karatzas & Shreve (1998), normalized at the initial time to one. An indication for the potential nonexistence of an equivalent risk neutral probability measure for the real market is given by the following important observation:

If the GOP is proxied by a diversified world stock index, as we shall suggest in the next chapter, then one can observe the benchmarked savings account for the world market and, thus, the candidate Radon-Nikodym derivative of its hypothetical risk neutral measure. We show in Fig. 9.4.2 the candidate Radon-Nikodym derivative of the hypothetical risk neutral measure of the world stock market with respect to the US dollar as domestic currency when using the Morgan Stanley capital weighted world stock accumulation index (MSCI) as proxy for the GOP. The path of this process seems to trend systematically downward, which is not typical for a martingale. However, for economic reasons the graph in Fig. 9.4.2 is rather typical, as we shall discuss below. In the long run the benchmarked savings account must be expected to decline systematically in reality. Otherwise, investors have no reason to invest in the stock market. This has been empirically confirmed by Dimson, Marsh & Staunton (2002), who showed that the market capitalization weighted world stock index, when discounted by the US dollar savings account, showed an annually discretely compounded net growth rate of about 0.049 over the last century. From economic reasoning it does not appear to be natural that the trajectory of the benchmarked savings account should form in reality a martingale. However, this martingale property is needed for the application of the Cameron-Martin Girsanov Theorem.

The downward trending trajectory in Fig. 9.4.2 resembles more the path of a strict supermartingale. Of course, a single path cannot prove that the candidate Radon-Nikodym derivative of the hypothetical risk neutral measure is a strict supermartingale. However, based on the economic argument that stock market investments grow in the long term faster than a savings account, one should be prepared to acknowledge such possibility, when developing long term market models. Of course, even if we agree that the benchmarked savings account is not a martingale under the real world probability measure, this is insufficient to infer that no equivalent risk neutral probability measure exists. But it is certainly enough evidence for us to consider this possibility seriously, which we acknowledge by working under the benchmark approach with its real world pricing concept.

What we have just observed creates serious concerns about the practical applicability of the risk neutral pricing methodology that has been the prevailing approach in finance for several decades. Within this book we aim to provide with the benchmark approach a framework that allows to handle not only models that have an equivalent risk neutral probability measure but also models for which this is *not* the case. The real world pricing concept makes financial modeling, derivative pricing and calibration less complicated since a measure transformation is not required.

Under the benchmark approach some potential model risk is removed which could be caused by the fact that an equivalent risk neutral probability measure may not exist for the existing financial market.

## 9.5 Girsanov Transformation and Bayes Rule (*)

In the previous section, an equivalent probability measure transformation was applied, which is also known as *Girsanov transformation*. The following section describes such transformation more generally. It will also introduce *Bayes's Theorem*, which is needed to interpret conditional expectations under a given measure by using those defined under another measure. Both results are important for equivalent probability measure changes.

### Change of Probability Measure (*)

We denote by $\boldsymbol{W} = \{\boldsymbol{W}_t = (W_t^1, \ldots, W_t^m)^\top,\, t \in [0,T]\}$ an $m$-dimensional standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, as given in Sect. 5.1, with $\mathcal{A}_0$ being the trivial $\sigma$-algebra, augmented by the sets of zero probability. For an $\underline{\mathcal{A}}$-predictable $m$-dimensional stochastic process $\boldsymbol{\theta} = \{\boldsymbol{\theta}_t = (\theta_t^1, \ldots, \theta_t^m)^\top,\, t \in [0,T]\}$ with

$$\int_0^T \sum_{i=1}^m (\theta_t^i)^2 \, dt < \infty \tag{9.5.1}$$

almost surely, we assume that the strictly positive *Radon-Nikodym derivative process* $\Lambda_{\boldsymbol{\theta}} = \{\Lambda_{\boldsymbol{\theta}}(t),\, t \in [0,T]\}$, where

$$\Lambda_{\boldsymbol{\theta}}(t) = \exp\left\{-\int_0^t \boldsymbol{\theta}_s^\top \, d\boldsymbol{W}_s - \frac{1}{2}\int_0^t \boldsymbol{\theta}_s^\top \, \boldsymbol{\theta}_s \, ds\right\} < \infty \tag{9.5.2}$$

almost surely for $t \in [0, T]$ is an $(\underline{\mathcal{A}}, P)$-*martingale*. By the Itô formula (6.2.11) it follows from (9.5.2) that

$$\Lambda_{\boldsymbol{\theta}}(t) = 1 - \sum_{i=1}^{m} \int_0^t \Lambda_{\boldsymbol{\theta}}(s) \, \theta_s^i \, dW_s^i \tag{9.5.3}$$

for $t \in [0, T]$. Since $\Lambda_{\boldsymbol{\theta}}$ is by the above assumption an $(\underline{\mathcal{A}}, P)$-martingale we have

$$E\left(\Lambda_{\boldsymbol{\theta}}(t) \,\middle|\, \mathcal{A}_s\right) = \Lambda_{\boldsymbol{\theta}}(s) \tag{9.5.4}$$

for $t \in [0, T]$ and $s \in [0, t]$ and, in particular,

$$E\left(\Lambda_{\boldsymbol{\theta}}(t) \,\middle|\, \mathcal{A}_0\right) = \Lambda_{\boldsymbol{\theta}}(0) = 1. \tag{9.5.5}$$

Now, we define a measure $P_{\boldsymbol{\theta}}$ via the Radon-Nikodym derivative

$$\frac{dP_{\boldsymbol{\theta}}}{dP} = \Lambda_{\boldsymbol{\theta}}(T), \tag{9.5.6}$$

by setting

$$P_{\boldsymbol{\theta}}(A) = E(\Lambda_{\boldsymbol{\theta}}(T) \, \mathbf{1}_A) = E_{\boldsymbol{\theta}}(\mathbf{1}_A) \tag{9.5.7}$$

for $A \in \mathcal{A}_T$. Recall that $\mathbf{1}_A$ is the indicator function for $A$ and $E_{\boldsymbol{\theta}}$ means expectation with respect to $P_{\boldsymbol{\theta}}$.

Note that $P_{\boldsymbol{\theta}}$ is not just a measure but also a probability measure because

$$P_{\boldsymbol{\theta}}(\Omega) = E(\Lambda_{\boldsymbol{\theta}}(T)) = E\left(\Lambda_{\boldsymbol{\theta}}(T) \,\middle|\, \mathcal{A}_0\right) = \Lambda_{\boldsymbol{\theta}}(0) = 1 \tag{9.5.8}$$

due to the martingale property of $\Lambda_{\boldsymbol{\theta}}$. This indicates why the martingale property of the Radon-Nikodym derivative is so important. It guarantees that the resulting risk neutral measure is a probability measure.

If the Radon-Nikodym derivative for the candidate risk neutral measure is a strict supermartingale, then the equality (9.5.8) does not hold and $P_{\boldsymbol{\theta}}(\Omega)$ is strictly less than one. As we shall see, this case arises, for instance, under the MMM, see Fig. 13.3.2.

## Bayes's Theorem (*)

As seen in the risk neutral pricing formula (9.4.13), it is useful to be able to change the probability measure for conditional expectations. For a simple case this is indicated by formula (9.5.7). There exists a general tool, which is the following *Bayes rule*, that allows one to establish a relationship between conditional expectations with respect to different equivalent probability measures.

**Theorem 9.5.1.** (Bayes) *Assume that a given strictly positive Radon-Nikodym derivative process $\Lambda_{\boldsymbol{\theta}}$ is an $(\underline{\mathcal{A}}, P)$-martingale determining a corresponding equivalent probability measure $P_{\boldsymbol{\theta}}$. Then for any given stopping time $\tau \in [0, T]$ and any $\mathcal{A}_\tau$-measurable random variable $Y$, satisfying the integrability condition*

$$E_{\boldsymbol{\theta}}(|Y|) < \infty, \tag{9.5.9}$$

*one can apply the* Bayes rule

$$E_{\boldsymbol{\theta}}\left(Y \,\middle|\, \mathcal{A}_s\right) = \frac{E\left(\Lambda_{\boldsymbol{\theta}}(\tau)\, Y \,\middle|\, \mathcal{A}_s\right)}{E\left(\Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right)} \tag{9.5.10}$$

*for $s \in [0, \tau]$.*

### Proof of Bayes's Theorem (*)

We indicate here a proof of Bayes's Theorem. For a stopping time $\tau \in [s, T]$ and given fixed time $s \in [0, T]$ one can prove Bayes's theorem by using formula (9.5.7) for the probability $P_{\boldsymbol{\theta}}(A)$ together with the properties (1.3.63)–(1.3.66) of conditional expectations and the martingale property of $\Lambda_{\boldsymbol{\theta}}$. Then for each $\mathcal{A}_\tau$-measurable random variable $Y$ and a set $A \in \mathcal{A}_s$ with some fixed time $s \in [0, T]$ we can show that both sides of (9.5.10) are identical for any such set $A$, that is,

$$\begin{aligned}
\mathbf{1}_A\, E_{\boldsymbol{\theta}}\left(Y \,\middle|\, \mathcal{A}_s\right) &= E_{\boldsymbol{\theta}}\left(\mathbf{1}_A\, Y \,\middle|\, \mathcal{A}_s\right) = E\left(\mathbf{1}_A\, Y\, \Lambda_{\boldsymbol{\theta}}(T) \,\middle|\, \mathcal{A}_s\right) \\
&= E\left(\mathbf{1}_A\, Y\, \Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right) = E\left(\mathbf{1}_A\, E\left(Y\, \Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right) \,\middle|\, \mathcal{A}_s\right) \\
&= E\left(\Lambda_{\boldsymbol{\theta}}(s) \left(\frac{\mathbf{1}_A}{\Lambda_{\boldsymbol{\theta}}(s)}\, E\left(Y\, \Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right)\right) \,\middle|\, \mathcal{A}_s\right) \\
&= E_{\boldsymbol{\theta}}\left(\frac{\mathbf{1}_A}{\Lambda_{\boldsymbol{\theta}}(s)}\, E\left(Y\, \Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right) \,\middle|\, \mathcal{A}_s\right) \\
&= E_{\boldsymbol{\theta}}\left(\mathbf{1}_A\, \frac{E\left(\Lambda_{\boldsymbol{\theta}}(\tau)\, Y \,\middle|\, \mathcal{A}_s\right)}{E\left(\Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right)} \,\middle|\, \mathcal{A}_s\right) = \mathbf{1}_A\, \frac{E\left(\Lambda_{\boldsymbol{\theta}}(\tau)\, Y \,\middle|\, \mathcal{A}_s\right)}{E\left(\Lambda_{\boldsymbol{\theta}}(\tau) \,\middle|\, \mathcal{A}_s\right)}.
\end{aligned}$$

This proves Theorem 9.5.1. □

### Girsanov Theorem (*)

The following important result is known as *Girsanov Theorem* for which we shall indicate a proof at the end of the section. A simple version of the Girsanov Theorem has been already given with the Cameron-Martin Girsanov Theorem, see Theorem 9.4.2. The Girsanov Theorem allows us to perform a measure transformation, which transforms an $(\underline{\mathcal{A}}, P)$-drifted Wiener process, as given in (9.4.1), into a Wiener process under a new probability measure $P_{\boldsymbol{\theta}}$. Such a transformation is called Girsanov transformation.

**Theorem 9.5.2.** (Girsanov) *If for $T \in (0, \infty)$ a given strictly positive Radon-Nikodym derivative process $\Lambda_{\boldsymbol{\theta}}$ is an $(\underline{\mathcal{A}}, P)$-martingale, then the m-dimensional process $\boldsymbol{W}_{\boldsymbol{\theta}} = \{\boldsymbol{W}_{\boldsymbol{\theta}}(t),\, t \in [0,T]\}$, given by*

$$\boldsymbol{W}_{\boldsymbol{\theta}}(t) = \boldsymbol{W}_t + \int_0^t \theta_s \, ds \tag{9.5.11}$$

*for all $t \in [0,T]$, is an m-dimensional standard Wiener process on the filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P_{\boldsymbol{\theta}})$.*

Note that certain assumption needs to be satisfied before one can apply the above Girsanov Theorem. The sole key assumption is that $\Lambda_{\boldsymbol{\theta}}$ must be a strictly positive $(\underline{\mathcal{A}}, P)$-martingale. For instance, if the Radon-Nikodym derivative process is almost surely only a strictly positive local martingale, then this does not guarantee that $P_{\boldsymbol{\theta}}$ is a probability measure.

## Novikov Condition (*)

As just mentioned, a key assumption of the risk neutral approach is that $\Lambda_{\boldsymbol{\theta}}$ has to be a strictly positive $(\underline{\mathcal{A}}, P)$-martingale. A sufficient condition for the Radon-Nikodym derivative process $\Lambda_{\boldsymbol{\theta}}$ to be an $(\underline{\mathcal{A}}, P)$-martingale is the *Novikov condition*, see Novikov (1972), which requires that

$$E\left(\exp\left\{\frac{1}{2}\int_0^T \boldsymbol{\theta}_s^\top \, \boldsymbol{\theta}_s \, ds\right\}\right) < \infty. \tag{9.5.12}$$

This condition is fulfilled for the BS model, as was given in (9.1.1), since the market price of risk $\theta$, given in (9.1.16), is a constant. For the case, when $\Lambda_{\boldsymbol{\theta}}$ is already known to be a strictly positive $(\underline{\mathcal{A}}, P)$-local martingale, then some other sufficient conditions can potentially be applied. Some conditions of this kind are given in Lemma 5.2.2. Further conditions can be found in Revuz & Yor (1999).

## Proof of the Girsanov Theorem (*)

For simplicity, we only indicate the proof of Theorem 9.5.2 for the one-dimensional case, that is $m = 1$. Furthermore, we assume for simplicity that $\Lambda_\theta \theta, \Lambda_\theta, \Lambda_\theta W_\theta \theta,\, \Lambda_\theta (W_\theta)^2 \theta \in \mathcal{L}_T^2$ and that $P$ is equivalent to $P_\theta$. The general case is obtained by similar arguments, see Karatzas & Shreve (1991).

**1.** First, let us show that $P_\theta$ is a probability measure. It follows by application of the Itô formula (6.2.11) to the expression (9.5.2) that

$$d\Lambda_\theta(t) = -\Lambda_\theta(t)\, \theta_t \, dW_t \tag{9.5.13}$$

with $\Lambda_\theta(0) = 1$. For the strictly positive process $\Lambda_\theta$ we have a.s. the inequality $\Lambda_\theta(t) > 0$ and from equation (9.5.5) the property

$$E(\Lambda_\theta(t)) = 1 \tag{9.5.14}$$

for all $t \in [0, T]$. From equation (9.5.7) we conclude for any event $A \in \mathcal{A}$ that

$$P_\theta(A) = \int_\Omega \mathbf{1}_A(\omega)\, \Lambda_\theta(T)\, dP(\omega) \geq 0, \tag{9.5.15}$$

where $\mathbf{1}_A(\omega)$ is the indicator function for $\omega$ being in $A$. This combined with the property (9.5.14) shows that

$$P_\theta(\Omega) = \int_\Omega \Lambda_\theta(T)\, dP(\omega) = E(\Lambda_\theta(T)) = 1. \tag{9.5.16}$$

Therefore, $P_\theta(\cdot)$ is a well-defined probability measure on $(\Omega, \mathcal{A})$.

**2.** We now consider the product $\Lambda_\theta(t)\, W_\theta(t)$ and show that it forms a martingale. By the Itô formula (6.2.11) and equations (9.5.11) and (9.5.13) the SDE for $\Lambda_\theta W_\theta$ can be written in the form

$$\begin{aligned} d(\Lambda_\theta(t)\, W_\theta(t)) &= \Lambda_\theta(t)\, dW_\theta(t) + W_\theta(t)\, d\Lambda_\theta(t) + d[\Lambda_\theta, W_\theta]_t \\ &= \Lambda_\theta(t)\, dW_t + \Lambda_\theta(t)\, \theta_t\, dt - W_\theta(t)\, \Lambda_\theta(t)\, \theta_t\, dW_t - \Lambda_\theta(t)\, \theta_t\, dt \\ &= \Lambda_\theta(t)\, (1 - W_\theta(t)\, \theta_t)\, dW_t \end{aligned} \tag{9.5.17}$$

for $t \in [0, T]$. Thus, since $\Lambda_\theta(1 - W_\theta \theta) \in \mathcal{L}^2_T$ it follows by the martingale property (5.4.3) of Itô integrals that the process $\Lambda_\theta W_\theta$ is an $(\underline{\mathcal{A}}, P)$-martingale.

**3.** For $t \in [0, T]$ and $s \in [0, t]$, using the equivalence of $P_\theta$, we obtain with Theorem 9.5.1 from the martingale property of $\Lambda_\theta W_\theta$ the conditional expectation

$$\begin{aligned} E_\theta\left(W_\theta(t) \,\middle|\, \mathcal{A}_s\right) &= E\left(\Lambda_\theta(t)\, W_\theta(t) \,\middle|\, \mathcal{A}_s\right) \\ &= E\left(\Lambda_\theta(s)\, W_\theta(s) \,\middle|\, \mathcal{A}_s\right) \\ &= E_\theta\left(W_\theta(s) \,\middle|\, \mathcal{A}_s\right) = W_\theta(s). \end{aligned} \tag{9.5.18}$$

Note that $W_\theta$ is an $(\underline{\mathcal{A}}, P_\theta)$-martingale. Note that it is not only a martingale with respect to the filtration that it generates.

**4.** Let us now show that $W_\theta$ is under $P_\theta$ a continuous square integrable martingale. Note that we obtain from (9.5.17) and (9.5.11) by the Itô formula

$$d\left(\Lambda_\theta(t)\, (W_\theta(t))^2\right) = \Lambda_\theta(t)\, dt + \Lambda_\theta(t)\, (W_\theta(t))^2\, \theta_t\, dW_t \tag{9.5.19}$$

for $t \in [0, T]$. Now, the square integrability of $W_\theta$ under $P_\theta$ follows, so $(W_\theta)^2\, \Lambda_\theta\, \theta \in \mathcal{L}^2_T$. From (9.5.19) we can conclude that $W_\theta$ is a continuous, square integrable $(\underline{\mathcal{A}}, P_\theta)$-martingale, see (5.1.2) with $W_\theta(0) = 0$, see (9.5.11).

**5.** The quadratic variation process $[W_\theta] = \{[W_\theta]_t,\ t \in [0, T]\}$, see (5.2.2) and (5.2.8), of the continuous $(\underline{\mathcal{A}}, P_\theta)$-martingale $W_\theta$ is, according to (9.5.11), of the form

$$[W_\theta]_t = t \tag{9.5.20}$$

for $t \in [0, T]$. It then follows by Lévy's Theorem, see Theorem 6.5.1, that $W_\theta$ is a standard Wiener process on the probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P_\theta)$. □

## 9.6 Change of Numeraire (*)

It became clear in our previous discussion on real world pricing and risk neutral pricing that there exist equivalent ways of obtaining derivative prices as conditional expectations under certain probability measures by using corresponding numeraires. This has been formalized in Geman, El Karoui & Rochet (1995). Each of these alternative choices of numeraires result in corresponding SDEs for the prices. Often, different numeraires can be used to characterize the same derivative price. Some numeraire choices can provide significant analytic or computational advantages. The expectations involved are simply different ways of representing the same integral value. What actually happens in a numeraire change is a change of variables in an integration. We emphasize that certain conditions have to be satisfied to perform a numeraire change. This is analogous to the well-known fact that not all changes of variables are feasible for certain integrations.

### Benchmarked PDE (*)

To illustrate the change of numeraire technique, let us recall from the real world pricing formula (9.1.34) that the benchmarked option price can be expressed as conditional expectation of the benchmarked payoff. We shall now show for the BS model, as introduced in Sect. 9.1, that the benchmarked pricing function $\hat{V} : [0,T] \times (0,\infty) \times (0,\infty) \to [0,\infty)$, obtained as $\hat{V}(t, S_t, S_t^{\delta_*}) = \hat{V}(t)$, can be expressed as a PDE solution.

With a view on (9.1.32)–(9.1.34) let us determine whether there exists a sufficiently often differentiable benchmarked pricing function $\hat{V}(\cdot,\cdot,\cdot)$ such that

$$\hat{V}(t) = \hat{V}(t, S_t, S_t^{\delta_*}) = E\left( \frac{H(S_T)}{S_T^{\delta_*}} \,\middle|\, \mathcal{A}_t \right), \tag{9.6.1}$$

for $t \in [0,T]$ with $S_t$ and $S_t^{\delta_*}$ satisfying the SDEs (9.1.1) and (9.1.15), respectively.

Application of the Itô formula to the function $\hat{V}(t, S, S^{\delta_*})$ yields, as in (9.1.31), the equation

$$\begin{aligned} \frac{H(S_T)}{S_T^{\delta_*}} &= \hat{V}(T, S_T, S_T^{\delta_*}) \\ &= \hat{V}(t, S_t, S_t^{\delta_*}) + \int_t^T \tilde{L}^0\, \hat{V}(s, S_s, S_s^{\delta_*})\, ds \\ &\quad + \int_t^T \left( \frac{\partial \hat{V}(s, S_s, S_s^{\delta_*})}{\partial S}\, \sigma_s\, S_s + \frac{\partial \hat{V}(s, S_s, S_s^{\delta_*})}{\partial S^{\delta_*}}\, \theta_s\, S_s^{\delta_*} \right) dW_s \end{aligned} \tag{9.6.2}$$

with operator

$$\tilde{L}^0\,\hat{V}(t,S,S^{\delta_*}) = \frac{\partial \hat{V}(t,S,S^{\delta_*})}{\partial t} + a_t\,S\,\frac{\partial \hat{V}(t,S,S^{\delta_*})}{\partial S} + \frac{1}{2}\,\sigma_t^2\,S^2\,\frac{\partial^2 \hat{V}(t,S,S^{\delta_*})}{\partial S^2}$$
$$+ (r_t + \theta_t^2)\,S^{\delta_*}\,\frac{\partial \hat{V}(t,S,S^{\delta_*})}{\partial S^{\delta_*}} + \frac{1}{2}\,\theta_t^2\,(S^{\delta_*})^2\,\frac{\partial^2 \hat{V}(t,S,S^{\delta_*})}{\partial (S^{\delta_*})^2}$$
$$+ \sigma_t\,\theta_t\,S\,S^{\delta_*}\,\frac{\partial^2 \hat{V}(t,S,S^{\delta_*})}{\partial S\,\partial S^{\delta_*}} \tag{9.6.3}$$

for $t \in [0,T)$ and $S, S^{\delta_*} \in (0,\infty)$.

Since the process $\hat{V} = \{\hat{V}(t,S_t,S_t^{\delta_*}),\ t \in [0,T]\}$ is an $(\underline{\mathcal{A}},P)$-martingale, see (9.6.1), it follows from (9.6.2) that we obtain the *benchmarked PDE*

$$\tilde{L}^0\,\hat{V}(t,S,S^{\delta_*}) = 0 \tag{9.6.4}$$

for $(t,S,S^{\delta_*}) \in (0,T)\times(0,\infty)\times(0,\infty)$ with *benchmarked terminal condition*

$$\hat{V}(T,S,S^{\delta_*}) = \frac{H(S)}{S^{\delta_*}} \tag{9.6.5}$$

for $(S,S^{\delta_*}) \in (0,\infty)\times(0,\infty)$. Note that we have linked the conditional expectation (9.6.1) to the PDE (9.6.4)–(9.6.5). Such a relationship is generally known as a Feynman-Kac formula, which we shall describe in the next section. In the above case the numeraire at time $t$ is the GOP $S_t^{\delta_*}$ and the pricing measure is the real world probability measure $P$.

### Recovering the BS-PDE (*)

Now, we use a transformation of variables to confirm that the benchmarked PDE (9.6.4)–(9.6.5) is for the given BS model simply a transformation of the BS-PDE (8.2.23)–(8.2.24). Using the formula (9.1.25), we obtain

$$\hat{V}(t,S,S^{\delta_*}) = \frac{V(t,S)}{S^{\delta_*}} \tag{9.6.6}$$

for $(t,S,S^{\delta_*}) \in (0,T)\times(0,\infty)\times(0,\infty)$. Then the PDE (9.6.4)–(9.6.5) becomes

$$\frac{1}{S^{\delta_*}}\left(\frac{\partial V(t,S)}{\partial t} + a_t\,S\,\frac{\partial V(t,S)}{\partial S} + \frac{1}{2}\,\sigma_t^2\,S^2\,\frac{\partial^2 V(t,S)}{\partial S^2}\right.$$
$$\left.-\,(r_t+\theta_t^2)\,V(t,S) + \theta_t^2\,V(t,S) - \sigma_t\,\theta_t\,S\,\frac{\partial V(t,S)}{\partial S}\right) = 0 \tag{9.6.7}$$

for $(t,S,S^{\delta_*}) \in (0,T)\times(0,\infty)\times(0,\infty)$ with terminal condition

$$V(T,S) = H(S) \tag{9.6.8}$$

for $S \in (0,\infty)$. Consequently, by (9.1.16) and (9.6.7), the function $V(t,S)$ must satisfy the PDE

$$\frac{\partial V(t,S)}{\partial t} + r_t\, S\, \frac{\partial V(t,S)}{\partial S} + \frac{1}{2}\,\sigma_t^2\, S^2\, \frac{\partial^2 V(t,S)}{\partial S^2} - r_t\, V(t,S) = 0 \tag{9.6.9}$$

for $(t, S) \in (0, T) \times (0, \infty)$ with terminal condition (9.6.8). This recovers the BS-PDE (8.2.23) with terminal condition (8.2.24). It confirms that the benchmark approach provides an alternative way of obtaining the BS-PDE for the pricing function of a European option.

### Risk Neutral PDE (*)

By using the savings account $B$ as numeraire in the BS model, let us now recall what we obtained under the risk neutral probability measure $P_\theta$. We have established through the risk neutral pricing formula a link between the conditional expectation (9.4.13) under $P_\theta$ and the BS-PDE (8.2.21)–(8.2.22).

By similar arguments that provided (9.4.13), it holds for the discounted option price $\bar{V}(t, \bar{S}_t)$ that

$$\bar{V}(t, \bar{S}_t) = E_\theta\left(\frac{H(\bar{S}_T\, B_T)}{B_T}\,\middle|\, \mathcal{A}_t\right) \tag{9.6.10}$$

for $t \in [0, T]$. On the other hand, we obtain for the discounted pricing function $\bar{V}(t, \bar{S})$ by (8.2.21)–(8.2.22) the, so-called, *risk neutral* PDE

$$\frac{\partial \bar{V}(t,\bar{S})}{\partial t} + \frac{1}{2}\,\sigma_t^2\, \bar{S}^2\, \frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} = 0 \tag{9.6.11}$$

for $(t, \bar{S}) \in [0, T) \times (0, \infty)$ with terminal condition

$$\bar{V}(T, \bar{S}) = \frac{H(\bar{S}\, B_T)}{B_T} \tag{9.6.12}$$

for $\bar{S} \in (0, \infty)$. As we shall see in Sect. 9.7, also the conditional expectation (9.6.10) refers to a Feynman-Kac formula, here under the risk neutral probability measure $P_\theta$. In the above case the numeraire is the savings account $B$ and the pricing measure is the risk neutral probability measure $P_\theta$.

### Change of Numeraire Technique (*)

The above discussed possibility to use various strictly positive portfolios as numeraire to compute option prices, provides theoretical and computational freedom for finding convenient ways of derivative pricing. This has been observed by practitioners and researchers who realized that the risk neutral probability measure is not necessarily the most convenient probability measure for pricing certain payoffs. Geman et al. (1995) developed this into a general technique which is called the *change of numeraire technique*.

In general, a *numeraire* $S^{\bar{\delta}} = \{S_t^{\bar{\delta}},\, t \in [0, T]\}$ is in this book a strictly positive portfolio process with a corresponding strategy $\bar{\delta} = \{\bar{\delta}_t,\, t \in [0, T]\}$. Intuitively, a numeraire is used as a reference to normalize all other portfolios with respect to it. By choosing a numeraire $S^{\bar{\delta}}$ one considers the relative price of a portfolio $\frac{S_t^\delta}{S_t^{\bar{\delta}}}$.

### Self-Financing under Numeraire Change (*)

Now, we shall show that *self-financing* portfolios remain self-financing after a numeraire change. This is a desirable but not obvious feature of continuous time financial market models. We have seen an example of this kind in (8.2.28). To illustrate this property more generally, consider under the given BS model a numeraire $S^{\bar{\delta}}$ and a portfolio $S^{\delta}$. Then we have by (9.1.3)

$$S_t^{\delta} = \delta_t^0 \, B_t + \delta_t^1 \, S_t \tag{9.6.13}$$

and by (9.1.4)

$$dS_t^{\delta} = \delta_t^0 \, dB_t + \delta_t^1 \, dS_t \tag{9.6.14}$$

and

$$dS_t^{\bar{\delta}} = \bar{\delta}_t^0 \, dB_t + \bar{\delta}_t^1 \, dS_t \tag{9.6.15}$$

for $t \in [0, T]$. By the Itô formula it follows for the ratio $\frac{S_t^{\delta}}{S_t^{\bar{\delta}}}$ that

$$d\left(\frac{S_t^{\delta}}{S_t^{\bar{\delta}}}\right) = \frac{1}{S_t^{\bar{\delta}}} \, dS_t^{\delta} + S_t^{\delta} \, d\left(\frac{1}{S_t^{\bar{\delta}}}\right) + d\left[\frac{1}{S^{\bar{\delta}}}, S^{\delta}\right]_t. \tag{9.6.16}$$

By (9.6.14) and (9.6.13) we obtain

$$\begin{aligned} d\left(\frac{S_t^{\delta}}{S_t^{\bar{\delta}}}\right) &= \delta_t^0 \left(\frac{1}{S_t^{\bar{\delta}}} \, dB_t + B_t \, d\left(\frac{1}{S_t^{\bar{\delta}}}\right)\right) \\ &\quad + \delta_t^1 \left(\frac{1}{S_t^{\bar{\delta}}} \, dS_t + S_t \, d\left(\frac{1}{S_t^{\bar{\delta}}}\right) + d\left[\frac{1}{S^{\bar{\delta}}}, S\right]_t\right). \end{aligned} \tag{9.6.17}$$

Application of the Itô formula to the ratios $\frac{B_t}{S_t^{\bar{\delta}}}$ and $\frac{S_t}{S_t^{\bar{\delta}}}$ allows us to conclude that

$$d\left(\frac{S_t^{\delta}}{S_t^{\bar{\delta}}}\right) = \delta_t^0 \, d\left(\frac{B_t}{S_t^{\bar{\delta}}}\right) + \delta_t^1 \, d\left(\frac{S_t}{S_t^{\bar{\delta}}}\right) \tag{9.6.18}$$

for $t \in [0, T]$. This confirms that the portfolio $S^{\delta}$, when denominated in units of the numeraire $S^{\bar{\delta}}$, is changing its value only due to the gains from trade in $\frac{B}{S^{\bar{\delta}}}$ and $\frac{S}{S^{\bar{\delta}}}$. Thus, the portfolio is also in the denomination of another numeraire $S^{\bar{\delta}}$ a self-financing portfolio. By using the Itô formula this property can be shown to hold generally for any model that we consider.

### Numeraire Pairs (*)

When presenting the above pricing rules we always have considered *numeraire pairs* $(S^{\bar{\delta}}, P_{\theta_{\bar{\delta}}})$. This means, when we selected a numeraire $S^{\bar{\delta}}$, then there was also a corresponding candidate for a related pricing measure $P_{\theta_{\bar{\delta}}}$. In the real world pricing formula (9.1.34) this pair consists of the GOP $S^{\delta_*}$ as numeraire and the real world probability measure $P$ as pricing measure, thus, resulting in the numeraire pair $(S^{\delta_*}, P)$. This is the only case where we are always sure that the pricing measure is an equivalent probability measure because there is no measure change involved.

In the derivation of the risk neutral measure $P_\theta$ in Sect. 9.4 we used the savings account $B$ as numeraire, which yields the numeraire pair $(B, P_\theta)$. This is just another possible choice for a numeraire. Note that we have to make sure that $P_\theta$ is an equivalent probability measure when using this numeraire pair.

There can be also other numeraires that are convenient for the pricing of certain classes of derivatives, for instance, for the computation of interest rate term structure derivatives.

The following result provides a useful tool for the construction of numeraire pairs. From the real world pricing formula (9.1.34) it follows that

$$\frac{V(t)}{S_t^{\delta_*}} = E\left(\frac{H(S_T)}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t\right) \tag{9.6.19}$$

for all $t \in [0, T]$. We now introduce a strictly positive portfolio $S^{\bar{\delta}}$, which we use as numeraire. The numeraire, when benchmarked and normalized to the initial value one, has the form

$$\Lambda_{\theta_{\bar{\delta}}}(t) = \frac{\hat{S}_t^{\bar{\delta}}}{\hat{S}_0^{\bar{\delta}}} = \frac{S_t^{\bar{\delta}}}{S_t^{\delta_*}} \frac{S_0^{\delta_*}}{S_0^{\bar{\delta}}} \tag{9.6.20}$$

for $t \in [0, T]$. Then we can write by using (9.6.19) and (9.6.20)

$$\frac{V(t)}{S_t^{\bar{\delta}}} = E\left(\frac{S_t^{\delta_*}}{S_t^{\bar{\delta}}} \frac{S_T^{\bar{\delta}}}{S_T^{\delta_*}} \frac{H(S_T)}{S_T^{\bar{\delta}}} \,\bigg|\, \mathcal{A}_t\right) = E\left(\frac{\Lambda_{\theta_{\bar{\delta}}}(T)}{\Lambda_{\theta_{\bar{\delta}}}(t)} \frac{H(S_T)}{S_T^{\bar{\delta}}} \,\bigg|\, \mathcal{A}_t\right). \tag{9.6.21}$$

The benchmarked numeraire $\Lambda_{\theta_{\bar{\delta}}}(t)$ satisfies by (9.1.37) the SDE

$$d\Lambda_{\theta_{\bar{\delta}}}(t) = d\left(\frac{\hat{S}_t^{\bar{\delta}}}{\hat{S}_0^{\bar{\delta}}}\right) = \Lambda_{\theta_{\bar{\delta}}}(t)\left(\pi_{\bar{\delta}}^1(t)\,\sigma_t - \theta_t\right) dW_t \tag{9.6.22}$$

for $t \in [0, T]$. Note by Lemma 5.4.1 that $\Lambda_{\theta_{\bar{\delta}}}$ is an $(\underline{\mathcal{A}}, P)$-local martingale because the SDE (9.6.22) is driftless. Assume now that we have chosen a numeraire $S^{\bar{\delta}}$ such that $\Lambda_{\theta_{\bar{\delta}}}$ is an $(\underline{\mathcal{A}}, P)$-martingale. This allows us to show that $P_{\theta_{\bar{\delta}}}$ is a probability measure when defined via the Radon-Nikodym derivative

$$\frac{dP_{\theta_{\bar{\delta}}}}{dP} = \Lambda_{\theta_{\bar{\delta}}}(T). \tag{9.6.23}$$

We then can introduce the drifted Wiener process $W_{\theta_{\bar{\delta}}} = \{W_{\theta_{\bar{\delta}}}(t),\, t \in [0, T]\}$ with

$$dW_{\theta_{\bar{\delta}}}(t) = dW_t + \theta_{\bar{\delta}}(t)\, dt, \tag{9.6.24}$$

where

$$\theta_{\bar{\delta}}(t) = \theta_t - \pi^1_{\bar{\delta}}(t)\, \sigma_t \tag{9.6.25}$$

for $t \in [0, T]$. Now, we are in a position to apply Theorem 9.5.2 to conclude that by the Girsanov transformation (9.6.24) $W_{\theta_{\bar{\delta}}}$ is a standard Wiener process under the probability measure $P_{\theta_{\bar{\delta}}}$. This provides us, rather generally, with the numeraire pair $(S^{\bar{\delta}}, P_{\theta_{\bar{\delta}}})$.

Obviously, there is no measure transformation involved if we choose the GOP $S^{\delta_*}$ as numeraire since in this case we have from (9.6.25)

$$\theta_{\delta_*}(t) = 0$$

for all $t \in [0, T]$.

If we use the savings account $B$ as numeraire, then $\pi^1_{\bar{\delta}}(t) = 0$ and we obtain from (9.6.25)

$$\theta_{\bar{\delta}}(t) = \theta_t.$$

This is the risk neutral measure change, where the probability measure $P_{\theta_{\bar{\delta}}} = P_\theta$ equals the risk neutral probability measure.

We could also use, for instance, the underlying security $S$ as numeraire, where $\pi^1_{\bar{\delta}}(t) = 1$ and we obtain by (9.6.25)

$$\theta_{\bar{\delta}}(t) = \theta_t - \sigma_t.$$

This also would provide under the above BS model an appropriate measure transformation.

Note however, the situation is different, if we choose the unfair portfolio $S^{\bar{\delta}}_t = S^{\delta_*}_t Z_t$ given in (9.1.42)–(9.1.43). Obviously, by (9.1.44) this numeraire, when benchmarked is not an $(\underline{\mathcal{A}}, P)$-martingale. By (8.7.23) it is a strict supermartingale. The pricing measure $P_{\theta_{\bar{\delta}}}$ is in this case *not* a probability measure. In particular, we have

$$P_{\theta_{\bar{\delta}}}(\Omega) = E\left(\Lambda_{\theta_{\bar{\delta}}}(T) \,\middle|\, \mathcal{A}_0\right) < \Lambda_{\theta_{\bar{\delta}}}(0) = 1.$$

Consequently, the Girsanov Theorem cannot be applied.

### Change of Numeraire Pricing Formula (*)

Using a strictly positive numeraire $S^{(\underline{\delta})}$ and noting that $\Lambda_{\theta_{\bar{\delta}}}(0) = 1$ we can always rewrite the real world pricing formula (9.1.34) in the form

$$V(0, S_0) = E\left(\frac{S^{\delta_*}_0}{S^{\delta_*}_T}\, H(S_T)\, \mathcal{A}_0\right) = E\left(\Lambda_{\theta_{\bar{\delta}}}(T)\, \frac{H(S_T)}{S^{\bar{\delta}}_T} \,\middle|\, \mathcal{A}_0\right). \tag{9.6.26}$$

Note that the quantity

$$\frac{S_0^{\delta_*}}{S_T^{\delta_*}} = \frac{\Lambda_{\theta_{\bar{\delta}}}(T)}{S_T^{\bar{\delta}}} \tag{9.6.27}$$

remains *numeraire invariant* under all above discussed numeraire changes. Let us compute the expectation on the right hand side of (9.6.26) by application of Bayes's Theorem and formula (9.5.10). The required corresponding conditional expectation is of the form

$$E\left(\Lambda_{\theta_{\bar{\delta}}}(T)\,\frac{H(S_T)}{S_T^{\bar{\delta}}}\,\bigg|\,\mathcal{A}_0\right) = E_{\theta_{\bar{\delta}}}\left(\frac{H(S_T)}{S_T^{\bar{\delta}}}\,\bigg|\,\mathcal{A}_0\right),$$

where $E_{\theta_{\bar{\delta}}}$ denotes expectation under $P_{\theta_{\bar{\delta}}}$. For this formula to be valid it is necessary that the assumptions of the Girsanov Theorem and the Bayes Theorem can be verified. This requires $\Lambda_{\theta_{\bar{\delta}}}$ to form an $(\underline{\mathcal{A}}, P)$-martingale to guarantee that $P_{\theta_{\bar{\delta}}}$ is an equivalent probability measure. If this is the case, then we obtain the *change of numeraire pricing formula*

$$V(0, S_0) = E_{\theta_{\bar{\delta}}}\left(\frac{H(S_T)}{S_T^{\bar{\delta}}}\,\bigg|\,\mathcal{A}_0\right). \tag{9.6.28}$$

We learned from our previous discussion and example (9.1.38)–(9.1.43) in Sect. 9.1 that not all benchmarked numeraires form $(\underline{\mathcal{A}}, P)$-martingales. This indicates that the change of numeraire pricing formula (9.6.28) may fail to hold in certain cases. One needs to check carefully the assumptions that are needed for choosing a numeraire pair. Otherwise, an inappropriate numeraire choice, like the unfair portfolio in (9.1.42), may lead to wrong prices.

In the risk neutral case the Radon-Nikodym derivative process $\Lambda_\theta$ for the candidate risk neutral measure $P_\theta$ needs to be an $(\underline{\mathcal{A}}, P)$-martingale to provide the risk neutral pricing formula (9.4.13). Consequently, by (9.6.20) it is necessary that the benchmarked savings account $\frac{B_t}{S_t^{\delta_*}}$ forms an $(\underline{\mathcal{A}}, P)$-martingale to allow the use of the standard risk neutral approach.

## 9.7 Feynman-Kac Formula (*)

As previously shown, several of the existing pricing approaches can be expressed via pricing formulas that have the form of conditional expectations. These conditional expectations lead to pricing functions that satisfy certain PDEs, which are usually *Kolmogorov backward equations*, as was shown for real world pricing and for risk neutral pricing. The link between the conditional expectations and respective PDEs can be interpreted as an application of the, so-called, *Feynman-Kac formula*. In this section we formulate the Feynman-Kac formula under rather general assumptions, allowing also first exit times and jump diffusions. For a wide range of models this formula provides the Kolmogorov backward PDEs that characterize pricing functions of derivatives.

### SDE for Factor Process (*)

At first we consider a fixed time horizon $T \in (0, \infty)$ and a $d$-dimensional Markov process $\boldsymbol{X}^{t,\boldsymbol{x}} = \{\boldsymbol{X}^{t,\boldsymbol{x}}_s,\ s \in [t, T]\}$ describing some factors, which satisfies the vector SDE

$$d\boldsymbol{X}^{t,\boldsymbol{x}}_s = \boldsymbol{a}(s, \boldsymbol{X}^{t,\boldsymbol{x}}_s)\, ds + \sum_{k=1}^{m} \boldsymbol{b}^{\boldsymbol{k}}(s, \boldsymbol{X}^{t,\boldsymbol{x}}_s)\, dW^k_s \tag{9.7.1}$$

for $s \in [t, T]$ with initial value $\boldsymbol{X}^{t,x}_t = \boldsymbol{x} \in \Re^d$ at time $t \in [0, T]$, see (7.8.1)–(7.8.4). The process $\boldsymbol{W} = \{\boldsymbol{W}_t = (W^1_t, \ldots, W^m_t)^\top,\ t \in [0, T]\}$ is assumed to represent an $m$-dimensional standard Wiener process on the filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. One can show, similarly as in the proof of Theorem 7.8.2, that under appropriate assumptions, which will be described below, the process $\boldsymbol{X}^{t,\boldsymbol{x}}$ is a diffusion process with drift coefficient $\boldsymbol{a}(\cdot, \cdot)$ and diffusion coefficients $\boldsymbol{b}^{\boldsymbol{k}}(\cdot, \cdot)$, $k \in \{1, 2, \ldots, m\}$. In general, $\boldsymbol{a} = (a^1, \ldots, a^d)^\top$ and $\boldsymbol{b}^{\boldsymbol{k}} = (b^{1,k}, \ldots, b^{d,k})^\top$, $k \in \{1, 2, \ldots, m\}$ represent vector valued functions on $[0, T] \times \Re^d$ into $\Re^d$, such that a pathwise unique solution of the SDE (9.7.1) exists. Usually, the components of the SDE (9.7.1) are the factors in a financial market model.

### Terminal Payoff Function (*)

Let us describe the case for a European option, where we have a terminal payoff $H(\boldsymbol{X}^{t,\boldsymbol{x}}_T)$ at the maturity date $T$ with some given payoff function $H : \Re^d \to [0, \infty)$ such that

$$E(|H(\boldsymbol{X}^{t,\boldsymbol{x}}_T)|) < \infty. \tag{9.7.2}$$

We can then introduce the pricing function $u : [0, T] \times \Re^d \to [0, \infty)$

$$u(t, \boldsymbol{x}) = E\left(H(\boldsymbol{X}^{t,\boldsymbol{x}}_T) \,\middle|\, \mathcal{A}_t\right) \tag{9.7.3}$$

for $(t, \boldsymbol{x}) \in [0, T] \times \Re^d$. The *Feynman-Kac formula* for this payoff structure refers to the fact that under sufficient regularity of $\boldsymbol{a}, \boldsymbol{b}^{\boldsymbol{1}}, \ldots, \boldsymbol{b}^{\boldsymbol{m}}$ and $H$ the function $u : (0, T) \times \Re^d \to [0, \infty)$ satisfies the PDE

$$\begin{aligned} L^0\, u(t, \boldsymbol{x}) &= \frac{\partial u(t, \boldsymbol{x})}{\partial t} + \sum_{i=1}^{d} a^i(t, \boldsymbol{x}) \frac{\partial u(t, \boldsymbol{x})}{\partial x^i} \\ &\qquad + \frac{1}{2} \sum_{i,k=1}^{d} \sum_{j=1}^{m} b^{i,j}(t, \boldsymbol{x})\, b^{k,j}(t, \boldsymbol{x}) \frac{\partial^2 u(t, \boldsymbol{x})}{\partial x^i \partial x^k} \\ &= 0 \end{aligned} \tag{9.7.4}$$

for $(t, \boldsymbol{x}) \in (0, T) \times \Re^d$ with terminal condition

$$u(T, \boldsymbol{x}) = H(\boldsymbol{x}) \tag{9.7.5}$$

for $\boldsymbol{x} \in \Re^d$. This type of European payoff will be covered by a general version of the Feynman-Kac formula that we present later in this section. For instance, it can be applied to determine the discounted pricing function for risk neutral pricing with zero interest rate when the expectation is taken for the discounted payoff with respect to the equivalent risk neutral probability measure. Under the real world pricing of the benchmark approach the above version of the Feynman-Kac formula would allow the calculation of the benchmarked pricing function under the real world probability measure.

### Discounted Payoff Function (*)

Let us now generalize the above payoff function by discounting it with a given *discount rate process* $r$, which is obtained as a function of the given vector diffusion process $\boldsymbol{X}^{t,\boldsymbol{x}}$, that is $r : [0,T] \times \Re^d \to \Re$. For instance, in a risk neutral setting the discount rate is given by the short term interest rate.

Over the period $[t, T]$ we obtain for the *discounted payoff*

$$\exp\left\{-\int_t^T r(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds\right\} H(\boldsymbol{X}_T^{t,\boldsymbol{x}})$$

the pricing function

$$u(t, \boldsymbol{x}) = E\left(\exp\left\{-\int_t^T r(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds\right\} H(\boldsymbol{X}_T^{t,\boldsymbol{x}}) \,\Big|\, \mathcal{A}_t\right) \tag{9.7.6}$$

for $(t, \boldsymbol{x}) \in [0, T] \times \Re^d$. Under conditions that we shall specify below, it follows that the pricing function $u$ satisfies the PDE

$$L^0\, u(t, \boldsymbol{x}) = r(t, \boldsymbol{x})\, u(t, \boldsymbol{x}) \tag{9.7.7}$$

for $(t, \boldsymbol{x}) \in (0, T) \times \Re^d$ with terminal condition

$$u(T, \boldsymbol{x}) = H(\boldsymbol{x}) \tag{9.7.8}$$

for $\boldsymbol{x} \in \Re^d$, where the PDE operator $L^0$ is given in (9.7.4). Also this version of the Feynman-Kac formula is covered by a more general result that follows later.

### Terminal Payoff and Payoff Rate (*)

Now, we add to the above discounted payoff structure some payoff stream, which continuously pays with a *payoff rate* $g : [0, T] \times \Re^d \to [0, \infty)$ some amount per unit of time. This can model, for instance, an income stream in a company, continuous dividend payments for a share or continuous interest payments. The corresponding *discounted payoff with payoff rate* is then at time $t \in [0,T]$ of the form

$$\exp\left\{-\int_t^T r(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds\right\} H(\boldsymbol{X}_T^{t,\boldsymbol{x}}) + \int_t^T \exp\left\{-\int_t^s r(z, \boldsymbol{X}_z^{t,\boldsymbol{x}})\, dz\right\} g(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds.$$

This leads to the pricing function

$$u(t,\boldsymbol{x}) = E\Bigg(\exp\left\{-\int_t^T r(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds\right\} H(\boldsymbol{X}_T^{t,\boldsymbol{x}}) + \int_t^T \exp\left\{-\int_t^s r(z, \boldsymbol{X}_z^{t,\boldsymbol{x}})\, dz\right\} g(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds \,\Bigg|\, \mathcal{A}_t\Bigg) \tag{9.7.9}$$

for $(t,\boldsymbol{x}) \in [0,T] \times \Re^d$. As we show below, this pricing function satisfies the PDE

$$L^0\, u(t,\boldsymbol{x}) + g(t,\boldsymbol{x}) = r(t,\boldsymbol{x})\, u(t,\boldsymbol{x}) \tag{9.7.10}$$

for $(t,\boldsymbol{x}) \in (0,T) \times \Re^d$ with terminal condition

$$u(T,\boldsymbol{x}) = H(\boldsymbol{x}) \tag{9.7.11}$$

for $\boldsymbol{x} \in \Re^d$.

### SDE with Jumps (*)

We consider now jump diffusions. Let $\Gamma$ denote an open connected subset of $\Re^d$ and $T \in (0,\infty)$ a fixed time horizon. We consider for a $d$-dimensional process $\boldsymbol{X}^{t,\boldsymbol{x}} = \{\boldsymbol{X}_s^{t,\boldsymbol{x}},\ s \in [t,T]\}$, see (6.4.19), the vector SDE

$$d\boldsymbol{X}_s^{t,\boldsymbol{x}} = \boldsymbol{a}(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, ds + \sum_{k=1}^m \boldsymbol{b}^{\boldsymbol{k}}(s, \boldsymbol{X}_s^{t,\boldsymbol{x}})\, dW_s^k + \sum_{j=1}^{\ell} \int_{\mathcal{E}} \boldsymbol{c}^{\boldsymbol{j}}(v, s-, \boldsymbol{X}_{s-}^{t,\boldsymbol{x}})\, p_{\varphi_j}^j(dv, ds) \tag{9.7.12}$$

for $t \in [0,T]$, $s \in [t,T]$ and $\boldsymbol{x} \in \Gamma$ with value

$$\boldsymbol{X}_t^{t,\boldsymbol{x}} = \boldsymbol{x} \tag{9.7.13}$$

at time $t$, see (7.6.23). Here $\boldsymbol{W} = \{\boldsymbol{W}_t = (W_t^1, \ldots, W_t^m)^\top,\ t \in [0,T]\}$ is again an $m$-dimensional standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ as introduced in Sect. 5.1. Furthermore, $p_{\varphi_j}^j(\cdot,\cdot)$ denotes a Poisson measure, $j \in \{1,2,\ldots,\ell\}$, as introduced in Sect. 3.5, satisfying condition (3.5.14). Here $\boldsymbol{a} = (a^1, \ldots, a^d)^\top$ and $\boldsymbol{b}^{\boldsymbol{k}} = (b^{1,k}, \ldots, b^{d,k})^\top$, $k \in \{1,2,\ldots,m\}$, are vector valued functions from $[0,T] \times \Gamma$ into $\Re^d$ and $\boldsymbol{c}^{\boldsymbol{j}} = (c^{1,j}, \ldots, c^{d,j})^\top$, $j \in \{1,2,\ldots,\ell\}$, is a vector valued function on $\mathcal{E} \times [0,T] \times \Gamma$, $\mathcal{E} = \Re \backslash \{0\}$.

### Feynman-Kac Formula with Jumps (*)

For the above payoff structure with discounted terminal payoff and a given payoff rate, we can form the pricing function

$$u(t,\boldsymbol{x}) = E\left(\exp\left\{-\int_t^T r(s,\boldsymbol{X}_s^{t,\boldsymbol{x}})\,ds\right\} H(\boldsymbol{X}_T^{t,\boldsymbol{x}}) + \int_t^T \exp\left\{-\int_t^s r(z,\boldsymbol{X}_z^{t,\boldsymbol{x}})\,dz\right\} g(s,\boldsymbol{X}_s^{t,\boldsymbol{x}})\,ds \,\middle|\, \mathcal{A}_t\right) \tag{9.7.14}$$

for $t \in [0,T] \times \Re^d$. It turns out under appropriate conditions, as will be described below, that $u$ satisfies the *partial integro differential equation* (PIDE)

$$L^0\, u(t,\boldsymbol{x}) + g(t,\boldsymbol{x}) = r(t,\boldsymbol{x})\, u(t,\boldsymbol{x}) \tag{9.7.15}$$

for $(t,\boldsymbol{x}) \in (0,T)$ with terminal condition

$$u(T,\boldsymbol{x}) = H(\boldsymbol{x}) \tag{9.7.16}$$

for $\boldsymbol{x} \in \Re^d$. Here the operator $L^0$ is given in the form

$$L^0 u(t,\boldsymbol{x}) = \sum_{i=1}^d a^i(t,\boldsymbol{x})\,\frac{\partial u(t,\boldsymbol{x})}{\partial x^i} + \frac{1}{2}\sum_{i,k=1}^d \sum_{j=1}^m b^{i,j}(t,\boldsymbol{x})\,b^{k,j}(t,\boldsymbol{x})\,\frac{\partial^2 u(t,\boldsymbol{x})}{\partial x^i\,\partial x^k} + \frac{\partial u(t,\boldsymbol{x})}{\partial t} + \sum_{j=1}^{\ell}\int_{\mathcal{E}} \left[u(s,x^1 + c^{1,j}(v,s,\boldsymbol{x}),\ldots,x^d + c^{d,j}(v,s,\boldsymbol{x})) - u(s,x^1,\ldots,x^d)\right]\,\varphi_j(dv), \tag{9.7.17}$$

where we abuse slightly the notation by writing $u(s,(x^1,\ldots,x^d)^\top) = u(s,x^1,\ldots,x^d)$. Note that an extra integral term is generated by the jumps as a consequence of the Itô formula with jumps, see (6.4.11) and (6.4.20).

### Functional with First Exit Time (*)

Assume that there is a, so-called, *continuation region* $\Phi$, which is an open connected subset of $[0,T] \times \Gamma$. We continue to receive payments as long as the process $\boldsymbol{X}^{t,\boldsymbol{x}}$ stays in the continuation region in $\Phi$. For instance, in the case of a, so-called, *knock-out-barrier option* this would mean that $\boldsymbol{X}_s^{t,\boldsymbol{x}}$ has to stay below a given critical barrier to receive the terminal payment. Then we define the *first exit time* $\tau_\Phi^t$ from $\Phi$ after $t$ as

$$\tau_\Phi^t = \inf\{s \in [t,T] :\, (s,\boldsymbol{X}_s^{t,\boldsymbol{x}}) \notin \Phi\}, \tag{9.7.18}$$

which is a stopping time, see (5.1.13).

To characterize a general payoff structure we use a *terminal payoff function* $H : (0,T] \times \Gamma \to [0,\infty)$ for payments at time $\tau^t_\Phi$, a *payoff rate* $g : [0,T] \times \Gamma \to [0,\infty)$ for incremental payments during the time period $[t, \tau^t_\Phi)$ and a *discount rate* $r : [0,T] \times \Gamma \to \Re$. These quantities are all assumed to be measurable functions. Assume that the process $\boldsymbol{X}^{t,\boldsymbol{x}}$ does not explode or leave $\Gamma$ before time $T$. We then define the *pricing function* $u : \Phi \to [0,\infty)$ by

$$u(t,\boldsymbol{x}) = E\left( H(\tau^t_\Phi, \boldsymbol{X}^{t,\boldsymbol{x}}_{\tau^t_\Phi}) \exp\left\{ -\int_t^{\tau^t_\Phi} r(s, \boldsymbol{X}^{t,\boldsymbol{x}}_s)\, ds \right\} \right.$$
$$\left. + \int_t^{\tau^t_\Phi} g(s, \boldsymbol{X}^{t,\boldsymbol{x}}_s) \exp\left\{ -\int_t^s r(z, \boldsymbol{X}^{t,\boldsymbol{x}}_u)\, dz \right\} ds \,\middle|\, \mathcal{A}_t \right) \tag{9.7.19}$$

for $(t,\boldsymbol{x}) \in \Phi$.

**General Feynman-Kac Formula (*)**

For the formulation of the PIDE for the function $u$ we use the operator $L^0$ given in (9.7.17). Under sufficient regularity of $\Phi$, $\boldsymbol{a}$, $\boldsymbol{b}^{\boldsymbol{1}}$, ..., $\boldsymbol{b}^{\boldsymbol{m}}$, $\boldsymbol{c}^{\boldsymbol{1}}, \ldots, \boldsymbol{c}^{\boldsymbol{\ell}}$, $H$, $g$, $\varphi_1$, ..., $\varphi_\ell$ and $r$ one can show by application of the Itô formula (6.4.11) that the pricing function $u$ satisfies the PIDE

$$L^0 u(t,\boldsymbol{x}) + g(t,\boldsymbol{x}) = r(t,\boldsymbol{x})\, u(t,\boldsymbol{x}) \tag{9.7.20}$$

for $(t,\boldsymbol{x}) \in \Phi$ with boundary condition

$$u(t,\boldsymbol{x}) = H(t,\boldsymbol{x}) \tag{9.7.21}$$

for $(t,\boldsymbol{x}) \in ((0,T] \times \Gamma)\backslash\Phi$. This result links the functional (9.7.19) to the PIDE (9.7.20)–(9.7.21) and can again be called a Feynman-Kac formula.

The above Feynman-Kac formula also holds for a partly negative terminal payoff function $H$ and payoff rate $g$. One can split these payoffs into their negative and positive parts, where each can be separately handled by the above result. The Feynman-Kac formula can be conveniently derived by application of the Itô formula (6.4.20). Due to the complexity of boundary conditions that one has to deal with, such a derivation is useful, in principle, only for particular classes of asset price models and functionals. Therefore, we do not state here an extremely general and, consequently, very technical theorem that formulates a fully general Feynman-Kac formula for SDEs with jump component. However, it is clear that under similar conditions, as we formulate for the already rather general case below, that one obtains the Feynman-Kac formula also in the case with jumps by using the smoothness of the PIDE solution, the Itô formula and the martingale property of the resulting functional.

**Conditions for the Feynman-Kac Formula (*)**

For the case $\Phi = (0,T) \times \Gamma$ and assuming no jumps, that is $\boldsymbol{c}^{\boldsymbol{1}} = \ldots = \boldsymbol{c}^{\boldsymbol{\ell}} = 0$ and $\tau^t_\Phi = T$, let us now formulate some technical conditions that ensure that the Feynman-Kac formula holds.

(A) The drift coefficient $\boldsymbol{a}$ and diffusion coefficients $\boldsymbol{b}^{\boldsymbol{k}}$, $k \in \{1,2,\ldots,m\}$, are assumed to be on $[0,T] \times \Gamma$ locally Lipschitz-continuous in $\boldsymbol{x}$, uniformly in $t$. That is, for each compact subset $\Gamma^1$ of $\Gamma$ there exists a constant $K_{\Gamma^1} < \infty$ such that

$$|\boldsymbol{a}(t,\boldsymbol{x}) - \boldsymbol{a}(t,\boldsymbol{y})| + \sum_{k=1}^{m} |\boldsymbol{b}^{\boldsymbol{k}}(t,\boldsymbol{x}) - \boldsymbol{b}^{\boldsymbol{k}}(t,\boldsymbol{y})| \le K_{\Gamma^1}\, |\boldsymbol{x} - \boldsymbol{y}| \qquad (9.7.22)$$

for all $t \in [0,T]$ and $\boldsymbol{x}, \boldsymbol{y} \in \Gamma^1$.

(B) For all $(t,\boldsymbol{x}) \in [0,T) \times \Gamma$ the solution $\boldsymbol{X}^{t,\boldsymbol{x}}$ of (9.7.12) neither explodes nor leaves $\Gamma$ before $T$, that is

$$P\left(\sup_{t\le s\le T} |\boldsymbol{X}^{t,\boldsymbol{x}}_s| < \infty\right) = 1 \qquad (9.7.23)$$

and

$$P(\boldsymbol{X}^{t,\boldsymbol{x}}_s \in \Gamma \text{ for all } s \in [t,T]) = 1. \qquad (9.7.24)$$

(C) There exists an increasing sequence $(\Gamma_n)_{n\in\mathcal{N}}$ of bounded, open and connected domains of $\Gamma$ such that $\cup_{n=1}^{\infty}\Gamma_n = \Gamma$, and for each $n \in \mathcal{N}$ the PDE

$$L^0 u_n(t,\boldsymbol{x}) + g(t,\boldsymbol{x}) = r(t,\boldsymbol{x})\, u_n(t,\boldsymbol{x}) \qquad (9.7.25)$$

has a unique solution $u_n$, see Friedman (1975), on $(0,T)\times\Gamma_n$ with boundary condition

$$u_n(t,\boldsymbol{x}) = u(t,\boldsymbol{x}) \qquad (9.7.26)$$

on $((0,T) \times \partial\Gamma_n) \cup (\{T\} \times \Gamma_n)$, where $\partial\Gamma_n$ denotes the boundary of $\Gamma_n$.

(D) The process $b^{i,k}(\cdot,\boldsymbol{X}_\cdot)\,\frac{\partial u(\cdot,\boldsymbol{X}_\cdot)}{\partial x^i}$ is from $\mathcal{L}^2_T$ for all $i \in \{1,2,\ldots,d\}$ and $k \in \{1,2,\ldots,m\}$.

For the following theorem, which is similar to a result in Heath & Schweizer (2000), we shall give a proof at the end of the section.

**Theorem 9.7.1.** *In the case without jumps under the conditions (A), (B), (C) and (D), the function u given by* (9.7.19) *is the unique solution of the PDE* (9.7.20) *with boundary condition* (9.7.21)*, where u is differentiable with respect to t and twice differentiable with respect to the components of* $\boldsymbol{x}$.

Condition (A) is satisfied if, for instance, $\boldsymbol{a}$ and $\boldsymbol{b} = (\boldsymbol{b}^{\boldsymbol{1}},\ldots,\boldsymbol{b}^{\boldsymbol{m}})$ are differentiable in $\boldsymbol{x}$ on the open set $(0,T)\times\Gamma$ with derivatives that are continuous on $[0,T]\times\Gamma$.

To establish condition (B) one needs to exploit specific properties of the process $\boldsymbol{X}^{t,\boldsymbol{x}}$ given by the SDE (9.7.12).

Condition (C) can be shown to be implied by the following assumptions:

(C1) There exists an increasing sequence $(\Gamma_n)_{n\in\mathcal{N}}$ of bounded, open and connected subdomains of $\Gamma$ with $\Gamma_n \cup \partial\Gamma_n \subset \Gamma$ such that $\cup_{n=1}^{\infty}\Gamma_n = \Gamma$, and each $\Gamma_n$ has a twice differentiable boundary $\partial\Gamma_n$.

(C2) For each $n \in \mathcal{N}$ the functions $\boldsymbol{a}$ and $\boldsymbol{b}\boldsymbol{b}^\top$ are uniformly Lipschitz-continuous on $[0,T] \times (\Gamma_n \cup \partial\Gamma_n)$.

(C3) For each $n \in \mathcal{N}$ the function $\boldsymbol{b}(t,\boldsymbol{x})\boldsymbol{b}(t,\boldsymbol{x})^\top$ is uniformly elliptic on $\Re^d$ for $(t,\boldsymbol{x}) \in [0,T] \times \Gamma_n$, that is there exists a $\delta_n > 0$ such that

$$\boldsymbol{y}^\top \boldsymbol{b}(t,\boldsymbol{x})\,\boldsymbol{b}(t,\boldsymbol{x})^\top \boldsymbol{y} \geq \delta_n\,|\boldsymbol{y}|^2 \tag{9.7.27}$$

for all $\boldsymbol{y} \in \Re^d$.

(C4) For each $n \in \mathcal{N}$ the functions $r$ and $g$ are uniformly *Hölder-continuous* on $[0,T]\times(\Gamma_n \cup \partial\Gamma_n)$, that is there exists a constant $\bar{K}_n$ and an exponent $q_n > 0$ such that

$$|r(t,\boldsymbol{x}) - r(t,\boldsymbol{y})| + |g(t,\boldsymbol{x}) - g(t,\boldsymbol{y})| \leq \bar{K}_n\,|\boldsymbol{x} - \boldsymbol{y}|^{q_n} \tag{9.7.28}$$

for $t \in [0,T]$ and $\boldsymbol{x},\boldsymbol{y} \in (\Gamma_n \cup \partial\Gamma_n)$.

(C5) For each $n \in \mathcal{N}$ the function $u$ is finite and continuous on $([0,T] \times \partial\Gamma_n) \cup (\{T\} \times (\Gamma_n \cup \partial\Gamma_n))$.

Condition (D) is satisfied when

$$\int_0^T E\left(\left(b^{i,k}(t,\boldsymbol{X}_t)\,\frac{\partial u(t,\boldsymbol{X}_t)}{\partial x^i}\right)^2\right) dt < \infty$$

for all $i \in \{1,2,\ldots,d\}$ and $k \in \{1,2,\ldots,m\}$. This condition ensures that the process $u(\cdot,\boldsymbol{X}.)$ is a martingale and the PDE (9.7.20)–(9.7.21) has a unique solution.

### On the Proof of Theorem 9.7.1 (*)

Let us now indicate the proof of Theorem 9.7.1. It follows from condition (A) that (9.7.12) has a unique solution up to an explosion time, see Theorem II.5.2 in Kunita (1984). Due to (B) this explosion time has to be greater than $T$ almost surely so that the stochastic process $\boldsymbol{X}^{t,\boldsymbol{x}}$ is well defined on $[t,T]$. The expectation in (9.7.19) is then also well-defined with values in $[0,\infty)$ because $H$ and $g$ are nonnegative. Condition (C) implicitly contains the assumption that for all $n \in \mathcal{N}$ and $(t,\boldsymbol{x}) \in ((0,\infty) \times \partial\Gamma_n) \cup (\{T\} \times \Gamma_n)$ the function $u(t,\boldsymbol{x})$ is finite, that is $u(t,\boldsymbol{x}) < \infty$. For fixed $(t,\boldsymbol{x}) \in (0,T) \times \Gamma$ the condition (C) allows us then to find an $n \in \mathcal{N}$ such that $\boldsymbol{x} \in \Gamma_n$.

Let us denote by

$$\tau^t_{\Gamma_n} = \inf\{s \in [t,T] :\, \boldsymbol{X}^{t,\boldsymbol{x}}_s \notin \Gamma_n\} \tag{9.7.29}$$

the first exit time of $(s,\boldsymbol{X}^{t,\boldsymbol{x}}_s)$ from $[t,T) \times \Gamma_n$, see (9.7.18). Due to the continuity of $\boldsymbol{X}^{t,\boldsymbol{x}}$ it is

$$\left(\tau^t_{\Gamma_n}, \boldsymbol{X}^{t,\boldsymbol{x}}_{\tau^t_{\Gamma_n}}\right) \in ((0,T) \times \partial\Gamma_n) \cup (\{T\} \times \Gamma_n)$$

such that

$$u\left(\tau^t_{\Gamma_n}, \boldsymbol{X}^{t,\boldsymbol{x}}_{\tau^t_{\Gamma_n}}\right) < \infty.$$

We then have by application of the Itô formula (6.2.11) to $u_n$, conditions (9.7.25) and (9.7.26) that

$$u_n(t,\boldsymbol{x}) = E\left(u\left(\tau^t_{\Gamma_n}, \boldsymbol{X}^{t,\boldsymbol{x}}_{\tau^t_{\Gamma_n}}\right) \,\Big|\, \mathcal{A}_t\right), \tag{9.7.30}$$

where the appearing Itô integral is, due to the boundedness of $\Gamma_n$, an $(\underline{\mathcal{A}}, P)$-martingale.

Because of (A) and (B) it follows that $\boldsymbol{X}^{t,\boldsymbol{x}}$ is a strong Markov process, see Theorem IV.2.3 and the remark after Theorem IV.6.1 in Ikeda & Watanabe (1989). This means that the Markov property still holds when the present time is chosen to be a stopping time. These results are stated for $\boldsymbol{a}$ and $\boldsymbol{b}$ not depending on $t$ and $\boldsymbol{x}$ from $\Re^d$, but the condition (B) allows us to replace $\Re^d$ by $\Gamma$. Then the results can be shown to hold for time dependent $\boldsymbol{a}$ and $\boldsymbol{b}$, as in Chap. 6 of Stroock & Varadhan (1982). Therefore, by the strong Markov property we obtain

$$E\Bigg(H(T,\boldsymbol{X}^{t,\boldsymbol{x}}_T)\,\exp\left\{-\int_t^T r(s,\boldsymbol{X}^{t,\boldsymbol{x}}_s)\,ds\right\}$$
$$-\int_t^T g(s,\boldsymbol{X}^{t,\boldsymbol{x}}_s)\,\exp\left\{-\int_t^s r(u,\boldsymbol{X}^{t,\boldsymbol{x}}_u)\,du\right\}ds \,\Bigg|\, \mathcal{A}_{\tau^t_{\Gamma_n}}\Bigg) = u\left(\tau^t_{\Gamma_n}, \boldsymbol{X}^{t,\boldsymbol{x}}_{\tau^t_{\Gamma_n}}\right)$$

and, thus, by (9.7.19) and (9.7.30)

$$u(t,\boldsymbol{x}) = E\left(u\left(\tau^t_{\Gamma_n}, \boldsymbol{X}^{t,\boldsymbol{x}}_{\tau^t_{\Gamma_n}}\right) \,\Big|\, \mathcal{A}_t\right) = u_n(t,\boldsymbol{x}).$$

Hence for all $n \in \mathcal{N}$ the functions $u$ and $u_n$ coincide on $(0,T)\times\Gamma_n$. This implies by (C) that $u$ satisfies (9.7.20) on $(0,T) \times \Gamma$. From (9.7.12) and (9.7.19) we obtain then the boundary condition (9.7.21) and also the uniqueness of $u$, if we exploit the fact that $u(\cdot, \boldsymbol{X}_\cdot)$ is a martingale due to (D). □

## 9.8 Exercises for Chapter 9

**9.1.** Prove for the BS model with constant volatility $\sigma_t > 0$ appreciation rate $a$ and short rate $r$ that the domestic savings account discounted stock price process $\bar{S}$ is an $(\underline{\mathcal{A}}, P_\theta)$-martingale under the risk neutral probability measure $P_\theta$.

**9.2.** Show that the discounted European call option price process for the BS model with constant parameters is a martingale under the risk neutral probability measure.

**9.3.** Formulate the SDE for the European put option price for the BS model with constant parameters under the risk neutral probability measure $P_\theta$ and under the original probability measure $P$.

**9.4.** Starting from the risk neutral SDE for the stock price verify that the benchmarked stock price for the BS model is an $(\underline{\mathcal{A}}, P)$-martingale.

**9.5.** Compute the European call option price as an expectation under the risk neutral probability measure for the BS model.

**9.6.** (*) Write down for the BS model the Itô SDE for the Radon-Nikodym derivative process of the risk neutral measure.

**9.7.** (*) Use under the BS model with constant interest rate $r$ the zero coupon bond price $P(t,T)$ with maturity $T$ as numeraire, $t \in [0,T]$. Describe the corresponding numeraire pair. What is the relationship of the resulting pricing measure with the risk neutral probability measure?

**9.8.** (*) Apply for the BS model the Feynman-Kac formula to compute the PDE for the price $V(t,S_t)$ at time $t$ of the payoff $H(S_T) = S_T^2$ of the square of the underlying security at maturity $T$. Can you explicitly solve the corresponding PDE?

# 10

# Continuous Financial Markets

This and the following three chapters present a range of new concepts and ideas that do not fit under presently prevailing approaches. They derive a general, unified framework for modeling continuous financial markets.

Since the middle of the last century there have been substantial developments in portfolio theory and derivative pricing. Some of these developments did not take much notice of the others. This chapter introduces the *benchmark approach* which attempts to unify and generalize these seemingly different theories and approaches. As we shall see, the benchmark approach suggests a change in the common practice of derivative pricing using the more general real world pricing concept. This allows a wider class of financial market models that cover realistic models. Since this may appear to be quite radical for some readers, we first demonstrate the benchmark approach for the case of a general continuous financial market. In particular, in the next chapter we emphasize a deep link between derivative pricing and portfolio optimization.

At the end of this chapter a Diversification Theorem shows that diversified portfolios approximate the *growth optimal portfolio* (GOP). One can show that the GOP exists in any reasonable financial market model. The benchmark approach uses this remarkable portfolio as a benchmark in several ways and employs it as central building block in financial modeling. In portfolio optimization it will be used as a classical benchmark for fund management. In derivative pricing it will be selected as the numeraire, allowing us to use the real world probability measure as pricing measure.

## 10.1 Primary Security Accounts and Portfolios

### Trading Uncertainty

For the modeling of a financial market over an infinite time interval $[0, \infty)$ we rely on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. The filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t\in[0,\infty)}$ is assumed to satisfy the usual conditions, see Sect. 5.1. If not otherwise stated,

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

$\mathcal{A}_0$ is assumed to be the trivial sigma-algebra. The filtration $\underline{\mathcal{A}}$ describes the structure of information entering the market, in the sense that the sigma-algebra $\mathcal{A}_t$ expresses the information relevant to the market at time $t$. For simplicity, we restrict ourselves in this chapter to markets with continuous security prices. *Trading uncertainty* is expressed by the independent standard $(\underline{\mathcal{A}}, P)$-Wiener processes $W^k = \{W_t^k,\, t \in [0,\infty)\}$, for $k \in \{1,2,\ldots,d\}$ and $d \in \mathcal{N}$. Note that there may be additional nontraded uncertainty present in the market. Such uncertainty can model randomness, for instance, in volatilities, appreciation rates, short rates or other quantities.

### Primary Security Accounts

We consider a market comprising $d+1$ *primary security accounts*. These include a *savings account* $S^0 = \{S_t^0,\, t \in [0,\infty)\}$, which is a *locally riskless* primary security account whose value at time $t$ is given by

$$S_t^0 = \exp\left\{\int_0^t r_s\, ds\right\} < \infty \tag{10.1.1}$$

for $t \in [0,\infty)$, where $r = \{r_t,\, t \in [0,\infty)\}$ denotes the adapted *short rate*, also called the short term interest rate. For an interpretation of the savings account as the limit of a roll-over short term bond account we refer to the end of Sect. 10.4.

The market also includes $d$ nonnegative, risky primary security account processes $S^j = \{S_t^j,\, t \in [0,\infty)\}$, $j \in \{1,2,\ldots,d\}$, each of which contains units of one type of security. It is important to note that in a primary security account all proceeds are reinvested. Typically, these securities are stocks with all dividends reinvested. However, foreign savings accounts, bonds and, possibly even, derivatives may also form primary security accounts.

To specify the dynamics of continuous primary securities in the given market, we assume very generally that the $j$th primary security account value $S_t^j$, $j \in \{1,2,\ldots,d\}$, satisfies the SDE

$$dS_t^j = S_t^j \left( a_t^j\, dt + \sum_{k=1}^d b_t^{j,k}\, dW_t^k \right) \tag{10.1.2}$$

for $t \in [0,\infty)$ with $S_0^j > 0$. Here the process $b^{j,k} = \{b_t^{j,k},\, t \in [0,\infty)\}$ is the *volatility* of the $j$th primary security account with respect to the $k$th Wiener process $W^k$. Suppose that $b^{j,k}$ is a given predictable process that satisfies the integrability condition

$$\int_0^T \sum_{j=1}^d \sum_{k=1}^d \left(b_t^{j,k}\right)^2 dt < \infty \tag{10.1.3}$$

almost surely, for all $j, k \in \{1, 2, \ldots, d\}$ and $T \in [0, \infty)$. Furthermore, we assume that the *appreciation rate* $a^j = \{a^j_t,\ t \in [0, \infty)\}$, $j \in \{1, 2, \ldots, d\}$, is a predictable process such that

$$\int_0^T \sum_{j=0}^{d} |a^j_s|\, ds < \infty \tag{10.1.4}$$

almost surely, for all $T \in [0, \infty)$. For instance, a Black-Scholes (BS) model, see (7.5.4), is obtained if one assumes the appreciation rates, the short rate and the volatilities to be constants. The dynamics of the above primary security accounts can be very general, because appreciation rates and volatilities can be chosen quite freely as predictable stochastic processes.

### Market Price of Risk

We use the same number $d$ of Wiener processes for the modeling of trading uncertainty as there are risky primary security accounts. If the number of securities were greater than the number of Wiener processes, then we would have redundant securities that can be removed from the set of primary security accounts. Alternatively, if there were fewer risky securities than Wiener processes, then the market would be in some sense incomplete with respect to trading uncertainty. The core analysis of this chapter is then still valid, although, some additional considerations arise. We shall discuss certain aspects of *incomplete markets* in Sect. 11.5.

The following assumption avoids redundant primary security accounts. It is also the key assumption which the existence of a GOP secures in our market and, therefore, the absence of arbitrage, as we shall see later.

**Assumption 10.1.1.** *The volatility matrix* $\boldsymbol{b}_t = [b^{j,k}_t]^d_{j,k=1}$ *is invertible for Lebesgue-almost every* $t \in [0, \infty)$*, with inverse matrix* $\boldsymbol{b}^{-1}_t = [b^{-1\,j,k}_t]^d_{j,k=1}$.

Let $\boldsymbol{S} = \{\boldsymbol{S}_t = (S^0_t, S^1_t, \ldots, S^d_t)^\top,\ t \in [0, \infty)\}$ denote the vector of primary security account processes. Let $\boldsymbol{a} = \{\boldsymbol{a}_t = (a^1_t, \ldots, a^d_t)^\top,\ t \in [0, \infty)\}$ denote the vector of appreciation rate processes. Let $\boldsymbol{b} = \{\boldsymbol{b}_t,\ t \in [0, \infty)\}$ denote the volatility matrix process and let $\boldsymbol{r} = \{\boldsymbol{r}_t = (r_t, r^1_t, r^2_t, \ldots, r^d_t)^\top,\ t \in [0, \infty)\}$ denote the vector of short rate and dividend rate processes with respect of the primary security accounts. All of the above processes are assumed to be $\underline{\mathcal{A}}$-adapted. Note that $r_t$ is the short rate of the domestic currency, while $r^j_t$ represents the adapted dividend rate of the $j$th stock.

**Definition 10.1.2.** *We call the above market* $\mathcal{S}^C_{(d)} = (\boldsymbol{S}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}, \underline{\mathcal{A}}, P)$*, with d risky primary security accounts, a* continuous financial market *(CFM) when it satisfies Assumption 10.1.1.*

Assumption 10.1.1 allows us in a CFM to introduce the $k$th *market price of risk* $\theta_t^k$ with respect to the $k$th trading uncertainty, which is the $k$th Wiener process $W^k$, via the equation

$$\theta_t^k = \sum_{j=1}^{d} b_t^{-1\,j,k} \left(a_t^j - r_t\right), \tag{10.1.5}$$

which we can write conveniently by using the unit vector $\mathbf{1} = (1, \ldots, 1)^\top$ as

$$\boldsymbol{\theta}_t = \left(\theta_t^1, \ldots, \theta_t^d\right)^\top = \boldsymbol{b}_t^{-1} \left(\boldsymbol{a}_t - r_t\, \mathbf{1}\right) \tag{10.1.6}$$

for $t \in [0, \infty)$ and $k \in \{1, 2, \ldots, d\}$. Recall that $\boldsymbol{b}_t^{-1}$ is the inverse of the volatility matrix $\boldsymbol{b}_t$. We shall see that the market prices of risk are central invariants in a CFM, where they are uniquely determined. This uniqueness is a consequence of Assumption 10.1.1. Since the market price of risk only needs to be determined with respect to each source of trading uncertainty, it is sufficient to use the same number of Wiener processes as there are primary security accounts. Uncertainty that does not appear as trading uncertainty does not have a market price of risk until it becomes securitized.

Now, we can rewrite the SDE (10.1.2) for the $j$th primary security account in the form

$$dS_t^j = S_t^j \left( r_t\, dt + \sum_{k=1}^{d} b_t^{j,k} \left(\theta_t^k\, dt + dW_t^k\right)\right) \tag{10.1.7}$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. We emphasize that (10.1.7) describes the same dynamics as the SDE (10.1.2). We have just reparameterized the latter equation in terms of the market prices of risk.

### Self-Financing Strategies and Portfolios

We call a predictable stochastic process $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \delta_t^1, \ldots, \delta_t^d)^\top,\ t \in [0, \infty)\}$ a *strategy*, if for each $j \in \{0, 1, \ldots, d\}$ the Itô integral

$$I_{\delta^j, S^j}(t) = \int_0^t \delta_s^j\, dS_s^j \tag{10.1.8}$$

exists, see Sect. 5.3. Here $\delta_t^j$, $j \in \{0, 1, \ldots, d\}$, is the number of units of the $j$th primary security account that are held at time $t \in [0, \infty)$ in the corresponding *portfolio* $S^\delta = \{S_t^\delta,\ t \in [0, \infty)\}$. The value $S_t^\delta$ of this portfolio at time $t$ is given by

$$S_t^\delta = \sum_{j=0}^{d} \delta_t^j\, S_t^j. \tag{10.1.9}$$

A strategy $\delta$ and the corresponding portfolio $S^\delta$ are said to be *self-financing* if

$$dS_t^\delta = \sum_{j=0}^{d} \delta_t^j \, dS_t^j \tag{10.1.10}$$

for $t \in [0, \infty)$. This means that all changes in portfolio value are due to gains or losses from trade in the primary security accounts. We make the following standing assumption, which may be interpreted as a *law of conservation of value*.

**Assumption 10.1.3.** *All strategies and portfolios are self-financing.*

Due to this assumption we shall simply omit the phrase "self-financing" in the remainder of the book. By (10.1.10) and (10.1.7) the value of a portfolio $S_t^\delta$ satisfies the SDE

$$dS_t^\delta = S_t^\delta \, r_t \, dt + \sum_{k=1}^{d} \sum_{j=0}^{d} \delta_t^j \, S_t^j \, b_t^{j,k} \left( \theta_t^k \, dt + dW_t^k \right) \tag{10.1.11}$$

for $t \in [0, \infty)$. Note that we do not impose any restrictions on the sign of the portfolio value. In general, a portfolio can become zero or negative in value. In what follows, we denote by $\mathcal{V}^+$ the set of all *strictly positive portfolios*.

### Fractions

Let $S^\delta \in \mathcal{V}^+$ be a strictly positive portfolio process. In this case it is often convenient to introduce the *fraction* $\pi_{\delta,t}^j$ of $S_t^\delta$ that is invested in the $j$th primary security account $S_t^j$, $j \in \{0, 1, \ldots, d\}$, at time $t$. This fraction is given by the expression

$$\pi_{\delta,t}^j = \delta_t^j \, \frac{S_t^j}{S_t^\delta} \tag{10.1.12}$$

for $j \in \{0, 1, \ldots, d\}$. Note that fractions can be negative, but must always sum to one, that is

$$\sum_{j=0}^{d} \pi_{\delta,t}^j = 1 \tag{10.1.13}$$

for $t \in [0, \infty)$. Applying (10.1.10), (10.1.7) and (10.1.12) to (10.1.11) gives the SDE

$$dS_t^\delta = S_t^\delta \left( r_t \, dt + \sum_{k=1}^{d} \sum_{j=1}^{d} \pi_{\delta,t}^j \, b_t^{j,k} \left( \theta_t^k \, dt + dW_t^k \right) \right) \tag{10.1.14}$$

for $t \in [0, \infty)$.

## 10.2 Growth Optimal Portfolio

Kelly (1956) discovered an important portfolio that will be the central object of study in the benchmark approach presented in this book, see Definition 9.1.1. Later in Latané (1959), Breiman (1960), Thorp (1961), Markowitz (1976) and Long (1990) it was applied to gambling, portfolio optimization and derivative pricing. This portfolio is the *growth optimal portfolio* (GOP). It may be characterized as the portfolio maximizing the expected log-utility from terminal wealth, that is the quantity $E(\ln(S^\delta_T))$, for any $T \in [0, \infty)$ over all strictly positive portfolios $S^\delta$.

The GOP has been a fascinating object for theoreticians and practitioners alike, because it possesses a number of remarkable properties. For instance, it is the portfolio that has the maximal expected growth rate over any time horizon. This strictly positive portfolio almost surely outperforms any other strictly positive portfolio over a sufficiently long time horizon. A review on the GOP can be found, for instance, in Hakansson & Ziemba (1995). The paper by Kelly (1956) was motivated by questions arising in information theory. It derived the striking result that there is an optimal gambling strategy, see Thorp (1961), that almost surely accumulates finally more wealth than any other strategy. This is the GOP strategy which maximizes the geometric mean of weighted primary security accounts in a portfolio. In this respect it was already pointed out in Williams (1936) that one should concentrate on maximizing the geometric mean.

### GOP in a CFM

To identify the GOP in a CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$ consider $S^\delta \in \mathcal{V}^+$ and apply the Itô formula to obtain the SDE for $\ln(S^\delta_t)$ in the form

$$d\ln(S^\delta_t) = g^\delta_t\,dt + \sum_{k=1}^{d}\sum_{j=1}^{d} \pi^j_{\delta,t}\, b^{j,k}_t\, dW^k_t \tag{10.2.1}$$

with *growth rate*

$$g^\delta_t = r_t + \sum_{k=1}^{d}\left(\sum_{j=1}^{d} \pi^j_{\delta,t}\, b^{j,k}_t\, \theta^k_t - \frac{1}{2}\left(\sum_{j=1}^{d} \pi^j_{\delta,t}\, b^{j,k}_t\right)^2\right) \tag{10.2.2}$$

at time $t \in [0, \infty)$, see Exercise 10.1. This allows us to define a GOP in a CFM, see Definition 9.1.1.

**Definition 10.2.1.** *In a CFM $\mathcal{S}^{C}_{(d)}$ a strictly positive portfolio process $S^{\delta_*} = \{S^{\delta_*}_t,\ t \in [0, \infty)\} \in \mathcal{V}^+$ is called a* GOP*, if for all $t \in [0, \infty)$ and all strictly positive portfolios $S^\delta \in \mathcal{V}^+$, the growth rates satisfy the inequality*

$$g_t^{\delta_*} \geq g_t^{\delta} \tag{10.2.3}$$

*almost surely.*

We shall see later that this is a convenient definition for the case of a CFM, but there are alternative ways to characterize a GOP. These alternatives allow to guarantee, for instance, the existence of a GOP in a general semimartingale market, see Platen (2004a). Let us now identify the GOP for the given CFM.

**SDE of a GOP**

By using the first order conditions obtained from differentiating the growth rate $g_t^\delta$ in (10.2.2) with respect to the fractions $\pi_{\delta,t}^j$ of the risky primary security accounts, we obtain

$$0 = \sum_{k=1}^{d} b_t^{j,k} \left( \theta_t^k - \sum_{\ell=1}^{d} \pi_{\delta,t}^{\ell}\, b_t^{\ell,k} \right) \tag{10.2.4}$$

for all $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. From (10.2.4) one can conclude that the *optimal fractions* satisfy the vector equation

$$\boldsymbol{\theta}_t = \boldsymbol{b}_t^\top\, \boldsymbol{\pi}_{\delta_*,t}. \tag{10.2.5}$$

This means that the $j$th optimal fraction has the form

$$\pi_{\delta_*,t}^j = \sum_{k=1}^{d} \theta_t^k\, b_t^{-1\,j,k} \tag{10.2.6}$$

for $j \in \{1, 2, \ldots, d\}$, which we can write with (10.1.6) in vector form as

$$\boldsymbol{\pi}_{\delta_*,t} = \left(\boldsymbol{b}_t^{-1}\right)^\top \boldsymbol{\theta}_t = \left(\boldsymbol{b}_t^{-1}\right)^\top \boldsymbol{b}_t^{-1} \left(\boldsymbol{a}_t - r_t\, \mathbf{1}\right) \tag{10.2.7}$$

for all $t \in [0, \infty)$. By substituting these optimal fractions into the SDE (10.1.14) it is straightforward to show that $S_t^{\delta_*}$ satisfies the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*} \left( r_t\, dt + \sum_{k=1}^{d} \theta_t^k \left( \theta_t^k\, dt + dW_t^k \right) \right) \tag{10.2.8}$$

for $t \in [0, \infty)$. Obviously, the GOP is uniquely determined up to its initial value $S_0^{\delta_*} > 0$. In the given CFM the value of the GOP, as well as its fractions, are uniquely determined. In more general models with redundant securities one can show that the value process for the GOP is still uniquely determined when maximizing the growth rate. However, there are then several ways of selecting the fractions. By choosing a set of primary security accounts one also ensures uniqueness of the fractions.

We shall see later that the structure of the SDE (10.2.8) of the GOP is of crucial importance for understanding the typical dynamics of the market. It is remarkable that as a result of the above optimization of the growth rate, the risk premium, see (9.3.1), of the GOP is simply the square of its volatility.

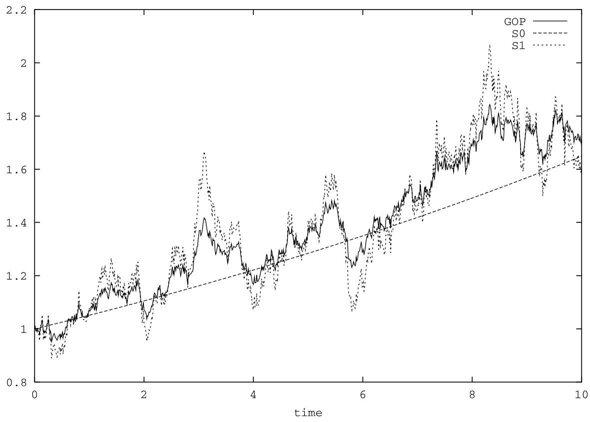



**Fig. 10.2.1.** Two asset BS model, $S^0$, $S^1$ and GOP

## Example of a Two Asset Black-Scholes Model

For illustration let us consider a very basic CFM with $d = 1$ risky primary security account, constant appreciation rates, constant volatilities and a constant short rate. This is then a BS model. The savings account satisfies here the differential equation

$$dS_t^0 = S_t^0 \, r \, dt \tag{10.2.9}$$

for $t \in [0, \infty)$ and $S_0^0 = 1$. We assume that the risky primary security account $S_t^1$ satisfies the SDE

$$dS_t^1 = S_t^1 \left( a^1 \, dt + b^{1,1} \, dW_t^1 \right), \tag{10.2.10}$$

with constant, deterministic volatility $b^{1,1} > 0$, and we set $S_0^1 = 1$. In Fig. 10.2.1 we show simulated paths of $S_t^0$ and $S_t^1$, where $r = 0.05$, $b^{1,1} = 0.2$ and $a^1 = 0.07$.

The volatility $b_t = b^{1,1}$ is invertible and so Assumption 10.1.1 is satisfied. By (10.2.8) the GOP $S_t^{\delta_*}$ is determined by the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*} \left( r \, dt + \theta^1 \left( \theta^1 \, dt + dW_t^1 \right) \right) \tag{10.2.11}$$

for $t \in [0, \infty)$ and $S_0^{\delta_*} > 0$ with market price of risk

$$\theta^1 = \frac{a^1 - r}{b^{1,1}}, \tag{10.2.12}$$

see (10.1.5). In Fig. 10.2.1 we also chart the path of the GOP. Its volatility is $\theta^1 = 0.1$, and we have chosen $S_0^{\delta_*} = 1$. The figure shows that the GOP $S^{\delta_*}$ and the primary security account $S^1$ exhibit perfectly correlated fluctuations in this simple BS model. This, of course, will be no longer the case for a typical example with $d > 1$.

## 10.3 Supermartingale Property

### Benchmarking

In this book we employ the GOP as a benchmark for problems in derivative pricing and portfolio optimization, and refer to the resulting methodology as the *benchmark approach*. The GOP serves as a benchmark in several ways. In the previous chapter we used it as a numeraire portfolio for the pricing of derivatives. In Sect. 10.5 we shall see that it is the best performing portfolio according to several performance measures and, thus, an ideal benchmark for investment management.

As in Sect. 9.1, we call any security, when expressed in units of the GOP, a benchmarked security and refer to this procedure as benchmarking. The benchmarked portfolio value of a portfolio $S^\delta$ is given by the ratio

$$\hat{S}_t^\delta = \frac{S_t^\delta}{S_t^{\delta_*}} \tag{10.3.1}$$

at time $t$. By the Itô formula it satisfies the SDE

$$d\hat{S}_t^\delta = \sum_{j=0}^{d} \delta_t^j \, \hat{S}_t^j \sum_{k=1}^{d} \left( b_t^{j,k} - \theta_t^k \right) dW_t^k \tag{10.3.2}$$

for $t \in [0, \infty)$, see Exercise 10.2. Note that this SDE holds in general, even for portfolios that can become zero or negative.

### Benchmarked Portfolios as Supermartingales

Since (10.3.2) is a driftless SDE it follows from Lemma 5.4.1 that any benchmarked portfolio is an $(\underline{\mathcal{A}}, P)$-local martingale. The set of *nonnegative portfolios*, denoted by $\mathcal{V}$, is of particular importance. For instance, the portfolios of total tradable wealth of investors are nonnegative. Since any nonnegative local martingale is an $(\underline{\mathcal{A}}, P)$-supermartingale according to Lemma 5.2.3, we obtain the following fundamental statement directly.

**Theorem 10.3.1.** *In a CFM $\mathcal{S}_{(d)}^C$ it follows that $\hat{S}^\delta$ is an $(\underline{\mathcal{A}}, P)$-supermartingale for any nonnegative portfolio $S^\delta \in \mathcal{V}$.*

The supermartingale property of benchmarked nonnegative securities is fundamental. A number of important results follow directly from it. We shall derive some of these in the sequel.

### Arbitrage

For a financial market model to be viable one needs to exclude some basic form of arbitrage. This is of particular importance for the portfolios of total tradable wealth of investors. The investors, sometimes also called market participants, are the only agents who could exploit potential arbitrage opportunities in the market. The nonnegativity of their portfolios is legally enforced by their *limited liability*. Since we consider the total tradable wealth, limited liability does not permit a market participant to trade any longer if her or his total portfolio of tradable wealth becomes negative. She or he has to declare bankruptcy in such a case. This is an extremely important feature of any real market which should be reflected in a market model. The limited liability of investors is also essential for the functioning of the market. We conclude from Theorem 10.3.1 that the benchmarked portfolio of total tradable wealth of each investor is a supermartingale. This fact will play an important role in ensuring the absence of arbitrage. The definition of arbitrage given below focuses on nonnegative portfolios. These are the only portfolios that we have to consider, due to the limited liability of investors. We introduce the following rather weak notion of arbitrage, as described in Platen (2002).

**Definition 10.3.2.** *A nonnegative portfolio $S^\delta \in \mathcal{V}$ is an* arbitrage *if it starts at zero, that is $S^\delta_0 = 0$ almost surely, and is strictly positive with strictly positive probability at a later bounded stopping time $\tau \in (0, \infty)$, that is,*

$$P(S^\delta_\tau > 0) > 0. \tag{10.3.3}$$

We shall apply this notion of arbitrage generally throughout the book, not only for CFMs. Note that we do not consider arbitrage for negative portfolios or portfolios that can be positive and negative. Such portfolios cannot represent the total tradable wealth of investors, who are the only ones who could exploit potential arbitrage opportunities. A similar arbitrage definition as above was used in Loewenstein & Willard (2000).

### No Arbitrage in a CFM

Using Theorem 10.3.1 and Definition 10.3.2, we now prove an important result.

**Corollary 10.3.3.** *A CFM $\mathcal{S}^C_{(d)}$ does not allow arbitrage with any of its nonnegative portfolios.*

**Proof**: For any nonnegative portfolio $S^\delta \in \mathcal{V}$, with $S^\delta_0 = 0$ almost surely, it follows from the supermartingale property of $\hat{S}^\delta$, see Theorem 10.3.1, and the Optional Sampling Theorem, see (5.1.9), that

$$0 = \hat{S}^\delta_0 \geq E\left(\hat{S}^\delta_\tau \,\middle|\, \mathcal{A}_0\right) = E\left(\hat{S}^\delta_\tau\right) \geq 0$$

for any bounded stopping time $\tau \in [0, \infty)$. Therefore, due to the fact that $\hat{S}^\delta$ is nonnegative and $S^{\delta_*}$ strictly positive it follows that

$$P(S^\delta_\tau > 0) = P(\hat{S}^\delta_\tau > 0) = 0. \tag{10.3.4}$$

□

This demonstrates that all trajectories of the portfolio of tradable wealth of an investor are absorbed at zero when reaching zero.

### A Classical Form of Arbitrage (*)

As we shall see in the next chapter, due to the possibility of allowing benchmarked savings accounts to be strict supermartingales, the benchmark approach provides a richer modeling framework than the classical no-arbitrage approach. The classical *arbitrage pricing theory* (APT) is described, for instance, in Harrison & Kreps (1979), Harrison & Pliska (1981) and Delbaen & Schachermayer (1994, 1998, 2006). The *no free lunch with vanishing risk* (NFLVR) concept provides the most general formulation of the APT and is linked to the *fundamental theorem of asset pricing* in Delbaen & Schachermayer (1998, 2006). This theorem states essentially that NFLVR is equivalent to the existence of an equivalent risk neutral probability measure.

Note that the no-arbitrage criterion resulting from Definition 10.3.2 is weaker than the NFLVR condition. This means that some financial market models that exclude arbitrage in the sense of Definition 10.3.2 may in fact not admit any equivalent risk neutral probability measure. This creates no problem from the viewpoint of the benchmark approach. As argued in Loewenstein & Willard (2000) the real economic content of a no-arbitrage condition lies in the existence of a competitive equilibrium in the sense that an investor who prefers more to less should find an optimal trading strategy if and only if arbitrage opportunities similar to those in Definition 10.3.2 are excluded. We shall see later that the existence of the GOP allows the investor to find a strategy that optimizes her or his wealth for a range of different objectives and models. However, when the GOP explodes, then portfolio optimization does not make sense and there is arbitrage. In this case we no longer have a reasonable model.

In the APT negative portfolios are not excluded in its typical definition of arbitrage. Furthermore, under the APT one guarantees NFLVR in the denomination of, say, a given currency with some quite subjectively chosen lower bound. It seems not trivial to make this choice invariant under different denominations of the securities. Theoretically and also practically this may be not completely satisfactory. The arbitrage concept in Definition 10.3.2 does not have such a problem. The lower bound for the total tradable wealth of an investor is zero, which expresses her or his limited liability. This lower bound is also zero in any currency or other denomination.

The extra modeling freedom that the benchmark approach provides will be essential for accommodating alternative models that may be realistic descriptions of the typical market dynamics. We shall demonstrate that financial market modeling, derivative pricing, hedging and portfolio optimization can be conveniently performed under the real world probability measure without assuming the existence of an equivalent risk neutral probability measure. The existence of such a probability measure is not relevant from an economic point of view. However, in the historical development of the APT it has been a convenient and very helpful assumption. As we shall see, this assumption seems to be too restrictive for the construction of realistic models for the long term dynamics of observed markets.

We may summarize the above discussion by saying that the existence of an equivalent risk neutral probability measure is a mathematical convenience, but not a necessity. What is important in a market model is the existence of a GOP. This provides the supermartingale property of the benchmarked portfolios of total tradable wealth of investors. As we shall see, the GOP is crucial for determining optimal investment strategies and for pricing and hedging of derivatives.

## 10.4 Real World Pricing

### Fair Prices are Minimal Prices

In Definition 9.1.2 we introduced the key notion of a fair security price process. According to this definition, the GOP is used for benchmarking, and benchmarked securities that form martingales are called fair. By its martingale property the benchmarked value of a fair price is the best forecast of its future benchmarked values. Let $H$ denote an $\mathcal{A}_\tau$-measurable *payoff* payable at a bounded stopping time $\tau$, with $E(\frac{H}{S^{\delta_*}_\tau}) < \infty$. Then the real world pricing formula (9.1.30) expresses the fair price $U_H(t \wedge \tau)$ of $H$ at time $t \in [0, \infty)$ as

$$U_H(t \wedge \tau) = S^{\delta_*}_{t\wedge\tau}\, E\left(\frac{H}{S^{\delta_*}_\tau} \,\middle|\, \mathcal{A}_t\right), \tag{10.4.1}$$

where $t \wedge \tau = \min(t, \tau)$. If the CFM represents a Markovian system of stochastic processes, then the Feynman-Kac formula, see Sect. 9.7, allows us to calculate the benchmarked fair pricing function by using the real world pricing formula (9.1.34). We have seen such calculations in Sect. 9.1 for options under the BS model. In the next chapter we shall study real world pricing for other models.

We recall from Theorem 10.3.1 that any nonnegative benchmarked portfolio in a CFM is a supermartingale. Let us compare a nonnegative supermartingale $X = \{X_t,\, t \in [0, \infty)\}$ and a nonnegative martingale $M = \{M_t,\, t \in [0, \infty)\}$ that both have almost surely the same random value at some bounded stopping time $\tau \in (0, \infty)$, that is $P(X_\tau = M_\tau) = 1$. It follows from the supermartingale inequality (5.1.7), the martingale equality (5.1.2) and the Optional Sampling Theorem, see (5.1.19), that the initial value of the martingale cannot be greater than that of the supermartingale. This means, that for $t \in [0, \infty)$ one has

$$M_{t\wedge\tau} = E\left(M_\tau \,\middle|\, \mathcal{A}_t\right) = E\left(X_\tau \,\middle|\, \mathcal{A}_t\right) \leq X_{t\wedge\tau}. \tag{10.4.2}$$

Let us summarize this important result in the following lemma.

**Lemma 10.4.1.** *A martingale is the minimal nonnegative supermartingale that reaches at a bounded stopping time a given nonnegative integrable value.*

The fundamental fact that a nonnegative martingale is the minimal nonnegative supermartingale that matches a given random variable at a bounded stopping time allows us to draw the following conclusion.

**Corollary 10.4.2.** *Consider a nonnegative fair portfolio $S^\delta \in \mathcal{V}$ in a CFM and let $S^{\delta'} \in \mathcal{V}$ be a second nonnegative portfolio such $S^\delta_\tau = S^{\delta'}_\tau$ almost surely at some bounded stopping time $\tau$. Then at $t \in [0, \infty)$ we have*

$$S^\delta_{t\wedge\tau} \leq S^{\delta'}_{t\wedge\tau} \tag{10.4.3}$$

*almost surely.*

As we shall see later, there are, in general, also other portfolios that can generate the same future payoff. However, when these other portfolios are benchmarked, they will turn out to be strict supermartingales and will have, therefore, an initial value above that of the fair portfolio. It follows that fair prices are minimal in the above sense. Accordingly, it is clear that a rational investor, who wants to obtain a certain future payoff should always form a fair portfolio with her or his total tradable wealth.

By these arguments it becomes also clear that in a competitive market the fair price of a hedgable derivative is the economically correct price. Therefore, we shall price hedgable derivatives in this way. As we shall see in Sect. 11.4 the fair price is consistent with utility indifference pricing for nonhedgable payoffs.

## Risk Neutral Pricing

Consider for the moment in this subsection the case when an equivalent risk neutral probability measure $P_\theta$ exists with Radon-Nikodym derivative

$$\Lambda_\theta(t) = \frac{dP_\theta}{dP}\Big|_{\mathcal{A}_t} = \frac{\hat{S}^0_t}{\hat{S}^0_0} = \frac{S^0_t\, S^{\delta_*}_0}{S^{\delta_*}_t\, S^0_0}, \tag{10.4.4}$$

see Sect. 9.4. By the Girsanov Theorem and the Bayes Rule, described in Sect. 9.5, the risk neutral derivative price of an $\mathcal{A}_T$-measurable payoff $H$ satisfies the risk neutral pricing formula

$$U_H(t) = E_\theta \left( \frac{S_t^0}{S_T^0} \, H \, \Big| \, \mathcal{A}_t \right)$$

for $t \in [0,T]$. By application of Bayes's Theorem, see (9.5.10), with Radon-Nikodym derivative (10.4.4), we see that the risk neutral price equals the fair price

$$\begin{aligned} U_H(t) &= E\left( \frac{\Lambda_\theta(T)}{\Lambda_\theta(t)} \, \frac{S_t^0}{S_T^0} \, H \, \Big| \, \mathcal{A}_t \right) \\ &= E\left( \frac{S_T^0 \, S_t^{\delta_*}}{S_T^{\delta_*} \, S_t^0} \, \frac{S_t^0}{S_T^0} \, H \, \Big| \, \mathcal{A}_t \right) \\ &= S_t^{\delta_*} \, E\left( \frac{H}{S_T^{\delta_*}} \, \Big| \, \mathcal{A}_t \right) \end{aligned} \tag{10.4.5}$$

for $t \in [0,T]$. Consequently, risk neutral pricing is a special case of real world pricing in a CFM.

In the literature it is common practice to specify a model directly under an equivalent risk neutral probability measure, thereby begging the question about the existence of such an object. In our view this is not satisfactory, since the risk neutral approach breaks down when the Radon-Nikodym derivative process is not an $(\underline{\mathcal{A}}, P)$-martingale. We shall demonstrate this later in Chap. 12 for realistic models. Under the benchmark approach, the Radon-Nikodym derivative process becomes an object for explicit modeling.

### Zero Coupon Bond and Actuarial Pricing

Let us discuss an important example of real world pricing. Further issues on pricing and hedging will be discussed in Sects. 11.4 and 11.5. The simplest derivative payoff is that of a zero coupon bond, which pays one unit of the domestic currency at maturity $T \in [0,\infty)$.

Using the real world pricing formula (10.4.1), the fair price $P(t,T)$ at time $t$ of such an instrument is given by the equation

$$P(t,T) = S_t^{\delta_*} \, E\left( \frac{1}{S_T^{\delta_*}} \, \Big| \, \mathcal{A}_t \right) \tag{10.4.6}$$

for $t \in [0,T]$. Furthermore, for any $\mathcal{A}_T$-measurable payoff $H$ that is independent of $S_T^{\delta_*}$ we obtain from the real world pricing formula (10.4.1) and (10.4.6) the pricing formula

$$U_H(t) = S_t^{\delta_*} \, E\left( \frac{1}{S_T^{\delta_*}} \, \Big| \, \mathcal{A}_t \right) E\left(H \,\big|\, \mathcal{A}_t\right) = P(t,T) \, E\left(H \,\big|\, \mathcal{A}_t\right) \tag{10.4.7}$$

for $t \in [0,T]$ and $T \in [0,\infty)$, see (9.2.6). This is known as the *actuarial pricing formula* or net present value formula and is widely used in many areas, including insurance and accounting.

### Zero Coupon Bond SDE

If we benchmark the fair zero coupon bond price (10.4.6), then we obtain

$$\hat{P}(t,T) = \frac{P(t,T)}{S_t^{\delta_*}} = E\left(\frac{1}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t\right) \tag{10.4.8}$$

for $t \in [0,T]$. The benchmarked zero coupon bond price process $\hat{P}(\cdot,T) = \{\hat{P}(t,T),\ t \in [0,T]\}$ is an $(\underline{\mathcal{A}}, P)$-martingale. For particular models, we shall calculate benchmarked zero coupon bond prices explicitly in later chapters. Since these models will be Markovian, the Feynman-Kac formula, see Sect. 9.7, is applicable for this purpose.

By assuming that bonds are traded securities, it is reasonable to assume that for each $t \in [0,T]$ and $k \in \{1,2,\ldots,d\}$, there exists a unique, predictable *benchmarked bond volatility process* $\sigma^k(\cdot,T) = \{\sigma^k(t,T),\ t \in [0,T]\}$, such that

$$d\hat{P}(t,T) = -\hat{P}(t,T) \sum_{k=1}^{d} \sigma^k(t,T)\, dW_t^k \tag{10.4.9}$$

and thus

$$\hat{P}(t,T) = -\hat{P}(0,T)\, \exp\left\{ -\sum_{k=1}^{d} \left( \int_0^t \frac{(\sigma^k(s,T))^2}{2}\, ds + \int_0^t \sigma^k(s,T)\, dW_s^k \right) \right\} \tag{10.4.10}$$

for $t \in [0,T]$. Then by (10.4.8), (10.4.9) and (10.2.8) and application of the Itô formula we obtain the SDE

$$dP(t,T) = P(t,T) \left( r_t\, dt + \sum_{k=1}^{d} \left(\theta_t^k - \sigma^k(t,T)\right) \left(\theta_t^k\, dt + dW_t^k\right) \right) \tag{10.4.11}$$

for $t \in [0,T]$.

### Forward Rate Equation

Zero coupon bond prices are not the most common way to express the *interest rate term structure*, that is, the information about fixed term borrowing and lending of domestic currency. A more convenient representation is in terms of the family of forward rates or the forward rate curve. Yield curves, discretely compounded forward rates or, so-called, LIBOR rates are also common. These representations are all closely related, see Musiela & Rutkowski (2005) and Brigo & Mercurio (2005).

The time $t$ *forward rate* $f(t,T)$ for the maturity date $T \in [0,\infty)$ is defined as

$$f(t,T) = -\frac{\partial}{\partial T}\, \ln(P(t,T)). \tag{10.4.12}$$

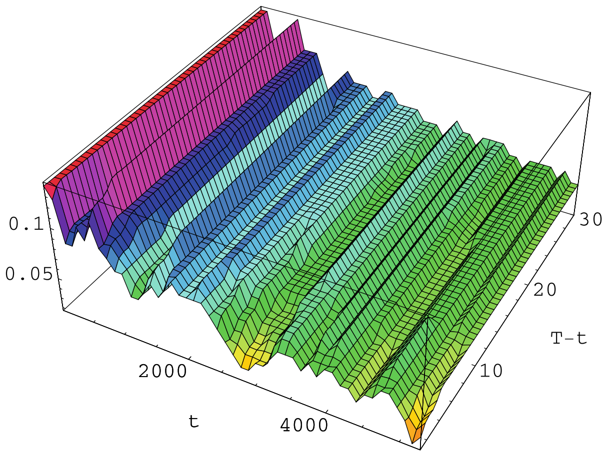



**Fig. 10.4.1.** Forward rate curve evolution from 1982 until 2002

This yields

$$P(t,T) = \exp\left\{-\int_t^T f(t,u)\,du\right\} \tag{10.4.13}$$

for all $t \in [0,T]$. Figure 10.4.1 plots the US dollar forward rate curve out to 30 years over the period from 1982 until 2002. We see that the forward rates are stochastic.

From (10.4.10) and (10.4.12) and an application of the Itô formula and the generalized Fubini theorem one obtains the *forward rate equation*

$$\begin{aligned} f(t,T) &= -\frac{\partial}{\partial T}\,\ln(\hat{P}(t,T)) \\ &= -\frac{\partial}{\partial T}\left(\ln(\hat{P}(0,T)) + \sum_{k=1}^{d}\left(-\frac{1}{2}\int_0^t (\sigma^k(s,T))^2\,ds - \int_0^t \sigma^k(s,T)\,dW_s^k\right)\right) \\ &= f(0,T) + \sum_{k=1}^{d}\int_0^t \left(\frac{\partial}{\partial T}\,\sigma^k(s,T)\right)\left(\sigma^k(s,T)\,ds + dW_s^k\right) \end{aligned} \tag{10.4.14}$$

for $t \in [0,T]$, see Platen (2002). Here, and in what follows, we shall use $\frac{\partial}{\partial T} f(\cdot,\cdot)$ to denote the partial derivative of $f(\cdot,\cdot)$ with respect to the second variable.

According to (10.4.14) the drift and diffusion coefficients of the forward rates depend only on the volatilities of the benchmarked zero coupon bonds. Note that $W^1,\ldots,W^d$ are Wiener processes under the real world probability measure.

The shortest forward rate is the short rate. We obtain the following expression for the short rate $r_t$ from (10.4.14):

$$r_t = f(t,t) = f(0,t) + \sum_{k=1}^{d} \int_0^t \left( \frac{\partial}{\partial T} \sigma^k(s,t) \right) (\sigma^k(s,t)\, ds + dW_s^k) \quad (10.4.15)$$

for $t \in [0,\infty)$. By noting that a zero coupon bond close to maturity has a very small volatility it follows from (10.4.11) that

$$\sigma^k(t,t) = \theta_t^k \quad (10.4.16)$$

for $t \in [0,\infty)$ and $k \in \{1,2,\dots,d\}$. Therefore, the SDE for the short rate is

$$dr_t = \frac{\partial}{\partial T} f(0,t)\, dt + \sum_{k=1}^{d} \left( \frac{\partial}{\partial T} \sigma^k(t,t) \right) (\theta_t^k\, dt + dW_t^k) \quad (10.4.17)$$

for $t \in [0,\infty)$. This SDE is interesting, because it tells us that the slope $\frac{\partial}{\partial T} f(0,t)$ of the initial forward rate curve and the slope $\frac{\partial}{\partial T} \sigma^k(t,t)$ of the volatility of the benchmarked short term bond are essential ingredients for modeling the short rate dynamics.

### Heath-Jarrow-Morton Equation (*)

Equation (10.4.14) resembles the *Heath-Jarrow-Morton* (HJM) *equation*, see Heath, Jarrow & Morton (1992), which was derived under the assumption that an equivalent risk neutral probability measure $P_\theta$ exists. In a CFM the candidate risk neutral measure $P_\theta$ may be not equivalent to $P$. This does not create a problem for the benchmark approach, since we use real world pricing.

It is sometimes suggested in the literature that we are free to choose the drift coefficients of the forward rates under the real world probability measure $P$. Under the benchmark approach, as (10.4.14) shows, this is not the case. The drift and diffusion coefficients are determined by the volatilities of benchmarked zero coupon bonds.

Let us now recover the HJM equation from (10.4.14). For this purpose we introduce the *bond volatility*

$$\tilde{\sigma}^k(t,T) = \sigma^k(t,T) - \theta_t^k \quad (10.4.18)$$

for $t \in [0,T]$, $T \in [0,\infty)$ and $k \in \{1,2,\dots,d\}$, see (10.4.11). The HJM equation follows directly from (10.4.14) and (10.4.18) in the form

$$f(t,T) = f(0,T) + \sum_{k=1}^{d} \int_0^t \left( \frac{\partial}{\partial T} \tilde{\sigma}^k(s,T) \right) \left( \tilde{\sigma}^k(s,T)\, ds + d\tilde{W}_t^k \right) \quad (10.4.19)$$

with

$$d\tilde{W}_t^k = \theta_t^k\, dt + dW_t^k \quad (10.4.20)$$

for $t \in [0,T]$. If there exists an equivalent risk neutral probability measure $P_\theta$, then $\tilde{W}^k$ is a Wiener process under $P_\theta$. We have shown in (10.4.14) that a HJM-type equation holds in a more general setting than considered in Heath et al. (1992). In particular, (10.4.14) does not require the existence of an equivalent risk neutral probability measure. We shall see in Sect. 13.3 that the removal of this assumption allows the total market price of risk to impact the forward rate curve, which is economically reasonable. Most continuous interest rate term structure models in the literature, see, for instance, Musiela & Rutkowski (2005) and Sect. 4.3, fit into the framework above.

### Annuity and Savings Account (*)

Market conventions determine the maturity dates applicable to different contracts. These can have, for instance, a daily, weekly, monthly or yearly pattern. Typically, the maturity is selected from a set of possible dates $\{T_0, T_1, \ldots\}$. For simplicity, let us assume that the possible maturity dates are equally spaced with

$$T_i = i\,h, \tag{10.4.21}$$

for $i \in \{0, 1, \ldots\}$, where $h > 0$. Given such a set of possible maturity dates let

$$i_t = \max\{i \in \mathcal{N} : T_i \le t\} \tag{10.4.22}$$

denote the index of the last maturity date before $t \in [0, \infty)$.

Now, consider the *annuity* $B_h = \{B_h(t),\, t \in [0, \infty)\}$, which is an account of short term bonds that are rolled over as soon as they mature, that is,

$$B_h(t) = P(t, T_{i_t+1}) \prod_{k=1}^{i_t+1} \frac{1}{P(T_{k-1}, T_k)} \tag{10.4.23}$$

for $t \in [0, \infty)$, see Jamshidian (1989). This expresses the evolution of a roll-over zero coupon bond account. Here one starts with one dollar and continually invests in the zero coupon bond with the shortest maturity date. At maturity one is rolling over the wealth into the next zero coupon bond.

Since $h$ is small in practice, typically one day, it is theoretically convenient to work with the almost sure limit $S^0_t$ of the annuity $B_h(t)$. To formalize this, we assume that $B_h(t)$ converges almost surely to some limit $S^0_t$ as $h \to 0$ for all $t \in [0, \infty)$, see (2.7.1). Now, define the savings account

$$S^0_t \stackrel{\text{a.s.}}{=} \lim_{h \to 0} B_h(t) \tag{10.4.24}$$

at time $t \in [0, \infty)$. This also shows that the value $S^0_t$ of a savings account at time $t$ has to be interpreted as a limit, which is obtained from a sequence of roll-over short term fair zero coupon bond accounts. Assume that the resulting domestic savings account value $S^0_t$ forms a continuous, adapted stochastic process and recall from (10.4.15) that the shortest forward rate equals the short rate $r_t = f(t, t)$. From (10.4.23) and (10.4.24) it follows that the savings account satisfies the differential equation

$$dS_t^0 = S_t^0\, r_t\, dt \tag{10.4.25}$$

for $t \in [0, \infty)$ with initial condition $S_0^0 = 1$. This coincides with formula (10.1.1) as expected.

The domestic savings account process $S^0 = \{S_t^0,\, t \in [0, \infty)\}$ provides a theoretical characterization of the *time value of money*. For convenience we interpret the savings account as a primary security account. This also means by (10.4.23) that, in reality, one has with $S^0$ an annuity which evolves over short time periods, say daily, like a fair bond. We shall see later for reasonable models that the savings account is not always a fair price process.

Finally, we mention that one can define an annuity similar to that in (10.4.23) in the denomination of units of a given primary security, for instance, in units of a foreign currency. In the case of a foreign currency the limit of the annuity yields the corresponding foreign savings account. In the case of a stock as primary security a corresponding share savings account is obtained.

### Forward Contract (*)

Let us introduce a *forward contract* on a given portfolio $S^\delta$, written at time $t \in [0, T)$ with delivery date $T \in (0, \infty)$. The *forward price* $F^\delta(t, T)$ at time $t \in [0, T]$ is the $\mathcal{A}_t$-measurable value that makes the fair value at time $t$ of the payoff $H = S_T^\delta - F^\delta(t, T)$ to zero. This means that for fixed $t \in [0, T)$ and $T \in (0, \infty)$ by the real world pricing formula (10.4.1) that

$$0 = S_t^{\delta_*}\, E\left(\frac{H}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) = S_t^{\delta_*}\, E\left(\frac{S_T^\delta - F^\delta(t, T)}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right). \tag{10.4.26}$$

By using the fair zero coupon bond price it follows the forward price

$$F^\delta(t, T) = \frac{S_t^{\delta_*}\, E\left(\hat{S}_T^\delta \,\big|\, \mathcal{A}_t\right)}{S_t^{\delta_*}\, E\left(\frac{1}{\hat{S}_T^\delta} \,\big|\, \mathcal{A}_t\right)} = \frac{S_t^\delta\, E\left(\frac{\hat{S}_T^\delta}{\hat{S}_t^\delta} \,\big|\, \mathcal{A}_t\right)}{P(t, T)}. \tag{10.4.27}$$

For a fair portfolio $S^\delta$ this provides the forward price

$$F^\delta(t, T) = \frac{S_t^\delta}{P(t, T)}. \tag{10.4.28}$$

In the case when we have an explicit formula for the zero coupon bond the forward pricing formula is explicit, as long as $S^\delta$ is fair. In the case when $S^\delta$ is not fair, then there is still a fair price $U_{S_T^\delta}(t)$ at time $t$ for the payoff $S_T^\delta$, where by the real world pricing formula we obtain

$$U_{S_T^\delta}(t) = S_t^{\delta_*}\, E\left(\hat{S}_T^\delta \,\big|\, \mathcal{A}_t\right). \tag{10.4.29}$$

Thus, we obtain from (10.4.27) the forward price

$$F^{\delta}(t,T) = \frac{U_{S^{\delta}_T}(t)}{P(t,T)} \tag{10.4.30}$$

for $t \in [0,T]$.

## 10.5 GOP as Best Performing Portfolio

We shall now show that the GOP is not only the numeraire of choice for real world pricing, it is also a natural benchmark in portfolio optimization. We shall see that the GOP is the best investment portfolio under several objectives. This is also in line with views expressed, for instance, in Markowitz (1959, 1976), Latané (1959), Breiman (1961), Hakansson (1971), Thorp (1972), Rubinstein (1976), Cover (1991), Ziemba & Mulvey (1998), Browne (1999) and Stutzer (2000). We prove here such properties by using the fundamental supermartingale property of benchmarked, nonnegative portfolios.

### Outperforming Growth Rate and Expected Return

The GOP can be considered to be the best performing portfolio in various ways. In the following, we describe some mathematical manifestations of this fact in the setting of a CFM. Here appreciation rates, volatilities and short rates are allowed to be very flexible stochastic processes. Almost all of the following results hold also for more general financial markets, as we shall see for the case of jump diffusion markets in Sect. 14.1.

By inequality (10.2.3) it follows for any strictly positive portfolio process $S^{\delta} \in \mathcal{V}^{+}$ that its growth rate $g^{\delta}_t$ at any time $t$ is never greater than the growth rate $g^{\delta_*}_t$ of the GOP. This yields a first characterization of best performance shown in terms of growth rates.

From Theorem 10.3.1 we know that any strictly positive benchmarked portfolio $\hat{S}^{\delta}$ forms an $(\underline{\mathcal{A}}, P)$-supermartingale. This means that the expected return

$$E\left( \frac{\hat{S}^{\delta}_{t+h} - \hat{S}^{\delta}_t}{\hat{S}^{\delta}_t} \,\middle|\, \mathcal{A}_t \right) \leq 0 \tag{10.5.1}$$

is always nonpositive over any time period $[t, t+h] \subseteq [0,\infty)$ with length $h > 0$. This provides a second characterization of outperformance where the GOP, when used as benchmark, does not allow any strictly positive benchmarked portfolio to generate expected returns greater than zero.

### Outperforming the Long Term Growth Rate

Define for a strictly positive portfolio $S^{\delta} \in \mathcal{V}^{+}$ in a CFM its *long term growth rate* $\tilde{g}^{\delta}$ as the almost sure upper limit

$$\tilde{g}^{\delta} \overset{\text{a.s.}}{=} \limsup_{T\to\infty} \frac{1}{T} \ln\left(\frac{S_T^{\delta}}{S_0^{\delta}}\right), \tag{10.5.2}$$

assuming that this limit exists for the GOP.

Note that this pathwise defined quantity does not involve any expectation. For example, in the special case of a BS model, as discussed in Chap. 9 with constant short rate $r$ and constant total market price of risk $|\theta|$, the GOP has by the Law of Large Numbers, see Sect. 2.1, the long term growth rate $\tilde{g}^{\delta_*} = r + \frac{|\theta|^2}{2}$.

The following result presents the fascinating property of the GOP that after sufficient long time its trajectory attains almost surely pathwise a value not less than that of any other strictly positive portfolio.

**Theorem 10.5.1.** *In a CFM $\mathcal{S}_{(d)}^C$ the GOP $S^{\delta_*}$ attains almost surely the greatest long term growth rate compared with all other strictly positive portfolios $S^{\delta} \in \mathcal{V}^+$, that is*

$$\tilde{g}^{\delta_*} \geq \tilde{g}^{\delta}, \tag{10.5.3}$$

*almost surely.*

**Proof**: Similarly to, for instance, Karatzas & Shreve (1998) we consider a strictly positive portfolio $S^{\delta} \in \mathcal{V}^+$ with

$$S_0^{\delta} = S_0^{\delta_*}. \tag{10.5.4}$$

By Theorem 10.3.1 the strictly positive benchmarked portfolio $\hat{S}^{\delta}$ is an $(\underline{\mathcal{A}}, P)$-supermartingale. As a supermartingale it satisfies by (10.5.4) and (5.1.20) the inequality

$$\exp\{\varepsilon\, k\}\, P\left(\sup_{k\le t<\infty} \hat{S}_t^{\delta} > \exp\{\varepsilon\, k\} \,\Big|\, \mathcal{A}_0\right) \leq E\left(\hat{S}_k^{\delta} \,\Big|\, \mathcal{A}_0\right) \leq \hat{S}_0^{\delta} = 1 \tag{10.5.5}$$

for all $k \in \mathcal{N}$ and $\varepsilon \in (0,1)$. For fixed $\varepsilon \in (0,1)$ one finds

$$\sum_{k=1}^{\infty} P\left(\sup_{k\le t<\infty} \ln\left(\hat{S}_t^{\delta}\right) > \varepsilon\, k \,\Big|\, \mathcal{A}_0\right) \leq \sum_{k=1}^{\infty} \exp\{-\varepsilon\, k\} < \infty. \tag{10.5.6}$$

Now, the Borel-Cantelli Lemma, see Sect. 2.7, implies the existence of a random variable $K_\varepsilon$ such that

$$\ln\left(\hat{S}_t^{\delta}\right) \leq \varepsilon\, k \leq \varepsilon\, t$$

for all $k \geq K_\varepsilon$ and $t \geq k$, almost surely. Thus, one has almost surely

$$\sup_{T\geq k} \frac{1}{T} \ln\left(\hat{S}_T^{\delta}\right) \leq \varepsilon$$

for all $k \geq K_\varepsilon$, and therefore

$$\limsup_{T\to\infty} \frac{1}{T} \ln\left(\frac{S_T^\delta}{S_0^\delta}\right) \leq \limsup_{T\to\infty} \frac{1}{T} \ln\left(\frac{S_T^{\delta_*}}{S_0^{\delta_*}}\right) + \varepsilon \tag{10.5.7}$$

almost surely. Noting that relation (10.5.7) holds for all $\varepsilon \in (0,1)$ the inequality (10.5.3) follows by (10.5.2). $\square$

This provides a third characterization of best performance via long term growth rates. It expresses a most desirable pathwise feature of the GOP for a long term investor. Since there is only one path of the GOP that the world market ever experiences, Theorem 10.5.1 makes the strong statement that the GOP provides the portfolio that almost surely delivers the best outcome after a sufficiently long time. An important practical problem then is: what time horizon is necessary so that one can benefit with reasonable probability from the above property of the GOP. This, of course, depends on the dynamics of the underlying market.

## Systematic Outperformance

For an investor it is of interest to know also over short time periods whether there exists no other portfolio that can *systematically outperform* the best performing portfolio with some strictly positive probability. If there were no such best performing portfolio, then one could say that the market permits some form of arbitrage, in the sense that there would be always a better performing portfolio, if compared to any given candidate of a best performing portfolio. To formulate rigorously this fourth characterization of best performance of the GOP, we introduce the following definition, see Platen (2004a).

**Definition 10.5.2.** *A strictly positive portfolio $S^\delta \in \mathcal{V}^+$ is said to* systematically outperform *another strictly positive portfolio $S^{\bar{\delta}} \in \mathcal{V}^+$ if for some bounded stopping times $\tau \in [0,\infty)$ and $\sigma \in [\tau,\infty)$ with*

$$S_\tau^\delta = S_\tau^{\bar{\delta}} \tag{10.5.8}$$

*and*

$$S_\sigma^\delta \geq S_\sigma^{\bar{\delta}} \tag{10.5.9}$$

*almost surely, it holds that*

$$P\left(S_\sigma^\delta > S_\sigma^{\bar{\delta}} \,\middle|\, \mathcal{A}_\tau\right) > 0. \tag{10.5.10}$$

According to Definition 10.5.2, if a nonnegative portfolio systematically outperforms the GOP, then, with some strictly positive probability, it can generate wealth that is strictly greater than that accrued via the GOP over some period. We can now prove the following result.

**Corollary 10.5.3.** *In a CFM $\mathcal{S}_{(d)}^C$ no strictly positive portfolio systematically outperforms the GOP.*

**Proof**: Consider a benchmarked, nonnegative portfolio $\hat{S}^\delta = \{\hat{S}^\delta_t, t \in [0,\infty)\}$ with benchmarked value

$$\hat{S}^\delta_\tau = 1 \tag{10.5.11}$$

almost surely at a stopping time $\tau \in [0,\infty)$. Assume for a bounded stopping time $\sigma \in [\tau,\infty)$ the inequality

$$\hat{S}^\delta_\sigma \geq 1, \tag{10.5.12}$$

almost surely. From the supermartingale property of $\hat{S}^\delta$ that is provided by Theorem 10.3.1, the Optional Sampling Theorem, see (5.1.19), and from (10.5.11) and (10.5.9) it follows that

$$0 \geq E\left(\hat{S}^\delta_\sigma - \hat{S}^\delta_\tau \,\middle|\, \mathcal{A}_\tau\right) = E\left(\hat{S}^\delta_\sigma - 1 \,\middle|\, \mathcal{A}_\tau\right) \geq 0. \tag{10.5.13}$$

Due to (10.5.13) and (10.5.12), the benchmarked value $\hat{S}^\delta_\sigma$ cannot be strictly greater than $\hat{S}^\delta_\tau = 1$ with any strictly positive probability. Thus, it follows by (10.5.12) that $\hat{S}^\delta_\sigma = 1$ almost surely, which means that $S^\delta_\sigma = S^{\delta_*}_\sigma$. This proves the statement of Corollary 10.5.3 by Definition 10.5.2. □

Note that in Fernholz, Karatzas & Kardaras (2005) a notion of *relative arbitrage* has been introduced that turns out to be similar to the above notion of systematic outperformance, see Platen (2004a). Another notion of relative arbitrage in the Fernholz et al. (2005) paper is related to the long term growth rates of portfolios, as discussed in Theorem 10.5.1 and Platen (2004a). It has been shown in Fernholz et al. (2005) that for particular models certain proposed portfolios can systematically outperform the *market portfolio* (MP). This is consistent with Corollary 10.5.3 as long as the MP does not equal the GOP. However, if the MP equals the GOP, then by Corollary 10.5.3 this is not possible.

Finally, we remark that Delbaen & Schachermayer (1995) used an interesting notion of a *maximal element*, which is related to that of a portfolio which cannot be systematically outperformed. The GOP turns out to be a maximal element in the sense of Delbaen & Schachermayer (1995).

## 10.6 Diversified Portfolios in CFMs

### Sequence of CFMs

Diversification is an ancient concept in portfolio optimization that has been successfully applied for centuries. To demonstrate its power we consider a market with a large number of primary security accounts.

Given our discussion on the ICAPM in Sect. 9.3, it would be desirable to extract a result that links the GOP closely to the *market portfolio* (MP) of tradable wealth. We shall show in this section that one can establish asymptotically such a result in a sequence of markets by exploiting diversification.

Without particular modeling assumptions it will be shown under some regularity condition that *diversified portfolios* (DPs) approximate the GOP. This will provide a robustness property for sequences of DPs and will identify these as proxies of the GOP, see Platen (2004c, 2005b).

For the construction of a sequence of CFMs we rely again on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ with filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t\in[0,\infty)}$, satisfying the usual conditions, see Sect. 5.1. The trading uncertainty is modeled by independent standard Wiener processes $W^k = \{W_t^k,\, t \in [0,\infty)\}$, $k \in \mathcal{N}$.

In what follows, we consider a sequence of CFMs $(\mathcal{S}^{\mathrm{C}}_{(d)})_{d\in\mathcal{N}}$ indexed by the number $d$ of risky primary security accounts in $\mathcal{S}^{\mathrm{C}}_{(d)}$, see Definition 10.1.2. For a given value of $d \in \mathcal{N}$, the corresponding $d$th CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$ comprises $d+1$ primary security accounts. These include a savings account $S^0_{(d)} = \{S^0_{(d)}(t),\, t \in [0,\infty)\}$, which is locally riskless and at time $t$ given as

$$S^0_{(d)}(t) = \exp\left\{\int_0^t r_s\, ds\right\} \tag{10.6.1}$$

for $t \in [0,\infty)$. Here $r = \{r_t,\, t \in [0,\infty)\}$ denotes the short rate process, which is assumed to be adapted. For simplicity, we keep the short rate process for all CFMs in our sequence the same. The $d$th CFM also includes $d$ nonnegative, risky primary security account processes $S^j_{(d)} = \{S^j_{(d)}(t),\, t \in [0,\infty)\}$, $j \in \{1,2,\ldots,d\}$, for which we shall rewrite the general SDE (10.1.7) in the form

$$dS^j_{(d)}(t) = S^j_{(d)}(t)\left(r_t\, dt + \sum_{k=1}^d \left(\sigma^{0,k}_{(d)}(t) - \sigma^{j,k}_{(d)}(t)\right)\left(\sigma^{0,k}_{(d)}(t)\, dt + dW_t^k\right)\right) \tag{10.6.2}$$

using the volatility

$$b^{j,k}_{(d)}(t) = \sigma^{0,k}_{(d)}(t) - \sigma^{j,k}_{(d)}(t) \tag{10.6.3}$$

and the market price of risk

$$\theta^k_{(d)}(t) = \sigma^{0,k}_{(d)}(t) \tag{10.6.4}$$

for $j, k \in \{1,2,\ldots,d\}$, $d \in \mathcal{N}$ and $t \in [0,\infty)$.

### General and Specific Market Risk

We shall see that the above representation of volatilities turns out to be convenient in what follows. One may call the market price of risk $\theta_t^k$ the *general market volatility* with respect to the $k$th Wiener process. Since $\theta_t^k$ is the $k$th volatility of the GOP one can interpret it as a measure of the $k$th *general market risk*, that is the fluctuations of the market as a whole caused by $W^k$. We call the predictable process $\sigma^{j,k} = \{\sigma^{j,k}_{(d)}(t),\, t \in [0,\infty)\}$ the $(j,k)$th *specific volatility*, since it is the volatility of the GOP when denominated in units of the $j$th primary security account with respect to the $k$th Wiener process, $j \in \{0, 1, \ldots, d\}$, $k \in \{1, 2, \ldots, d\}$. According to (10.3.2) this is the negative $k$th volatility of the benchmarked $j$th primary security account, that is, when $\delta_t^j = 1$ and $\delta_t^i = 0$ for $i \neq j$. Since $\sigma^{j,k}$ measures the specific fluctuations of the $j$th primary security account caused by $W^k$ that are not captured by the GOP one can say that it measures the $k$th *specific market risk* with respect to $W^k$, see Platen & Stahl (2003).

Now, let us introduce the following definition.

**Definition 10.6.1.** *We call a sequence* $(\mathcal{S}_{(d)}^C)_{d \in \mathcal{N}}$ *a* sequence of CFMs *if in* $\mathcal{S}_{(d)}^C$ *the primary security accounts satisfy SDEs of the type* (10.6.2)*, the* $d \times d$ *volatility matrix*

$$\boldsymbol{b}_{(d)}(t) = [b_{(d)}^{j,k}(t)]_{j,k=1}^d = [\sigma_{(d)}^{0,k}(t) - \sigma_{(d)}^{j,k}(t)]_{j,k=1}^d \tag{10.6.5}$$

*is invertible for Lebesgue-almost every* $t \in [0, \infty)$*, and for all* $j \in \{0, 1, \ldots\}$*,* $k \in \mathcal{N}$ *and* $t \in [0, \infty)$ *the* $(j,k)$*th specific volatility* $\sigma_{(d)}^{j,k}(t)$ *converges for all* $t \in [0, \infty)$ *almost surely to a finite limit* $\sigma^{j,k}(t)$ *as* $d \to \infty$*, that is*

$$\lim_{d \to \infty} \sigma_{(d)}^{j,k}(t) \overset{a.s.}{=} \sigma^{j,k}(t) < \infty. \tag{10.6.6}$$

### Key Relations in CFMs

Before we go any further, let us adapt our notation to the given setting of a sequence of CFMs. In our extended setup we recall some of the key notations and relationships in a CFM by using a slightly refined notation. For $d \in \mathcal{N}$ we call a predictable process $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \delta_t^1, \ldots, \delta_t^d)^\top, t \in [0, \infty)\}$ a strategy under the $d$th CFM $\mathcal{S}_{(d)}^{\mathrm{C}}$ with portfolio

$$S_{(d)}^\delta(t) = \sum_{j=0}^d \delta_t^j \, S_{(d)}^j(t), \tag{10.6.7}$$

see (10.1.9). Note that the strategy $\delta$ depends on $d$, however, for simpler notation we shall suppress this dependence at the beginning of our discussion until it becomes necessary. For a given strategy $\delta$ we use the $j$th fraction

$$\pi_{\delta,t}^j = \delta_t^j \, \frac{S_{(d)}^j(t)}{S_{(d)}^\delta(t)} \tag{10.6.8}$$

for $t \in [0, \infty)$ and $j \in \{0, 1, \ldots, d\}$, see (10.1.12).

As shown in Sect. 10.1, there exists in the $d$th CFM $\mathcal{S}_{(d)}^{\mathrm{C}}$ a strictly positive portfolio process $S_{(d)}^{\delta_*} = \{S_{(d)}^{\delta_*}(t), t \in [0, \infty)\}$, the $d$th GOP, such that any portfolio $S_{(d)}^\delta(t)$ in $\mathcal{S}_{(d)}^{\mathrm{C}}$, when expressed in units of $S_{(d)}^{\delta_*}(t)$, yields a corresponding benchmarked portfolio $\hat{S}_{(d)}^\delta = \{\hat{S}_{(d)}^\delta(t), t \in [0, \infty)\}$, defined by

$$\hat{S}^{\delta}_{(d)}(t) = \frac{S^{\delta}_{(d)}(t)}{S^{\delta_*}_{(d)}(t)} \tag{10.6.9}$$

for $t \in [0, \infty)$. When strictly positive, this value satisfies by (10.3.2) and (10.6.8) the driftless SDE

$$d\hat{S}^{\delta}_{(d)}(t) = -\hat{S}^{\delta}_{(d)}(t) \sum_{j=0}^{d} \pi^{j}_{\delta,t} \sum_{k=1}^{d} \sigma^{j,k}_{(d)}(t)\, dW^k_t \tag{10.6.10}$$

for $t \in [0, \infty)$ with $\hat{S}^{\delta}_{(d)}(0) > 0$.

The $d$th GOP value $S^{\delta_*}_{(d)}(t)$ solves by (10.2.8) and (10.6.4) the SDE

$$dS^{\delta_*}_{(d)}(t) = S^{\delta_*}_{(d)}(t) \left( r_t\, dt + \sum_{k=1}^{d} \sigma^{0,k}_{(d)}(t) \left( \sigma^{0,k}_{(d)}(t)\, dt + dW^k_t \right) \right) \tag{10.6.11}$$

for $t \in [0, \infty)$. Here we assume, for simplicity, that $\hat{S}^{0}_{(d)}(0) = 1$, so that by (10.6.9) and (10.6.1) we obtain the initial value $S^{\delta_*}_{(d)}(0) = 1$. The value $S^{\delta}_{(d)}(t)$ at time $t$ of a strictly positive portfolio in the $d$th CFM satisfies by (10.1.14) and (10.6.3) the SDE

$$dS^{\delta}_{(d)}(t) = S^{\delta}_{(d)}(t) \left( r_t\, dt + \sum_{k=1}^{d} \sum_{j=0}^{d} \pi^{j}_{\delta,t} \left( \sigma^{0,k}_{(d)}(t) - \sigma^{j,k}_{(d)}(t) \right) \left( \sigma^{0,k}_{(d)}(t)\, dt + dW^k_t \right) \right) \tag{10.6.12}$$

for $t \in [0, \infty)$.

By (10.6.12) and applications of the Itô formula, the logarithm $\ln(S^{\delta}_{(d)}(t))$ of a strictly positive portfolio $S^{\delta}_{(d)}(t)$ at time $t$ satisfies then the SDE

$$d \ln \left( S^{\delta}_{(d)}(t) \right) = g^{\delta}_{(d)}(t)\, dt + \sum_{k=1}^{d} \sum_{j=0}^{d} \pi^{j}_{\delta,t} \left( \sigma^{0,k}_{(d)}(t) - \sigma^{j,k}_{(d)}(t) \right) dW^k_t \tag{10.6.13}$$

with *growth rate*

$$g^{\delta}_{(d)}(t) = r_t + \sum_{k=1}^{d} \left[ \sum_{j=1}^{d} \pi^{j}_{\delta,t} \left( \sigma^{0,k}_{(d)}(t) - \sigma^{j,k}_{(d)}(t) \right) \sigma^{0,k}_{(d)}(t) \right.$$
$$\left. - \frac{1}{2} \left( \sum_{j=1}^{d} \pi^{j}_{\delta,t} \left( \sigma^{0,k}_{(d)}(t) - \sigma^{j,k}_{(d)}(t) \right) \right)^2 \right] \tag{10.6.14}$$

for all $t \in [0, \infty)$. In the $d$th CFM the corresponding $d$th GOP $S^{\delta_*}_{(d)}$ achieves the maximum growth rate $g^{\delta_*}_{(d)}(t)$ for all $t \in [0, \infty)$. Thus, by (10.2.2), (10.2.6) and (10.6.4) one obtains the inequality

$$g^{\delta}_{(d)}(t) \leq g^{\delta_*}_{(d)}(t) = r_t + \sum_{k=1}^{d} \frac{1}{2} \left(\sigma^{0,k}_{(d)}(t)\right)^2 \tag{10.6.15}$$

for all $t \in [0, \infty)$ and all strictly positive portfolios $S^{\delta}_{(d)}$ in $\mathcal{S}^{\mathrm{C}}_{(d)}$.

### Sequences of Diversified Portfolios

For a sequence of CFMs we aim now to identify a class of corresponding sequences of portfolios that approximate the corresponding sequence of GOPs. From now on we shall indicate the dependence of a strategy $\delta_d$ on the number $d$ when formed in the $d$th CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$. Let us introduce the notion of a sequence of *diversified portfolios* (DPs). These are portfolios with vanishing fractions as $d \to \infty$.

**Definition 10.6.2.** *For a sequence of CFMs $(\mathcal{S}^{C}_{(d)})_{d\in\mathcal{N}}$ we call a corresponding sequence $(S^{\delta_d}_{(d)})_{d\in\mathcal{N}}$ of strictly positive portfolio processes $S^{\delta_d}_{(d)}$ with $S^{\delta_d}_{(d)}(0) = 1$* a sequence of diversified portfolios *(DPs) if some constants $K_1, K_2 \in (0, \infty)$ and $K_3 \in \mathcal{N}$ exist, independent of $d$, such that for $d \in \{K_3, K_3 + 1, \ldots\}$ one has the estimate*

$$\left|\pi^{j}_{\delta_d,t}\right| \leq \frac{K_2}{d^{\frac{1}{2}+K_1}} \tag{10.6.16}$$

*almost surely for all $j \in \{0, 1, \ldots, d\}$ and $t \in [0, \infty)$.*

This means that the fraction $\pi^{j}_{\delta_d,t}$ of the value of a DP, which is invested at time $t$ in the $j$th primary security account, $j \in \{0, 1, \ldots\}$, vanishes sufficiently fast as $d$ tends to infinity. More precisely, the fraction needs to decrease slightly faster than $d^{-\frac{1}{2}}$. For example, condition (10.6.16) is satisfied by a sequence of equally weighted portfolios, where $\pi^{i}_{\delta_d,t} = \pi^{j}_{\delta_d,t}$ for all $d \in \mathcal{N},\ t \in [0, \infty)$ and $i, j \in \{0, 1, \ldots, d\}$. For a sequence $(S^{\delta_d}_{(d)})_{d\in\mathcal{N}}$ of DPs the fractions may still vary considerably across the holdings of $S^{\delta_d}_{(d)}$ in primary security accounts. For instance, they may increase up to a multiple $\frac{a}{d+1}$ of the average fraction $\frac{1}{d+1}$ with some factor $a \geq 1$. From a practical point of view, condition (10.6.16) requires that the absolute fractions in a sequence of DPs are not of extreme magnitude when compared to the value $\frac{1}{d+1}$.

### Regular Sequence of CFMs

Consider the $d$th CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$ for fixed $d \in \mathcal{N}$ as an element of a sequence of CFMs $(\mathcal{S}^{\mathrm{C}}_{(d)})_{d\in\mathcal{N}}$. Then the $j$th benchmarked primary security account process $\hat{S}^{j}_{(d)} = \{\hat{S}^{j}_{(d)}(t),\ t \in [0, \infty)\}$, with

$$\hat{S}^j_{(d)}(t) = \frac{S^j_{(d)}(t)}{S^{\delta_*}_{(d)}(t)}, \tag{10.6.17}$$

satisfies by (10.6.10) and (10.6.8) the SDE

$$d\hat{S}^j_{(d)}(t) = -\hat{S}^j_{(d)}(t) \sum_{k=1}^{d} \sigma^{j,k}_{(d)}(t)\, dW^k_t \tag{10.6.18}$$

for $t \in [0, \infty)$ and $j \in \{0, 1, \ldots, d\}$. One can say that the $(j, k)$th specific volatility $\sigma^{j,k}_{(d)}(t)$ of the benchmarked $j$th primary security account $\hat{S}^j_{(d)}(t)$ measures its $k$th specific market risk at time $t \in [0, \infty)$, see the comments after (10.6.4). It is the specific risk with respect to the $k$th Wiener process for $k \in \{1, 2, \ldots, d\}$. This is the risk that measures the fluctuations of $S^j_{(d)}$ against the market as a whole, see Basle (1995) and Platen & Stahl (2003).

In order to obtain reasonable limits for sequences of DPs, some condition needs to be imposed on a given sequence of CFMs. For this purpose let us introduce for all $t \in [0, \infty)$, $d \in \mathcal{N}$ and $k \in \{1, 2, \ldots, d\}$ the $k$th *total specific volatility* for the $d$th CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$ in the form

$$\hat{\sigma}^k_{(d)}(t) = \sum_{j=0}^{d} |\sigma^{j,k}_{(d)}(t)| \tag{10.6.19}$$

$k \in \{1, 2, \ldots, d\}$. This quantity sums the absolute values of the specific volatilities with respect to the $k$th Wiener process in $\mathcal{S}^{\mathrm{C}}_{(d)}$. If the total specific volatility is small for some $k$, then only a few benchmarked primary security accounts have a larger specific volatility with respect to the $k$th Wiener process.

**Definition 10.6.3.** *A sequence of CFMs $(\mathcal{S}^C_{(d)})_{d \in \mathcal{N}}$ is called* regular *if there exists a constant $K_5 \in (0, \infty)$, independent of $d$, such that*

$$E\left(\left(\hat{\sigma}^k_{(d)}(t)\right)^2\right) \le K_5 \tag{10.6.20}$$

*for all $t \in [0, \infty)$, $d \in \mathcal{N}$ and $k \in \{1, 2, \ldots, d\}$.*

Condition (10.6.20) expresses the property that, for a regular sequence of CFMs, each of the independent sources of trading uncertainty influences only a restricted group of benchmarked primary security accounts, see (10.6.17) and (10.6.18). This condition appears to provide a reasonable assumption for the existing global financial market.

### Sequence of Approximate GOPs

For the application of the benchmark approach it is useful to be able to identify in practical terms, at least approximately, the GOP as a tradable portfolio in the market. To determine the fractions of the GOP exactly one needs an accurate model and accurate estimates of the volatilities and market prices of risk. It is a challenge to provide such a precise model and the corresponding required estimates from the available historical data. We propose here an alternative route where we construct proxies for the GOP.

For given $d \in \mathcal{N}$ let us consider in the $d$th CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$ a strictly positive portfolio process $S^{\delta_d}_{(d)}$ with strategy $\boldsymbol{\delta}_d = \{\boldsymbol{\delta}_d(t) = (\delta^0_d(t), \delta^1_d(t), \ldots, \delta^d_d(t))^\top$, $t \in [0,\infty)\}$ and

$$S^{\delta_*}_{(d)}(0) = 1. \tag{10.6.21}$$

We introduce the *tracking rate* $R^{\delta_d}_{(d)}(t)$ at time $t$ for the portfolio $S^{\delta_d}_{(d)}$ by setting

$$R^{\delta_d}_{(d)}(t) = \sum_{k=1}^{d} \left( \sum_{j=0}^{d} \pi^j_{\delta_d,t}\, \sigma^{j,k}_{(d)}(t) \right)^2 \tag{10.6.22}$$

for $t \in [0,\infty)$. By (10.6.3), (10.6.4) and (10.2.7) it follows for all $k \in \{1, 2, \ldots, d\}$ that

$$\sum_{j=0}^{d} \pi^j_{\delta_*,d,t}\, \sigma^{j,k}_{(d)}(t) = 0.$$

Therefore, the tracking rate in (10.6.22) is for the GOP $S^{\delta_*}_{(d)}$ zero, that is $R^{\delta_*}_{(d)}(t) = 0$ for all $t \in [0,\infty)$. By (10.6.10) it follows that the benchmarked portfolio $\hat{S}^{\delta_d}_{(d)}$ has quadratic variation

$$\left[\hat{S}^{\delta_d}_{(d)}\right]_t = \int_0^t R^{\delta_d}_{(d)}(s) \left(\hat{S}^{\delta_d}_{(d)}(s)\right)^2 ds \tag{10.6.23}$$

for $t \in [0,\infty)$. One notes that a portfolio $S^{\delta_d}_{(d)}$ with $S^{\delta_d}_{(d)}(0) = 1$ equals the $d$th GOP almost surely if and only if

$$R^{\delta_d}_{(d)}(t) = 0 \tag{10.6.24}$$

almost surely for all $t \in [0,\infty)$. Taking this into account, a given portfolio process $S^{\delta_d}_{(d)}$ can be expected to approximate the GOP $S^{\delta_*}_{(d)}$ if the tracking rate $R^{\delta_d}_{(d)}(t)$ becomes small for $d \to \infty$ and all $t \in [0,\infty)$. Let us formalize this observation.

**Definition 10.6.4.** *For a sequence of CFMs $(\mathcal{S}^C_{(d)})_{d\in\mathcal{N}}$ we call a sequence $(S^{\delta_d}_{(d)})_{d\in\mathcal{N}}$ of strictly positive portfolio processes, which start at the value one, a sequence of* approximate GOPs *if for all $t \in [0,\infty)$ the corresponding sequence of tracking rates vanishes in probability. That is, we have*

$$\lim_{d\to\infty} R^{\delta_d}_{(d)}(t) \stackrel{P}{=} 0 \tag{10.6.25}$$

*for all $t \in [0,\infty)$.*

Definition 10.6.4 will be exploited below.

### Diversification Theorem

Now, we shall see that without any major assumptions on the dynamics of the market, any sequence of DPs converges towards the sequence of GOPs. This reveals a fundamental robustness property of DPs and provides a theoretical basis for the intuitively known phenomenon of diversification in investment management. We emphasize that this property is model independent. Only a regularity condition for the market is required in the following *Diversification Theorem*, see Platen (2005b).

**Theorem 10.6.5.** *For a regular sequence of CFMs* $(\mathcal{S}^C_{(d)})_{d\in\mathcal{N}}$*, any sequence* $(S^{\delta_d}_{(d)})_{d\in\mathcal{N}}$ *of DPs is a sequence of approximate GOPs.*

**Proof**: First, let us estimate, for a given sequence $(S^{\delta_d}_{(d)})_{d\in\mathcal{N}}$ of DPs and for $d \in \{K_3, K_3+1, \ldots\}$ the tracking rate $R^{\delta_d}_{(d)}(t)$ for $S^{\delta_d}_{(d)}(t)$, see (10.6.22), at time $t \in [0,\infty)$. By (10.6.22) we see that

$$R^{\delta_d}_{(d)}(t) \leq \sum_{k=1}^{d} \left( \sum_{j=0}^{d} |\pi^j_{\delta_d,t}|\, |\sigma^{j,k}_{(d)}(t)| \right)^2 ,$$

which leads by (10.6.16) and (10.6.19) to the inequality

$$E\left(R^{\delta_d}_{(d)}(t)\right) \leq \frac{(K_2)^2}{d^{1+2K_1}} \sum_{k=1}^{d} E\left(\left(\hat{\sigma}^k_{(d)}(t)\right)^2\right) \leq \frac{(K_2)^2}{d^{2K_1}}\, K_5 \tag{10.6.26}$$

for all $t \in [0,\infty)$. Consequently, since by Definition 10.6.2 we have $K_1 > 0$, the expected tracking rate vanishes as $d \to \infty$ for all $t \in [0,\infty)$. Thus, by the Markov inequality (1.3.57) and Definition 10.6.4, a sequence of DPs is a sequence of approximate GOPs for any given regular sequence of CFMs. □

The above Diversification Theorem has some resemblance with the Central Limit Theorem (CLT) that we introduced in Sect. 2.1. If random variables are sufficiently independent and have, say, second moments, then under the CLT their equally weighted sum converges in distribution to a limit which has a Gaussian distribution. The particular type of distribution of the independent random variables in this weighted sum is not relevant. In this sense the CLT is model independent. Similarly, the above Diversification Theorem identifies the GOP as the limit of a sequence of approximately equally weighted portfolios and does not require to specify the distribution for the benchmarked primary security account processes. As long as the market is regular, Theorem 10.6.5 ensures that DPs converge towards the GOP. This is consistent with the well-known fact that global market indices behave and fluctuate very similarly.

The additional insight we obtain from Theorem 10.6.5 is that DPs approximate the GOP. In reality this still leaves room for some differences between the GOP and diversified portfolios, since the real market has a finite number of securities. The Diversification Theorem provides a basis for exploiting the closeness of DPs with the GOP.

### A Sequence of Black-Scholes Models

To illustrate the effect of diversification stated by the above Diversification Theorem we consider the following BS type CFM. It is obtained by choosing a constant short rate $r_t = r$ for all $t \in [0, \infty)$, as well as constant volatilities. Furthermore, for $d \in \mathcal{N}$ we choose the following volatility specification

$$\sigma^{j,k}_{(d)}(t) = \begin{cases} \frac{\sigma}{\sqrt{d}} & \text{for} \quad j = 0 \\ 0 & \text{for} \quad j \neq k \\ -\sigma & \text{for} \quad j = k > 1 \end{cases} \tag{10.6.27}$$

for $j \in \{0, 1, \ldots, d\}$ and $k \in \{1, 2, \ldots, d\}$ with $\sigma > 0$. This yields the market price of risk $\sigma^{0,k}_{(d)}(t) = \frac{\sigma}{\sqrt{d}}$ with respect to the $k$th Wiener process for $t \in [0, \infty)$ and $k \in \{1, 2, \ldots, d\}$. When setting $\pi^j_{\delta,t} = 1$ and $\pi^i_{\delta,t} = 0$ for $i \neq j$ in (10.6.12) the $j$th primary security account in the $d$th CFM is obtained, which satisfies the SDE

$$dS^j_{(d)}(t) = S^j_{(d)}(t) \left[ \left( r + \sigma^2 \left( 1 + \frac{1}{\sqrt{d}} \right) \right) dt + \frac{\sigma}{\sqrt{d}} \sum_{k=1}^{d} dW^k_t + \sigma \, dW^j_t \right] \tag{10.6.28}$$

for $t \in [0, \infty)$, $j \in \{1, 2, \ldots, d\}$.

Note that in the SDE (10.6.28) the noise term $\frac{\sigma}{\sqrt{d}} \sum_{k=1}^{d} dW^k_t$ is common to all stocks and reflects the general market risk. We shall see that it drives the noise of the GOP. The GOP can typically be interpreted as the portfolio that models the movements of the market as a whole. Thus, it models the general market risk. The remaining noise term in (10.6.28), namely $\sigma dW^j_t$, generates the specific market risk of the $j$th primary security account. We see also in (10.6.27) that the $(j, j)$th specific volatility equals $-\sigma$.

We can show for the given volatility structure with (10.6.27), where $\theta^k_t = \frac{\sigma}{\sqrt{d}}$, that the fractions of the form

$$\pi^j_{\delta_{*,d},t} = \left( \sqrt{d} \left( 1 + \sqrt{d} \right) \right)^{-1} \tag{10.6.29}$$

satisfy the equation (10.2.5). From (10.1.13) we then infer that

$$\pi^0_{\delta_{*,d},t} = \left( 1 + \sqrt{d} \right)^{-1} \tag{10.6.30}$$

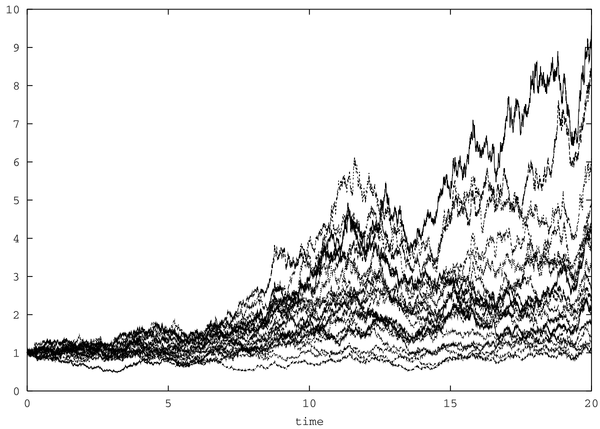



**Fig. 10.6.1.** Simulated primary security accounts

for $t \in [0, \infty)$. Recall that the GOP satisfies the SDE (10.6.11) with market price of risk $\sigma^{0,k}_{(d)}(t) = \theta^k_t = \frac{\sigma}{\sqrt{d}}$ given in (10.6.27). That is, we obtain

$$dS^{\delta_*}_{(d)}(t) = S^{\delta_*}_{(d)}(t)\Big(\big(r + \sigma^2\big)\, dt + \sigma\, dW_t\Big), \tag{10.6.31}$$

where

$$dW_t = \frac{1}{\sqrt{d}} \sum_{k=1}^{d} dW^k_t$$

for $t \in [0, \infty)$.

To visualize some properties of the given sequence of BS models we simulate the risky primary security accounts of such a market for $d = 50$ over a period of $T = 20$ years. Here we choose the volatility parameter $\sigma = 0.15$, the initial values $S^j_{(d)}(0) = 1$ for $j \in \{1, 2, \ldots, 50\}$ and use the constant short rate $r = 0.05$. Figure 10.6.1 displays the simulated sample paths of the first twenty risky primary security accounts by using the explicit solution of the BS model. The corresponding simulated GOP is shown in Fig. 10.6.3. As previously discussed, its fluctuations characterize the general market risk. It follows from (10.6.19) and (10.6.27) that the $k$th total specific volatility is bounded, so that the sequence of CFMs is regular, see Exercise 10.3.

The corresponding benchmarked primary security accounts, see (10.6.17), are shown in Fig. 10.6.2. By Definition 10.6.3 it follows that the given sequence of CFMs is regular and Theorem 10.6.5 can be applied. This means that for the given regular sequence of BS models a corresponding sequence of DPs is a sequence of approximate GOPs.

### Examples of Sequences of Approximate GOPs

Below we shall give examples for sequences of DPs which qualify as sequences of approximate GOPs. First, we consider for $d \in \mathcal{N}$ the *equi-value weighted*

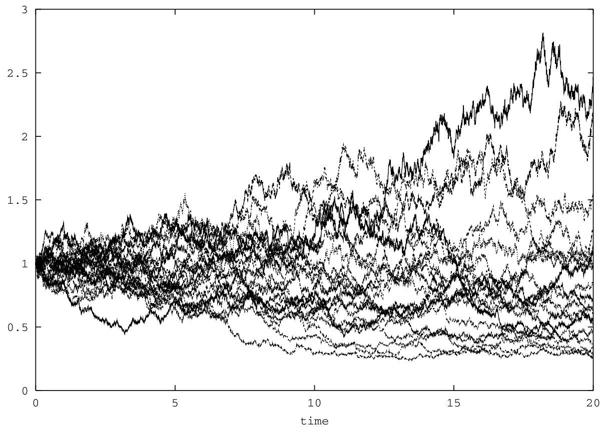



**Fig. 10.6.2.** Simulated benchmarked primary security accounts

*index* (EWI) $S^{\delta_d}_{(d)}$. It is the portfolio with value $S^{\delta_d}_{(d)}(t)$ at time $t$ that keeps equal fractions of its value in each of the primary security accounts at all times, that is,

$$\pi^j_{\underline{\delta}_d}(t) = \frac{1}{d+1} \tag{10.6.32}$$

for all $t \in [0,\infty)$ and $j \in \{0,1,\ldots,d\}$. This means that the holdings of the EWI are continuously reallocated, such that always an equal fraction of wealth is invested in each primary security account. This simple, theoretically highly important index is also called the *equal value index* or *value line index*. Its return is the arithmetic average of the returns of the underlying securities.

Let us consider the EWI for the above sequence of BS models. In this case the tracking rate of the EWI can be computed directly. According to (10.6.22) and (10.6.28)–(10.6.31), the tracking rate satisfies the inequality

$$\begin{aligned} R^{\delta}_{(d)}(t) &= \sum_{k=1}^{d} \left( \frac{1}{d+1} \, \left( \sigma^{0,k}_{(d)}(t) + \sigma^{k,k}_{(d)}(t) \right) \right)^2 \\ &= d \left( \frac{1}{d+1} \left( \frac{\sigma}{\sqrt{d}} - \sigma \right) \right)^2 \leq \frac{\sigma^2}{(d+1)} \end{aligned} \tag{10.6.33}$$

for each $d \in \mathcal{N}$. Obviously, we have a vanishing tracking rate. Therefore, according to Definition 10.6.4, the sequence of EWIs forms a sequence of approximate GOPs.

For the case $d = 50$, Fig. 10.6.3 displays the corresponding simulated GOP and the EWI. One notes that both portfolios behave in a very similar manner, as suggested by the Diversification Theorem. This shows that the convergence can already be observed for relatively small values of $d$. The differences between the portfolios are, in fact, difficult to detect by visual inspection if one chooses $d$ significantly larger than 50.

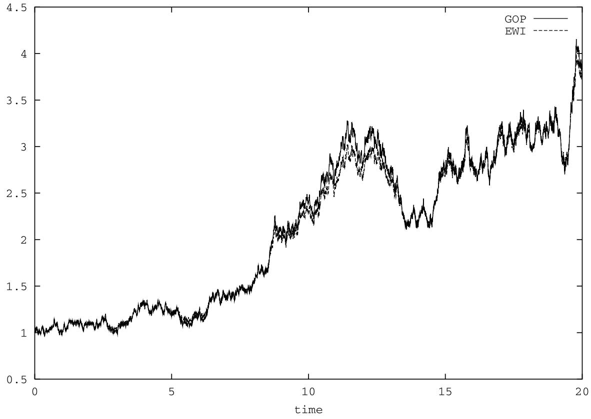



**Fig. 10.6.3.** Simulated GOP and EWI for $d = 50$

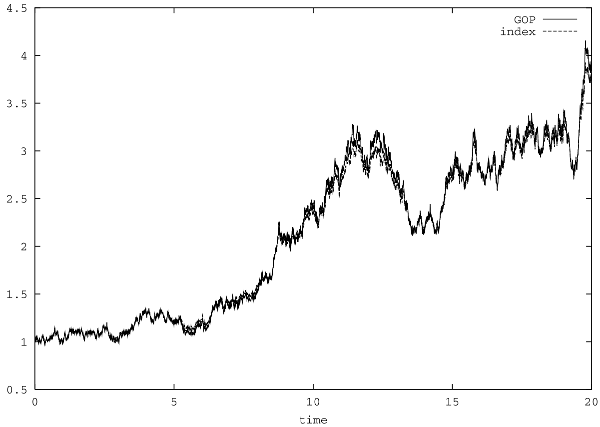



**Fig. 10.6.4.** Simulated accumulation index and GOP

Another possible candidate for an approximate GOP is an *accumulation index*. To establish the property that a sequence of diversified accumulation indices forms a sequence of approximate GOPs, one needs to prove that the underlying portfolios have a vanishing tracking rate and that the given sequence of CFMs is regular. An illustrative example of an accumulation index is given in Fig. 10.6.4 for the above BS model with $d = 50$. It uses the same trajectories of primary security accounts that were previously simulated and holds one unit of each primary security account for the entire period. In this case, the fractions of the portfolio value invested in each security are continuously changing. Figure 10.6.4 illustrates for our example that also the accumulation index provides a good approximation of the GOP.

### Approximate GOP for the World Stock Market

For regular sequences of CFMs it has been demonstrated that sequences of diversified portfolios approximate the sequence of GOPs. We underline that sequences of approximate GOPs do not depend on particular model assumptions. From this point of view, even if one may never find a perfectly accurate model for the world stock market dynamics, a diversified world stock index, when interpretable as a DP, represents a natural candidate for a proxy of the GOP.

By assuming that the fluctuations of most tradable wealth are reflected in stocks, this provides us with an argument why a diversified stock market index or a diversified market portfolio (MP) of stocks are potentially good proxies for the GOP. This argument does not depend on any particular model assumptions or the requirement that all investors optimize their portfolios in a certain manner. The diversified nature of the MP already makes it a reasonable proxy for the GOP in a regular sequence of CFMs.

By the power of diversification, all DPs are asymptotically similar in a regular sequence of markets. The above Diversification Theorem emphasizes the fact that for a sequence of CFMs corresponding sequences of DPs are rather robust and yield portfolios that are close to each other. However, if a global portfolio contains also a few stocks with relatively large market capitalization, then the above Diversification Theorem suggests for a regular sequence of markets that the GOP is likely to have a smaller fraction in these stocks. This points into the direction of interesting empirical and theoretical results in Fernholz (2002), where it is suggested that one reduces the fractions of the large stocks in an MP and enlarges those of small stocks. This yields a better diversified stock portfolio, which in Fernholz (2002) was shown to outperform the MP.

Accumulation indices, also called, total return indices represent self-financing portfolios with variations in their weights due to the method of construction. The most common market capitalization weighted index is a market capitalization weighted world stock index. Here regional market capitalization weighted stock indices are weighted according to their corresponding market capitalization. This construction leads to what we shall call a *world stock index* (WSI). For the graphs to be shown the market capitalization weights are adjusted annually, with weights and daily data provided by Thompson Financial. The resulting WSI for the period from 1974 until 2004 is shown in Fig. 10.6.5 together with the corresponding EWI.

The disadvantage of this way of construction is that certain regional markets can obtain in the WSI too much weight. The WSI may then no longer be interpretable as a DP. This is likely to happen when there is a country wide asset bubble as, for instance, was the case for Japanese stocks between 1984 and 1990. The upper graph in Fig. 10.6.5 was during this period the WSI. A similar effect was observed around 2000 where again the WSI was above the EWI. The long term growth rate of the EWI appears to be still as large as that of the WSI. It was in 2004 above the WSI. The Diversification Theorem

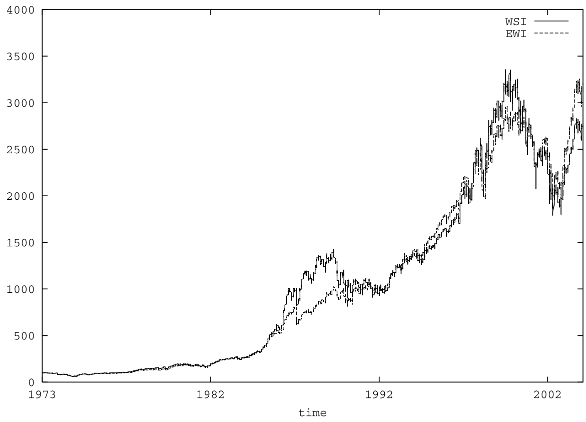



**Fig. 10.6.5.** WSI and EWI

provides support for this observation. It suggests to select a DP as a robust choice for a good proxy of the GOP in favor of an index that allows very large fractions. The EWI is a DP since it has equal fractions. Forthcoming work will present constructions of DPs that can serve as proxies for the GOP, see Le & Platen (2006). For many tasks with short time horizon it is not relevant which particular approximation of the GOP one selects. The robustness and model independence provided by the Diversification Theorem is sufficiently powerful in such cases.

## 10.7 Exercises for Chapter 10

**10.1.** In a CFM derive the form of the growth rate $g_t^\delta$ of a strictly positive portfolio $S^\delta$ satisfying the SDE

$$dS_t^\delta = S_t^\delta \left( r_t\,dt + \sum_{k=1}^{d} \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\,(\theta_t^k\,dt + dW_t^k) \right).$$

**10.2.** For a nonnegative portfolio $S_t^\delta$ with SDE as in Exercise 10.1 and the GOP $S_t^{\delta_*}$ satisfying the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*} \left( r_t\,dt + \sum_{k=1}^{d} \theta_t^k\,(\theta_t^k\,dt + dW_t^k) \right),$$

show for the benchmarked portfolio value $\hat{S}_t^\delta = \frac{S_t^\delta}{S_t^{\delta_*}}$ its SDE. If $\hat{S}^\delta$ is square integrable does this SDE, in general, imply that $\hat{S}^\delta$ is a submartingale, martingale or supermartingale?

**10.3.** Show that the sequence of Black-Scholes models in Sect. 10.6 is a regular sequence of CFMs.

# 11

# Portfolio Optimization

This chapter derives and extends a range of classical results from portfolio optimization and derivative pricing in incomplete markets in the context of a CFM. First, we consider the question of how wealth should be optimally transferred into the future given the preferences of an investor. This is a central question in economics and finance and leads into the area of portfolio optimization. We shall advocate the GOP as the best long term investment. This is consistent with views formulated in Latané (1959), Breiman (1961), Hakansson (1971) and Thorp (1972).

For the case when the investment horizon is short it was pointed out in Samuelson (1963, 1969, 1979) that one should not use the GOP as the only investment. We shall show that the optimal portfolio of an investor, who maximizes an expected utility from discounted terminal wealth, can be separated into two funds, the savings account and the GOP. This generalizes earlier results in Tobin (1958b) and Sharpe (1964) to the continuous market case. Such an optimal portfolio, which invests only into the GOP and the savings account, turns out to be an efficient portfolio in a mean-variance sense, see Markowitz (1959). It has always the maximum Sharpe ratio in the sense of Sharpe (1964, 1966).

Furthermore, we generalize the intertemporal capital asset pricing model (ICAPM) derived in Merton (1973a) under very weak assumptions. Under the assumption that the fundamental relationships in the market are invariant under changes of currency denomination, it is demonstrated that the GOP matches the market portfolio.

Real world pricing emerges as the natural pricing concept when deriving for a nonreplicable payoff its utility indifference price. The resulting benchmarked prices are martingales, independent of the underlying utility of the investor. This provides a fundamental relationship between portfolio optimization and derivative pricing. The GOP is selected as numeraire and the real world probability measure is the pricing measure. The existence of an equivalent risk neutral probability measure is not required.

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

## 11.1 Locally Optimal Portfolios

Within this section we aim to derive and generalize under weak assumptions classical results on portfolio selection, these include Sharpe ratio maximization, two fund separation, the Markowitz efficient frontier and the ICAPM.

### Discounted Portfolios

Suppose investors select portfolios for the investment of their total tradable wealth, which perform better than other portfolios in a sense specified below. We aim to clarify when the MP approximates the GOP if all investors perform some form of portfolio optimization.

We assume that an investor adjusts for the time value of money by considering *discounted portfolios*, where the savings account is used for discounting. She or he can always invest in the locally riskless asset, which is the savings account, without facing short term fluctuations. When accepting short term fluctuations an investor expects a "better" portfolio performance than is provided by the savings account. Below we specify what we mean by "better" performance. Given a strictly positive portfolio $S^\delta \in \mathcal{V}^+$, its *discounted value*

$$\bar{S}^\delta_t = \frac{S^\delta_t}{S^0_t} \tag{11.1.1}$$

satisfies by (10.1.1), (10.1.14) and an application of the Itô formula the SDE

$$d\bar{S}^\delta_t = \sum_{k=1}^{d} \psi^k_{\delta,t} \left(\theta^k_t \, dt + dW^k_t\right) \tag{11.1.2}$$

with *kth diffusion coefficient*

$$\psi^k_{\delta,t} = \sum_{j=1}^{d} \delta^j_t \, \bar{S}^j_t \, b^{j,k}_t \tag{11.1.3}$$

for $k \in \{1, 2, \ldots, d\}$ and $t \in [0, \infty)$. Note that $\psi^k_{\delta,t}$ makes sense also in the case when $\bar{S}^\delta_t$ equals zero.

Obviously, by (11.1.2) and (11.1.3) the discounted portfolio process $\bar{S}^\delta$ has *discounted drift*

$$\alpha^\delta_t = \sum_{k=1}^{d} \psi^k_{\delta,t} \, \theta^k_t \tag{11.1.4}$$

at time $t \in [0, \infty)$, which measures its trend at time $t$. One can say that the discounted drift models the increase per unit of time of the *underlying value* of $\bar{S}^\delta$ at time $t$. This can be interpreted as the fundamental economic value of the portfolio, which would be visible if one were able to remove the speculative fluctuations

$$\bar{M}_t = \sum_{k=1}^{d} \int_0^t \psi_{\delta,s}^k \, dW_s^k$$

from the discounted portfolio value

$$\bar{S}_t^\delta = \bar{S}_0^\delta + \int_0^t \alpha_s^\delta \, ds + \bar{M}_t.$$

From an economic point of view $\alpha^\delta = \{\alpha_t^\delta, \, t \in [0, \infty)\}$ is a highly relevant parameter process, since it describes the average discounted wealth that underpins the market. It provides a natural link to the macro economy. We shall use the underlying value in Chap. 13 to derive a parsimonious market model.

The magnitude of the trading uncertainty of a discounted portfolio $\bar{S}^\delta$ at time $t \in [0, \infty)$ can be measured by its *aggregate diffusion coefficient*

$$\gamma_t^\delta = \sqrt{\sum_{k=1}^{d} \left(\psi_{\delta,t}^k\right)^2} \tag{11.1.5}$$

or equivalently by its *aggregate volatility*

$$b_t^\delta = \frac{\gamma_t^\delta}{\bar{S}_t^\delta} \tag{11.1.6}$$

for $\bar{S}_t^\delta > 0$. The square $(\gamma_t^\delta)^2$ of the aggregate diffusion coefficient measures the variance per unit of time of the fluctuating increments of $\bar{S}^\delta$.

## Locally Optimal Portfolios

Let us identify the typical SDE of a family of portfolios that capture the objective of investors who locally in time on average prefer a larger discounted wealth increase for the same risk level. This means that these investors prefer a higher mean for the same variance. To characterize such a portfolio, which performs "better" than others in the above sense, we introduce the following definition, similar to those in Platen (2002, 2004a) and Christensen & Platen (2007).

**Definition 11.1.1.** *In a CFM $\mathcal{S}_{(d)}^C$ we call a strictly positive portfolio $S^{\tilde{\delta}} \in \mathcal{V}^+$* locally optimal*, if for all $t \in [0, \infty)$ and all strictly positive portfolios $S^\delta \in \mathcal{V}^+$ with given aggregate diffusion coefficient value*

$$\gamma_t^\delta = \gamma_t^{\tilde{\delta}} \tag{11.1.7}$$

*it has the largest discounted drift, that is,*

$$\alpha_t^\delta \leq \alpha_t^{\tilde{\delta}} \tag{11.1.8}$$

*almost surely.*

This type of local optimality can be interpreted as a continuous time generalization of *mean-variance optimality* in the sense of Markowitz (1952, 1959). Indeed, we shall see later that a locally optimal portfolio can be shown to be an *efficient portfolio* in a generalized Markowitz mean-variance sense. A discounted, locally optimal portfolio exhibits at all times the largest trend in comparison with all other discounted strictly positive portfolios with the same aggregate diffusion coefficient and, thus, with the same risk level.

### Sharpe Ratio

An important investment characteristic is the *Sharpe ratio* $s^\delta_t$, see Sharpe (1964, 1966). It is defined for any strictly positive portfolio $S^\delta \in \mathcal{V}^+$ with positive aggregate volatility $b^\delta_t > 0$ at time $t$ as the ratio of the *risk premium*

$$p_{S^\delta}(t) = \frac{\alpha^\delta_t}{\bar{S}^\delta_t} \tag{11.1.9}$$

over its aggregate volatility $b^\delta_t$, see (11.1.6), that is,

$$s^\delta_t = \frac{p_{S^\delta}(t)}{b^\delta_t} = \frac{\alpha^\delta_t}{\gamma^\delta_t} \tag{11.1.10}$$

for $t \in [0, \infty)$, see (11.1.4)–(11.1.6). We observe that the Sharpe ratio equals the ratio of the discounted drift over the aggregate diffusion coefficient. Under the mean-variance approach of Markowitz, investors aim to maximize the Sharpe ratio, which in a CFM corresponds by Definition 11.1.1 to the selection of a locally optimal portfolio. Below we shall analyze Sharpe ratios of locally optimal portfolios. We show that these are greater or equal to the Sharpe ratios of other portfolios.

### Portfolio Selection Theorem

In preparation for the Portfolio Selection Theorem, which we present below, let us introduce the *total market price of risk*

$$|\boldsymbol{\theta}_t| = \sqrt{\sum_{k=1}^{d} \left(\theta^k_t\right)^2} \tag{11.1.11}$$

at time $t \in [0, \infty)$, which is by (10.2.8) and (11.1.6) the aggregate volatility of the GOP. If the total market price of risk is zero, then all discounted drifts are zero and all strictly positive portfolios are, by Definition 11.1.1, locally optimal. To avoid such unrealistic dynamics we introduce the following assumption.

**Assumption 11.1.2.** *Assume in a CFM $\mathcal{S}^C_{(d)}$ for all $t \in [0, \infty)$ that the total market price of risk is strictly greater than zero and finite almost surely, with*

$$0 < |\boldsymbol{\theta}_t| < \infty, \tag{11.1.12}$$

*and the fraction of the GOP wealth that is invested in the savings account does not equal one, that is,*

$$\pi^0_{\delta_*,t} \neq 1 \tag{11.1.13}$$

*almost surely.*

We now formulate a *Portfolio Selection Theorem*, see Platen (2002), which generalizes some classical results, for instance, given in Markowitz (1959), Sharpe (1964), Merton (1973a) and Khanna & Kulldorff (1999), to the case of a CFM.

**Theorem 11.1.3.** (Portfolio Selection Theorem) *Consider a CFM $\mathcal{S}^C_{(d)}$ satisfying Assumption 11.1.2. For any strictly positive portfolio $S^\delta \in \mathcal{V}^+$ with nonzero aggregate diffusion coefficient and aggregate volatility $b^\delta_t$, see (11.1.6), its Sharpe ratio $s^\delta_t$ satisfies the inequality*

$$s^\delta_t \leq |\boldsymbol{\theta}_t| \tag{11.1.14}$$

*for all $t \in [0, \infty)$, where equality arises when $S^\delta$ is locally optimal. Furthermore, the value $\bar{S}^\delta_t$ at time $t$ of a discounted, locally optimal portfolio satisfies the SDE*

$$d\bar{S}^\delta_t = \bar{S}^\delta_t \frac{b^\delta_t}{|\boldsymbol{\theta}_t|} \sum_{k=1}^{d} \theta^k_t \left(\theta^k_t \, dt + dW^k_t\right), \tag{11.1.15}$$

*with fractions*

$$\pi^j_{\delta,t} = \frac{b^\delta_t}{|\boldsymbol{\theta}_t|} \pi^j_{\delta_*,t} \tag{11.1.16}$$

*for all $j \in \{1, 2, \ldots, d\}$ and $t \in [0, \infty)$. Each discounted portfolio that satisfies an SDE of the type (11.1.15) is a locally optimal portfolio.*

The proof of this theorem is given at the end of this section and can be found in Platen (2002). It exploits the fact that at any time $t$ the fractions of the family of discounted, locally optimal portfolios $\bar{S}^\delta$ can be parameterized by the aggregate volatility $b^\delta_t$. Obviously, for $b^\delta_t = 0$ one obtains the savings account as locally optimal portfolio, whereas in the case $b^\delta_t = |\boldsymbol{\theta}_t|$ it is the GOP that arises.

Note that we would have obtained equivalent results if we searched for the family of portfolios that minimizes the aggregate diffusion coefficient for given discounted drift. Similarly, we could have minimized the aggregate volatility for given risk premium. Furthermore, we shall show in Sect. 11.3 that also expected utility maximization leads to locally optimal portfolios. This robustness of portfolio optimization in a CFM is very satisfying, because it demonstrates the equivalence of several seemingly different objectives.

### Two Fund Separation and Fractional Kelly Strategies

By analyzing the structure of the fractions of a locally optimal portfolio, as given in (11.1.16), and applying (10.1.13) we obtain for the fraction of wealth held in the GOP the expression

$$\frac{b_t^\delta}{|\boldsymbol{\theta}_t|} = \frac{1-\pi^0_{\delta,t}}{1-\pi^0_{\delta_*,t}} \tag{11.1.17}$$

for $t \in [0,\infty)$. This leads directly to the following result.

**Corollary 11.1.4.** *Under the assumptions of Theorem 11.1.3, any locally optimal portfolio $S^\delta \in \mathcal{V}^+$ can be decomposed at time $t$ into a fraction of wealth $\frac{b_t^\delta}{|\boldsymbol{\theta}_t|}$ that is invested in the GOP and a remaining fraction that is held in the savings account. In particular, one has*

$$\pi^0_{\delta,t} = 1 - \frac{b_t^\delta}{|\boldsymbol{\theta}_t|}\left(1-\pi^0_{\delta_*,t}\right) \tag{11.1.18}$$

*for all $t \in [0,\infty)$.*

Theorem 11.1.3 can be interpreted as a *Two Fund Separation Theorem*, since only the two funds; the GOP and the savings account, are involved when forming locally optimal portfolios. Such an investment strategy is also known as a *fractional Kelly strategy*, see Kelly (1956), Latané (1959), Thorp (1972) and Hakansson & Ziemba (1995). When all wealth is invested in the GOP, then this corresponds to the *Kelly strategy*. Results on two fund separation go back to Tobin (1958b), Breiman (1960), Sharpe (1964), Merton (1973a), Khanna & Kulldorff (1999) and Nielsen & Vassalou (2004). An investor, who forms with her or his total tradable wealth a locally optimal portfolio, has according to Corollary 11.1.4 to choose the volatility $b_t^\delta$ of the portfolio and then invests the fraction of wealth $\frac{b_t^\delta}{|\boldsymbol{\theta}_t|}$ at time $t$ in the GOP. The remainder of her or his wealth is held in the savings account. We emphasize that only these two funds are needed to form locally optimal portfolios. We shall see in the next section that two fund separation also arises if an investor aims to maximize expected utility from discounted terminal wealth.

### Risk Aversion Coefficient

We can interpret

$$J_t^\delta = \frac{1-\pi^0_{\delta_*,t}}{1-\pi^0_{\delta,t}} = \frac{|\boldsymbol{\theta}_t|}{b_t^\delta} \tag{11.1.19}$$

as a *risk aversion coefficient* similar as in the sense of Pratt (1964) and Arrow (1965). The risk aversion coefficient for obtaining the GOP equals one, and when investing only in the savings account it equals infinity. The latter corresponds to being infinitely risk averse. According to (11.1.15), (11.1.19) and (11.1.17) a discounted locally optimal portfolio $\bar{S}_t^\delta$ satisfies then the SDE

$$d\bar{S}_t^\delta = \bar{S}_t^\delta \, \frac{1}{J_t^\delta} \, |\boldsymbol{\theta}_t| \, (|\boldsymbol{\theta}_t| \, dt + dW_t), \tag{11.1.20}$$

where

$$dW_t = \sum_{k=1}^{d} \frac{\theta_t^k}{|\boldsymbol{\theta}_t|} \, dW_t^k \tag{11.1.21}$$

for $t \in [0, \infty)$. From the SDEs (11.1.20) and (10.2.8) it follows that the fraction of wealth invested in the GOP is $\frac{1}{J_t^\delta}$. This fraction is, therefore, the fraction that characterizes at time $t$ a fractional Kelly strategy.

### Capital Market Line

Note that the *expected rate of return* or *appreciation rate* $a_t^\delta$ of a portfolio $S^\delta$ is at time $t$ the sum of short rate and risk premium and, thus, given by the expression

$$a_t^\delta = r_t + p_{S^\delta}(t) \tag{11.1.22}$$

for $t \in [0, \infty)$.

One can visualize the relationship (11.1.22) by using (11.1.14) and (11.1.10) for the family of locally optimal portfolios by the *capital market line*, see Sharpe (1964). This line shows the expected return $a_t^\delta$, given in (11.1.22), of a locally optimal portfolio $S^\delta$ in dependence on its aggregate volatility, see (11.1.6). That is, by (11.1.10) and (11.1.14) we obtain the fundamental linear relationship

$$a_t^\delta = r_t + |\boldsymbol{\theta}_t| \, b_t^\delta \tag{11.1.23}$$

for $t \in [0, \infty)$. Consequently, the slope of the capital market line equals the total market price of risk, which is, in general, a fluctuating stochastic process. The expected return for zero aggregate volatility is according to (11.1.23) the short rate. It follows from (11.1.19) that a portfolio process $S^\delta$ at the capital market line with volatility $b_t^\delta$ has at time $t$ the fraction $\frac{1}{J_t^\delta} = \frac{b_t^\delta}{|\boldsymbol{\theta}_t|}$ invested in the GOP, which characterizes its fractional Kelly strategy.

### Markowitz Efficient Frontier

For a locally optimal portfolio process $S^\delta$ it follows from the SDE (11.1.15) and (11.1.17) that at a given time $t$ its aggregate volatility, see (11.1.6), equals

$$b_t^\delta = \frac{1 - \pi_{\delta,t}^0}{1 - \pi_{\delta_*,t}^0} \, |\boldsymbol{\theta}_t| \tag{11.1.24}$$

and its *risk premium* $p_{S^\delta}(t)$ is

$$p_{S^\delta}(t) = b_t^\delta \, |\boldsymbol{\theta}_t| \tag{11.1.25}$$

for $t \in [0, \infty)$.

Note that the risk premium, see (11.1.9), of a portfolio $S^\delta$ is the appreciation rate of the corresponding discounted portfolio $\bar{S}^\delta$. By analogy with the one period mean-variance approach in Markowitz (1959), one can introduce in a CFM a family of *efficient portfolios*, which is parameterized by the squared aggregate volatility. When using formula (11.1.22) for the expected rate of return this leads to the following definition:

**Definition 11.1.5.** *In a CFM satisfying Assumption 11.1.2, an* efficient portfolio $S^\delta \in \mathcal{V}^+$ *is one whose expected rate of return* $a_t^\delta$*, as a function of its squared volatility* $(b_t^\delta)^2$*, lies on the* efficient frontier $a_t^\delta$*, defined as*

$$a_t^\delta = r_t + \sqrt{(b_t^\delta)^2} \, |\boldsymbol{\theta}_t| \tag{11.1.26}$$

*for all times* $t \in [0, \infty)$.

By exploiting relations (11.1.25) and (11.1.26), the following result can be directly obtained.

**Corollary 11.1.6.** *Under the assumptions of Theorem 11.1.3 any locally optimal portfolio* $S^\delta \in \mathcal{V}^+$ *is also an efficient portfolio.*

The relationship (11.1.26) can be interpreted as a generalization of the Markowitz efficient frontier to the continuous time setting. It holds for locally optimal portfolios under rather weak assumptions. Due to the inequality (11.1.14) in the Portfolio Selection Theorem and relation (11.1.10) it is not possible to form a strictly positive portfolio that generates an expected rate of return above the efficient frontier.

Each optimal portfolio $S^\delta$ has an expected rate of return $a_t^\delta$ that is located at the efficient frontier given in (11.1.26). Note that the efficient frontier moves randomly up and down over time in dependence on the fluctuations of the short rate $r_t$. Its slope also changes over time according to the total market price of risk $|\boldsymbol{\theta}_t|$, which is, generally, stochastic. For a fixed time instant $t \in [0, \infty)$ the Fig. 11.1.1 shows the efficient frontier's dependence on the squared volatility $|b_t^\delta|^2$ of a locally optimal portfolio, where the parameter values $r_t = 0.05$ and $|\boldsymbol{\theta}_t|^2 = 0.04$ are chosen. This graph also includes the tangent of the efficient frontier with slope $\frac{1}{2}$ at the point $|b_t^\delta|^2 = |\boldsymbol{\theta}_t|^2$ that corresponds to the squared volatility of the GOP. The reason why the mean-variance approach holds generally in a CFM is that, due to the assumed continuity of asset prices the asset dynamics resembles, locally in time, that of a one period model with Gaussian log-returns.

### Efficient Growth Rates

As we have seen in Theorem 10.5.1, the focus of the long term investor should be the growth rate of her or his portfolio of tradable wealth. For illustration,

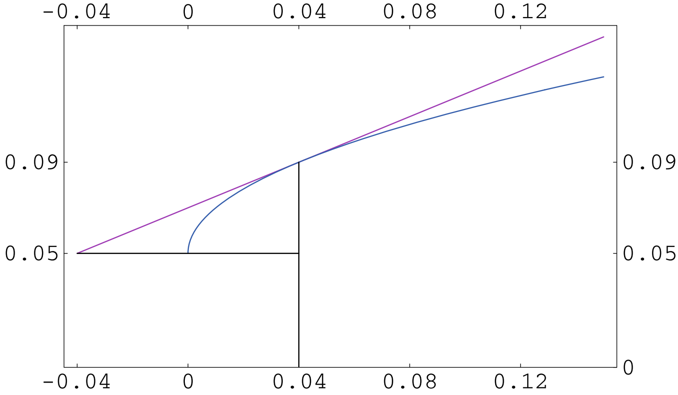



**Fig. 11.1.1.** Efficient frontier

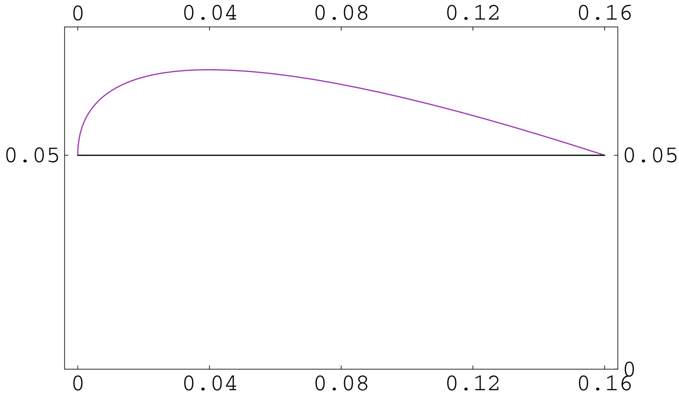



**Fig. 11.1.2.** Efficient growth rates

Fig. 11.1.2 shows for given $t \in [0, \infty)$ how the growth rate of a locally optimal portfolio $S^\delta$ depends on its squared volatility $|b_t^\delta|^2$, when using the same default parameters as in Fig. 11.1.1. One could call these growth rates the *efficient growth rates*. The corresponding frontier in dependence on the squared portfolio volatilities one can call the *efficient growth rate frontier*. The efficient growth rates satisfy the expression

$$g_t^\delta = r_t + \sqrt{|b_t^\delta|^2}\,|\boldsymbol{\theta}_t| - \frac{1}{2}\,|b_t^\delta|^2 = r_t + \frac{|\boldsymbol{\theta}_t|^2}{J_t^\delta}\left(1 - \frac{1}{2\,J_t^\delta}\right), \qquad (11.1.27)$$

see (10.2.2), (11.1.16), (10.2.6) and (11.1.11). Note that for the value of the squared volatility $|b_t^\delta|^2 = |\boldsymbol{\theta}_t|^2$, that is $J_t^\delta = 1$, the efficient growth rates achieve their maximum, yielding the growth rate of the GOP

$$g_t^{\delta_*} = r_t + \frac{1}{2}\,|\boldsymbol{\theta}_t|^2. \qquad (11.1.28)$$

For the volatility value $|b_t^\delta| = 2\,|\boldsymbol{\theta}_t|$ the efficient growth rate equals the short rate. As we have seen in Sect. 10.5, the GOP is the best performing portfolio under various criteria, in particular, for long term growth. By choosing a volatility value $|b_t^\delta| > |\boldsymbol{\theta}_t|$ one is, in principle, *overbetting*. This means that one faces larger fluctuations, which are more risky than those of the GOP due to a short position in the savings account. Such a fractional Kelly strategy does not perform as well as the Kelly strategy in the long term. Overbetting diminishes the long term growth rate. However, some investors may achieve by luck spectacular growth over some short period when overbetting but others may fail dramatically.

We have seen that the GOP is a central object in a CFM which facilitates the intertemporal generalization of the classical Markowitz-Tobin-Sharpe static mean-variance portfolio analysis, see Markowitz (1959), Tobin (1958a) and Sharpe (1964). Due to two fund separation the GOP is also a highly important benchmark for fund management. Two fund separation is equivalent to some kind of a fractional Kelly strategy. In Theorem 10.5.1 it was shown that the GOP almost surely outperforms pathwise any other portfolio after a sufficiently long time. Furthermore, Corollary 10.5.3 showed that even over any short time period it cannot be systematically outperformed by any other portfolio.

### Lagrange Multipliers and Optimization (*)

As we shall see, the proof of the Portfolio Selection Theorem uses only standard multivariate calculus and a basic understanding of stochastic calculus. Before we give the proof of Theorem 11.1.3 let us mention a standard result on Lagrange multipliers and optimization.

Let $U : \Re^n \to \Re$ and $\boldsymbol{g} : \Re^k \to \Re^k$ be differentiable functions. Furthermore, assume that $U$ is strictly concave and that $\boldsymbol{g}$ is convex. Under these assumptions we consider the problem of solving the optimization problem to find the maximum

$$U(\boldsymbol{x}_*) = \max_{\boldsymbol{x} \in \Re^n} U(\boldsymbol{x}) \tag{11.1.29}$$

such that

$$g^i(\boldsymbol{x}_*) = 0 \tag{11.1.30}$$

for all $i \in \{1, 2, \ldots, k\}$ and $\boldsymbol{x}_* \in \Re^n$. This problem is equivalent to finding a zero of the gradient of the corresponding Lagrangian

$$\mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda}) = U(\boldsymbol{x}) - \boldsymbol{\lambda}^\top \boldsymbol{g}(\boldsymbol{x}) \tag{11.1.31}$$

for $\boldsymbol{x} = (x^1, x^2, \ldots, x^n)^\top \in \Re^n$ and $\boldsymbol{\lambda} = (\lambda^1, \lambda^2, \ldots, \lambda^k)^\top \in \Re^k$, see Luenberger (1969). More precisely, if the pair $(\boldsymbol{x}_*, \boldsymbol{\lambda}_*) \in \Re^n \times \Re^n$ solves the system of first order conditions

$$0 = \frac{\partial \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial x^i} = \frac{\partial U(\boldsymbol{x})}{\partial x^i} - \sum_{\ell=1}^{k} \lambda^\ell \, \frac{\partial g^\ell(\boldsymbol{x})}{\partial x^i} \tag{11.1.32}$$

for $i \in \{1, 2, \ldots, n\}$ and

$$0 = \frac{\partial \mathcal{L}(\boldsymbol{x}, \boldsymbol{\lambda})}{\partial \lambda^i} = g^i(\boldsymbol{x}) \tag{11.1.33}$$

for $i \in \{1, 2, \ldots, k\}$, then $\boldsymbol{x}_*$ is the unique maximizer of the optimization problem.

In the case when the vector $\boldsymbol{\lambda}_*$ of the Lagrangian multipliers consists only of nonnegative components, then $\boldsymbol{x}_*$ is also the unique maximizer of the optimization problem

$$U(\boldsymbol{x}_*) = \max_{\boldsymbol{x} \in \Re^n} U(\boldsymbol{x}) \tag{11.1.34}$$

such that

$$g^i(\boldsymbol{x}_*) \leq 0 \tag{11.1.35}$$

for all $i \in \{1, 2, \ldots, k\}$.

**Proof of Theorem 11.1.3 (*)**

To prove the Portfolio Selection Theorem we follow essentially the proof given in Platen (2002). To identify a discounted, locally optimal portfolio, as described in Definition 11.1.1, we maximize locally in time the drift (11.1.4), subject to the constraint (11.1.7). For this purpose we use the Lagrange multiplier $\lambda$, as described in the above subsection, and consider the function

$$\mathcal{L}(\psi_\delta^1, \ldots, \psi_\delta^d, \lambda) = \sum_{k=1}^d \psi_\delta^k \, \theta^k + \lambda \left( \left( \gamma^{\tilde{\delta}} \right)^2 - \sum_{k=1}^d \left( \psi_\delta^k \right)^2 \right) \tag{11.1.36}$$

by suppressing time dependence. For $\psi_\delta^1, \psi_\delta^2, \ldots, \psi_\delta^d$ to provide a maximum for $\mathcal{L}(\psi_\delta^1, \ldots, \psi_\delta^d, \lambda)$ it is necessary that the first-order conditions

$$\frac{\partial \mathcal{L}(\psi_\delta^1, \ldots, \psi_\delta^d, \lambda)}{\partial \psi_\delta^k} = \theta^k - 2\,\lambda\,\psi_\delta^k = 0 \tag{11.1.37}$$

are satisfied for all $k \in \{1, 2, \ldots, d\}$ as well as

$$\frac{\partial \mathcal{L}(\psi_\delta^1, \ldots, \psi_\delta^d, \lambda)}{\partial \lambda} = \left( \gamma^{\tilde{\delta}} \right)^2 - \sum_{k=1}^d \left( \psi_\delta^k \right)^2 = 0. \tag{11.1.38}$$

Consequently, a locally optimal portfolio $S^{(\tilde{\delta})}$, which maximizes the discounted drift, must satisfy the relation

$$\psi_{\tilde{\delta}}^k = \frac{\theta^k}{2\,\lambda} \tag{11.1.39}$$

for all $k \in \{1, 2, \ldots, d\}$. Furthermore, by (11.1.38) we must have

$$\sum_{k=1}^{d} \left(\psi_{\tilde{\delta}}^{k}\right)^2 = \left(\gamma^{\tilde{\delta}}\right)^2. \tag{11.1.40}$$

We can now use the constraint (11.1.7), together with (11.1.40), (11.1.5) and (11.1.11), to obtain from (11.1.39) the relation

$$\left(\gamma^{\tilde{\delta}}\right)^2 = \sum_{k=1}^{d} \left(\psi_{\tilde{\delta}}^{k}\right)^2 = \frac{\sum_{k=1}^{d} (\theta^k)^2}{4\,\lambda^2}. \tag{11.1.41}$$

By (11.1.12) we have $|\boldsymbol{\theta}| = \sqrt{\sum_{k=1}^{d} (\theta^k)^2} > 0$ and obtain by (11.1.39) and (11.1.41) the equation

$$\psi_{\tilde{\delta}}^{k} = \frac{\gamma^{\tilde{\delta}}}{|\boldsymbol{\theta}|}\, \theta^k \tag{11.1.42}$$

for all $k \in \{1, 2, \ldots, d\}$. This yields at time $t$ by (11.1.4) for a locally optimal portfolio $S^{\tilde{\delta}}$ the discounted drift

$$\alpha_t^{\tilde{\delta}} = \gamma_t^{\tilde{\delta}}\, \frac{|\boldsymbol{\theta}_t|^2}{|\boldsymbol{\theta}_t|} = \gamma_t^{\tilde{\delta}}\, |\boldsymbol{\theta}_t|. \tag{11.1.43}$$

This leads, by (11.1.10), to the equality in (11.1.14). Due to the above optimization the inequality in (11.1.14) follows for any strictly positive portfolio with nonzero aggregate diffusion coefficient.

Equation (11.1.42), when substituted into (11.1.2), provides by (11.1.5) the SDE

$$d\bar{S}_t^{\tilde{\delta}} = \gamma_t^{\tilde{\delta}} \sum_{k=1}^{d} \frac{\theta_t^k}{|\boldsymbol{\theta}_t|}\, (\theta_t^k\, dt + dW_t^k). \tag{11.1.44}$$

Using (11.1.6) this yields the SDE (11.1.15). Furthermore, it follows for $k \in \{1, 2, \ldots, d\}$ from (11.1.3), (10.1.12), (11.1.42) and (11.1.6) that

$$\psi_{\tilde{\delta},t}^{k} = \sum_{j=1}^{d} \tilde{\delta}_t^j\, \bar{S}_t^j\, b_t^{j,k} = \bar{S}_t^{(\tilde{\delta})} \sum_{j=1}^{d} \pi_{\tilde{\delta},t}^{j}\, b_t^{j,k} = \frac{\gamma_t^{\tilde{\delta}}}{|\boldsymbol{\theta}_t|}\, \theta_t^k = \bar{S}_t^{(\tilde{\delta})}\, b_t^{\tilde{\delta}}\, \frac{\theta_t^k}{|\boldsymbol{\theta}_t|}. \tag{11.1.45}$$

Using the invertibility of the volatility matrix one obtains, see Assumption 10.1.1, the fraction

$$\pi_{\tilde{\delta},t}^{j} = \frac{b_t^{\tilde{\delta}}}{|\boldsymbol{\theta}_t|} \sum_{k=1}^{d} \theta_t^k\, b_t^{-1\,j,k} \tag{11.1.46}$$

and, thus, by (10.2.6) the equation (11.1.16) for all $j \in \{1, 2, \ldots, d\}$ and $t \in [0, \infty)$. Using the SDE of the discounted GOP one notes that an SDE of the form (11.1.15) belongs to a discounted portfolio, which has for its given aggregate diffusion coefficient the maximum discounted drift. Thus, by Definition 11.1.1 the corresponding portfolio is a locally optimal portfolio. □

## 11.2 Market Portfolio and GOP

In Sect. 9.3 we already considered a version of the intertemporal capital asset pricing model (ICAPM). The capital asset pricing model (CAPM) was developed in one and multiperiod discrete time settings by Sharpe (1964), Lintner (1965) and Mossin (1966). Its continuous time analog, the ICAPM, was established for continuous markets in Merton (1973a) as an equilibrium model of exchange using utility maximization and equilibrium arguments. Most of the following results are established in Platen (2005c, 2006a, 2006b).

### Intertemporal Capital Asset Pricing Model

By using a locally optimal portfolio as a reference portfolio, we shall now derive the ICAPM for a CFM. For this purpose let us consider a strictly positive, risky, locally optimal portfolio $S^{\underline{\delta}} \in \mathcal{V}^+$. Then by (11.1.9), (10.1.14), (10.2.1) and (11.1.15) the risk premium $p_{S^\delta}(t)$ of a strictly positive portfolio $S^\delta \in \mathcal{V}^+$ can be expressed as

$$p_{S^\delta}(t) = \sum_{k=1}^{d} \sum_{j=1}^{d} \pi^j_{\delta,t} \, b^{j,k}_t \, \theta^k_t = \frac{d[\ln(S^\delta), \ln(S^{\underline{\delta}})]_t}{dt} \, \frac{|\boldsymbol{\theta}_t|}{b^{\underline{\delta}}_t} \tag{11.2.1}$$

at time $t$. Here $[\ln(S^\delta), \ln(S^{\underline{\delta}})]_t$ denotes the covariation at time $t$ of the stochastic processes $\ln(S^\delta)$ and $\ln(S^{\underline{\delta}})$, see Sect. 5.2. The time derivative of the covariation is the local, in time, analogue of the covariance of log-returns for continuous time processes.

For a strictly positive portfolio $S^\delta \in \mathcal{V}^+$ the *systematic risk parameter* $\beta_{S^\delta}(t)$, also called the *beta*, is defined as the ratio of the covariations

$$\beta_{S^\delta}(t) = \frac{\frac{d[\ln(S^\delta), \ln(S^{\underline{\delta}})]_t}{dt}}{\frac{d[\ln(S^{\underline{\delta}})]_t}{dt}}, \tag{11.2.2}$$

for $t \in [0, \infty)$, where $S^{\underline{\delta}}$ denotes again a strictly positive, risky, locally optimal portfolio. This allows us to deduce by (11.2.1) and (11.2.2) the core relationship of the ICAPM.

**Theorem 11.2.1.** *Under the assumptions of Theorem 11.1.3, for any strictly positive portfolio $S^\delta \in \mathcal{V}^+$ the portfolio beta with respect to a strictly positive, risky, locally optimal portfolio $S^{\underline{\delta}} \in \mathcal{V}^+$, with nonzero aggregate volatility, has the form*

$$\beta_{S^\delta}(t) = \frac{p_{S^\delta}(t)}{p_{S^{\underline{\delta}}}(t)} \tag{11.2.3}$$

*for $t \in [0, \infty)$.*

The above expression for the portfolio beta is exactly what the ICAPM suggests if the market portfolio (MP) is a locally optimal portfolio. In this case, Theorem 11.2.1 already proves the ICAPM in a general CFM setting. This raises the question: When is the MP a locally optimal portfolio?

## Market Portfolio

Let us assume the existence of $n \in \mathcal{N}$ investors who hold all tradable wealth in the market, which is the total sum of all units of primary security accounts. The portfolio of tradable wealth of the $\ell$th investor is denoted by $S^{\delta_\ell}$, $\ell \in \{1, 2, \ldots, n\}$. Due to the limited liability of investors $S^{\delta_\ell} \in \mathcal{V}$ is nonnegative. The total portfolio $S_t^{\delta_{\mathrm{MP}}}$ of the tradable wealth of all investors is then the MP, which is given by the sum

$$S_t^{\delta_{\mathrm{MP}}} = \sum_{\ell=1}^{n} S_t^{\delta_\ell} \tag{11.2.4}$$

at time $t \in [0, \infty)$. We have seen in the previous section that Sharpe ratio maximizing investors form locally optimal portfolios. We shall see in Sect. 11.3 that also expected utility maximizing investors form locally optimal portfolios. Therefore, it is natural to make the following assumption.

**Assumption 11.2.2.** *Each investor forms a nonnegative, locally optimal portfolio with her or his total tradable wealth.*

Since the sum of locally optimal portfolios is again a locally optimal portfolio we can prove the following result.

**Theorem 11.2.3.** *For a CFM, where each investor holds a locally optimal portfolio with respect to the domestic currency denomination, the MP is a locally optimal portfolio.*

**Proof**: The discounted MP $\bar{S}_t^{\delta_{\mathrm{MP}}} = \frac{S_t^{\delta_{\mathrm{MP}}}}{S_t^0}$ at time $t$ is under the assumptions of the theorem by (11.1.15), (11.1.17) and (11.2.4) determined by the SDE

$$\begin{aligned} d\bar{S}_t^{\delta_{\mathrm{MP}}} &= \sum_{\ell=1}^{n} d\bar{S}_t^{\delta_\ell} = \sum_{\ell=1}^{n} \frac{\left(\bar{S}_t^{\delta_\ell} - \delta_\ell^0\right)}{\left(1 - \pi_{\delta_*,t}^0\right)} \sum_{k=1}^{d} \theta_t^k \left(\theta_t^k\, dt + dW_t^k\right) \\ &= \bar{S}_t^{\delta_{\mathrm{MP}}} \, \frac{\left(1 - \pi_{\delta_{\mathrm{MP}},t}^0\right)}{\left(1 - \pi_{\delta_*,t}^0\right)} \sum_{k=1}^{d} \theta_t^k \left(\theta_t^k\, dt + dW_t^k\right) \end{aligned} \tag{11.2.5}$$

for $t \in [0, \infty)$. This shows, by (11.1.15), that the MP $S_t^{\delta_{\mathrm{MP}}}$ has the SDE of a locally optimal portfolio. This proves by Theorem 11.1.3 that the MP is a locally optimal portfolio. □

It is straightforward to draw the following conclusion from Theorem 11.2.1 and Theorem 11.2.3.

**Corollary 11.2.4.** *Under the assumptions of Theorem 11.2.3 the ICAPM relationship* (11.2.3) *holds when using the market portfolio as reference portfolio.*

This proves the ICAPM under the assumptions of Theorem 11.2.3. It is important to emphasize the fact that the derivation of this result does not require any assumptions about expected utility maximization, equilibrium or Markovianity, as typically imposed in the literature. Note also that no matter what locally optimal portfolio the investor holds, the ICAPM follows with the MP as reference portfolio.

### Market Portfolio and GOP

It is reasonable to discuss the following invariance of a financial market model. By invariance we mean here the property of the market that relationships that hold for one currency denomination apply also for another currency denomination. This can be expressed by the following assumption.

**Assumption 11.2.5.** *The fundamental relationships in the market are invariant under a change of currency denomination.*

As the following theorem shows, this assumption has interesting consequences.

**Theorem 11.2.6.** *In a CFM where a strictly positive portfolio is locally optimal in at least two currency denominations this portfolio must be a GOP.*

This theorem will be derived at the end of this section. It allows us to draw interesting conclusions. If one assumes that the investors optimize their tradable wealth in two currency denominations by forming an MP that is a locally optimal portfolio in each of the two currencies, then by Theorem 11.2.6 the MP is the GOP. Of course, the investors will never exactly form an MP that is a perfect locally optimal portfolio in two currency denominations. However, the reality may come close to this situation. This then allows the conclusion that the MP may be not too far from the GOP.

In Sect. 10.6 we concluded under some regularity condition on the market that a portfolio approximates the GOP purely on the basis of the assumption that it is a diversified portfolio. The above optimal portfolio selection leads to a complementing result, as long as the sequence of CFMs $(\mathcal{S}^{\mathrm{C}}_{(d)})_{d\in\mathcal{N}}$ is regular and the corresponding sequence of MPs is that of diversified portfolios.

For the given world market the MP is, in principle, observable. For instance, a potential proxy is given by the daily *MSCI*, which essentially reflects the stock portfolio of the developed markets. For illustration, in Fig. 11.2.1 we show the MSCI in units of the US dollar savings account for the period from 1970 until 2003. We have alternatively studied the WSI and EWI in Sect. 10.6 as potential proxies of the GOP. We have already seen that the differences between all these proxies of the GOP are minor from a practical point of view.

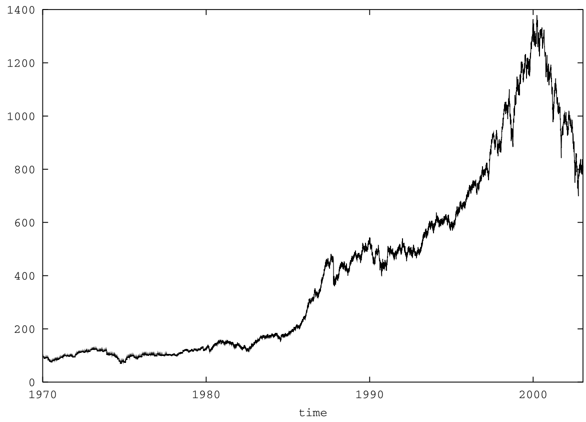



**Fig. 11.2.1.** Discounted MSCI

### Proof of Theorem 11.2.6 (*)

Let us denote, the GOP at time $t$ in the $i$th currency by $S_i^{\delta_*}(t)$ for $i \in \{0, 1\}$. This satisfies by (10.2.8) the SDE

$$dS_i^{\delta_*}(t) = S_i^{\delta_*}(t) \left( r_t^i \, dt + \sum_{k=1}^{d} \theta_i^k(t) \left( \theta_i^k(t) \, dt + dW_t^k \right) \right) \qquad (11.2.6)$$

for $t \in [0, \infty)$. Here $r_t^i$ is the $i$th short rate for the $i$th currency denomination and $\theta_i^k(t)$ the market price of risk for the $i$th currency denomination with respect to the $k$th Wiener process. Furthermore, we denote by $S_i^j(t)$ the $j$th savings account at time $t$, denominated in the $i$th currency, $i, j \in \{0, 1\}$.

A locally optimal portfolio $S_0^{\tilde{\delta}}(t)$ at time $t$, when denominated in units of the 0th currency, satisfies by Theorem 11.1.3, see (11.1.15), the SDE

$$dS_0^{\tilde{\delta}}(t) = S_0^{\tilde{\delta}}(t) \left( r_t^0 \, dt + \frac{\left(1 - \pi^0_{\tilde{\delta},t}\right)}{\left(1 - \pi^0_{\delta_*,t}\right)} \sum_{k=1}^{d} \theta_0^k(t) \left( \theta_0^k(t) \, dt + dW_t^k \right) \right) \qquad (11.2.7)$$

for $t \in [0, \infty)$.

The exchange rate $X_t^{1,0}$ from the 0th into the first currency at time $t$ can be written as

$$X_t^{1,0} = \frac{S_1^{\delta_*}(t)}{S_0^{\delta_*}(t)}. \qquad (11.2.8)$$

It satisfies by (11.2.6) and an application of the Itô formula the SDE

$$dX_t^{1,0} = X_t^{1,0} \left( (r_t^1 - r_t^0) \, dt + \sum_{k=1}^{d} \left( \theta_1^k(t) - \theta_0^k(t) \right) \left( \theta_1^k(t) \, dt + dW_t^k \right) \right) \qquad (11.2.9)$$

for $t \in [0, \infty)$.

Denominating now the locally optimal portfolio $S^{\tilde{\delta}}$ in units of the first currency yields by the Itô formula, (11.2.6) and (11.2.9) the SDE

$$
\begin{aligned}
dS_1^{\tilde{\delta}}(t) &= d\left(S_0^{\tilde{\delta}}(t)\, X_t^{1,0}\right) \\
&= S_1^{\tilde{\delta}}(t)\Bigg(r_t^1\, dt + \sum_{k=1}^{d}\left[\frac{\left(1-\pi^0_{\tilde{\delta},t}\right)}{\left(1-\pi^0_{\delta_*,t}\right)}\left(\theta_0^k(t)\right)^2 + \left(\theta_1^k(t)-\theta_0^k(t)\right)\theta_1^k(t)\right. \\
&\qquad\left. + \frac{\left(1-\pi^0_{\tilde{\delta},t}\right)}{\left(1-\pi^0_{\delta_*,t}\right)}\theta_0^k(t)\left(\theta_1^k(t)-\theta_0^k(t)\right)\right] dt \\
&\qquad + \sum_{k=1}^{d}\left[\frac{\left(1-\pi^0_{\tilde{\delta},t}\right)}{\left(1-\pi^0_{\delta_*,t}\right)}\theta_0^k(t)+\theta_1^k(t)-\theta_0^k(t)\right] dW_t^k\Bigg) \\
&= S_1^{\tilde{\delta}}(t)\left(r_t^1\, dt + \sum_{k=1}^{d}\left(\theta_1^k(t)-\theta_0^k(t)\left(\frac{\pi^0_{\delta_*,t}-\pi^0_{\tilde{\delta},t}}{1-\pi^0_{\delta_*,t}}\right)\right)\left(\theta_1^k(t)\, dt + dW_t^k\right)\right).
\end{aligned}
$$

For $S_1^{\tilde{\delta}}(t)$ to satisfy the SDE of a locally optimal portfolio in the first currency denomination requires by (11.1.15) and (11.1.17) for $t \in [0,\infty)$ the equality

$$
\theta_1^k(t) - \theta_0^k(t)\left(\frac{\pi^0_{\delta_*,t}-\pi^0_{\tilde{\delta},t}}{1-\pi^0_{\delta_*,t}}\right) = \frac{\left(1-\pi^1_{\tilde{\delta},t}\right)}{\left(1-\pi^1_{\delta_*,t}\right)}\,\theta_1^k(t).
$$

To achieve this equality one needs to satisfy the equation

$$
\pi^0_{\delta_*,t} = \pi^0_{\tilde{\delta},t} \tag{11.2.10}
$$

for all $t \in [0,\infty)$. This demonstrates by (11.2.7) and (11.2.6) that $S^{\tilde{\delta}}$ is under the assumptions of Theorem 11.2.6 a GOP. $\square$

## 11.3 Expected Utility Maximization

Utility functions, as introduced in von Neumann & Morgenstern (1953), have been widely used in portfolio optimization and economic modeling, see Merton (1973a). We study now the type of portfolio that an expected utility maximizer forms in a CFM. We shall show under appropriate assumptions that this will again be a locally optimal portfolio. As a consequence of Corollary 11.1.4 a two fund separation theorem holds also for expected utility maximization. Therefore, when some investors maximize expected utility, others maximize

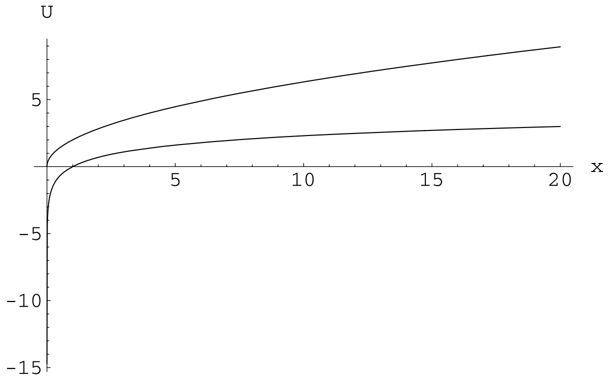



**Fig. 11.3.1.** Examples for power utility (upper graph) and log-utility (lower graph)

Sharpe ratios and the rest maximizes the growth rate for a given portfolio volatility, then the MP is still a locally optimal portfolio and, thus, a combination of the GOP and the savings account. As already mentioned, this can also be interpreted as a fractional Kelly strategy, see Hakansson & Ziemba (1995). Some of the following results appear in Platen (2006a, 2006c).

### Utility Functions

A *utility function* is a real valued function $U(\cdot)$ which allocates a real number to any nonnegative level of wealth. Once a utility function is chosen, then all alternative wealth levels are ranked by evaluating their expected utility values. It turns out that the following class of utility functions can express the personal preferences of market participants.

**Definition 11.3.1.** *A* utility function $U : [0,\infty) \to [-\infty,\infty)$ *is a real valued, twice differentiable, strictly increasing and strictly concave function, where* $U'(0) = \infty$ *and* $U'(\infty) = 0$.

Examples of utility functions are given by the *power utility*

$$U(x) = \frac{1}{\gamma}\, x^{\gamma} \tag{11.3.1}$$

for $\gamma \neq 0$ and $\gamma < 1$ and the *log-utility*

$$U(x) = \ln(x) \tag{11.3.2}$$

for $x \in [0,\infty)$, where $\ln(0)$ is set to minus infinity. In Fig. 11.3.1 we show with the upper graph an example for a power utility function with $\gamma = \frac{1}{2}$, together with the log-utility displayed as the lower curve. The properties of a utility function given in Definition 11.3.1 have economic interpretations.

The strict monotonicity reflects the natural preference of an investor for more rather than less wealth. In this sense investors are *nonsatiable*. The concavity of $U(x)$ implies that $U'(x)$ is decreasing in $x$. This models the fact that a typical investor has some *risk aversion*, which may depend on her or his level of total tradable wealth. Note that the derivative $U'$ of a utility function has an inverse function $U'^{-1}$, which will be of importance in our analysis below.

### Expected Utility Maximization

We aim to identify the portfolio which an expected utility maximizer constructs. Let us consider a utility function $U : [0, \infty) \to [-\infty, \infty)$ and fix a terminal time horizon $T \in [0, \infty)$.

An investor can always compare her or his investment strategy $\delta$ with the one where all wealth is invested in the locally riskless security, that is, the savings account $S^0$. Therefore, we shall take the time value of money into account by discounting with the savings account $S^0$. This means, we shall consider an investor who maximizes expected utility from discounted terminal wealth. Furthermore, we assume that the investor maximizes only over fair portfolios, since according to Corollary 10.4.2, these are the portfolios that require the minimal initial investment to reach a desired future payoff. This payoff is in our case the utility of discounted terminal wealth. It is not rational to invest in an unfair portfolio, because there exists then a cheaper fair portfolio that provides exactly the same utility.

**Definition 11.3.2.** *Define the set $\bar{\mathcal{V}}^+_{S_0}$ of strictly positive, savings account discounted, fair portfolios $\bar{S}^\delta$ with given initial value $\bar{S}^\delta_0 = S_0 > 0$.*

We maximize now the expected utility

$$v^{\tilde{\delta}} = \max_{\bar{S}^\delta \in \bar{\mathcal{V}}^+_{S_0}} v^\delta \tag{11.3.3}$$

with

$$v^\delta = E\left(U\left(\bar{S}^\delta_T\right) \,\middle|\, \mathcal{A}_0\right), \tag{11.3.4}$$

where the maximum is taken over the set $\bar{\mathcal{V}}^+_{S_0}$ and is assumed to exist.

Furthermore, to obtain a tractable solution of the expected utility maximization problem, we assume in this section, for simplicity, that the discounted GOP $\bar{S}^{\delta_*}$ itself is a strictly positive Markov process with

$$d\bar{S}^{\delta_*}_t = \bar{S}^{\delta_*}_t \, \theta(t, \bar{S}^{\delta_*}_t) \left(\theta(t, \bar{S}^{\delta_*}_t)\, dt + dW_t\right) \tag{11.3.5}$$

for $t \in [0, \infty)$ and given volatility function $\theta : [0, \infty) \times (0, \infty) \to (0, \infty)$. In Chap. 13 we shall demonstrate by deriving the minimal market model that this is an acceptable assumption. This assumption can be relaxed in many ways yielding slightly more complex but similar results.

The following theorem describes the structure of the optimal portfolio of the expected utility maximizer. Its derivation follows the, so-called, martingale approach in portfolio optimization as described, for instance, in Korn (1997), Karatzas & Shreve (1998) and Zhao & Ziemba (2003). The theorem is derived at the end of the section, see also Platen (2006c).

**Theorem 11.3.3.** *Consider a CFM that satisfies the Assumption 11.1.2 and has a Markovian, strictly positive discounted GOP* $\bar{S}^{\delta_*}$*, satisfying* (11.3.5)*. Then the discounted, strictly positive, fair portfolio* $\bar{S}^{\tilde{\delta}} \in \bar{\mathcal{V}}^+_{S_0}$*, which maximizes the given utility function* $U(\cdot)$*, is a locally optimal portfolio in the sense of Definition 11.1.1 and satisfies the SDE*

$$d\bar{S}^{\tilde{\delta}}_t = \bar{S}^{\tilde{\delta}}_t \, \frac{1}{J^{\tilde{\delta}}_t} \, \theta(t, \bar{S}^{\delta_*}_t) \, \Big( \theta(t, \bar{S}^{\delta_*}_t) \, dt + dW_t \Big) \tag{11.3.6}$$

*with risk aversion coefficient*

$$J^{\tilde{\delta}}_t = \frac{1}{1 - \frac{\hat{S}^0_t}{\hat{u}(t, \hat{S}^0_t)} \, \frac{\partial \hat{u}(t, \hat{S}^0_t)}{\partial \hat{S}^0}}, \tag{11.3.7}$$

*and benchmarked fair portfolio value*

$$\hat{S}^{\tilde{\delta}}_t = \hat{u}(t, \hat{S}^0_t) = E\left( U'^{-1}\left( \lambda \, \hat{S}^0_T \right) \hat{S}^0_T \,\Big|\, \mathcal{A}_t \right) \tag{11.3.8}$$

*at time* $t \in [0, T]$*. The constant* $\lambda$ *follows from the match of the initial value*

$$S_0 = \hat{u}(0, \hat{S}^0_0) \, S^{\delta_*}_0. \tag{11.3.9}$$

Note in (11.3.7) that the risk aversion coefficient is the inverse of the fraction of $S^{\tilde{\delta}}$ that is invested in the GOP. We notice from (11.3.8) that the problem of maximizing expected utility from discounted terminal wealth has been transformed into that of hedging a particular payoff of the type

$$H = U'^{-1}(\lambda \hat{S}^0_T) S^0_T.$$

This demonstrates that there is a deep link between expected utility maximization and hedging. We shall discuss hedging issues in more detail in the next section. Due to the Markovianity of $\bar{S}^{\delta_*}$ one can in the given case calculate $\hat{u}(\cdot, \cdot)$ and replicate the payoff $H$ by a fair, locally optimal portfolio. More precisely, one can apply the Feynman-Kac formula, see Sect. 9.7, to obtain the function $\hat{u}(\cdot, \cdot)$ as the solution of a PDE. From $\hat{u}(\cdot, \cdot)$ one can then determine the fraction of wealth to be held in the GOP and the remaining fraction that has to be invested in the savings account. Note that if $\bar{S}^{\delta_*}$ is driven by $n \in \mathcal{N}$ tradable factors that form together a Markov process, then one obtains $n + 1$ fund separation for the resulting optimal portfolios. However, as we will see in Chap. 13 the MMM suggests in reality two fund separation.

**Examples on Expected Utility Maximization**

A disadvantage of the expected utility approach is that only in rare cases one can provide explicit results. To illustrate the above theorem we discuss two simple examples.

**1.** In the first example we consider the log-utility function $U(x) = \ln(x)$. Its derivative is $U'(x) = \frac{1}{x}$, which has the inverse $U'^{-1}(y) = \frac{1}{y}$. Since the second derivative $U''(x) = -\frac{1}{x^2}$ is negative, the utility function is concave, as required in Definition 11.3.1. We recall that maximizing expected logarithmic utility is equivalent to selecting the Kelly criterion for portfolio optimization, see Kelly (1956) and Hakansson & Ziemba (1995).

According to (11.3.8) we obtain for $t \in [0, \infty)$ the conditional expectation

$$\hat{u}(t, \hat{S}^0_t) = E\left(U'^{-1}\left(\lambda\, \hat{S}^0_T\right) \hat{S}^0_T \,\Big|\, \mathcal{A}_t\right) = E\left(\frac{1}{\lambda\, \hat{S}^0_T}\, \hat{S}^0_T \,\Big|\, \mathcal{A}_t\right) = \frac{1}{\lambda} \tag{11.3.10}$$

for $t \in [0, T]$. By equation (11.3.9) we obtain the Lagrange multiplier

$$\lambda = \frac{\bar{S}^{\delta_*}_0}{S_0}. \tag{11.3.11}$$

By formula (11.3.7) the risk aversion coefficient equals the constant

$$J^{\tilde{\delta}}_t = 1, \tag{11.3.12}$$

which shows that the corresponding expected log-utility maximizing portfolio $S^{\tilde{\delta}}$ is a GOP, see (11.1.19). This allows us to interpret the GOP as the portfolio which maximizes expected log-utility. Therefore, we could have defined earlier the GOP as the strictly positive portfolio which maximizes expected log-utility from discounted terminal wealth. Indeed, this idea has been followed in Platen (2004a) in the case of other asset price dynamics, since such a definition is generally applicable beyond the setting of a CFM. Note that in the relationships of this example the particular dynamics of the GOP did not play any role. We obtain the expected log-utility in the form

$$v^{\tilde{\delta}} = E\left(\ln\left(\bar{S}^{\delta_*}_T\right) \,\Big|\, \mathcal{A}_0\right) = \ln(\lambda) + \ln(S_0) + \frac{1}{2}\int_0^T E\left(\left(\theta(s, \bar{S}^{\delta_*}_s)\right)^2 \,\Big|\, \mathcal{A}_0\right) ds$$

if the local martingale part in the SDE for $\ln(\bar{S}^{\delta_*}_t)$ forms a martingale, see Exercise 11.1.

**2.** Our second example uses the power utility $U(x) = \frac{1}{\gamma} x^\gamma$ for $\gamma < 1$ and $\gamma \neq 0$. Its derivative is $U'(x) = x^{\gamma-1}$ and the corresponding inverse has the form $U'^{-1}(y) = y^{\frac{1}{\gamma-1}}$. The second derivative $U''(x) = (\gamma-1)x^{\gamma-2}$ is negative, which makes $U(\cdot)$ a suitable concave function.

According to (11.3.8) we have

$$\hat{u}(t, \hat{S}_t^0) = E\left(\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\right)^{\frac{1}{\gamma-1}} \frac{1}{\bar{S}_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) = \lambda^{\frac{1}{\gamma-1}}\, E\left(\left(\bar{S}_T^{\delta_*}\right)^{\frac{\gamma}{1-\gamma}} \,\Big|\, \mathcal{A}_t\right). \tag{11.3.13}$$

If there are analytic formulas for the conditional moments of the discounted GOP $\bar{S}_T^{\delta_*}$, then one can write down an explicit expression for the value of $\hat{u}(t, \hat{S}_t^0)$. Since $\bar{S}^{\delta_*}$ is in Theorem 11.3.3 assumed to be Markovian, one can apply the Feynman-Kac formula, see Sect. 9.7, to obtain the function $\hat{u}(\cdot, \cdot)$.

For simplicity, let us consider here the case where $\bar{S}^{\delta_*}$ is a geometric Brownian motion with $\theta(t, \bar{S}_t^{\delta_*}) = \theta > 0$. Thus, we obtain from (11.3.13) the expression

$$\begin{aligned}\hat{u}(t, \hat{S}_t^0) &= \lambda^{\frac{1}{\gamma-1}} \left(\bar{S}_t^{\delta_*}\right)^{\frac{\gamma}{1-\gamma}} E\left(\exp\left\{\frac{\gamma}{1-\gamma}\left(\frac{\theta^2}{2}\,(T-t) + \theta\,(W_T - W_t)\right)\right\} \,\Big|\, \mathcal{A}_t\right) \\ &= \lambda^{\frac{1}{\gamma-1}} \left(\bar{S}_t^{\delta_*}\right)^{\frac{\gamma}{1-\gamma}} \exp\left\{\frac{\theta^2}{2}\,\frac{\gamma}{(1-\gamma)^2}\,(T-t)\right\} \end{aligned} \tag{11.3.14}$$

for $t \in [0, T]$. By using (11.3.9) we obtain the Lagrange multiplier

$$\lambda = S_0^{\gamma-1} \left(\bar{S}_0^{\delta_*}\right)^{\gamma} \exp\left\{\frac{\theta^2}{2}\,\frac{\gamma}{1-\gamma}\,T\right\}. \tag{11.3.15}$$

Furthermore, from (11.3.14) by noting that $\hat{S}_t^0 = (\bar{S}_t^{\delta_*})^{-1}$ we obtain, see (10.3.1), the partial derivative

$$\frac{\partial \hat{u}(t, \hat{S}_t^0)}{\partial \hat{S}^0} = \frac{\hat{u}(t, \hat{S}_t^0)}{\hat{S}_t^0}\,\frac{\gamma}{\gamma - 1}. \tag{11.3.16}$$

Therefore, by (11.3.7) for the power utility under the BS model we obtain the risk aversion coefficient

$$J_t^{\tilde{\delta}} = 1 - \gamma \tag{11.3.17}$$

and the expected utility

$$v^{\tilde{\delta}} = E\left(\frac{1}{\gamma}\left(\bar{S}_T^{\tilde{\delta}}\right)^{\gamma} \,\Big|\, \mathcal{A}_0\right) = \frac{1}{\gamma}\,\exp\left\{\frac{\theta^2}{2}\,\frac{\gamma}{1-\gamma}\,T\right\}(S_0)^{\gamma},$$

see Exercise 11.2.

This recovers well-known results derived in Merton (1973a). One notes that as $\gamma \to 0$, the above risk aversion coefficient converges to one, which selects asymptotically the GOP as the expected utility maximizing portfolio. Note that for a power utility the particular dynamics of the discounted GOP are relevant. In this special case we have then also the constant fraction $\frac{1}{J_t^{\tilde{\delta}}} = \frac{1}{1-\gamma}$ of wealth invested in the GOP and the remainder in the savings account. This is again a fractional Kelly strategy.

**Proof of Theorem 11.3.3 (*)**

**1.** Since we only consider fair portfolios, we have a constrained optimization problem. Let us apply in the following the, so-called, martingale approach, see Karatzas & Shreve (1998). We express the constrained optimization problem (11.3.3) by using a Lagrange multiplier $\lambda \in \Re$, see Sect. 11.1, and maximize the functional

$$v^\delta = E\left(U\left(\bar{S}^\delta_T\right) \Big| \mathcal{A}_0\right) - \lambda \left(E\left(\frac{S^\delta_T}{S^{\delta_*}_T} \Big| \mathcal{A}_0\right) - \frac{S_0}{S^{\delta_*}_0}\right) \tag{11.3.18}$$

over the set $\bar{\mathcal{V}}^+_{S_0}$ of strictly positive, discounted, fair portfolios $\bar{S}^\delta$ starting with $\bar{S}^\delta_0 = S_0$. Then (11.3.18) can be rewritten as

$$v^\delta = E\left(U\left(\bar{S}^\delta_T\right) - \lambda\left(\frac{\bar{S}^\delta_T}{\bar{S}^{\delta_*}_T} - \frac{S_0}{\bar{S}^{\delta_*}_0}\right) \Big| \mathcal{A}_0\right). \tag{11.3.19}$$

This means, we seek a discounted portfolio $\bar{S}^{\tilde{\delta}} \in \bar{\mathcal{V}}^+_{S_0}$ so that

$$\begin{aligned} v^{\tilde{\delta}} = \max_{\bar{S}^\delta \in \bar{\mathcal{V}}^+_{S_0}} v^\delta &\leq E\left(\max_{\bar{S}^\delta \in \bar{\mathcal{V}}^+_{S_0}} \left\{U\left(\bar{S}^\delta_T\right) - \lambda\left(\frac{\bar{S}^\delta_T}{\bar{S}^{\delta_*}_T} - \frac{S_0}{\bar{S}^{\delta_*}_0}\right)\right\} \Big| \mathcal{A}_0\right) \\ &\leq E\left(\max_{\bar{S}^\delta_T > 0} \left\{U\left(\bar{S}^\delta_T\right) - \lambda\left(\frac{\bar{S}^\delta_T}{\bar{S}^{\delta_*}_T} - \frac{S_0}{\bar{S}^{\delta_*}_0}\right)\right\} \Big| \mathcal{A}_0\right). \end{aligned} \tag{11.3.20}$$

First let us solve a static optimization problem. This is an optimization that maximizes in (11.3.20), the expression under the conditional expectation on the right hand side of the last inequality, with respect to $\bar{S}^\delta_T$. One can read off the corresponding first order condition

$$U'\left(\bar{S}^\delta_T\right) - \frac{\lambda}{\bar{S}^{\delta_*}_T} = 0, \tag{11.3.21}$$

which for $\lambda > 0$ characterizes a maximum since $U$ is concave, $U'(0) = \infty$ and $U'(\infty) = 0$. Note that due to the strict concavity of $U$ its derivative $U'$ has an inverse function $U'^{-1}$. By applying the inverse function $U'^{-1}$ of $U'$ it follows from (11.3.21) that the value

$$\bar{S}^\delta_T = U'^{-1}\left(\frac{\lambda}{\bar{S}^{\delta_*}_T}\right) \tag{11.3.22}$$

is the candidate for the optimal value of the discounted portfolio at time $T$ that an expected utility maximizer should replicate. It is very important to realize that this candidate value turns out to be a function of the discounted GOP value. In principle, we face now a hedging problem that replicates via $\bar{S}^\delta$ the payoff given in (11.3.22).

**2.** Since $U'^{-1} : [0,\infty] \to [0,\infty]$, it makes only sense to consider in the following strictly positive values of $\lambda$. Since $S^\delta$ is assumed to be a fair portfolio one needs by (11.3.22) to choose the constant $\lambda \in (0,\infty)$ such that

$$\frac{S_0}{S_0^{\delta_*}} = \frac{S_0^\delta}{S_0^{\delta_*}} = E\left( \frac{S_T^\delta}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_0 \right) = E\left( \frac{\bar{S}_T^\delta}{\bar{S}_T^{\delta_*}} \,\Big|\, \mathcal{A}_0 \right) = E\left( U'^{-1}\left( \frac{\lambda}{\bar{S}_T^{\delta_*}} \right) \frac{1}{\bar{S}_T^{\delta_*}} \,\Big|\, \mathcal{A}_0 \right). \tag{11.3.23}$$

Due to the properties of $U$ given in Definition 11.3.1, it follows that there exists a $\lambda \in (0,\infty)$ such that (11.3.23) holds. Note that for very small $\lambda > 0$ one obtains extremely large payoffs $U'^{-1}(\frac{\lambda}{\bar{S}_T^{\delta_*}})\frac{1}{\bar{S}_T^{\delta_*}}$. With (11.3.23) we have identified a candidate value for an expected utility maximizing portfolio.

**3.** We now show that there is a strategy $\tilde{\delta}$ that replicates with its benchmarked portfolio value $\hat{S}_T^{\tilde{\delta}}$ the payoff $U'^{-1}(\frac{\lambda}{\bar{S}_T^{\delta_*}})\frac{1}{\bar{S}_T^{\delta_*}}$ in (11.3.22), such that $\bar{S}^{\tilde{\delta}} \in \bar{\mathcal{V}}_{S^0}^+$. Since the benchmarked savings account $\hat{S}_t^0 = (\bar{S}_t^{\delta_*})^{-1}$ forms a Markov process we obtain the $(\underline{\mathcal{A}}, P)$-martingale $\hat{u}(\cdot, \hat{S}^0_\cdot) = \{\hat{u}(t, \hat{S}_t^0),\, t \in [0,T]\}$ with

$$\hat{u}(t, \hat{S}_t^0) = \hat{S}_t^{\tilde{\delta}} = E\left( U'^{-1}\left( \frac{\lambda}{\bar{S}_T^{\delta_*}} \right) \frac{1}{\bar{S}_T^{\delta_*}} \,\Big|\, \mathcal{A}_t \right) = E\left( U'^{-1}\left( \lambda\, \hat{S}_T^0 \right) \hat{S}_T^0 \,\Big|\, \mathcal{A}_t \right) \tag{11.3.24}$$

for $t \in [0,T]$. Here $\hat{u}(t, \hat{S}_t^0)$ is a function of $t$ and $\hat{S}_t^0$ only, which can be identified via the Feynman-Kac formula (9.7.3). By application of the Itô formula and using the martingale property of $\hat{u}(\cdot, \hat{S}^0_\cdot)$ we obtain

$$d\hat{u}(t, \hat{S}_t^0) = \frac{\partial \hat{u}(t, \hat{S}_t^0)}{\partial \hat{S}^0}\, d\hat{S}_t^0$$

for $t \in [0,\infty)$. Hence, one can form the locally optimal portfolio $S^{\tilde{\delta}}$ by investing at time $t$ in $\hat{u}(t, \hat{S}_t^0) - \hat{S}_t^0 \frac{\partial \hat{u}(t, \hat{S}_t^0)}{\partial \hat{S}^0}$ units of the GOP and investing the remaining wealth in $\frac{\partial \hat{u}(t, \hat{S}_t^0)}{\partial \hat{S}^0}$ units of the savings account. Note that we have

$$\bar{S}_t^{\tilde{\delta}} = \frac{\hat{S}_t^{\tilde{\delta}}}{\hat{S}_t^0} = \frac{\hat{u}(t, \hat{S}_t^0)}{\hat{S}_t^0}.$$

Consequently, the discounted, locally optimal portfolio $\bar{S}_t^{\tilde{\delta}}$ satisfies the SDE

$$
\begin{aligned}
d\bar{S}_t^{\tilde{\delta}} &= \hat{u}(t,\hat{S}_t^0)\, d\bar{S}_t^{\delta_*} + \bar{S}_t^{\delta_*}\, d\hat{u}(t,\hat{S}_t^0) + d[\bar{S}^{\delta_*},\hat{u}]_t \\
&= \left(\hat{u}(t,\hat{S}_t^0) - \hat{S}_t^0\, \frac{\partial \hat{u}(t,\hat{S}_t^0)}{\partial \hat{S}^0}\right) d\bar{S}_t^{\delta_*} \\
&= \bar{S}_t^{\tilde{\delta}} \left(\hat{u}(t,\hat{S}_t^0) - \hat{S}_t^0\, \frac{\partial \hat{u}(t,\hat{S}_t^0)}{\partial \hat{S}^0}\right) \frac{\bar{S}_t^{\delta_*}}{\bar{S}_t^{\tilde{\delta}}}\, \theta(t,\bar{S}_t^{\delta_*}) \left(\theta(t,\bar{S}_t^{\delta_*})\, dt + dW_t\right) \\
&= \bar{S}_t^{\tilde{\delta}} \left(J_t^{\tilde{\delta}}\right)^{-1} \theta(t,\bar{S}_t^{\delta_*}) \left(\theta(t,\bar{S}_t^{\delta_*})\, dt + dW_t\right),
\end{aligned}
$$

where $\hat{u}(t,\hat{S}_t^0) = \frac{\bar{S}_t^{\tilde{\delta}}}{\bar{S}_t^{\delta_*}}$, with risk aversion coefficient

$$
J_t^{\tilde{\delta}} = \frac{\hat{u}(t,\hat{S}_t^0)}{\hat{u}(t,\hat{S}_t^0) - \hat{S}_t^0\, \frac{\partial \hat{u}(t,\hat{S}_t^0)}{\partial \hat{S}^0}} = \left(1 - \frac{\hat{S}_t^0}{\hat{u}(t,\hat{S}_t^0)}\, \frac{\partial \hat{u}(t,\hat{S}_t^0)}{\partial \hat{S}^0}\right)^{-1}
$$

for $t \in [0,\infty)$.

**4.** It follows from (11.3.24) that $\hat{S}^{\tilde{\delta}}$ is a martingale. Furthermore, we note by the nonnegativity of $U'^{-1}$ that $S^{\tilde{\delta}}$ is nonnegative. The solution that has been obtained must be shown to belong to the set $\bar{\mathcal{V}}_{S_0}^+$. For this purpose it suffices to show that equality holds in (11.3.20). This is achieved by observing that for positive $\lambda$, satisfying (11.3.9), one has

$$
\begin{aligned}
&E\left(\max_{\bar{S}_T^{\delta}>0} \left\{U\left(\bar{S}_T^{\delta}\right) - \lambda\left(\frac{\bar{S}_T^{\delta}}{\bar{S}_T^{\delta_*}} - \frac{S_0}{\bar{S}_0^{\delta_*}}\right)\right\} \,\middle|\, \mathcal{A}_0\right) \\
&\quad = E\left(U\left(U'^{-1}\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\right)\right) - \lambda\left(\frac{U'^{-1}\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\right)}{\bar{S}_T^{\delta_*}} - \frac{S_0}{\bar{S}_0^{\delta_*}}\right) \,\middle|\, \mathcal{A}_0\right) \\
&\quad = E\left(U\left(\bar{S}_T^{\tilde{\delta}}\right) - \lambda\left(\frac{\bar{S}_T^{\tilde{\delta}}}{\bar{S}_T^{\delta_*}} - \frac{S_0}{\bar{S}_0^{\delta_*}}\right) \,\middle|\, \mathcal{A}_0\right) \\
&\quad = E\left(U\left(\bar{S}_T^{\tilde{\delta}}\right) \,\middle|\, \mathcal{A}_0\right) = v^{\tilde{\delta}}.
\end{aligned}
$$

□

## 11.4 Pricing Nonreplicable Payoffs

This section addresses the problem of pricing nonreplicable payoffs. These are payoffs that cannot be replicated by a fair portfolio of primary security accounts. By utility indifference pricing we shall show that the concept of real world pricing naturally applies to the pricing of nonreplicable payoffs.

### Utility Indifference Price

In the following we shall continue to use our notation of the previous sections, in particular Sect. 11.3, which considered expected utility maximization in the framework of a CFM. Let us assume that the investor uses the utility function $U$ with time horizon $T \in (0, \infty)$, as defined in Definition 11.3.1. The investor has the total tradable wealth $S_t^{\tilde{\delta}}$ accumulated at time $t \in [0, \infty)$, which she or he invests according to an expected utility maximizing strategy $\tilde{\delta}$, see Sect. 11.1.

We consider now the problem that the investor has to price a random, discounted, nonnegative payoff $\bar{H}$ that is $\mathcal{A}_T$-measurable and delivered at the same time $T$ which determines the time horizon for the expected utility function. We allow $\bar{H}$ to be nonreplicable. This means that the discounted payoff $\bar{H}$ or parts of it cannot be replicated by a fair portfolio of primary security accounts. Let us assume that the total face value of the discounted payoff that the investor wants to purchase is vanishing small, that is, it amounts to $\varepsilon \bar{H}$ where $\varepsilon \ll 1$ is a very small real number.

We aim to identify a consistent price for the above payoff at time $t = 0$ from the viewpoint of the expected utility maximizer. For this purpose we apply the concept of *utility indifference pricing*. This is a classical economic concept that has been generating renewed interest in continues time finance due to the important work in Davis (1997). The utility indifference price is the price at which the investor is indifferent between buying the contract that provides the discounted payoff $\varepsilon \bar{H}$, or not accepting the price when taking her or his expected utility maximization objective into account.

Consider now a contract that can be purchased for a hypothetical price $V$ at time $t = 0$ and which delivers the discounted payoff $\bar{H}$ at maturity $T \in (0, \infty)$. Assume that the investor buys a vanishing fraction $\varepsilon \ll 1$ of the contract at time $t = 0$ for the amount $\varepsilon V$. This corresponds to the price $V$ at time $t = 0$ per total contract. She or he continues to invest the bulk of the wealth with her or his locally optimal strategy $\tilde{\delta}$, determined by the expected utility maximization for the utility function $U(\cdot)$ with time horizon $T$. Similarly to (11.3.3)–(11.3.4) we introduce the expected utility function

$$v_{\varepsilon,V}^{\tilde{\delta}} = E\left( U\left( (S_0 - \varepsilon\, V)\, \frac{\bar{S}_T^{\tilde{\delta}}}{S_0} + \varepsilon \bar{H} \right) \Bigg| \, \mathcal{A}_0 \right) \tag{11.4.1}$$

for $\varepsilon \geq 0$. Here $S_0 - \varepsilon V$ is invested at time $t = 0$ in a portfolio which starts at one and follows the locally optimal strategy $\tilde{\delta}$. At the delivery date $T$ the discounted payoff $\varepsilon \bar{H}$ is added to the discounted payoff $(S_0 - \varepsilon V) \frac{\bar{S}_T^{\tilde{\delta}}}{S_0}$ of the investment in the locally optimal portfolio. Note that the purchasing price $\varepsilon V$ is at time $t = 0$ subtracted from the locally optimal portfolio value. This allows us to formulate the following definition of a utility indifference price.

**Definition 11.4.1.** *In the above framework the value $V$ is called the* utility indifference price *for the discounted payoff $\bar{H}$ if*

$$\lim_{\varepsilon \to 0} \frac{v^{\tilde{\delta}}_{\varepsilon,V} - v^{\tilde{\delta}}_{0,V}}{\varepsilon} = 0 \qquad (11.4.2)$$

*almost surely.*

This means that the maximized expected utility of the investor changes only by a small amount for prices that are in the neighborhood of the utility indifference price. To see the structure of the resulting expected utility more clearly, let us derive from (11.4.1), by a Taylor expansion, the representation

$$\begin{aligned} v^{\tilde{\delta}}_{\varepsilon,V} &\approx E\left(U\left(\bar{S}^{\tilde{\delta}}_T\right) + U'\left(\bar{S}^{\tilde{\delta}}_T\right)\varepsilon\left(\bar{H} - V\,\frac{\bar{S}^{\tilde{\delta}}_T}{S_0}\right)\,\Big|\,\mathcal{A}_0\right) \\ &= v^{\tilde{\delta}}_{0,V} + \varepsilon\, E\left(U'\left(\bar{S}^{\tilde{\delta}}_T\right)\left(\bar{H} - V\,\frac{\bar{S}^{\tilde{\delta}}_T}{S_0}\right)\,\Big|\,\mathcal{A}_0\right). \end{aligned} \qquad (11.4.3)$$

Here we neglect higher order terms in $\varepsilon$, assuming appropriate conditions. This expansion allows us to identify the utility indifference price. It is clear that for particular dynamics and specific utility functions, as well as payoffs, one needs to check whether the above expansion applies.

### Utility Indifference Pricing Formula

When appropriate conditions are imposed, one can derive for a given utility function, discounted payoff $\bar{H}$ and prescribed market dynamics a corresponding utility indifference price. What is needed in such a derivation are sufficient integrability and smoothness properties. For instance, for a BS model and power utility such properties are guaranteed. For the utility indifference price we derive its general formula heuristically by indicating the crucial steps for its derivation without formulating any assumptions. However, this can be done for particular classes of models, utilities and payoffs. The general result that we shall obtain below will always be the same.

We obtain from the expansion (11.4.3) the relation

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\left(v^{\tilde{\delta}}_{\varepsilon,V} - v^{\tilde{\delta}}_{0,V}\right) = E\left(U'\left(\bar{S}^{\tilde{\delta}}_T\right)\left(\bar{H} - V\,\frac{\bar{S}^{\tilde{\delta}}_T}{S_0}\right)\,\Big|\,\mathcal{A}_0\right). \qquad (11.4.4)$$

We emphasize that $S^{\tilde{\delta}}$ is here the locally optimal portfolio that maximizes the given expected utility when $\varepsilon$ is set to zero. From equation (11.4.4) and Definition 11.4.1 we obtain then the *utility indifference pricing formula* in the form

$$V = \frac{E\left(U'\left(\bar{S}^{\tilde{\delta}}_T\right)\bar{H}\,\big|\,\mathcal{A}_0\right)}{E\left(U'\left(\bar{S}^{\tilde{\delta}}_T\right)\frac{\bar{S}^{\tilde{\delta}}_T}{S_0}\,\big|\,\mathcal{A}_0\right)}. \qquad (11.4.5)$$

This formula holds rather generally. It allows us to determine the utility indifference price for a given utility and given discounted payoff $\bar{H}$. We emphasize that the payoff is possibly not replicable. If it were replicable, then the minimal price for replicating this payoff is the fair price, which is given by the real world pricing formula.

### Real World Pricing of Nonreplicable Payoffs

For a general payoff $H$ and a general utility function $U(\cdot)$ we obtain under the assumptions of Theorem 11.3.3 by (11.3.8) that

$$\bar{S}_T^{\tilde{\delta}} = U'^{-1}\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\right). \tag{11.4.6}$$

It follows from formula (11.4.5) that in a surprisingly simple way $U'$ and $U'^{-1}$ offset each other in the following calculation

$$V = \frac{E\left(U'\left(U'^{-1}\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\right)\right)\bar{H}\,\Big|\,\mathcal{A}_0\right)}{E\left(U'\left(U'^{-1}\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\right)\right)\frac{\bar{S}_T^{\tilde{\delta}}}{S_0}\,\Big|\,\mathcal{A}_0\right)} = \frac{E\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\,\bar{H}\,\Big|\,\mathcal{A}_0\right)}{E\left(\frac{\lambda}{\bar{S}_T^{\delta_*}}\,\frac{\bar{S}_T^{\tilde{\delta}}}{S_0}\,\Big|\,\mathcal{A}_0\right)}.$$

Therefore, we obtain with (11.3.8) for $V$ the expression

$$V = \frac{E\left(\frac{H}{S_T^{\delta_*}}\,\Big|\,\mathcal{A}_0\right)}{\frac{1}{S_0}\,E\left(\hat{S}_T^{\tilde{\delta}}\,\big|\,\mathcal{A}_0\right)} = \frac{E\left(\frac{H}{S_T^{\delta_*}}\,\Big|\,\mathcal{A}_0\right)}{\frac{1}{S_0}\,\hat{u}(0,\hat{S}_0^0)} = \frac{E\left(\frac{H}{S_T^{\delta_*}}\,\Big|\,\mathcal{A}_0\right)}{\frac{S_0}{S_0\,S_0^{\delta_*}}}. \tag{11.4.7}$$

For the utility indifference price this yields by (11.3.9) the relation

$$V = S_0^{\delta_*}\,E\left(\frac{H}{S_T^{\delta_*}}\,\Big|\,\mathcal{A}_0\right). \tag{11.4.8}$$

We observe that this is the real world pricing formula (9.1.30). This means that under utility indifference pricing payoffs, which are not replicable by a fair portfolio of primary security accounts, are priced according to the real world pricing formula. Most importantly, we see that the utility indifference price does *not* depend on the utility function of the investor.

This is a very satisfying result not only from the theoretical but also from the practical point of view. It extends real world pricing naturally to the case of general nonreplicable payoffs. From a practical viewpoint it gives the buyer and the seller an acceptable price for any nonreplicable payoff.

## 11.5 Hedging

One important feature of a market is the possibility to hedge future uncertainties. In this section we study the hedging of uncertain payoffs.

### Hedge Portfolios

In the following we consider a CFM $\mathcal{S}^{\mathrm{C}}_{(d)}$, as defined in Sect. 10.1, and discuss the problem of hedging. Let $\tau \in (0, \infty)$ be a bounded stopping time and $H$ a nonnegative payoff that is paid at $\tau$. By generalizing (8.2.8), we say that a portfolio $S^{\delta}$ *replicates* a nonnegative payoff $H_\tau$ paid at a stopping time $\tau$ if

$$S^{\delta}_{\tau} = H \tag{11.5.1}$$

almost surely. Note that a general payoff can always be decomposed into its nonnegative and its negative part and considering nonnegative payoffs is therefore no restriction. As previously, a nonnegative payoff is replicable if there exists a nonnegative, replicating fair portfolio. We shall demonstrate later that there may exist several self-financing portfolios in a CFM that replicate a given nonnegative payoff. In the case of nonnegative replicating portfolios it follows from Corollary 10.4.2 that for a nonnegative payoff $H$ the replicating, fair portfolio $S^{\delta_H}$ is the minimal portfolio that replicates $H$. Note that this portfolio process is uniquely determined as a value process. However, there may be different securities that can be used for hedging.

### Tradable Martingale Representation

For a nonnegative replicable payoff the real world pricing formula provides the minimal nonnegative price process. From an economic point of view it is in a competitive market the correct price process. We shall determine below the strategy of the fair portfolio which hedges a given replicable nonnegative payoff. Recall that the benchmarked fair price process forms an $(\underline{\mathcal{A}}, P)$-martingale. It is of primary interest to find a representation for this martingale process. There are various methods that can be used to find the martingale representation of a benchmarked nonnegative payoff.

For instance, under the standard BS model, which we used for illustration in Chap. 9, we obtained in (9.1.31) a corresponding martingale representation for the benchmarked European call option payoff. It was derived from the real world pricing formula together with an application of the Itô formula to the benchmarked pricing function.

More generally, in the case when the market dynamics can be expressed via a set of Markovian factor processes, then one can apply the Feynman-Kac formula, see Sect. 9.7. This yields the benchmarked, fair pricing function of a corresponding replicable payoff. The corresponding benchmarked, fair price process forms then a martingale and similarly to (9.1.31), a real world martingale representation. We shall not present in this section any particular example for such a martingale representation. However, the following two chapters will discuss several such examples.

In a CFM not all Wiener processes which drive volatility processes and short rates need to represent trading uncertainty. Therefore, in general, not all payoffs are replicable. Under rather general assumptions one can usually establish a real world martingale representation for reasonable payoffs in a CFM. The particular structure of a martingale representation depends strongly on the model dynamics and can become quite complex for certain payoffs. As indicated, in a Markovian setting one can explicitly derive, via the Feynman-Kac formula, martingale representations for nonreplicable payoffs.

The following definition of a *tradable martingale representation* allows us to formulate general results on pricing and hedging of particular payoffs without specifying the dynamics of the CFM.

**Definition 11.5.1.** *We say that a given $\mathcal{A}_\tau$-measurable, nonnegative payoff $H$, which matures at a bounded stopping time $\tau$, has a* tradable martingale representation *if there exists a predictable vector process $\boldsymbol{x}_H = \{\boldsymbol{x}_H(t) = (x_H^1(t), \ldots, x_H^d(t))^\top,\ t \in [0, \tau]\}$, where*

$$\int_0^\tau \sum_{k=1}^d \left(x_H^k(s)\right)^2 ds < \infty \tag{11.5.2}$$

*almost surely such that*

$$\frac{H}{S_\tau^{\delta_*}} = \hat{U}_H(t) + \sum_{k=1}^d \int_t^\tau x_H^k(s)\, dW_s^k \tag{11.5.3}$$

*almost surely with*

$$\hat{U}_H(t) = E\left(\frac{H}{S_\tau^{\delta_*}} \,\middle|\, \mathcal{A}_t\right) < \infty \tag{11.5.4}$$

*for all $t \in [0, \tau]$.*

Note that the above tradable martingale representation (11.5.3) is expressed with respect to trading uncertainty, that is with respect to the Wiener processes $W^1, \ldots, W^d$. There are, in general, other sources of uncertainty in the market that are not securitized and therefore not tradable. Consequently, there exist, in general, nonnegative payoffs which do not have a tradable martingale representation. We shall see below that such payoffs are not fully replicable.

## Hedging Strategy

By using the above notion of a tradable martingale representation we prove the following result on the hedging of derivatives. In the corresponding proof, which is given at the end of this section, we use the SDE (10.3.2) of a benchmarked portfolio together with (11.5.3).

**Theorem 11.5.2.** *For a nonnegative payoff $H$ with a tradable martingale representation there exists a replicating, fair portfolio $S^{\delta_H}$, which satisfies at time $t \in [0,\tau]$ the real world pricing formula*

$$S_t^{\delta_H} = S_t^{\delta_*}\, \hat{U}_H(t) \tag{11.5.5}$$

*with $\hat{U}_H(t)$ given in (11.5.4). This portfolio has the vector of fractions*

$$\boldsymbol{\pi}_{\delta_H}(t) = \left(\boldsymbol{b}_{\delta_H}(t)^\top\, \boldsymbol{b}_t^{-1}\right)^\top, \tag{11.5.6}$$

*where the vector $\boldsymbol{b}_{\delta_H}(t) = (b^1_{\delta_H}(t), \ldots, b^d_{\delta_H}(t))^\top$ of portfolio volatilities has kth component*

$$b^k_{\delta_H}(t) = \sum_{j=1}^{d} \frac{\delta^j_H(t)\, \hat{S}^j_t}{\hat{U}_H(t)}\, b^{j,k}_t = \frac{x^k_H(t)}{\hat{U}_H(t)} + \theta^k_t \tag{11.5.7}$$

*for $t \in [0,\tau]$ and $k \in \{1,2,\ldots,d\}$.*

Theorem 11.5.2 states that a nonnegative payoff with tradable martingale representation can be replicated. It also characterizes the minimal hedge portfolio. We emphasize here again that for a CFM, which is built as a Markovian factor model, one can obtain by the Feynman-Kac formula for each integrable benchmarked payoff a corresponding martingale representation. This makes it advisable to prefer Markovian factor models if one aims to construct computationally tractable CFMs.

### Martingale Representation Theorem (*)

By the following result we shall see that payoffs can be decomposed into the sum of their hedgable part and their unhedgable part. Let us mention a *Martingale Representation Theorem*, for a proof see Karatzas & Shreve (1991), which is convenient for establishing a real world martingale representation for payoffs in a wide range of CFMs.

**Theorem 11.5.3.** (Martingale Representation Theorem) *For $T \in [0,\infty)$ assume that in a CFM $\mathcal{S}^C_{(d)}$ with given filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ the filtration $\underline{\mathcal{A}}$ is the augmentation under $P$ of the natural filtration $\mathcal{A}^W$ generated by the vector $\boldsymbol{W} = \{\boldsymbol{W}_t = (W^1_t, \ldots, W^m_t)^\top,\ t \in [0,T]\}$ of Wiener processes, $m \in \{d, d+1, \ldots\}$. Then for any square integrable benchmarked fair price process $\hat{V}_t = \{\hat{V}_t = \frac{V_t}{S_t^{\delta_*}},\ t \in [0,T]\}$ there exists a predictable, measurable process $\boldsymbol{x}_{V_T} = \{\boldsymbol{x}_{V_T}(t) = (x^1_{V_T}(t), \ldots, x^d_{V_T}(t))^\top,\ t \in [0,T]\}$ such that*

$$E\left(\int_0^T \left(x^k_{V_T}(s)\right)^2 ds\right) < \infty \tag{11.5.8}$$

*for $k \in \{1,2,\ldots,d\}$ and*

$$\hat{V}_t = \hat{V}_0 + \sum_{k=1}^{m} \int_0^t x_{V_T}^k(s)\, dW_s^k \tag{11.5.9}$$

*for* $t \in [0,T]$, *where* $\hat{V}$ *is almost surely continuous. Furthermore, if* $\tilde{x}^k = \{\tilde{x}^k(t),\, t \in [0,T],\, k \in \{1,2,\dots,d\}$, *are any other predictable measurable processes satisfying* (11.5.8) *and* (11.5.9), *then*

$$\int_0^T \sum_{k=1}^{m} \left| x_{V_T}^k(s) - \tilde{x}^k(s) \right|^2 \, ds = 0 \tag{11.5.10}$$

*almost surely.*

### Real World Martingale Decomposition (*)

Note that under the assumptions of the above theorem one has for any square integrable, benchmarked payoff $\hat{H} = \frac{H}{S_T^{\delta_*}}$ paid at time $T$, the unique representation (11.5.9), where

$$\hat{H} = \hat{V}_T = \hat{V}_0 + \sum_{k=1}^{m} \int_0^t x_H^k(s)\, dW_s^k. \tag{11.5.11}$$

It is essential to realize that Theorem 11.5.3 assumes that only the $m$ Wiener processes $W^1, \dots, W^m$ generate the total uncertainty in the model. This is why we have chosen in Theorem 11.5.3 the filtration $\underline{\mathcal{A}}$ to be the augmentation of the natural filtration $\mathcal{A}^W$. The Wiener processes $W^1, \dots, W^d$ model the trading uncertainty.

The *real world martingale decomposition* of the nonnegative, square integrable benchmarked payoff $\hat{H}$ is then given by the sum

$$\hat{H} = \hat{H}_h + \hat{H}_u. \tag{11.5.12}$$

It consists of its *hedgable part* $\hat{H}_h = \hat{U}_{H_h}(T)$, which we obtain at time $t$ as

$$\hat{U}_{H_h}(t) = \hat{U}_{H_h}^{(0)} + \sum_{k=1}^{d} \int_0^t x_H^k(s)\, dW_s^k \tag{11.5.13}$$

and its *unhedgable part* $\hat{H}_u = \hat{U}_{H_u}(T)$, which at time $t \in [0,T]$ is

$$\hat{U}_{H_u}(t) = \sum_{k=d+1}^{m} \int_0^t x_H^k(s)\, dW_s^k. \tag{11.5.14}$$

The hedgable part $\hat{H}_h$ can be replicated according to Theorem 11.5.2. We use in (11.5.13) for the nonnegative payoff $H$ its benchmarked fair price $\hat{U}_{H_h}(0)$ at time $t = 0$, that is,

$$\hat{U}_{H_h}(0) = E\left(\hat{H} \,\middle|\, \mathcal{A}_0\right). \qquad (11.5.15)$$

Then the benchmarked value

$$\hat{V}_t = \hat{U}_{H_h}(t) + \hat{U}_{H_u}(t) \qquad (11.5.16)$$

corresponds to a fair process since it forms an $(\underline{\mathcal{A}}, P)$-martingale. As we have seen in Sect. 5.1, this martingale minimizes the expected least squares error of the benchmarked hedge. The choice of the real world pricing formula for the unhedgable part appears, therefore, as a projection in a least square sense. More precisely, the benchmarked fair price $\hat{V}_0$ can be interpreted as the projection of the benchmarked payoff into the space of $\mathcal{A}_0$-measurable, tradable portfolio values. Note that the benchmarked fair price $\hat{U}_{H_u}(0)$ of the unhedgable part is zero at time $t = 0$.

This means, when applying real world pricing for a payoff one is leaving its unhedgable part totally untouched. This is reasonable because any extra trading would create unnecessary uncertainty and potential costs. The benchmarked unhedgable part has according to (11.5.14) zero conditional expectation

$$E\left(\hat{U}_{H_u}(T) \,\middle|\, \mathcal{A}_0\right) = 0. \qquad (11.5.17)$$

In summary, we obtain from (11.5.12)–(11.5.17) for the benchmarked payoff $\hat{H}$ payable at time $T$ the real world martingale decomposition

$$\hat{H} = \hat{U}_{H_h}(0) + \sum_{k=1}^{d} \int_0^T x_H^k(s)\, dW_s^k + \sum_{k=d+1}^{m} \int_0^T x_H^k(s)\, dW_s^k. \qquad (11.5.18)$$

Let us indicate that by pooling a wide variety of independent unhedgable parts of payoffs, under appropriate integrability conditions, the Law of Large Numbers, see Sect. 2.1, makes their impact vanishing. For instance, the books of large investment banks and insurance companies pool substantial unhedgable payoffs and benefit from this effect.

### Föllmer-Schweizer Decomposition (*)

In the case when an equivalent risk neutral probability measure exists in a CFM, real world pricing coincides with the risk neutral pricing obtained under the, so-called, *minimal equivalent martingale measure* of Föllmer and Schweizer, see Föllmer & Schweizer (1991), Hofmann, Platen & Schweizer (1992) and Heath, Platen & Schweizer (2001).

It has been shown in Föllmer & Schweizer (1991), by assuming the existence of an equivalent risk neutral probability measure $P_\theta$, that the hedging of a payoff is linked to the existence of a corresponding martingale representation under $P_\theta$ for the discounted payoff. This important representation is known as the *Föllmer-Schweizer decomposition*, see Schweizer (1995). A similar decomposition exists for the benchmarked payoff in a general CFM, where one does not require the existence of an equivalent risk neutral probability measure.

To formulate explicitly the Föllmer-Schweizer decomposition we consider a CFM as assumed in Theorem 11.5.3 and multiply both sides of the representation (11.5.16) by the discounted GOP value and apply then the Itô formula. This provides the decomposition

$$\begin{aligned}\bar{H} &= \hat{H}\,\bar{S}_T^{\delta_*}\\ &= \hat{U}_{H_h}(0)\,\bar{S}_0^{\delta_*} + \sum_{k=1}^{d}\int_0^T \bar{S}_t^{\delta_*}\left(x_H^k(t) + \left(\hat{U}_{H_h}(t) + \hat{U}_{H_u}(t)\right)\theta_t^k\right)\\ &\qquad \times \left(\theta_t^k\,dt + dW_t^k\right) + \sum_{k=d+1}^{m}\int_0^T \bar{S}_t^{\delta_*}\,x_H^k(t)\,dW_t^k, \qquad (11.5.19)\end{aligned}$$

which is a Föllmer-Schweizer decomposition for the discounted payoff $\bar{H}$, see Schweizer (1995) and Exercise 11.3.

Note that in the case when a risk neutral probability measure $P_\theta$ exists, then the second term in the sum on the right hand side of (11.5.19) is a martingale under $P_\theta$. The third term is then a martingale under $P$ and under $P_\theta$. The, so-called, *minimal equivalent martingale measure*, see Schweizer (1995), changes only the drift of the Wiener processes that model trading uncertainty. The other sources of uncertainty remain unchanged.

## Complete Market (*)

In the literature one is often using the notion of a *complete market*, which we introduce now.

**Definition 11.5.4.** *A CFM where all integrable, benchmarked nonnegative payoffs have a tradable martingale representation in the sense of Definition 11.5.1, is called a* complete CFM. *Any other CFM we call* incomplete.

Note that in some literature a market is called complete when a unique equivalent risk neutral probability measure exists, see Harrison & Kreps (1979) and Harrison & Pliska (1981, 1983). As we have seen in Sect. 10.3, there is no economic necessity to insist on the existence of an equivalent risk neutral probability measure. Therefore, we defined here the completeness of a market in a more practical way. By Theorem 11.5.2 we obtain now directly the following result.

**Corollary 11.5.5.** *In a complete CFM all integrable, nonnegative payoffs can be perfectly replicated with the hedge portfolio characterized by relation (11.5.6). The price for setting up this replicating portfolio at some time $t \in [0,\tau)$ is obtained by the real world pricing formula (9.1.30). This price is the minimal price that permits the replication of a given payoff.*

This result emphasizes the fact that in a complete market all integrable payoffs can be replicated. An equivalent risk neutral probability measure is not required for the existence of a complete market. This is very important from a practical point of view when hedging derivatives for advanced models, as we shall see later. We have seen in our previous discussion, if the CFM is incomplete, then one can still perfectly replicate the hedgable part of a benchmarked payoff. Under real world pricing one leaves the unhedgable part as it is.

### Proof of Theorem 11.5.2 (*)

For a given payoff $H$, paid at a bounded stopping time $\tau \geq 0$, with tradable martingale representation we use the martingale representation (11.5.3). This leads us for a benchmarked hedge portfolio $\hat{S}^{\delta_H}$, see (10.3.2), to the replication condition

$$\begin{aligned} \frac{H}{S_\tau^{\delta_*}} - \hat{U}_H(t) &= \sum_{k=1}^{d} \int_t^\tau x_H^k(s)\, dW_s^k \\ &= \sum_{k=1}^{d} \int_t^\tau \hat{S}_s^{\delta_H} \left( b_{\delta_H}^k(s) - \theta_s^k \right) dW_s^k = \hat{S}_\tau^{\delta_H} - \hat{S}_t^{\delta_H} \end{aligned} \tag{11.5.20}$$

for $t \in [0, \tau]$. The formulas (10.3.2), (11.5.7) and (10.1.12) provide by direct comparison of the integrands in (11.5.20) the equation

$$\left( \boldsymbol{\pi}_{\delta_H}^\top(t)\, \boldsymbol{b}_t \right)^\top = \boldsymbol{b}_{\delta_H}(t)$$

for $t \in [0, \tau]$. By the invertibility of $\boldsymbol{b}_t$, see Assumption 10.1.1, this proves (11.5.6), and thus with (11.5.1) equation (11.5.5). □

## 11.6 Exercises for Chapter 11

**11.1.** Calculate the maximum expected log-utility for the BS model.

**11.2.** Compute the maximum expected power utility for the BS model.

**11.3.** (*) In the case when a risk neutral probability measure exists, derive by using the setup of Theorem 11.5.3 a representation for a discounted payoff $\bar{H} = \frac{H}{S_T^0}$, which is paid at time $T \in (0, \infty)$.

# 12

# Modeling Stochastic Volatility

This chapter introduces into the pricing and hedging of derivatives under stochastic volatility. The emphasis is on standard derivatives for various index models. We choose as underlying security a diversified index, which we interpret as GOP.

## 12.1 Stochastic Volatility

### Stochastic Volatility of an Index

Since a diversified accumulation index can be interpreted as a diversified portfolio we assume that its dynamics are closely approximated by that of a GOP. The value $S_t^{\delta_*}$ of a GOP at time $t$ satisfies by (10.2.8) the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*} \left(r_t \, dt + |\theta_t| \, (|\theta_t| \, dt + dW_t)\right) \tag{12.1.1}$$

for $t \in [0, \infty)$ with $S_0^{\delta_*} > 0$. Here $r = \{r_t, \, t \in [0, \infty)\}$ is the short term interest rate process, which we assume in this chapter, for simplicity, to be constant, such that $r_t = r \geq 0$ for all $t \in [0, \infty)$. Furthermore, $|\theta_t|$ denotes the volatility of the GOP at time $t$, which is the, in general, stochastic total market price of risk, see (11.1.11). Finally, $W = \{W_t, \, t \in [0, \infty)\}$ is a standard Wiener process on $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. Note that the volatility and the short rate characterize the dynamics of the GOP in the denomination of the domestic currency.

If we consider the logarithm of the GOP, then the SDE follows by the Itô formula and (12.1.1) in the form

$$d \ln \left(S_t^{\delta_*}\right) = \left(r + \frac{1}{2} \, |\theta_t|^2\right) dt + |\theta_t| \, dW_t \tag{12.1.2}$$

for $t \in [0, \infty)$, see Exercise 13.1. This allows us to obtain the GOP volatility $|\theta_t|$ at time $t$ by the volatility formula (5.2.14) as the time derivative of the quadratic variation of $\ln(S_t^{\delta_*})$, that is,

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

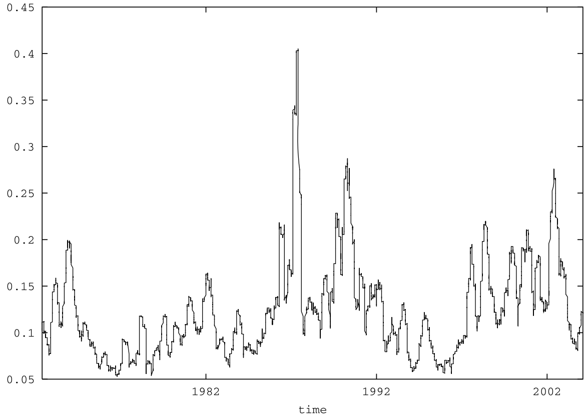



**Fig. 12.1.1.** Estimated volatility of WSI from 1973–2004

$$|\theta_t| = \sqrt{\frac{d}{dt}\left[\ln\left(S^{\delta_*}\right)\right]_t} \tag{12.1.3}$$

for $t \in [0, \infty)$. In Fig. 12.1.1 we plot for the WSI from Fig. 10.6.5, based on daily observations, the volatility which was obtained numerically by using the formula (12.1.3) for the period from 1973 until 2004. One observes that the volatility of this stock index is not a constant or a simple deterministic function of time. Obviously, it is a stochastic process with clusters of higher values. Taking this into account, the BS model is certainly not a perfect description of reality.

### Leverage Effect

To illustrate systematic deviations of an index dynamics from the BS model a study was undertaken by Kelly (1999). Using the standard BS model with constant volatility the P&L, that is the hedge error, was minimized when hedging a European call option with given strike $K$ and given time to maturity $T$, as described in Chap. 8. By using daily data from 1990 until 1998 from the S&P500 and a fixed volatility for the BS model the resulting average hedge errors are shown in Fig. 12.1.2 with dependence on moneyness. These hedge errors express the average P&L when exhausting all possible periods allowed by the data for the hedge analysis. One notes that there is a strong negative skew in the P&L from hedging European calls under the BS model. This indicates that an improved model for such an index needs to account for this stylized empirical feature, which one can also document for other time periods. Note that in this simple experiment no traded option prices from the market were involved.

The negative skew in Fig. 12.1.2 is a reflection of the *leverage effect*, see Black (1976), which expresses a negative correlation between the index and its volatility. When the index value increases the volatility decreases and vice

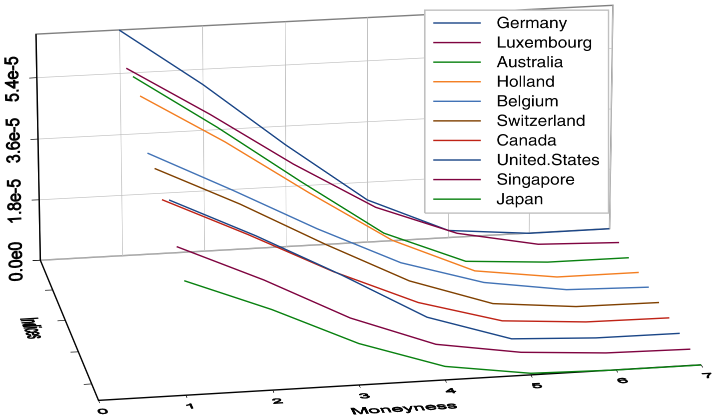



**Fig. 12.1.2.** Estimated hedge error for S&P500 under BS model

versa. To give an economic interpretation of the leverage effect let us interpret the index as a stock market index. If the index is relatively high, then the average market value of companies is rather high and the debt that these companies have, appears to be comparably low. Such a situation corresponds to low risk, which is reflected in low volatility of the index. On the other hand, there is a much higher risk associated with these companies if the market index is relatively low. For a low stock market index level the debt of the companies appears to be relatively high and the volatility, as a measure of risk, is therefore comparably high. This basic economic relationship explains, in principle, the observed negative correlation between the stock market index and its volatility. The leverage effect has been empirically documented in many ways, for instance, it was studied in Black (1976). It is a challenge for an advanced index model to explain and reflect this effect in a consistent and parsimonious manner, in particular, over long periods of time. A crucial step would be to reveal a potential functional dependence between index and volatility.

## Implied Volatilities

Stylized facts on stochastic volatility for traded index options are well documented in the econometrics and finance literature. For example Bollerslev, Chou & Kroner (1992) provide a survey using *autoregressive conditional heteroscedastic* (ARCH) models and Ghysels, Harvey & Renault (1996), Frey (1997) and Cont & Tankov (2004) provide reviews on stochastic volatility models.

In principle, the only parameter in the Black & Scholes (1973) option pricing formula, see (8.3.2), that cannot be directly observed is the volatility. Thus, by using certain given option prices the corresponding *implied volatility* can be obtained by inverting the Black-Scholes formula (8.3.2), as will be explained below.

There exists a liquid market for European call and put options on most stock indices. One can use the observed market prices to detect deviations from the BS model that traders, who survived successfully in the market, have learned to take into account. Let us denote by

$$c_{T,K}(0, S, \sigma, r) = c_{T,K}(0, S) \tag{12.1.4}$$

the European call option price at time $t = 0$ obtained from the Black-Scholes formula (8.3.2) when the volatility is $\sigma > 0$, the short rate $r \in [0, \infty)$, the time to maturity $T \in [0, \infty)$, the actual value of the underlying index $S > 0$ and the strike price equals $K > 0$. From the traded European call option price $V_{c,T,K}(0, S_0)$ with strike price $K$ and time to maturity $T$, which is observed in the market for an index with value $S_0$ at the time $t = 0$, one can deduce the *implied volatility* $\sigma_{\rm BS}^{\rm call}(0, S_0, T, K, r)$ by setting

$$V_{c,T,K}(0, S_0) = c_{T,K}\left(0, S_0, \sigma_{\rm BS}^{\rm call}(0, S_0, T, K, r), r\right). \tag{12.1.5}$$

There is no explicit solution to this equation and one needs to find the implied volatility $\sigma_{\rm BS}^{\rm call}(0, S_0, T, K, r)$ by some root finding method. For instance, the well-known Newton-Raphson iteration method can be used. Similarly, one finds implied volatilities for European puts. One can also price European call or put options according to a given model, for instance the MMM, and calculate the corresponding implied volatilities.

In the market it is often observed that away-from-the-money equity and exchange rate options have higher implied volatilities than at-the-money options. This phenomenon is commonly called the implied volatility *smile*, as for instance discussed in Rubinstein (1985), Clewlow & Xu (1994), Derman & Kani (1994a), Taylor & Xu (1994) or Platen & Schweizer (1998). For indices one observes a *negative skew* in the implied volatilities. This is also consistent with the pattern of hedge errors in Fig. 12.1.2 and is a manifestation of the leverage effect.

For the S&P500 in Fig. 12.1.3 we show implied volatilities for three months to maturity European options for the period from 1997 until 1998 in dependence on the moneyness $\frac{K}{S}$ of the strike over the underlying index value. One notes that the implied volatilities are not the same for different moneyness. Furthermore, one notes that the negatively skewed implied volatility curves evolve over time. In Fig. 12.1.4 we show implied volatilities of one year options, that is, for time to maturity $T = 1$, for the S&P500 during the same period. Note that the curvature of these implied volatility curves is less pronounced than that of the shorter dated options shown in Fig. 12.1.3. This leads to an implied volatility term structure.

More precisely, at a fixed time one can generate an *implied volatility surface* from observed option prices by interpolation over different maturities and strikes. Such a surface reflects the deviations of traded option prices from BS

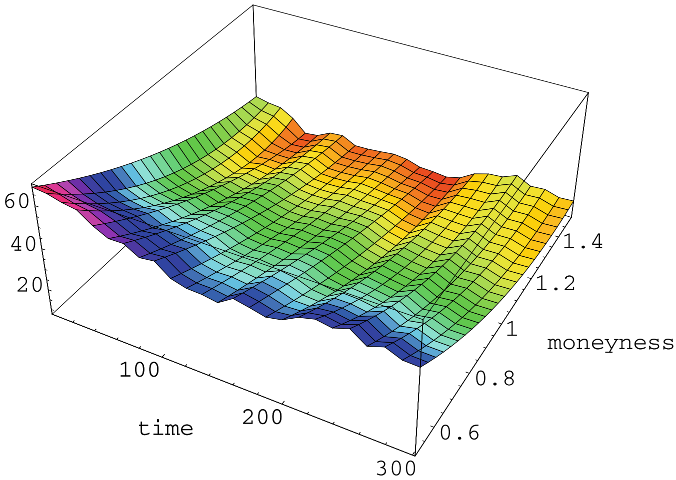



**Fig. 12.1.3.** Implied volatilities for S&P500 three month options

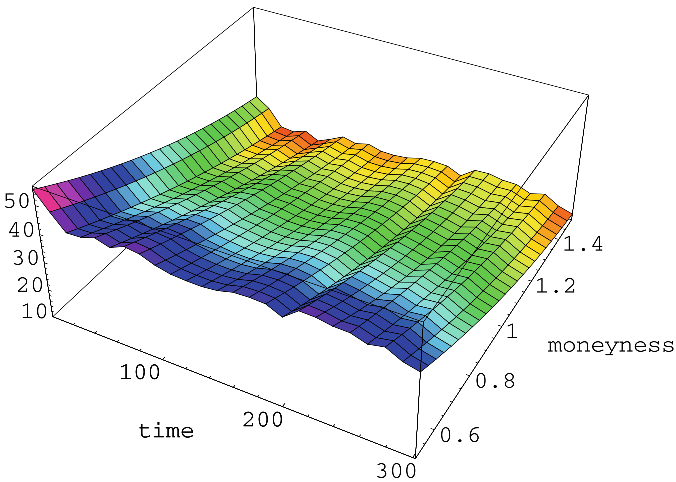



**Fig. 12.1.4.** Implied volatilities for S&P500 one year options

option prices. In Cont & da Fonseca (2002) an average shape of the implied volatility surface for European options on the S&P500 has been extracted. For the one year period from March 2000 until February 2001 we plot in Fig. 12.1.5 a graph that shows approximately the observed average shape of the implied volatility surface. This surface is negatively skewed with less curvature for larger times to maturity. For shorter times to maturity the implied volatility surface is more curved in a convex manner. At the money, the implied volatility shows a slight systematic increase over time. These stylized empirical facts should be explained by an advanced index model. Additionally, it is also observed that the implied volatility with fixed strike, say at-the-money, and fixed maturity, for instance one month, changes randomly over time.

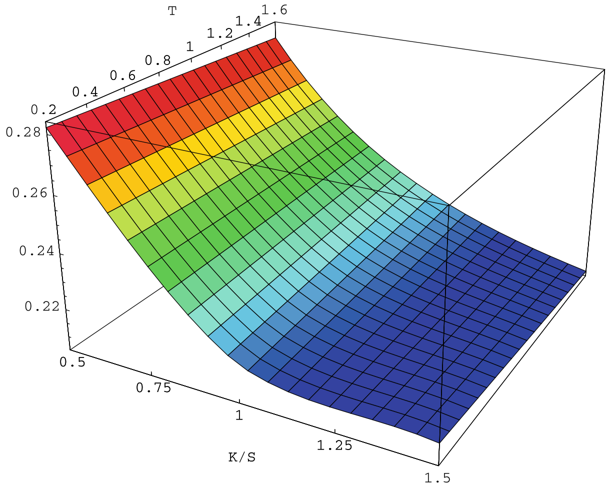



**Fig. 12.1.5.** Average S&P500 implied volatility surface

Consequently, the implied volatility term structure appears to be rather complex. Stylized facts on implied volatility surfaces are, for instance, documented in Dumas, Fleming & Whaley (1998), Schönbucher (1999), Ait-Sahalia & Lo (2000), Cont & da Fonseca (2002), Ledoit, Santa-Clara & Wolf (2003) and Le (2005). It would be highly desirable if an asset price model could also provide an economic interpretation for the volatility dynamics so that the trader can develop a reasoning behind this important market feature. In the following we shall discuss several volatility models which aim to match the type of implied volatility surface, as shown in Fig. 12.1.5.

## 12.2 Modified CEV Model

### CEV Model

We present in this section a modification of the well-known *constant elasticity of variance* (CEV) *model*, as suggested in Heath & Platen (2002a). The CEV model assumes constant elasticity of variance for log-returns. This means that the volatility is a power function. This type of model seems to have first appeared in Cox (1975) and Cox & Ross (1976). It is a natural one-factor extension of the BS model that provides nonconstant stochastic volatilities and, thus, nonconstant implied volatilities. It has been adapted and applied more recently, for instance, in Andersen & Andreasen (2000), Lewis (2000), Lo, Yuen & Hui (2000) and Brigo & Mercurio (2005).

The classical risk neutral approach to the pricing of derivatives under the CEV model is described, for instance, in Beckers (1980) and Schroder (1989). It should be emphasized that these classical formulations typically assume a risk neutral dynamics of the underlying security, which in the case of the CEV model may reach zero with strictly positive probability. This can lead to problems in the pricing of derivatives, see Lewis (2000) and Delbaen & Shirakawa (2002). The modified CEV model that we are going to consider does not have an equivalent risk neutral probability measure, as we shall see, and we apply real world pricing.

What makes the CEV type models attractive is that they easily generate a leverage effect, that is, a negative correlation between the index and its volatility, as discussed in Sect. 12.1.

### Modified CEV Model

We consider a CFM, as introduced in Chap. 10, with one source of uncertainty $W = \{W_t,\, t \in [0,\infty)\}$, modeled by a standard Wiener process under the real world probability measure $P$ on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. The deterministic savings account $S_t^0$ at time $t$ is given by the differential equation

$$dS_t^0 = r\, S_t^0\, dt \tag{12.2.1}$$

for $t \in [0,\infty)$ with $S_0^0 = 1$, where $r$ denotes the constant short rate. By introducing the total market price of risk process $|\theta| = \{|\theta_t|,\, t \in [0,\infty)\}$, the GOP $S_t^{\delta_*}$ satisfies the SDE (12.1.1). Recall that the total market price of risk $|\theta_t|$ appears as the volatility at time $t$ of the GOP. By introducing the drifted Wiener process $W_\theta = \{W_\theta(t),\, t \in [0,\infty)\}$ with

$$dW_\theta(t) = |\theta_t|\, dt + dW_t \tag{12.2.2}$$

we obtain from (12.1.1) for the GOP the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*}\,(r\, dt + |\theta_t|\, dW_\theta(t)) \tag{12.2.3}$$

for $t \in [0,\infty)$.

To illustrate the kind of problem that may arise if there is no risk neutral probability measure under the classical formulation of the *modified CEV* model, we consider the case where the GOP volatility $|\theta_t|$ is specified in the form

$$|\theta_t| = (S_t^{\delta_*})^{a-1}\, \psi \tag{12.2.4}$$

for $t \in [0,\infty)$ with *exponent* $a \in (-\infty,\infty)$ and *scaling parameter* $\psi > 0$. In this case the appreciation rate of the GOP, see (12.1.1), is stochastic as long as $a \neq 1$.

We then have for the GOP by (12.2.3) and (12.2.4) the dynamics

$$dS_t^{\delta_*} = S_t^{\delta_*}\, r\, dt + (S_t^{\delta_*})^a\, \psi\, dW_\theta(t) \tag{12.2.5}$$

for $t \in [0,\infty)$. The existence and uniqueness of a solution of the SDE (12.2.5) is, in general, not automatically guaranteed without extra conditions on the behavior of the process $S^{\delta_*}$ at zero, see Sect. 7.7 and Karatzas & Shreve (1991).

In (12.2.5), the process $W_\theta = \{W_\theta(t),\, t \in [0,\infty)\}$ is usually interpreted in the literature as a Wiener process under a risk neutral probability measure $P_\theta$. Let us follow this interpretation for the moment. However, it will be shown that for $a < 1$ there is a major problem with the application of the risk neutral methodology. The scaling parameter $\psi$ is, for simplicity, assumed to be constant.

Note that for the case $a = 1$ we have the BS model, which has a risk neutral probability measure $P_\psi$ and constant market price of risk $\psi$. As we shall see shortly, this is the only case where considering a risk neutral version of the model makes sense.

By (12.1.1) and (12.2.4) the GOP satisfies the SDE

$$dS_t^{\delta_*} = \left(S_t^{\delta_*}\, r + (S_t^{\delta_*})^{2a-1}\, \psi^2\right) dt + (S_t^{\delta_*})^a\, \psi\, dW_t \tag{12.2.6}$$

for $t \in [0,\infty)$. The above choice (12.2.4) of the market price of risk contrasts with what is used in the classical formulation of the CEV model, see Cox & Ross (1976) and Schroder (1989). For this reason we refer to (12.2.6) as *modified CEV model*, which has been studied in Heath & Platen (2002a). As we shall see, the real world dynamics of the GOP, governed by (12.2.6), remains for $a < 1$ strictly positive. This is not the case for its hypothetical risk neutral dynamics when $W_\theta$ is interpreted as a Wiener process under $P_\theta$, because $S^{\delta_*}$ may be absorbed at zero with strictly positive $P_\theta$ probability, as will become clear below.

### Squared Bessel Process

We shall now show that the modified CEV model is closely related to squared Bessel processes, see Sect. 8.7. By application of the Itô formula we obtain from (12.2.6) for the quantity

$$X_t = \left(S_t^{\delta_*}\right)^{2\,(1-a)} \tag{12.2.7}$$

the SDE

$$dX_t = \big(2(1-a)\, r\, X_t + \psi^2\, (1-a)\, (3-2a)\big)\, dt + 2\, \psi\, (1-a)\, \sqrt{X_t}\, dW_t \tag{12.2.8}$$

for $t \in [0,\infty)$ with $X_0 = (S_0^{\delta_*})^{2\,(1-a)} > 0$. It follows from Sect. 8.7 that $X = \{X_t,\, t \in [0,\infty)\}$ is a time transformed, squared Bessel process of dimension

$$\delta = \frac{3 - 2\,a}{1 - a} \tag{12.2.9}$$

for $a \neq 1$. Note that the SDE (12.2.8) has for $a \neq 1$ a nonnegative, unique strong solution, see Sect. 7.7, which for $a > 1$ we assume remains at zero when it reaches zero. By (12.2.7) the GOP can be expressed in the form

$$S_t^{\delta_*} = (X_t)^q, \tag{12.2.10}$$

where

$$q = \frac{1}{2\,(1-a)} \tag{12.2.11}$$

for $t \in [0, \infty)$ and $a \neq 1$. One notes that for extremely small $a < 1$ the dimension $\delta$ of the squared Bessel process $X$ equals approximately two, which yields strongly leptokurtic log-returns for the GOP. However, for the exponent $a$ when approaching one from below, the dimension $\delta$ tends to infinity, which yields lognormal dynamics for the GOP and, thus, Gaussian log-returns.

### Hypothetical Risk Neutral Measure Transformation

In this setting the candidate Radon-Nikodym derivative process $\Lambda_\theta = \{\Lambda_\theta(t), t \in [0, \infty)\}$, which determines the *hypothetical risk neutral measure* $P_\theta$ for the pricing of options with maturity $T$ with

$$\left.\frac{dP_\theta}{dP}\right|_{\mathcal{A}_T} = \Lambda_\theta(T), \tag{12.2.12}$$

is given by

$$\Lambda_\theta(t) = \frac{S_0^{\delta_*}}{S_t^{\delta_*}}\, S_t^0 \tag{12.2.13}$$

for $t \in [0, \infty)$, see (9.4.5). This means that by (12.2.10), the candidate Radon-Nikodym derivative equals the power of a time transformed, squared Bessel process of dimension $\delta$, that is

$$\Lambda_\theta(t) = S_t^0 \left(\frac{X_0}{X_t}\right)^q \tag{12.2.14}$$

for $t \in [0, \infty)$, where $q$ is given in (12.2.11).

Using (12.2.2) and (12.2.4) one can now rewrite the SDE (12.2.8) with respect to the drifted Wiener process $W_\theta$, see (12.2.2), in the form

$$dX_t = \big(2\,(1-a)\,r\,X_t + \psi^2\,(1-a)(1-2\,a)\big)\,dt + 2\,\psi\,(1-a)\,\sqrt{X_t}\,dW_\theta(t) \tag{12.2.15}$$

for $t \in [0, \infty)$. Consequently, if one interprets the process $X$ as a time transformed, squared Bessel process under a hypothetical risk neutral probability measure $P_\theta$, then it would have the dimension

$$\delta_\theta = \frac{1-2\,a}{1-a} \tag{12.2.16}$$

for $a \neq 1$. In Fig. 12.2.1 we show the dimensions $\delta$ and $\delta_\theta$, see (12.2.9) and (12.2.16), of the above discussed time transformed, squared Bessel processes

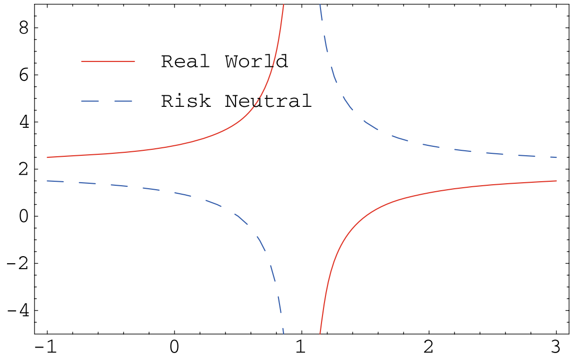



**Fig. 12.2.1.** Dimensions $\delta$ and $\delta_\theta$ as a function of the exponent $a$

as a function of the exponent $a$, see Heath & Platen (2002b) and also Lewis (2000).

Note from Fig. 12.2.1 that if $a < 1$ then $\delta > 2$ and $\delta_\theta < 2$, and if $a > 1$ then $\delta < 2$ and $\delta_\theta > 2$. These inequalities follow from (12.2.9) and (12.2.16). It is known, see (8.7.7), that a time transformed, squared Bessel process with a dimension greater than two remains strictly positive. However, a time transformed, squared Bessel process with a dimension less than two hits zero with some strictly positive probability, see (8.7.8).

## Hypothetical Risk Neutral Measure

In most of the previously mentioned literature one performs the modeling under a hypothetical risk neutral probability measure $P_\theta$. This allows one to express the hypothetical risk neutral probability $P_\theta(A)$ for an event $A$ in the form

$$P_\theta(A) = \int_A dP_\theta(\omega) = \int_A \frac{dP_\theta(\omega)}{dP(\omega)}\, dP(w) = \int_A \Lambda_\theta(T)\, dP(\omega), \tag{12.2.17}$$

where $\Lambda_\theta(T)$ is the candidate Radon-Nikodym derivative described in (12.2.14) at time $T \in (0, \infty)$. Then one can ask for the total risk neutral measure. By (12.2.17) one obtains

$$P_\theta(\Omega) = \int_\Omega \Lambda_\theta(T)\, dP(\omega) = E\left(\Lambda_\theta(T) \,\middle|\, \mathcal{A}_0\right). \tag{12.2.18}$$

If $\Lambda_\theta$ were an $(\underline{\mathcal{A}}, P)$-martingale, then $P_\theta(\Omega)$ would equal $\Lambda_\theta(0) = 1$ and $P_\theta$ would be a probability measure. However, for $a < 1$ it follows from (12.2.14) and Sect. 8.7 that $\Lambda_\theta$ is an $(\underline{\mathcal{A}}, P)$-strict local martingale, see (8.7.25), and, thus, by Lemma 5.2.3 an $(\underline{\mathcal{A}}, P)$-strict supermartingale. Consequently, we have for the total hypothetical risk neutral measure $P_\theta(\Omega) < 1$. This means that $P_\theta$ is for the given modified CEV model *not* a probability measure. Consequently, for this model there does not exist an equivalent risk neutral probability measure $P_\theta$ for $a < 1$.

In addition, for $a > 1$ the dimension $\delta$ is less than two and the exponent $q$ appearing in (12.2.10) is negative. It therefore follows from (8.7.8) that the GOP explodes for this parameter choice at some time with strictly positive $P$-probability. Consequently, the choice $a > 1$ does not lead to a viable model for the GOP. For this reason, we consider only the case $a < 1$ in the remainder of this section.

### Real World Pricing

As we have seen previously, with its real world pricing concept the benchmark approach provides a consistent pricing framework without requiring the existence of an equivalent risk neutral probability measure. For $T \in [0, \infty)$ let $H = H(S_T^{\delta_*})$ denote a nonnegative payoff with

$$E\left(\frac{H(S_T^{\delta_*})}{S_T^{\delta_*}}\right) < \infty. \tag{12.2.19}$$

Recall from Sect. 9.1 that a price process is fair if, when expressed in units of the GOP, it is an $(\underline{\mathcal{A}}, P)$-martingale. Then the fair, benchmarked price $\hat{U}_H(t)$ at time $t$ of this payoff is given by the conditional expectation

$$\hat{U}_H(t) = E\left(\frac{H(S_T^{\delta_*})}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) \tag{12.2.20}$$

for $t \in [0, T]$, see Definition 9.1.2. By using the transition density (8.7.9) of a squared Bessel process $X$ of dimension $\delta = \frac{3-2a}{1-a}$ one can for some given payoff $H(S_T^{\delta_*})$ explicitly calculate the benchmarked price $\hat{U}_H(t)$ for any $t \in [0, T]$.

The fair price $U_H(t)$ of the payoff $H$, when expressed in units of the domestic currency, is then obtained by the real world pricing formula

$$U_H(t) = S_t^{\delta_*}\, \hat{U}_H(t) \tag{12.2.21}$$

for $t \in [0, T]$, see (9.1.31) or (10.4.1).

### PDE for Benchmarked Pricing Function

As an alternative to the use of the transition density of the squared Bessel process $X$ one can exploit the Markovianity of the GOP $S^{\delta_*}$. This permits the application of the Feynman-Kac formula (9.7.3)–(9.7.4) to obtain the benchmarked price $\hat{U}_H(t) = \hat{u}_H(t, S_t^{\delta_*})$ as a function $\hat{u}_H : [0, T] \times [0, \infty) \to [0, \infty)$ of the time $t$ and the value $S_t^{\delta_*}$ of the GOP. To formulate this method of calculation we define the operator $L^0$ on a sufficiently smooth function $f : [0, T) \times (0, \infty) \to \Re$ by

$$L^0\, f(t,S) = \frac{\partial f(t,S)}{\partial t} + \left(r\, S + \psi^2\, S^{2a-1}\right) \frac{\partial f(t,S)}{\partial S} + \frac{1}{2}\, \psi^2\, S^{2a}\, \frac{\partial^2 f(t,S)}{\partial S^2} \tag{12.2.22}$$

for $(t,S) \in (0,T) \times (0,\infty)$.

Applying the above operator $L^0$ to the benchmarked pricing function $\hat{u}_H(\cdot,\cdot)$ by using (12.2.20), (12.2.6) and the Feynman-Kac formula, yields the PDE

$$L^0\, \hat{u}_H(t,S) = 0 \tag{12.2.23}$$

for $(t,S) \in (0,T) \times (0,\infty)$ with the terminal condition

$$\hat{u}_H(T,S) = \frac{H(S)}{S} \tag{12.2.24}$$

for $S \in (0,\infty)$. It remains to solve the PDE (12.2.23)–(12.2.24), which, for instance, can be achieved by a finite difference method, as will be described in Sect. 15.7.

## Martingale Representation

From the Feynman-Kac formula it follows that the benchmarked pricing function $\hat{u}_H : [0,T] \times [0,\infty) \to [0,\infty)$ is differentiable with respect to time $t$ and twice differentiable with respect to $S_t^{\delta_*}$ on $(0,T) \times (0,\infty)$. Consequently, by application of the Itô formula, using (12.2.6), we obtain the representation

$$\hat{u}_H(t, S_t^{\delta_*}) = \hat{u}_H(0, S_0^{\delta_*}) + \int_0^t (S_s^{\delta_*})^a\, \psi\, \frac{\partial \hat{u}_H(s, S_s^{\delta_*})}{\partial S^{\delta_*}}\, dW_s \tag{12.2.25}$$

for $t \in [0,T]$. This is the martingale representation of the benchmarked price, as discussed in Sect. 11.4. By Theorem 11.5.2 this representation provides the information about the hedge that enables one to replicate the payoff.

## Hedge Portfolio

Obviously, by (12.2.1) and (12.2.6), the benchmarked savings account $\hat{S}_t^0$ satisfies by the Itô formula the SDE

$$d\hat{S}_t^0 = -S_t^0\, \left(S_t^{\delta_*}\right)^{a-2}\, \psi\, dW_t \tag{12.2.26}$$

for $t \in [0,\infty)$. Trivially, we have $d\hat{S}_t^{\delta_*} = 0$ for $t \in [0,\infty)$.

For a given payoff function $H(S_T^{\delta_*})$ we can construct a hedge portfolio consisting of $\delta_H^0(t)$ units of the savings account $S_t^0$ and $\delta_H^1(t)$ units of the GOP $S_t^{\delta_*}$. As one can see from (12.2.25) and (12.2.26), to construct a replicating portfolio we need to choose the hedge ratios according to the prescription

$$\delta_H^0(t) = -\frac{(S_t^{\delta_*})^2}{\hat{S}_t^0}\, \frac{\partial \hat{u}_H(t, S_t^{\delta_*})}{\partial S^{\delta_*}} \tag{12.2.27}$$

and

$$\delta^1_H(t) = \hat{u}_H(t, S^{\delta_*}_t) - \delta^0_H(t)\, \hat{S}^0_t \tag{12.2.28}$$

for $t \in [0,T]$. This choice ensures that the value of the hedge portfolio, when measured in units of the domestic currency, equals the fair price $U_H(t) = u_H(t, S^{\delta_*}_t)$ at time $t \in [0,T]$. That is

$$u_H(t, S^{\delta_*}_t) = \delta^0_H(t)\, S^0_t + \delta^1_H(t)\, S^{\delta_*}_t \tag{12.2.29}$$

for $t \in [0,T]$. The benchmarked value of the hedge portfolio is, therefore, given by

$$\hat{u}_H(t, S^{\delta_*}_t) = \delta^0_H(t)\, \hat{S}^0_t + \delta^1_H(t) \tag{12.2.30}$$

for $t \in [0,T]$. By (12.2.20) this hedge portfolio replicates the payoff at the maturity date $T$. Since it is a fair portfolio it provides by Corollary 10.4.2 the minimal hedge.

### Benchmarked P&L

To illustrate the replication of the payoff, we define the *benchmarked* P&L $\hat{C}_H(t)$ for maintaining this hedge portfolio up to time $t \in [0,T]$. Similarly to (8.2.13) it equals the benchmarked value of the hedge portfolio minus the benchmarked gains from trade and the benchmarked initial value, that is,

$$\hat{C}_H(t) = \hat{u}_H(t, S^{\delta_*}_t) - \int_0^t \delta^0_H(s)\, d\hat{S}^0_s - \hat{u}_H(0, S^{\delta_*}_0) \tag{12.2.31}$$

for $t \in [0,T]$. By combining (12.2.31), (12.2.27), (12.2.25) and (12.2.22) we see that

$$\hat{C}_H(t) = 0 \tag{12.2.32}$$

for all $t \in [0,T]$. This means that the benchmarked P&L for maintaining the hedge portfolio is always zero. Consequently, the P&L

$$C_H(t) = \hat{C}_H(t)\, S^{\delta_*}_t = 0 \tag{12.2.33}$$

equals zero for all times $t \in [0,T]$. Thus, the payoff $H(S^{\delta_*}_T)$ can be perfectly hedged using the real world pricing formula (12.2.21) together with the hedging prescriptions (12.2.27) and (12.2.30).

### Hedge Ratio

By using (12.2.27), (12.2.30) and (12.2.25) the hedge ratio $\delta^1_H(t)$ can be rewritten in the form

$$\delta^1_H(t) = \hat{u}_H(t, S^{\delta_*}_t) + S^{\delta_*}_t \frac{\partial \hat{u}_H(t, S^{\delta_*}_t)}{\partial S^{\delta_*}} = \frac{\partial u_H(t, S^{\delta_*}_t)}{\partial S^{\delta_*}} \tag{12.2.34}$$

for $t \in [0,T]$. Therefore, the number of units $\delta^1_H(t)$ held in the GOP at time $t \in [0,T]$ equals the delta hedge ratio obtained by calculating the partial derivative of the fair price with respect to the value of the underlying, that is the GOP. This is entirely analogous to what we obtained in Chap. 8 under the BS model. It is also analogous to what one obtains in a classical risk neutral hedging framework when using a savings account for hedging, see, for instance, Karatzas & Shreve (1998). Note, however, that the above benchmark methodology still works when no equivalent risk neutral probability measure exists, as is the case for the given modified CEV model.

## Fair Zero Coupon Bond

The price $P_T(t, S^{\delta_*}_t)$ at time $t \in [0,T]$ for the fair zero coupon bond that pays one unit of the domestic currency at maturity $T$ is given as

$$P_T(t, S^{\delta_*}_t) = S^{\delta_*}_t \, \hat{P}_T(t, S^{\delta_*}_t) \tag{12.2.35}$$

for $t \in [0,T]$, see (9.1.34) and (10.4.1), with

$$\hat{P}_T(t, S^{\delta_*}_t) = E\left( \frac{1}{S^{\delta_*}_T} \,\Big|\, \mathcal{A}_t \right). \tag{12.2.36}$$

Note by (8.7.16) that the conditional expectation in (12.2.36) can be calculated explicitly. The following explicit bond pricing formula for the modified CEV model has been established in Miller & Platen (2008). By using the transition density (8.7.9) of a squared Bessel process of dimension $\delta > 2$ one obtains

$$P_T(t, S^{\delta_*}_t) = E\left( \frac{S^{\delta_*}_t}{S^{\delta_*}_T} \,\Big|\, \mathcal{A}_t \right) = \exp\{-r\,(T-t)\}\,\chi^2(\ell^*; \delta - 2) \tag{12.2.37}$$

for $t \in [0,T]$. Here $\chi^2(\cdot;\delta)$ is the central chi-square distribution function with $\delta$ degrees of freedom, see (1.2.11), where

$$\ell^* = \frac{2\,r\,(S^{\delta_*}_t)^{2(1-a)}}{\psi^2\,(1-a)\,(1-\exp\{-2(1-a)\,r\,(T-t)\})}. \tag{12.2.38}$$

It is clear that

$$P_T(t, S^{\delta_*}_t) < \exp\{-r\,(T-t)\} \tag{12.2.39}$$

for all $t \in [0,T)$ and $S^{\delta_*}_t > 0$, because $\chi^2(\ell^*; \delta - 2)$ is the value of a chi-square distribution.

We remark that the function $\hat{P}_T(\cdot,\cdot)$ of the benchmarked fair zero coupon bond price satisfies the PDE

$$L^0\,\hat{P}_T(t,S) = 0 \tag{12.2.40}$$

for $(t, S) \in [0, T) \times (0, \infty)$ with terminal condition

$$\hat{P}_T(T, S) = \frac{1}{S} \tag{12.2.41}$$

for $S \in (0, \infty)$, see (12.2.23)–(12.2.24), where the operator $L^0$ is given in (12.2.22).

### Savings Bond

The price process $u_H = \{u_H(t, S_t^{\delta_*}),\, t \in [0, \infty)\}$ in (12.2.28) is the only fair portfolio process, which perfectly replicates the payoff. However, as we shall see below, other nonnegative portfolio processes exist, which also perfectly replicate the payoff. As shown in Theorem 10.3.1, any benchmarked nonnegative portfolio process forms an $(\underline{\mathcal{A}}, P)$-supermartingale. From this it followed in Corollary 10.4.2 that the fair price, as obtained according to (12.2.29), yields the minimal price that permits perfect replication of the payoff. This will be verified for our case below.

Since we have assumed a constant short rate $r_t = r$ it is easy to introduce an artificial *savings bond* $P_T^* = \{P_T^*(t),\, t \in [0, \infty)\}$ with

$$P_T^*(t) = \frac{S_t^0}{S_T^0} = \exp\{-r\,(T - t)\} \tag{12.2.42}$$

for $t \in [0, T]$. By application of the Itô formula, it can be shown by (12.2.10), (12.2.8) and (12.2.42), that the benchmarked savings bond price process $\hat{P}_T^* = \{\hat{P}_T^*(t, S_t^{\delta_*}),\; t \in [0, T]\}$ with

$$\hat{P}_T^*(t, S_t^{\delta_*}) = \frac{P_T^*(t)}{S_t^{\delta_*}} = X_t^{-q}\, P_T^*(t) \tag{12.2.43}$$

satisfies the SDE

$$d\hat{P}_T^*(t, S_t^{\delta_*}) = -\hat{P}_T^*(t, S_t^{\delta_*})\, \frac{\psi}{\sqrt{X_t}}\, dW_t \tag{12.2.44}$$

for $t \in [0, T]$. Therefore, $\hat{P}_T^*$ forms an $(\underline{\mathcal{A}}, P)$-local martingale. Since a nonnegative, local martingale is an $(\underline{\mathcal{A}}, P)$-supermartingale, see Corollary (5.2.2), it follows that

$$\hat{P}_T^*(t, S_t^{\delta_*}) \geq E\left(\frac{1}{S_T^{\delta_*}}\,\Big|\,\mathcal{A}_t\right)$$

for $t \in [0, T]$. Therefore, by (12.2.35) and (12.2.43) we can deduce the inequality

$$P_T(t, S_t^{\delta_*}) = S_t^{\delta_*}\, E\left(\frac{1}{S_T^{\delta_*}}\,\Big|\,\mathcal{A}_t\right) \leq S_t^{\delta_*}\, \hat{P}_T^*(t, S_t^{\delta_*}) = P_T^*(t) \tag{12.2.45}$$

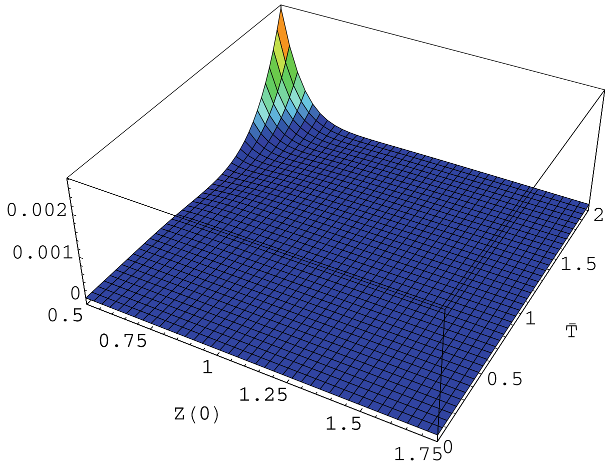



**Fig. 12.2.2.** Difference between savings and fair bond

for $t \in [0, T]$. This confirms our observation in (12.2.39) that the savings bond is at least as expensive as the fair zero coupon bond. In Fig. 12.2.2 we show the difference between the savings bond and the fair zero coupon bond for $a = -0.5$, $\psi = 0.2$ and $r = 0.04$. The savings bond is an unfair price process because when benchmarked it forms a strict supermartingale, see Exercise 12.6. However, it does not constitute an arbitrage in the sense of Definition 10.3.2.

## Free Snack from Savings Bond

The above relation (12.2.45) poses an obvious question about the potential existence of arbitrage. As shown in (12.2.27)–(12.2.28), there exists a trading strategy which hedges the fair zero coupon bond under consideration. One may now form a trading strategy $\delta$ consisting of the aforementioned hedge, which is funded by borrowing the amount $P_T(0, S_0^{\delta_*})$ from a savings account at initiation. The portfolio value $S_t^\delta$ at time $t \in [0, T]$ is then given by the expression

$$S_t^\delta = P_T(t, S_t^{\delta_*}) - P_T(0, S_0^{\delta_*}) \exp\{-r\, t\}. \tag{12.2.46}$$

We observe that $S_0^\delta = 0$ and

$$S_T^\delta = 1 - P_T(0, S_0^{\delta_*}) \exp\{-r\, T\} > 0, \tag{12.2.47}$$

as well as

$$S_t^\delta \geq -P_T(0, S_0^{\delta_*}) \exp\{-r\, t\} \tag{12.2.48}$$

almost surely for all $t \in [0, T]$. Thus, $\delta$ is a strategy with a wealth process that is uniformly bounded from below. Since $S_t^\delta$ may become negative this portfolio is not covered by our arbitrage concept given in Definition 10.3.2. However, it is covered by the concept of free lunch with vanishing risk, see

Delbaen & Schachermayer (2006). Since we have in the given case a free lunch with vanishing risk it follows by the fundamental theorem of asset pricing of Delbaen & Schachermayer (1998) that the modified CEV model does not admit an equivalent risk neutral probability measure. This confirms what we observed already when we studied the strict supermartingale property of the candidate Radon-Nikodym derivative for the hypothetical risk neutral probability measure.

In Loewenstein & Willard (2000) a portfolio of the above kind is called a *free snack*. As we have seen, it rules out the existence of an equivalent risk neutral probability measure. However, it does not constitute an economic reason for dismissing the given model.

### Benchmarked Savings Bond

Note that the pricing function of the *benchmarked savings bond* $\hat{P}^*_T(\cdot,\cdot)$ satisfies the PDE (12.2.23) with

$$L^0\, \hat{P}^*_T(t,S) = 0 \tag{12.2.49}$$

for $(t,S) \in [0,T) \times (0,\infty)$ and terminal condition

$$\hat{P}^*_T(T,S) = \frac{1}{S} \tag{12.2.50}$$

for $S \in (0,\infty)$. The PDE (12.2.40) with terminal condition (12.2.41) is the same as the one given in (12.2.49) and (12.2.50). Therefore, there is more than one solution to the PDE problem (12.2.49)–(12.2.50). This is related to the fact that the solution to this PDE is not fully determined without specification of its behavior along the spatial boundary at zero. From the absence of arbitrage in the sense of Definition 10.3.2 it follows from (10.3.4) that any nonnegative portfolio that reaches zero remains at zero after that time. For this reason the spatial boundary condition where $S$ reaches zero must be that of absorption.

The above savings bond provides a perfect hedge via a self-financing portfolio that replicates one monetary unit at maturity $T$. Note however that this is not the minimal possible hedge portfolio. The fair zero coupon bond portfolio, given by the price (12.2.35), provides the minimal hedge since its benchmarked value forms a martingale while the benchmarked savings bond is a strict supermartingale.

### European Call Option

For a *European call option* on the GOP with strike $K$ and maturity $T$ the benchmarked fair price $\hat{c}_{T,K}(t, S^{\delta_*}_t)$ at time $t$ is given by the formula

$$\hat{c}_{T,K}(t, S^{\delta_*}_t) = E\left(\frac{(S^{\delta_*}_T - K)^+}{S^{\delta_*}_T} \,\bigg|\, \mathcal{A}_t\right) = E\left(\left(1 - \frac{K}{S^{\delta_*}_T}\right)^+ \,\bigg|\, \mathcal{A}_t\right) \tag{12.2.51}$$

for $t \in [0, T]$. Note that the conditional expectation used in (12.2.51) is finite because the payoff $(1 - \frac{K}{S_T^{\delta_*}})^+$ is bounded. Thus, the inequality (12.2.19) for the European call payoff is satisfied. The corresponding fair price $c_{T,K}(t, S_t^{\delta_*})$, see (12.2.21), takes the form by the real world pricing formula (9.1.34) and (10.4.1)

$$c_{T,K}(t, S_t^{\delta_*}) = S_t^{\delta_*} \, \hat{c}_{T,K}(t, S_t^{\delta_*}) \tag{12.2.52}$$

for $t \in [0, T]$.

By (12.2.23) the function $\hat{c}_{T,K}(\cdot, \cdot)$ satisfies the PDE (12.2.23) with terminal condition

$$\hat{c}_{T,K}(T, S) = \hat{H}(S) = \left(1 - \frac{K}{S}\right)^+ \tag{12.2.53}$$

for $S \in (0, \infty)$, which can be solved numerically.

Alternatively, one can calculate the benchmarked European call price by exploiting the known transition density of the squared Bessel process $X$. This yields by (8.7.9), see Miller & Platen (2008), the explicit expression

$$\hat{c}_{T,K}(t, S_t^{\delta_*}) = (1 - \chi^2(u^*; \delta, \ell^*)) - \frac{K}{S_t^{\delta_*}} \exp\{-r\,(T-t)\}\, \chi^2(\ell^*; \delta - 2, u^*), \tag{12.2.54}$$

where

$$u^* = \frac{2\, r\, K^{2(1-a)}}{\psi^2\,(1-a)\,(\exp\{2(1-a)\, r\,(T-t)\} - 1)} \tag{12.2.55}$$

and $\ell^*$ is as in (12.2.38) for $t \in (0, T]$ and $\chi^2(\cdot; \delta, \cdot)$ is the non-central chi-square distribution (1.2.13) with degrees of freedom $\delta$. Now, when using the previous notation we obtain the explicit European call pricing formula

$$c_{T,K}(t, S_t^{\delta_*}) = S_t^{\delta_*} \left(1 - \chi^2(u^*; \delta, \ell^*)\right) - K \exp\{-r\,(T-t)\}\, \chi^2(\ell^*; \delta - 2, u^*) \tag{12.2.56}$$

for the modified CEV model, see Miller & Platen (2008). This explicit pricing formula is equivalent to similar CEV call option pricing formulas that one can find in Cox & Ross (1976), Beckers (1980), Schroder (1989), Cox (1996), Shaw (1998) and Delbaen & Shirakawa (2002). The important difference, however, is that an equivalent risk neutral probability measure does not exist for the modified CEV model. We shall discuss this issue further below.

According to (12.1.5) one can visualize a European call price efficiently by its implied volatility. For an exponent $a = -0.5$, that is with dimension $\delta \approx 2.67$, $\psi = 0.2$, a constant interest rate $r = 0.04$ and maturity dates of up to two years, Fig. 12.2.3 displays the corresponding implied volatility surface that results from the fair call option price using different values of the strike $K$ and time $t$ for fixed value of $S_t^{\delta_*} = 1$. In Fig. 12.2.3 we see negatively skewed implied volatilities. Note that we use here the fair zero coupon bond as discount factor for the inversion of the Black-Scholes formula when calculating implied volatilities. More precisely, we use the substitute short rate for a European call option with maturity $T$

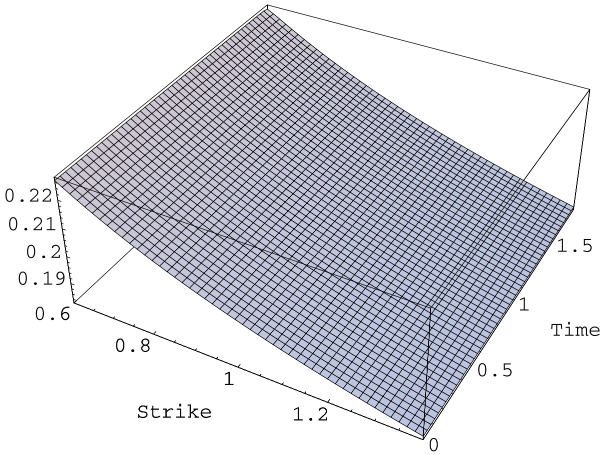



**Fig. 12.2.3.** Implied volatilities for fair European call prices

$$\hat{r} = -\frac{1}{T-t} \ln(P_T(t, S_t^{\delta_*})) = \frac{1}{T-t} \int_t^T f(t,s)ds \tag{12.2.57}$$

when calculating implied volatilities. Here $f(t,s)$ denotes the forward rate at time $t$ for the maturity $s$, see (10.4.12). We emphasize that it is important to make the above adjustment. Otherwise, implied put and call volatilities do not match.

## European Put Option

Similarly, one can also compute the fair European put option price

$$p_{T,K}(t, S_t^{\delta_*}) = S_t^{\delta_*} \, E\left(\left(\frac{K}{S_T^{\delta_*}} - 1\right)^+ \Bigg| \mathcal{A}_t\right) \tag{12.2.58}$$

for $t \in [0,T]$, which has by application of the transition density (8.7.9) the explicit form

$$p_{T,K}(t, S_t^{\delta_*}) = -S_t^{\delta_*} \chi^2(u^*; \delta, \ell^*) + K \exp\{-r\,(T-t)\} \\ \times \left(\chi^2(\ell^*; \delta-2) - \chi^2(\ell^*; \delta-2, u^*)\right) \tag{12.2.59}$$

for $t \in [0,T]$ with the notation (12.2.55)–(12.2.38). This explicit, fair European put pricing formula for the modified CEV model, see Miller & Platen (2008), is clearly different from the type of put pricing formulas that one would obtain from Cox & Ross (1976), Beckers (1980), Schroder (1989), Cox (1996) or Shaw (1998). The reason is that these authors priced a CEV model under the assumption that it has an equivalent risk neutral probability measure. Their benchmarked put prices are strict supermartingales. The modified

CEV model does not have an equivalent risk neutral probability measure and its benchmarked fair put prices are martingales.

In Lewis (2000) some rules are proposed that aim to account for the differences that arise when constructing some hypothetical risk neutral prices in models like the CEV model. Unfortunately, this approach appears to lead to conceptual problems when going beyond standard put and call options. The real world pricing concept of the benchmark approach also applies to the pricing under any reasonable model that has a GOP.

### Fair Put-Call Parity

By using the corresponding fair zero coupon bond price with maturity $T$ the *fair put-call parity* is satisfied, that is, the following relation holds

$$p_{T,K}(t, S_t^{\delta_*}) = c_{T,K}(t, S_t^{\delta_*}) - S_t^{\delta_*} + K\, P_T(t, S_t^{\delta_*}) \tag{12.2.60}$$

for $t \in [0, T]$. However, by (12.2.39) we have

$$p_{T,K}(t, S_t^{\delta_*}) < c_{T,K}(t, S_t^{\delta_*}) - S_t^{\delta_*} + K\, \exp\{-r\,(T-t)\} \tag{12.2.61}$$

for $t \in [0, T)$ and $S_t^{\delta_*} > 0$. This means, when using the savings bond instead of the fair bond in (12.2.60), put-call parity does not hold. Note that this effect arises here even in a model with constant interest rates.

As already indicated, since we use the fair zero coupon bond price as the discount factor for the computation of implied volatilities from the Black-Scholes formula, the implied volatilities of fair puts equal those of corresponding fair calls. However, if one would use the savings bond in such calculations as discount factor, then differences between the implied volatilities for puts and calls would emerge.

### Comparison to Hypothetical Risk Neutral Prices

Let us now compare the above results with those that one would obtain under formal application of the standard risk neutral pricing methodology. This means we are for a moment neglecting the fact that there does not exist an equivalent risk neutral probability measure for the given modified CEV model.

We define the hypothetical risk neutral price $c^*_{T,K}(t, S_t^{\delta_*})$ at time $t$ of a European call option on the GOP with strike $K$ and maturity $T$ by $c^*_{T,K}(t, S_t^{\delta_*})$ for $t \in [0, T]$. The benchmarked hypothetical risk neutral call price

$$\hat{c}^*_{T,K}(t, S_t^{\delta_*}) = \frac{c^*_{T,K}(t, S_t^{\delta_*})}{S_t^{\delta_*}} \tag{12.2.62}$$

forms an $(\underline{\mathcal{A}}, P)$-local martingale, as all benchmarked portfolio processes in a CFM. One notes the important fact that its benchmarked payoff is bounded.

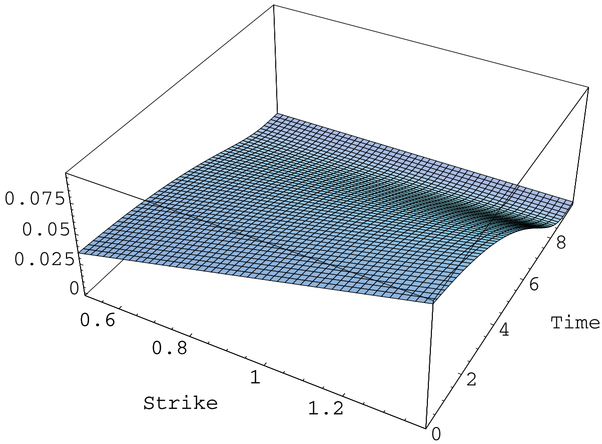



**Fig. 12.2.4.** Difference between hypothetical risk neutral and fair put prices

Therefore, the benchmarked hypothetical risk neutral price $\hat{c}^*_{T,K}(\cdot,\cdot)$ is uniformly bounded. By Lemma 5.2.2 (ii) it follows that bounded local martingales are martingales. Therefore, $\hat{c}^*_{T,K}$ forms a martingale such that

$$\hat{c}^*_{T,K}(t, S^{\delta_*}_t) = \hat{c}_{T,K}(t, S^{\delta_*}_t)$$

and thus

$$c^*_{T,K}(t, S^{\delta_*}_t) = c_{T,K}(t, S^{\delta_*}_t) \tag{12.2.63}$$

for all $t \in [0, T]$. This means that hypothetical risk neutral and fair call option prices coincide in the given case.

We now introduce the hypothetical risk neutral put price by the corresponding hypothetical risk neutral put-call parity relation

$$p^*_{T,K}(t, S^{\delta_*}_t) = c^*_{T,K}(t, S^{\delta_*}_t) - S^{\delta_*}_t + K\, P^*_T(t) \tag{12.2.64}$$

for $t \in [0, T]$, where $P^*_T(\cdot)$ is the savings bond, see (12.2.42). By applying (12.2.61) and (12.2.64) it can be inferred that

$$p_{T,K}(t, S^{\delta_*}_t) < p^*_{T,K}(t, S^{\delta_*}_t) \tag{12.2.65}$$

for all $t \in [0, T)$. This means that the fair put price is less than or equal to the hypothetical risk neutral put price.

Figure 12.2.4 shows the difference between the hypothetical risk neutral and the fair European put price as a function of the strike $K$ and time $t$ for the same parameter values used in Fig. 12.2.3. Note that these differences are always nonnegative, see (12.2.65). This visualizes again the fact that fair prices are the minimal prices that replicate a contingent claim, see Corollary 10.4.2.

### Difference in Asymptotic Put Prices

When considering the above analysis it becomes clear that differences between fair and hypothetical risk neutral prices arise when the payoff is not vanishing for vanishing GOP. In such a case the risk neutral pricing methodology suggests some prices that contradict economic reasoning. There always exists a corresponding fair price that allows a perfect hedge which is less or equal to the hypothetical risk neutral price.

Now, we shall demonstrate that the differences between fair prices and hypothetical risk neutral prices can become extreme if the underlying GOP value tends towards zero. One can show by the conditional moment estimate (8.7.16) for the benchmarked, fair zero coupon bond, see (12.2.36) and (12.2.10), when the GOP comes close to zero, that

$$\lim_{S_t^{\delta_*} \to 0} \hat{P}_T(t, S_t^{\delta_*}) \stackrel{\text{a.s.}}{=} \lim_{X_t \to 0} E\left( (X_T)^{-q} \,\middle|\, \mathcal{A}_t \right) < \infty \tag{12.2.66}$$

so that

$$\lim_{S_t^{\delta_*} \to 0} P_T(t, S_t^{\delta_*}) \stackrel{\text{a.s.}}{=} \lim_{S_t^{\delta_*} \to 0} S_t^{\delta_*} \, \hat{P}_T(t, S_t^{\delta_*}) = 0 \tag{12.2.67}$$

for $t \in [0,T]$ and $T \in (0, \infty)$. In addition, since

$$\lim_{S_t^{\delta_*} \to 0} c_{T,K}(t, S_t^{\delta_*}) \stackrel{\text{a.s.}}{=} \lim_{S_t^{\delta_*} \to 0} S_t^{\delta_*} \, \hat{c}_{T,K}(t, S_t^{\delta_*}) = 0, \tag{12.2.68}$$

by application of the fair put-call parity relation (12.2.60) we see for the fair put price that

$$\lim_{S_t^{\delta_*} \to 0} p_{T,K}(t, S_t^{\delta_*}) \stackrel{\text{a.s.}}{=} 0 \tag{12.2.69}$$

for $t \in [0,T]$. However, from the hypothetical risk neutral put-call parity relation (12.2.64) together with (12.2.63), (12.2.68) and (12.2.42) it can be seen that the corresponding hypothetical risk neutral put price satisfies for vanishing GOP the limit condition

$$\lim_{S_t^{\delta_*} \to 0} p^*_{T,K}(t, S_t^{\delta_*}) \stackrel{\text{a.s.}}{=} K \, \frac{S_t^0}{S_T^0} > 0 \tag{12.2.70}$$

for $t \in [0,T]$. By comparing (12.2.69) and (12.2.70), we note a difference between the behavior of the fair and hypothetical risk neutral put prices as the GOP comes close to zero. We emphasize again that the fair put price is, in economic terms, the correct price for this contingent claim as it can be perfectly hedged using the hedge ratios (12.2.27) and (12.2.30) and there is no lower put price which could be used for replication. We emphasize that in both cases the put payoff is replicated by a self-financing hedge portfolio.

The above study of the modified CEV model signals that one has to be very careful in the pricing and hedging under stochastic volatility. This is of particular relevance if an index model attempts to capture the leverage effect where volatility increases when the index value decreases. One can expect that this results in effects similar to those described above. For a realistic model the volatility has to become large when the index attains small values to be able to reflect the economically relevant risk involved. This suggests that realistic index models can be expected to face the above experienced problems when applying the risk neutral methodology.

## 12.3 Local Volatility Models

### LV Models

As we have seen in Sect. 12.1, the existence of implied volatility skews for options on indices is well documented. A natural one-factor extension of the BS model is obtained by introducing *local volatility* (LV) *models*. This means that the volatility is allowed to change as a function of the underlying and time. The resulting LV models have attracted the interest of many researchers and practitioners. They were pioneered by Dupire (1992, 1993, 1994) and Derman & Kani (1994a, 1994b) and have been widely used in practice. However, in this literature one typically assumes the existence of a risk neutral probability measure. This could be problematic since the modified CEV model, considered in the previous section, is a special case of an LV model. Another LV model is the MMM proposed in Platen (2001), which was mentioned in Sect. 7.5. We shall see in the next chapter that it does not have an equivalent risk neutral probability measure. Therefore, in this section we apply again real world pricing to obtain derivative prices.

In the first part of this section it will be our aim to estimate the real world transition density of the underlying index from observed call option prices and also its local volatility function. Typically, in the literature on LV models one extracts risk neutral transition densities from observed option prices, see Dupire (1994). This does not make sense for models that do not have an equivalent risk neutral probability measure. Therefore, it will be our aim to estimate the real world transition density without relying on the existence of an equivalent risk neutral probability measure.

### Local Volatility

Let us consider a CFM with GOP process $S^{\delta_*} = \{S^{\delta_*}_t, \, t \in [0,\infty)\}$, which we interpret, similarly to the previous section, as a diversified accumulation index. For simplicity, the short rate $r_t = r$ is assumed to be constant. We say that the GOP $S^{\delta_*}_t$ follows an LV model if it satisfies an SDE of the form

$$dS^{\delta_*}_t = S^{\delta_*}_t \left( \left( r + \sigma^2(t, S^{\delta_*}_t) \right) dt + \sigma(t, S^{\delta_*}_t) \, dW_t \right) \tag{12.3.1}$$

for $t \in [0, \infty)$, see (10.2.8). This formulation of the GOP dynamics incorporates the total market price of risk $|\theta_t|$, as a function of time $t$ and underlying security $S_t^{\delta_*}$, in the form of the *local volatility* (LV)

$$|\theta_t| = \sigma(t, S_t^{\delta_*}) \tag{12.3.2}$$

for $t \in [0, \infty)$. The specific structural assumption here is that the total market price of risk depends on the underlying security and time. The choice of the LV function characterizes the selected LV model. Here $W = \{W_t,\, t \in [0, \infty)\}$ denotes a standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, where $P$ is the real world probability measure. Furthermore, we assume that a unique strong solution of the SDE (12.3.1) exists, see Sect. 7.7, which is not trivial for certain classes of LV functions. In cases, where $S_t^{\delta_*}$ may reach zero, we choose zero as an absorbing boundary, similarly as in (7.7.18).

## LV Function

Under an LV model it is assumed that the volatility $\sigma(t, S_t^{\delta_*})$ is generated by a given *LV function* $\sigma : [0, \infty) \times [0, \infty) \to [0, \infty]$, which is a deterministic function of time and the underlying security.

If the *volatility process* $\sigma = \{\sigma(t, S_t^{\delta_*}),\, t \in [0, \infty)\}$ is deterministic, then we have a BS model for $S^{\delta_*}$. The modified CEV model, considered in the previous section, has as LV function the power function

$$\sigma(t, S_t^{\delta_*}) = (S_t^{\delta_*})^{a-1}\, \psi \tag{12.3.3}$$

for some exponent $a \in (-\infty, 1)$ and constant scaling parameter $\psi$.

Another LV model is obtained by the stylized version of the MMM, mentioned in Sect. 7.5. Here the LV function has the form

$$\sigma(t, S_t^{\delta_*}) = \sqrt{\frac{\alpha_0\, \exp\{(r+\eta)\, t\}}{S_t^{\delta_*}}}, \tag{12.3.4}$$

with constant net growth rate $\eta > 0$ and initial parameter $\alpha_0 > 0$. In this case the GOP can be modeled as

$$S_t^{\delta_*} = Y_t\, \alpha_0\, \exp\{(r+\eta)\, t\} \tag{12.3.5}$$

for $t \in [0, \infty)$ with parameters $\alpha_0, \eta, r > 0$. In (12.3.5) $Y = \{Y_t,\, t \in [0, \infty)\}$ is a square root process of dimension four, which satisfies the SDE

$$dY_t = (1 - \eta\, Y_t)\, dt + \sqrt{Y_t}\, dW_t \tag{12.3.6}$$

for $t \in [0, \infty)$, see (7.5.16), with $Y_0 = \frac{S_0^{\delta_*}}{\alpha_0} > 0$. One notes that the LV function (12.3.4) of the stylized MMM can be expressed simply as a function of the value of the square root process $Y$. That is, by (12.3.4) and (12.3.5) we can write

$$\sigma(t, S_t^{\delta_*}) = \frac{1}{\sqrt{Y_t}} \tag{12.3.7}$$

for $t \in [0, \infty)$. Consequently, the squared volatility is the inverse of a square root (SR) process. Such an SR process is known to have as stationary density a gamma density, see Sect. 4.5. Therefore, in the case of the stylized MMM the volatility has a stationary density and, thus, allows us to model some kind of an equilibrium.

### Benchmarked Savings Account

The benchmarked savings account process $\hat{S}^0 = \{\hat{S}_t^0,\, t \in [0, \infty)\}$ is again given by the ratio

$$\hat{S}_t^0 = \frac{S_t^0}{S_t^{\delta_*}}. \tag{12.3.8}$$

For the LV model it satisfies, by an application of the Itô formula together with (12.3.1), the driftless SDE

$$d\hat{S}_t^0 = -\hat{S}_t^0\, \sigma\left(t, \frac{S_t^0}{\hat{S}_t^0}\right) dW_t \tag{12.3.9}$$

for $t \in [0, \infty)$. Since a nonnegative, local martingale is a supermartingale, see Lemma 5.2.2 (i) and Theorem 10.3.1, the benchmarked savings account $\hat{S}^0$ is a supermartingale. Recall from Sect. 9.4 that the candidate Radon-Nikodym derivative process $\Lambda = \{\Lambda_t,\, t \in [0, \infty)\}$ of the hypothetical risk neutral probability measure is given by the normalized benchmarked savings account $\Lambda = \{\Lambda_t,\, t \in [0, \infty)\}$ with

$$\Lambda_t = \frac{\hat{S}_t^0}{\hat{S}_0^0} = \exp\left\{-\frac{1}{2}\int_0^t \sigma(s, S_s^{\delta_*})^2\, ds - \int_0^t \sigma(s, S_s^{\delta_*})\, dW_s\right\} \tag{12.3.10}$$

for $t \in [0, \infty)$. We have already seen for the modified CEV model that $\Lambda$ can become a strict $(\underline{\mathcal{A}}, P)$-supermartingale. Therefore, an equivalent risk neutral probability measure may not exist for a range of LV models.

### Real World Pricing under an LV Model

Let $H = H(S_T^{\delta_*})$ denote a nonnegative payoff with maturity date $T \in (0, \infty)$. Then its benchmarked fair price $\hat{U}_H(t)$ at time $t \in [0, T]$ is given by the conditional expectation

$$\hat{U}_H(t) = E\left(\frac{H(S_T^{\delta_*})}{S_T^{\delta_*}}\,\bigg|\, \mathcal{A}_t\right) \tag{12.3.11}$$

for $t \in [0, T]$, see Definition 9.1.2. The corresponding fair price $U_H(t)$ at time $t$, expressed in units of the domestic currency, is then

$$U_H(t) = S_t^{\delta_*}\, \hat{U}_H(t) \tag{12.3.12}$$

for $t \in [0, T]$, which is the real world pricing formula. Note that under an LV model $S^{\delta_*}$ is a diffusion process and, thus, Markovian. For a sufficiently smooth function $f : [0, T] \times (0, \infty) \to \Re$ define the operator $L^0$ by the expression

$$L^0\, f(t, S) = \frac{\partial f(t, S)}{\partial t} + (r + \sigma^2(t, S))\, S\, \frac{\partial f(t, S)}{\partial S} + \frac{1}{2}\, \sigma^2(t, S)\, S^2\, \frac{\partial^2 f(t, S)}{\partial S^2} \tag{12.3.13}$$

for $(t, S) \in (0, T) \times (0, \infty)$. Using (12.3.11) and (12.3.1) it follows by the Feynman-Kac formula (9.7.3)–(9.7.5) that the benchmarked fair pricing function $\hat{u}_H(\cdot, \cdot)$ with $\hat{u}_H(t, S_t^{\delta_*}) = \hat{U}_H(t)$ satisfies the PDE

$$L^0\, \hat{u}_H(t, S) = 0 \tag{12.3.14}$$

for $(t, S) \in (0, T) \times (0, \infty)$ with terminal condition

$$\hat{u}_H(T, S) = \frac{H(S)}{S} \tag{12.3.15}$$

for $S \in (0, \infty)$.

The benchmarked fair pricing function $\hat{u}_H(\cdot, \cdot)$ is uniquely determined by (12.3.11) and satisfies the PDE (12.3.14) with terminal condition (12.3.15) as its minimal solution. As we have noticed from the modified CEV model, one needs to be aware of the fact that for certain types of payoffs the solution to this PDE may not be unique. This was, for instance, the case for zero coupon bonds and European puts. These are payoffs with nonvanishing value when the GOP reaches zero. However, we emphasize that there is only one minimal solution to the PDE (12.3.14)–(12.3.15), which is given by the benchmarked fair pricing function. In this case the boundary for $S \to 0$ is absorbing. This ensures the absence of arbitrage in the sense of Definition 10.3.2, because benchmarked nonnegative portfolios that reach zero stay at zero in a CFM.

For the fair price of $H(S_T^{\delta_*})$ a corresponding self-financing hedge portfolio, which replicates the payoff, can be constructed, similarly as in the previous section. If the benchmarked fair pricing function $\hat{u}_H(\cdot, \cdot)$ is sufficiently smooth, then we have by application of the Itô formula a martingale representation of the form

$$\frac{H(S_T^{\delta_*})}{S_T^{\delta_*}} = \hat{u}_H(t, S_t^{\delta_*}) + \int_t^T \frac{\partial \hat{u}_H(s, S_s^{\delta_*})}{\partial S^{\delta_*}}\, S_s^{\delta_*}\, \sigma(s, S_s^{\delta_*})\, dW_s \tag{12.3.16}$$

for $t \in [0, T)$, see (11.5.3) and (12.2.25). One can form a hedge portfolio, see Theorem 11.5.2, consisting of $\delta_H^0(t)$ units of the domestic savings account and $\delta_H^1(t)$ units of the GOP at time $t$. By comparing (12.3.16) with the SDE for a benchmarked portfolio $\hat{S}^\delta$ one obtains the hedge ratios

$$\delta_H^0(t) = -\frac{(S_t^{\delta_*})^2}{S_t^0}\, \frac{\partial \hat{u}_H(t, S_t^{\delta_*})}{\partial S^{\delta_*}} \tag{12.3.17}$$

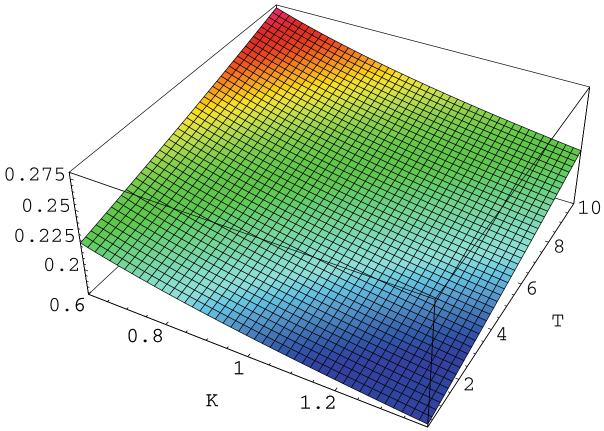



**Fig. 12.3.1.** Implied volatility surface for the stylized MMM

and

$$\delta_H^1(t) = \hat{u}_H(t, S_t^{\delta_*}) - \delta_H^0(t)\,\frac{S_t^0}{S_t^{\delta_*}} \tag{12.3.18}$$

for $t \in [0,T)$. These generalize the equations (12.2.27) and (12.2.28). This hedge portfolio replicates the payoff $H(S_T^{\delta_*})$. It provides perfect replication in the sense that the corresponding P&L remains zero, see (12.2.33). The portfolio value forms the minimal possible value since its benchmarked value is a martingale and coincides with the payoff at maturity $T$.

### European Calls

If we denote by $K$ the strike price of a European call option with maturity $T$, then at time $t$ the corresponding *fair call option price* $c(t, S_t^{\delta_*}, T, K)$ satisfies the relation

$$c(t, S_t^{\delta_*}, T, K) = S_t^{\delta_*}\, E\left(\left(1 - \frac{K}{S_T^{\delta_*}}\right)^+ \,\bigg|\, \mathcal{A}_t\right) \tag{12.3.19}$$

for $t \in [0,T]$.

Instead of the European call option prices their corresponding implied volatilities give a better view of the option market. We have shown in Fig. 12.2.3 the implied volatility surface for European calls, which results from the modified CEV model as a function of the strike and time to maturity. To provide another example we show in Fig. 12.3.1 the implied volatility surface's dependence on $T$ and $K$ for a fair call option under the stylized MMM given in Sect. 7.5, with $r = 0.04$, $\eta = 0.048$, $\alpha_0 = 0.03827$, and $S_0^{\delta_*} = 1$. In Fig. 12.3.1 we observe a pronounced negative skew. The term structure of implied volatility is characterized here by a gradual increase in at-the-money implied volatilities over time. Note that we take here, as in the previous section, the fair zero coupon bond as discount factor when calculating implied volatilities, see (12.2.57). This means that we adjust in the Black-Scholes formula the short rate to

$$\hat{r} = \frac{-1}{T-t}\, \ln(P(t,T)) \tag{12.3.20}$$

for calculating implied volatilities.

### Implied Transition Density of the GOP

We shall now demonstrate that it is, in principle, possible to estimate from observed option prices the transition probability density of the underlying GOP under an LV model. Let us denote by $p_{\hat{S}^0}(t, \hat{S}^0_t; T, \hat{S}^0_T)$ the transition density of the benchmarked savings account process $\hat{S}^0$ under the real world probability measure $P$. For convenient presentation we define the quantity

$$u(t, \hat{S}^0_t, T, \kappa) = \kappa\, \hat{c}_{T,K}(t, S^{\delta_*}_t) = \frac{S^0_T}{K\, S^{\delta_*}_t}\, c(t, S^{\delta_*}_t, T, K) \tag{12.3.21}$$

with the deterministic value

$$\kappa = \frac{S^0_T}{K}. \tag{12.3.22}$$

By (12.3.19) together with (12.3.8) this equation can be rewritten in the form

$$u(t, \hat{S}^0_t, T, \kappa) = E\left(\left(\kappa - \hat{S}^0_T\right)^+ \,\Big|\, \mathcal{A}_t\right). \tag{12.3.23}$$

Using an idea of Breeden & Litzenberger (1978), which was also applied by Dupire (1993) and Derman & Kani (1994b) in the risk neutral setting, it follows from (12.3.23) that

$$\begin{aligned}\frac{\partial}{\partial \kappa}\, u(t, \hat{S}^0_t, T, \kappa) &= \frac{\partial}{\partial \kappa} \int_0^\kappa (\kappa - y)\, p_{\hat{S}^0}(t, \hat{S}^0_t; T, y)\, dy \\ &= \int_0^\kappa p_{\hat{S}^0}(t, \hat{S}^0_t; T, y)\, dy. \end{aligned} \tag{12.3.24}$$

This allows us to express the real world transition density $p_{\hat{S}^0}$ in the form

$$p_{\hat{S}^0}(t, \hat{S}^0_t; T, \kappa) = \frac{\partial^2}{\partial \kappa^2}\, u(t, \hat{S}^0_t, T, \kappa) \tag{12.3.25}$$

for $t \in [0, T]$. By using (12.3.22) and (12.3.21) and calculating the partial derivative of $u$ in terms of partial derivatives of the call pricing function $c$ given in (12.3.19), the transition density $p_{\hat{S}^0}$ in (12.3.25) can be equivalently expressed in the form

$$p_{\hat{S}^0}(t, \hat{S}^0_t; T, \kappa) = \frac{K^3}{S^0_T\, S^{\delta_*}_t}\, \frac{\partial^2}{\partial K^2}\, c(t, S^{\delta_*}_t, T, K) \tag{12.3.26}$$

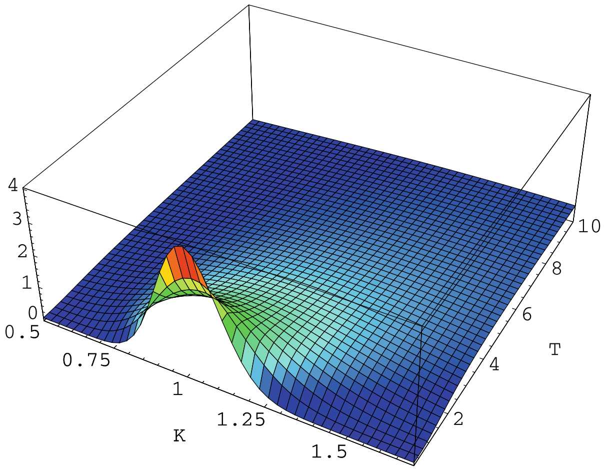



**Fig. 12.3.2.** Implied transition density obtained from CEV call option prices

for $t \in [0, T]$.

Let $p_{S^{\delta_*}}(t, S_t^{\delta_*}; T, K)$ denote the transition density for the GOP process $S^{\delta_*}$ under the real world probability measure $P$. Then the following result can be directly obtained by using the transformation (12.3.8) and formulas (12.3.25) and (12.3.26).

**Lemma 12.3.1.** *The transition density $p_{S^{\delta_*}}$ is of the form*

$$p_{S^{\delta_*}}(t, S_t^{\delta_*}; T, K) = \frac{K}{S_t^{\delta_*}} \frac{\partial^2}{\partial K^2} c(t, S_t^{\delta_*}, T, K) \tag{12.3.27}$$

*for* $t \in [0, T]$.

Consequently, by assuming the availability of a continuum of European call option prices with respect to strike and time to maturity we can theoretically infer the real world transition density of the GOP. This is different to most results in the literature where one infers risk neutral transition densities. As we have seen earlier, a corresponding equivalent risk neutral probability measure may, in general, not exist. Therefore, the derivation of risk neutral transition densities may not be that useful.

To illustrate the statement of Lemma 12.3.1, Fig. 12.3.2 displays a transition density of the GOP as a function of $K$ and $T$, which has been numerically computed by application of relation (12.3.27). As input we used the values of the European call options that were calculated earlier under the modified CEV model for obtaining the implied volatilities shown in Fig. 12.2.3. This means that Fig. 12.3.2 displays an inferred real world transition probability density $p_{S^{\delta_*}}$ for a GOP process $S^{\delta_*}$ for the case of the modified CEV model with $a = -\frac{1}{2}$, $\psi = 0.2$ and $r = 0.04$.

### Representation of the LV Function

We shall see under the LV model that, in principle, at any maturity date $T \in [0, \infty)$ and for any value $\kappa \in (0, \infty)$, the LV function value $\sigma(T, \kappa)$ can be recovered from a continuum of observed European call option prices. This is again similar to results described in Breeden & Litzenberger (1978), Dupire (1992, 1993, 1994) and Derman & Kani (1994a, 1994b). In our case the LV function is obtained without requiring the existence of an equivalent risk neutral probability measure, which is different to the approach taken in these references.

To derive the result conveniently let us make the following technical assumptions

$$\lim_{\kappa \to 0} \frac{1}{\kappa} \frac{\partial}{\partial T} u(t, \hat{S}_t^0, T, \kappa) = 0, \tag{12.3.28}$$

and

$$\lim_{\kappa \to 0} \sigma^2(T, K)\, \kappa \frac{\partial^2}{\partial \kappa^2} u(t, \hat{S}_t^0, T, \kappa) = 0. \tag{12.3.29}$$

These are reasonable conditions that apply to a wide range of LV models. They lead to the following result, which is derived in Heath & Platen (2006):

**Theorem 12.3.2.** *Under* (12.3.28) *and* (12.3.29) *has for fixed* $t \in [0, T]$ *and* $\hat{S}_t^0 > 0$ *the LV function the form*

$$\sigma(T, K) = \frac{\sqrt{2}}{\kappa} \left( \frac{\frac{\partial}{\partial T} u(t, \hat{S}_t^0, T, \kappa)}{\frac{\partial^2}{\partial \kappa^2} u(t, \hat{S}_t^0, T, \kappa)} \right)^{\frac{1}{2}} \tag{12.3.30}$$

*for* $(T, K) \in (0, \infty) \times (0, \infty)$*,* $t \in [0, T)$*, with* $\kappa$ *as given in* (12.3.22)*.*

### Dupire Formula

To express the LV function in terms of European call option prices one can use the transformations (12.3.22) and (12.3.21) to compute the corresponding partial derivatives. One then obtains the following result, which is equivalent to (12.3.30), see Heath & Platen (2006). It is known as the *Dupire formula*. Here it is obtained without relying on the existence of a risk neutral probability measure.

**Corollary 12.3.3.** (Dupire) *The LV function has the representation*

$$\sigma(T, K) = \frac{\sqrt{2}}{K} \sqrt{\frac{\frac{\partial}{\partial T} c(t, S_t^{\delta_*}, T, K) + K\, r\, \frac{\partial}{\partial K} c(t, S_t^{\delta_*}, T, K)}{\frac{\partial^2}{\partial K^2} c(t, S_t^{\delta_*}, T, K)}} \tag{12.3.31}$$

*for* $(T, K) \in (0, \infty) \times (0, \infty)$*,* $t \in [0, T)$*.*

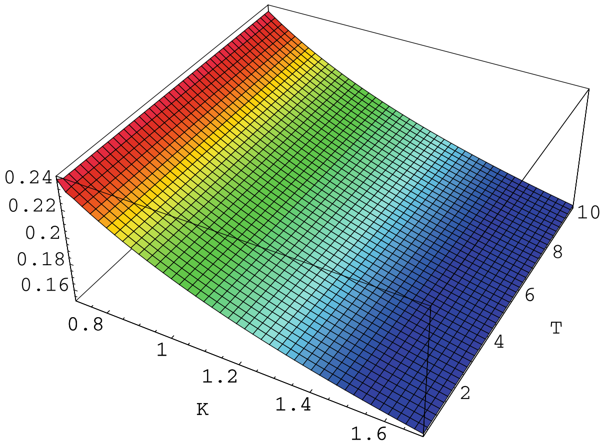



**Fig. 12.3.3.** LV function implied from modified CEV call option prices

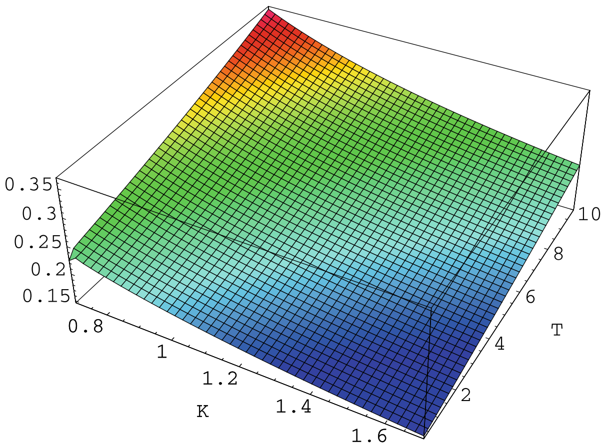



**Fig. 12.3.4.** LV function implied from MMM call option prices

For illustration, in Fig. 12.3.3 the LV function $\sigma(\cdot, \cdot)$ is displayed when obtained numerically via formula (12.3.31) from the European call option values that were used to compute the implied volatilities of the modified CEV model shown in Fig. 12.2.3. These results match, up to some negligible numerical errors, the corresponding LV function $\sigma(t, S) = S^{a-1}\psi$. Small errors in values are detectable in Fig. 12.3.3 for small $K$ and $T$, which are caused by the numerical implementation of the formula (12.3.31). These minor differences are explained by round-off and truncation errors from the discrete differentiations involved. Similarly we plot in Fig. 12.3.4 the LV function, numerically implied from call prices under the stylized MMM. Also we recover here, up to minor numerical errors for small $K$ and $T$, the LV function of the MMM.

It must be noted that a wide range of typically observed implied volatility surfaces can be calibrated via LV models. However, this does not mean that

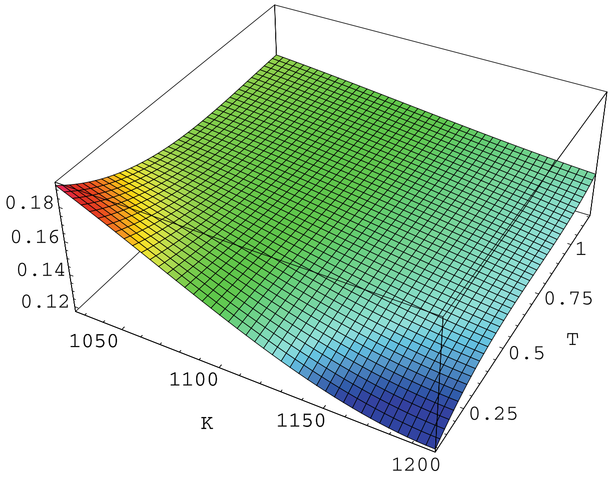



**Fig. 12.3.5.** Implied volatility surface for the S&P500 for 20 April 2004

the resulting LV model explains the dynamics of the underlying security. It only provides an LV function for European call and put options which allows us to match the observed option prices under the assumption of an LV model.

Finally, it is important to emphasize that implying a local volatility function from traded option prices is a difficult numerical task. Small deviations in prices can have a substantial effect on the implied LV function. This also creates a major drawback for the practical calibration of LV models. It would be valuable to have some economic reasoning behind the particular form of a selected LV function. The MMM, which we derive in the next chapter, provides such an economic explanation.

### Local Volatility Function of S&P500

To illustrate the above analysis further we consider observed index option prices for the S&P500 index. Due to the numerical sensitivity of the implied LV functions to small errors in option prices we work with smoothed data. Figure 12.3.5 shows a fit of the implied volatility surface for S&P500 European call options for 20 April 2004 as in Heath & Platen (2006). These implied volatilities were computed using prices obtained from the average of bid and ask prices using the short rate $r = 0.03$ and a dividend rate of $d = 0.01$. The corresponding closing price for the S&P500 index was $S_0^{\delta_*} = 1114$. A total of 83 option prices was used to obtain the displayed fit. A least squares fit, see Sect. 2.3, for the implied volatility surface was obtained using a set of two-dimensional cubic polynomials. The corresponding smoothed option prices are then used to calculate the real world transition densities according to formula (12.3.27). The resulting transition density function is displayed in Fig. 12.3.6.

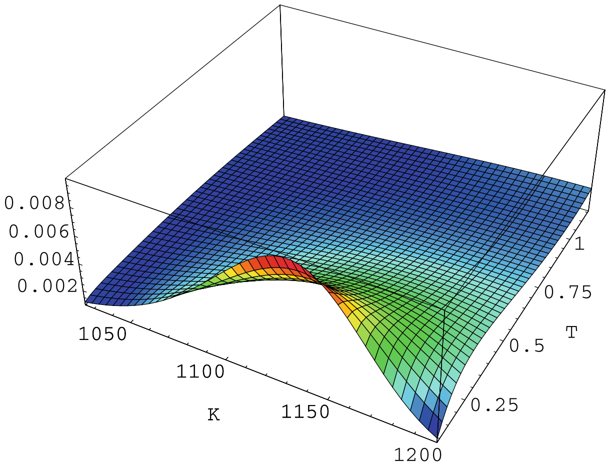



**Fig. 12.3.6.** Implied transition density for S&P500 for 20 April 2004

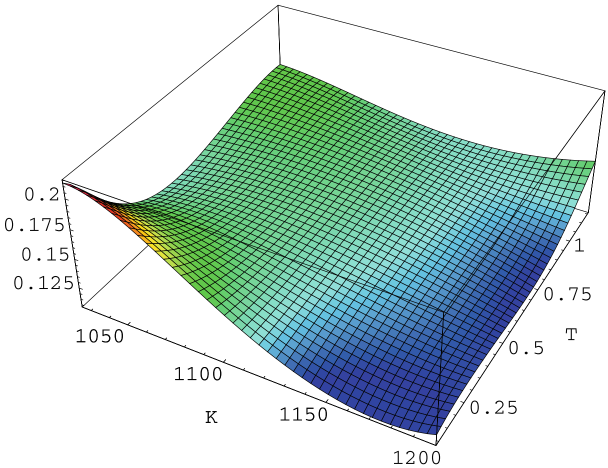



**Fig. 12.3.7.** LV function for S&P500 for 20 April 2004

The corresponding LV function is obtained by formula (12.3.31) and is displayed in Fig. 12.3.7. Because of the form of equation (12.3.31) and, in particular, the combination of first and second order partial derivatives, the shape of this surface turns out to be rather sensitive to the choice of the basis functions employed in the fitting procedure. Note that this LV function returns in our case exactly the implied volatility surface displayed in Fig. 12.3.5 and the corresponding smoothed S&P500 option prices. We observe a strong sensitivity of the LV function towards small deviations in option prices. Therefore, it is difficult to extract from observed data what the calibrated LV function of an index should be. The difficulties indicated above, in calibrating LV models in practice, emphasize the need for a better understanding of the nature of the volatility process itself. This should then provide a generic shape for the

LV function. In the next chapter we shall discuss this question further when deriving the MMM. It is interesting to note that the implied volatility surface in Fig. 12.3.5 is without any major curvature for times to maturity above six months. This is also the type of implied volatility surface that the MMM generates for this range of maturities, see Fig. 12.3.1.

### Proof of Theorem 12.3.2 (*)

From the SDE (12.3.9) and relation (12.3.8) it follows that the transition density $p_{\hat{S}^0}$ for $\hat{S}^0$ satisfies the Fokker-Planck equation

$$\frac{\partial}{\partial T} p_{\hat{S}^0}(t, \hat{S}^0_t; T, \kappa) - \frac{1}{2} \frac{\partial^2}{\partial \kappa^2} \left\{ \sigma^2(T, K)\, \kappa^2\, p_{\hat{S}^0}(t, \hat{S}^0_t; T, \kappa) \right\} = 0 \tag{12.3.32}$$

for $(T, \kappa) \in (0, \infty) \times (0, \infty)$ with initial condition

$$p_{\hat{S}^0}(t, \hat{S}^0_t; t, \kappa) = \delta(\hat{S}^0_t - \kappa), \tag{12.3.33}$$

where $\delta(\cdot)$ is the Dirac delta function, see (4.4.1). It, therefore, follows by using (12.3.25) that (12.3.32) can be rewritten in the form

$$\frac{\partial}{\partial T} \left( \frac{\partial^2}{\partial \kappa^2} u(t, \hat{S}^0_t, T, \kappa) \right) - \frac{1}{2} \frac{\partial^2}{\partial \kappa^2} \left\{ \sigma^2(T, K)\, \kappa^2\, \frac{\partial^2}{\partial \kappa^2} u(t, \hat{S}^0_t, T, \kappa) \right\} = 0$$

and hence

$$\frac{\partial^2}{\partial \kappa^2} \left\{ \frac{\partial}{\partial T} u(t, \hat{S}^0_t, T, \kappa) - \frac{1}{2} \sigma^2(T, K)\, \kappa^2\, \frac{\partial^2}{\partial \kappa^2} u(t, \hat{S}^0_t, T, \kappa) \right\} = 0. \tag{12.3.34}$$

Then there exist quantities $\beta_0(T)$ and $\beta_1(T)$ such that

$$\frac{\partial}{\partial T} u(t, \hat{S}^0_t, T, \kappa) - \frac{1}{2} \sigma^2(T, K)\, \kappa^2\, \frac{\partial^2}{\partial \kappa^2} u(t, \hat{S}^0_t, T, \kappa) = \beta_0(T) + \beta_1(T)\, \kappa. \tag{12.3.35}$$

From (12.3.35), (12.3.28) and (12.3.29) it follows that

$$\beta_0(T) = 0. \tag{12.3.36}$$

and

$$\beta_1(T) = 0. \tag{12.3.37}$$

Combining (12.3.35), (12.3.36) and (12.3.37) yields (12.3.30). □

## 12.4 Stochastic Volatility Models

### Modeling Volatility as a Separate Process

A number of continuous asset price models have been developed, which model the volatility process as a separate, possibly correlated, stochastic process.

This group of models includes the models by Hull & White (1987, 1988), Johnson & Shanno (1987), Scott (1987), Wiggins (1987), Chesney & Scott (1989), Melino & Turnbull (1990), Stein & Stein (1991), Hofmann et al. (1992) and Heston (1993) among others. In the following we provide a description of this type of model by applying results from Heath, Hurst & Platen (2001). The GOP models again an index which is interpreted as the underlying security.

The empirical results of Sect. 2.6 on the estimation of index log-returns indicate that for daily observations of stock index log-returns the Student $t$ distribution with about four degrees of freedom provides an excellent fit. We take this stylized empirical fact as motivation for the following study, which aims to construct stochastic volatility processes with a prescribed stationary density.

First, let us explain how this is linked to the results from the estimation of prescribed log-return densities. When considering small time steps, then a discounted GOP $\bar{S}^{\delta_*}$ with squared volatility $|\theta_t|^2$ generates at time $t$ approximately conditionally Gaussian distributed log-returns with a stochastic variance $|\theta_t|^2$ per unit of time. This means that for small time step size $h > 0$ one observes the conditionally Gaussian log-returns

$$\varDelta \ln(\bar{S}^{\delta_*}_t) = \ln\left(\frac{\bar{S}^{\delta_*}_{t+h}}{\bar{S}^{\delta_*}_t}\right) \sim \mathcal{N}\left(\frac{|\theta_t|^2}{2}\,h, |\theta_t|^2\,h\right). \tag{12.4.1}$$

Since we consider log-returns over a short time period $[t, t+h]$ the trend effect of the conditional mean in (12.4.1) can be neglected.

We now consider the case where the process $|\theta|^2$ has a given stationary density and the observation of the log-returns extends over a sufficiently long time period. This then results in the estimation of normal variance mixture log-returns, as described in Sect. 2.5, see also Fergusson & Platen (2006). For instance, when $\frac{1}{|\theta_t|^2}$ has as stationary density that of a gamma distributed random variable, then the estimated log-returns appear to be Student $t$ distributed, see Kessler (1997) and Prakasa Rao (1999). On the other hand, if the stationary density of $|\theta_t|^2$ is a gamma density, then the log-returns, when estimated, appear to be variance gamma distributed. We emphasize the fact that a stochastic volatility process needs a long observation period so that the squared volatility values have traversed reasonably often over the range of their typical values to generate the mixing effect of the random variance for the log-returns.

### A Class of Continuous Stochastic Volatility Models

Note from (12.1.1) that the discounted GOP

$$\bar{S}^{\delta_*}_t = \frac{S^{\delta_*}_t}{S^0_t} \tag{12.4.2}$$

satisfies the SDE

$$d\bar{S}_t^{\delta_*} = \bar{S}_t^{\delta_*} \left( |\theta_t|^2 \, dt + |\theta_t| \, dW_t \right) \tag{12.4.3}$$

for $t \in [0, \infty)$. Similarly to Heath, Hurst & Platen (2001) we now derive a class of continuous discounted GOP models with stochastic volatility processes that have stationary densities. To cover a wide range of stochastic volatility models let us consider the factor process $X = \{X_t, \, t \in [0, \infty)\}$ with

$$X_t = Y(|\theta_t|). \tag{12.4.4}$$

It involves a twice continuously differentiable function $Y(\cdot)$ that depends on the volatility process $|\theta| = \{|\theta_t|, \, t \in [0, \infty)\}$ of the GOP. The joint dynamics of the discounted GOP process $\bar{S}^{\delta_*}$ and the factor process $X$ are assumed to be governed by a time homogeneous system of SDEs

$$\begin{aligned} d\bar{S}_t^{\delta_*} &= \bar{S}_t^{\delta_*} \left( |\theta_t|^2 \, dt + |\theta_t| \left( \varrho \, d\bar{W}_t + \sqrt{1 - \varrho^2} \, d\tilde{W}_t \right) \right), \\ dX_t &= C(X_t) \, dt + D(X_t) \, d\bar{W}_t \end{aligned} \tag{12.4.5}$$

for $t \in [0, \infty)$. Here $\tilde{W}$ and $\bar{W}$ are independent standard Wiener processes under the real world probability measure $P$. In the SDE (12.4.5) the dynamics of the discounted GOP $\bar{S}^{\delta_*}$ involve the stochastic volatility process $|\theta|$. The functions $C(\cdot)$ and $D(\cdot)$ are assumed to satisfy appropriate conditions so that the SDE (12.4.5) admits a unique strong solution, see Sect. 7.7. The parameter $\varrho \in [-1, 1]$ is the correlation parameter.

We remark that, by application of the Itô formula to the function (12.4.4) of $X_t = Y(|\theta_t|)$, we obtain the SDE

$$dX_t = dY(|\theta_t|) = Y'(|\theta_t|) \, d|\theta_t| + \frac{1}{2} \, Y''(|\theta_t|) \, d[|\theta|]_t$$

and, thus, by rearranging this SDE for $d|\theta_t|$ with (12.4.5) the SDE

$$d|\theta_t| = \frac{1}{Y'(|\theta_t|)} \left( C(X_t) \, dt - \frac{1}{2} \, Y''(|\theta_t|) \, d[|\theta|]_t + D(X_t) \, d\bar{W}_t \right).$$

This leads to an SDE for the volatility process $|\theta| = \{|\theta_t| = Y^{-1}(X_t), \, t \in [0, \infty)\}$, which is of the form

$$d|\theta_t| = \left( \frac{C(Y(|\theta_t|))}{Y'(|\theta_t|)} - \frac{1}{2} \, \frac{D(Y(|\theta_t|))^2 \, Y''(|\theta_t|)}{Y'(|\theta_t|)^3} \right) dt + \frac{D(Y(|\theta_t|))}{Y'(|\theta_t|)} \, d\bar{W}_t \tag{12.4.6}$$

for $t \in [0, \infty)$. Here $Y'(\cdot)$ and $Y''(\cdot)$ denote the first and second derivatives of the function $Y(\cdot)$, respectively, and $Y^{-1}(\cdot)$ is the inverse function of $Y(\cdot)$ on $(0, \infty)$. The resulting stochastic volatility models differ according to different specifications of the functions $Y(\cdot)$, $C(\cdot)$ and $D(\cdot)$.

### Specific Stochastic Volatility Models

Let us now mention some well-known stochastic volatility models and explain how they fit into the above framework:

Hull & White (1988) proposed a model with mean reverting dynamics for the squared volatility process $|\theta|^2$. Here $X_t = Y(|\theta_t|) = |\theta_t|^2$, $C(x) = k\,(\bar{\theta}^2 - x)$ and $D(x) = \gamma\,\sqrt{x}$, where $k$, $\bar{\theta}$ and $\gamma$ are positive constants. This is also the dynamics used in Heston (1993). It provides a popular squared stochastic volatility model, the *Heston model*, which satisfies the SDE

$$d|\theta_t|^2 = k\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,\sqrt{|\theta_t|^2}\,d\bar{W}_t \tag{12.4.7}$$

for $t \in [0, \infty)$ with $|\theta_0|^2 > 0$. Note that one needs to have $\frac{\kappa}{\gamma^2} \geq \frac{1}{2}$ to obtain a stationary density for the stochastic volatility since the squared volatility is modeled by a square root process, see Sect. 8.7.

Scott (1987) and Stein & Stein (1991) used Ornstein-Uhlenbeck processes to model the volatility process $|\theta|$, where $X_t = Y(|\theta_t|) = |\theta_t|$, $C(x) = k\,(\bar{\theta} - x)$ and $D(x) = \gamma$. Here $k$, $\bar{\theta}$ and $\gamma$ are positive constants. The *Scott model* is defined by the SDE

$$d|\theta_t| = k\,(\bar{\theta} - |\theta_t|)\,dt + \gamma\,d\bar{W}_t \tag{12.4.8}$$

for $t \in [0, \infty)$, $|\theta_0| \in \Re$. Note that the squared volatility $|\theta_t|^2$ satisfies the SDE

$$d|\theta_t|^2 = 2\,k\left(\bar{\theta}\,|\theta_t| + \frac{\gamma^2}{2\,k} - |\theta_t|^2\right)dt + 2\,\gamma\,|\theta_t|\,d\bar{W}_t, \tag{12.4.9}$$

which resembles some generalized square root process.

Also Wiggins (1987), Chesney & Scott (1989) and Melino & Turnbull (1990) used the Ornstein-Uhlenbeck process, but for modeling the logarithm $\ln(|\theta_t|)$ of the volatility process. Here $X_t = Y(|\theta_t|) = \ln(|\theta_t|)$, $C(x) = k(\ln(\bar{\theta}) - x)$ and $D(x) = \gamma$, where, once again, $k$, $\bar{\theta}$ and $\gamma$ are positive constants. The *Wiggins model* satisfies then the SDE

$$d\ln(|\theta_t|) = k\left(\ln(\bar{\theta}) - \ln(|\theta_t|)\right)dt + \gamma\,d\bar{W}_t \tag{12.4.10}$$

for $t \in [0, \infty)$, $|\theta_0| > 0$. The squared volatility satisfies here the SDE

$$d|\theta_t|^2 = |\theta_t|^2\left(2\,\gamma^2 + k\,(\ln(\bar{\theta}^2) - \ln(|\theta_t|^2))\right)dt + 2\,\gamma\,|\theta_t|^2\,d\bar{W}_t, \tag{12.4.11}$$

which has multiplicative noise. Note that further stochastic volatility models can be expressed under the above framework, as we shall see below.

### Stationary Density

Let us now compute the stationary density for the process $X$ given by the SDE (12.4.5). The process $X$ is a time homogeneous diffusion process with transition densities depending only on the elapsed period of time. We, therefore, write $p(s,x;t,y)$ to denote the transition density of $X_t = y$ given $X_s = x$. The corresponding *Fokker-Planck equation*, see (4.4.1), is then given by

$$\frac{\partial p(s,x;t,y)}{\partial t} + \frac{\partial (C(y)\, p(s,x;t,y))}{\partial y} - \frac{1}{2}\, \frac{\partial^2 \left(D(y)^2\, p(s,x;t,y)\right)}{\partial y^2} = 0 \quad (12.4.12)$$

for all $t \in (s,\infty)$ and $s \in [0,\infty)$, with $(s,x)$ fixed.

Since $X$ is assumed to have a stationary density the transition density $p(s,x;t,y)$ approaches the stationary density function $\bar{p}$ as $t \to \infty$, that is

$$\bar{p}(y) = \lim_{t\to\infty} p(0,x;t,y), \quad (12.4.13)$$

for $x, y \in \Re$. It follows by the Fokker-Planck equation (12.4.12) that

$$C(y)\bar{p}(y) - \frac{1}{2}\frac{d(D(y)^2\, \bar{p}(y))}{dy} = \tilde{K}, \quad (12.4.14)$$

for all $y \in \Re$ and some constant $\tilde{K}$. Now, we assume that $\bar{p}(y) \to 0$ and $\frac{d\bar{p}(y)}{dy} \to 0$ as $|y| \to \infty$. Under these assumptions the constant $\tilde{K}$ must become zero. By direct integration we then obtain, as shown in (4.5.5),

$$\bar{p}(y) = \frac{A}{D(y)^2}\, \exp\left\{2\int_{y_0}^{y} \frac{C(u)}{D(u)^2}\, du\right\} \quad (12.4.15)$$

for $y \in \Re$, where $A$ is a normalizing constant such that

$$\int_{-\infty}^{\infty} \bar{p}(y)\, dy = 1.$$

Here $y_0$ is an appropriately chosen point in $(-\infty,\infty)$. Note that (12.4.15) gives the form of the stationary density function and accommodates a wide range of diffusions $X$ with stationary density.

### Inverse Gamma Density

As pointed out in Sect. 2.6, we obtain the Student $t$ distribution as normal variance mixture log-return distribution if the squared volatility has an inverse gamma distribution. Therefore, let us now introduce a class of squared volatility models, which have for

$$X_t = |\theta_t|^2 \quad (12.4.16)$$

an inverse gamma density as stationary density. As we shall see below, several diffusion processes can fulfill this requirement. The stationary density $\bar{p}_{\theta^2}$ for the squared volatility equals in this case

$$\bar{p}_{\theta^2}(y) = \frac{(\frac{1}{2}\,\delta)^{\frac{1}{2}\,\delta}}{\varepsilon^2\,\Gamma(\frac{1}{2}\,\delta)} \left(\frac{y}{\varepsilon^2}\right)^{-\frac{1}{2}\,\delta-1} \exp\left\{-\frac{\frac{1}{2}\,\delta\,\varepsilon^2}{y}\right\}, \tag{12.4.17}$$

for $y > 0$ with $\delta > 0$ degrees of freedom and scaling parameter $\varepsilon$, where $\Gamma(\cdot)$ denotes the gamma function, see (1.2.10). Note that we model here the density of the *inverse* of a random variable that is gamma distributed.

After rearrangement of (12.4.14) we obtain the formula

$$C(x) = \frac{1}{2\,\bar{p}_{\theta^2}(x)} \frac{d(D(x)^2 \bar{p}_{\theta^2}(x))}{dx}, \tag{12.4.18}$$

for $x > 0$. The function $C(\cdot)$ is therefore solely determined by the probability density function $\bar{p}_{\theta^2}(\cdot)$ for the stationary density of $|\theta|^2$ and the function $D(\cdot)$.

To be specific and obtain still a rich class of diffusions $X$ we let the diffusion coefficient function $D(\cdot)$ of $X$ have the form of a power function

$$D(x) = \gamma\,x^{\xi}, \tag{12.4.19}$$

for $x > 0$ with some positive constants $\gamma$ and $\xi$. This particular choice for the functional form of $D(\cdot)$ ensures that the diffusion coefficient of the squared volatility approaches zero when the squared volatility approaches zero. Furthermore, the exponent $\xi$ controls the feedback of the squared volatility on its diffusion coefficient. With this functional form for $D(\cdot)$ and the probability density function $\bar{p}_{\theta^2}$ in (12.4.17), the equation (12.4.18) provides for the squared volatility process $X$ the drift function

$$C(x) = k\,x^{2(\xi-1)}\,(\bar{\theta}^2 - x), \tag{12.4.20}$$

for $x > 0$, where $k = \frac{1}{4}\,\gamma^2\,(\delta + 2 - 4\,\xi)$, $\bar{\theta}^2 = \frac{\varepsilon^2\,\delta}{\delta+2-4\,\xi}$ and $\delta > 4\,\xi - 2$. In the special case $\delta + 2 - 4\xi = 0$ we set $k\bar{\theta}^2 = \frac{\gamma^2 \varepsilon^2 \delta}{4}$. The resulting family of discounted GOP models is, therefore, characterized by a squared volatility with SDE

$$d|\theta_t|^2 = k\,|\theta_t|^{4(\xi-1)}\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,|\theta_t|^{2\,\xi}\,d\bar{W}_t, \tag{12.4.21}$$

where $k, \bar{\theta}, \gamma$ and $\xi$ are all constants. Note that for a desired degree of freedom $\delta$ for the inverse gamma density (12.4.17) the parameters $k$, $\xi$, $\gamma$ and $\bar{\theta}^2$ cannot be chosen freely. In particular, we need to set

$$\delta = \frac{2\,(2\,\xi - 1)}{1 - \frac{\varepsilon^2}{\bar{\theta}^2}}. \tag{12.4.22}$$

It is important to note that several existing models are included in this class of squared volatility models. For the choice of the exponent $\xi = 1$ we obtain the *ARCH diffusion model*

$$d|\theta_t|^2 = k\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,|\theta_t|^2\,d\bar{W}_t, \tag{12.4.23}$$

which is the continuous time limit of the innovation process of the GARCH(1,1) and NGARCH(1,1) models described in Nelson (1990) and Frey (1997). The class of ARCH and GARCH time series models, which have many generalizations, was originally developed in Engle (1982). When taking the continuous time limit in a GARCH(1,1) model, see Nelson (1990), the underlying security and the squared volatility process appear to be driven by independent Wiener processes. The leverage effect, see Sect. 12.1, can be modeled in (12.4.23) when $\bar{W}$ and $W$ are assumed to be negatively correlated.

When the exponent is set to $\xi = \frac{3}{2}$ and $\varrho = -1$ we obtain the 3/2 *model*

$$d|\theta_t|^2 = k\,|\theta_t|^2\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,(|\theta_t|^2)^{\frac{3}{2}}\,d\bar{W}_t. \tag{12.4.24}$$

It corresponds to the squared volatility model suggested in Platen (1997) and covers the volatility structure of the *stylized MMM* mentioned in Sect. 7.5, see Platen (2001, 2002). We remark that in Lewis (2000) a version of a 3/2 model was studied among other models.

### ARCH Diffusion Model

Let us now investigate effects generated by the ARCH diffusion model considered in Hurst (1997), Lewis (2000) and Heath, Hurst & Platen (2001). This means, we consider the case $\xi = 1$ in (12.4.21).

In the following, the speed of adjustment is chosen to be $k = 2.0$ so that the half-life time of shocks, $\frac{\ln(2)}{k}$, is approximately eighteen weeks. The volatility of the squared volatility is set to $\gamma = 1.0$ so that the volatility of volatility is approximately 0.5. These choices for $k$ and $\gamma$ ensure a strong stochastic volatility effect. The initial squared volatility $|\theta_0|^2$ and the long term mean of the squared volatility $\bar{\theta}^2$ are both chosen to equal 0.04 so that initial and long term volatility are approximately 0.2. Furthermore, the correlation $\varrho$ is, for simplicity, first set to zero. The initial discounted GOP value is set to $\bar{S}_0^{\delta_*} = 100$ with a short rate of $r = 0.04$. The effect of changing each of these parameters, while keeping the others constant, is examined below.

Figure 12.4.1 displays the implied volatility surface for European call options with maturity dates, ranging from five weeks to one year and the strike $K$ ranging from 80 to 125. These kind of implied volatility surfaces are often observed for currency and equity options but not for index options. Note that the magnitude of the implied volatility smile or curvature decreases as the time to maturity increases. The smile effect for short dated options is very prominent but becomes less pronounced for longer dated options.

The effect of changing the correlation $\varrho$ on the implied volatility surface is now examined. Figure 12.4.2 shows implied volatilities for European call options with correlation $\varrho$ ranging from -0.5 to 0.5 and the strike $K$ ranging from 80 to 125, where the time to maturity is six months. When the correlation is negative it can be seen that out-of-the-money options have lower implied

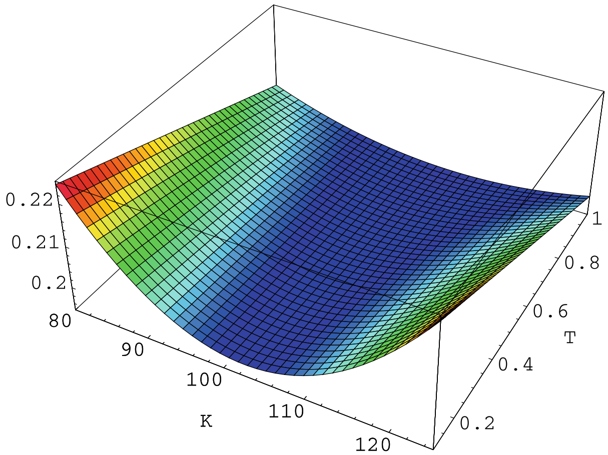



**Fig. 12.4.1.** Implied volatility surface for zero correlation

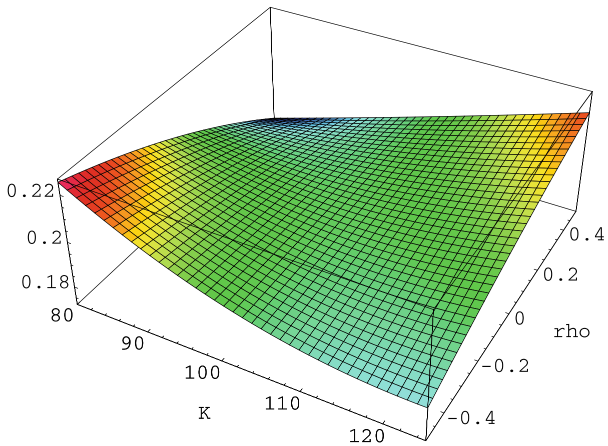



**Fig. 12.4.2.** Effect of changing correlation on implied volatilities

volatilities than in-the-money options. This effect is commonly called a negative implied volatility skew. Usually, for index options the implied volatilities are negatively skewed, see Fig. 12.1.5, reflecting the leverage effect created by negatively correlated index and volatility increments. Thus, with the choice $\varrho < 0$ the typical implied volatility curves for indices can be generated. Note that for $\varrho > 0$ a strong positively skewed implied volatility curve can be obtained.

Figure 12.4.3 displays implied volatilities for European call options with the speed of adjustment parameter $k$ ranging from 1 to 20, the strike $K$ ranging from 80 to 125 and where the time to maturity is six months. It can be observed that as the speed of adjustment parameter $k$ increases, the magnitude of the implied volatility smile decreases.

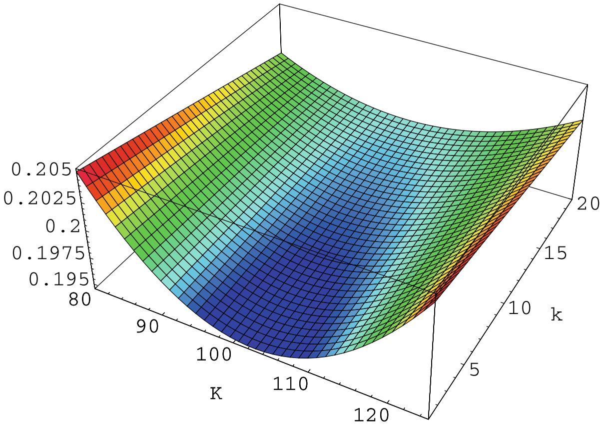



**Fig. 12.4.3.** Effect of changing speed of adjustment on implied volatilities

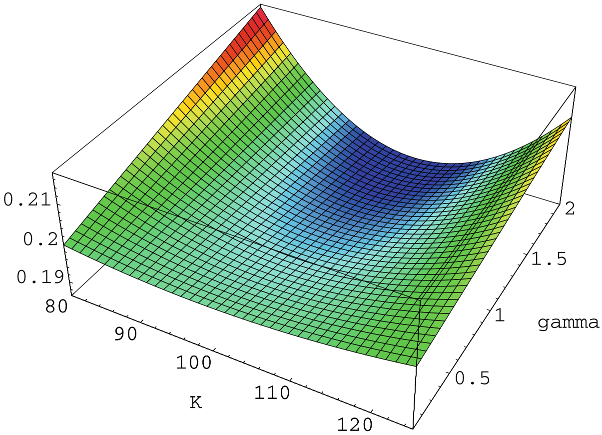



**Fig. 12.4.4.** Effect of changing volatility of the squared volatility on implied volatilities

Figure 12.4.4 depicts implied volatilities for European call options with the volatility $\gamma$ of squared volatility ranging from 0.1 to 2, or volatility of volatility ranging from approximately 0.5 to 1, the strike $K$ ranging from 80 to 125 and where the time to maturity is six months. Note that, as the volatility of squared volatility increases, the magnitude of the implied volatility smile increases. For $\gamma = 0$ we have a version of the BS model with no deformation or curvature in the implied volatility surface.

It is apparent that the ARCH diffusion model in (12.4.23) captures some of the typical properties of implied volatilities observed in index option markets, see Fig. 12.1.5. However, to generate such a negatively skewed implied volatility surface one needs to consider a rather strong negative correlation parameter $\varrho < 0$. As is evident from Fig. 12.4.2, the ARCH diffusion model with strong negative correlation can generate the negative skew pattern. Therefore, it can model some leverage effect. However, it requires a separate stochastic volatility process to achieve this. Other stochastic volatility models produce similar results to what has been demonstrated above for the ARCH diffusion model, see, for instance, Cont & Tankov (2004) and results on the MMM in the next chapter. This makes it difficult to decide which is potentially a better model.

An important drawback of the above stochastic volatility models is that they are genuine two-factor models, driven by two separate stochastic processes. This makes it a complex numerical task to value even standard index derivatives. A parsimonious, economically based one-factor model, which can generate similar skews and smiles in implied volatility surfaces, would be preferable. In particular, if it could explain the nature of the dynamics of the underlying index.

## 12.5 Exercises for Chapter 12

**12.1.** Prove that the ARCH diffusion model for squared volatility

$$d|\theta_t|^2 = \kappa\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,|\theta_t|^2\,dW_t$$

has an inverse gamma density as stationary density.

**12.2.** Show that the squared volatility of the model

$$d|\theta_t|^2 = \kappa\,|\theta_t|^2\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,|\theta_t|^3\,dW_t$$

has an inverse gamma density.

**12.3.** Compute the stationary density for the squared volatility for the Heston model

$$d|\theta_t|^2 = \kappa\,(\bar{\theta}^2 - |\theta_t|^2)\,dt + \gamma\,||\theta_t||\,dW_t.$$

**12.4.** Calculate the stationary density for the squared volatility $|\theta_t|^2$ of the Scott model, where

$$d|\theta_t| = \kappa\,(\bar{\theta} - |\theta_t|)\,dt + \gamma\,dW_t.$$

Characterize the type of the stationary density ?

**12.5.** Calculate the stationary density for the squared volatility $|\theta_t|^2$, which satisfies the SDE

$$d\ln(|\theta_t|^2) = \kappa\,(\bar{\xi} - |\theta_t|^2)\,dt + \gamma\,dW_t.$$

Which type of density is this ?

**12.6.** (*) Show under the modified CEV model that the benchmarked savings account and, thus, the benchmarked savings bond are strict local martingales.

# 13

# Minimal Market Model

This chapter derives an alternative model for the long term dynamics of the GOP from basic economic arguments. The discounted GOP drift, which models the long term trend of the economy, is chosen as the key parameter process. This leads to the minimal market model with the discounted GOP forming a time transformed squared Bessel process of dimension four. Its dynamics allows us to explain various empirical stylized facts and other properties relating to the long term behavior of a world stock index.

## 13.1 Parametrization via Volatility or Drift

### Volatility Parametrization

The market portfolio can be interpreted as an accumulation index, or total return index. Let us again assume that a diversified stock market index approximates the GOP. The SDE (10.2.8) of the GOP reveals a close link between its drift and diffusion coefficient. More precisely, the risk premium of the GOP equals the square of its volatility. To see this clearly, we rewrite the SDE (10.2.8) for the discounted GOP when assuming a CFM, see Definition 10.1.2, in the form

$$d\bar{S}_t^{\delta_*} = \bar{S}_t^{\delta_*} \, |\theta_t| \, (|\theta_t| \, dt + dW_t), \tag{13.1.1}$$

where

$$dW_t = \frac{1}{|\theta_t|} \sum_{k=1}^{d} \theta_t^k \, dW_t^k \tag{13.1.2}$$

is the stochastic differential of a standard Wiener process $W$. For the efficient modeling of the GOP it is important to find an appropriate parametrization. The SDE (13.1.1) uses the *volatility parametrization* of the GOP, which can be best identified after application of a logarithmic transformation to $\bar{S}_t^{\delta_*}$. By taking the logarithm of the discounted GOP $\bar{S}_t^{\delta_*}$ we obtain from (13.1.1) the SDE

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

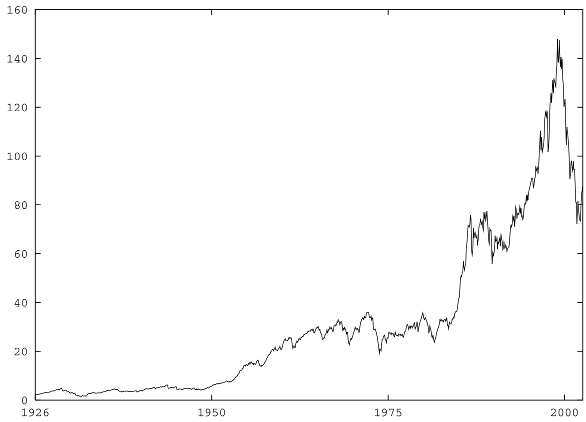



**Fig. 13.1.1.** Discounted WSI

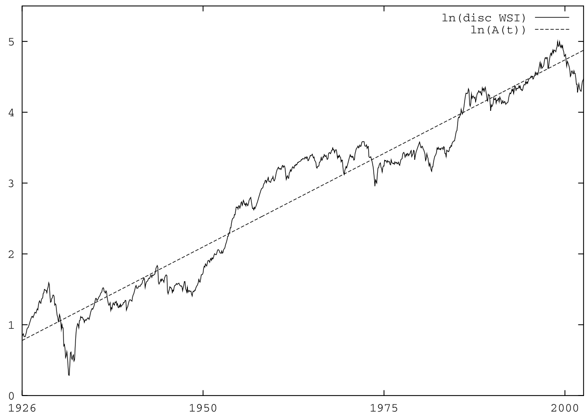



**Fig. 13.1.2.** Logarithm of discounted WSI

$$d \ln \left( \bar{S}_t^{\delta_*} \right) = \frac{1}{2} \left| \theta_t \right|^2 dt + \left| \theta_t \right| dW_t \tag{13.1.3}$$

for $t \in [0, \infty)$, see Exercise 13.1. On the right hand side of this equation only one parameter process appears in the drift and diffusion coefficients.

Figure 13.1.1 shows a discounted world stock index (WSI) observed in US dollars from 1926 until 2004. This index starts at $\bar{S}_0^{\delta_*} = 2.3$ in January 1926 and has been reconstructed from monthly data provided by Global Financial Data.

The logarithm of the above discounted WSI is displayed in Fig. 13.1.2. One notes that the logarithm of the discounted WSI increases on average linearly with some fluctuations and could be potentially related to some underlying stationary process. One possibility is to model the logarithm of the discounted WSI by a simple time transformed Wiener process, a Lévy process or another process with independent increments. However, the increasing variance of such a process over time would not match the dynamics that we observe. Therefore, there has to be some feedback effect modeled that drives the logarithm of the index back to its long-term average linear growth, consistent with stationary variance.

It is apparent that constant volatility is not compatible with stationary variance for the logarithm of the discounted WSI. Also if volatility is stochastic, stationary and independent of the driving noise of the WSI, then the variance of the logarithm of the WSI is going to increase over time. Thus, it would seem that some dependence between the WSI and its volatility needs to be established in a reasonable model for its long term dynamics.

Unfortunately, volatility does not have a major economic interpretation and is difficult to observe, see Corsi et al. (2001) and Barndorff-Nielsen & Shephard (2003). It simply emerges as a traditional parameter process in an attempt to model the random fluctuations or local risk of asset prices via the logarithmic transformation. The use of volatility as a parameter process grew historically from an early practice that employed geometric Brownian motion in the modeling of asset prices, see Osborne (1959), Samuelson (1971) and Black & Scholes (1973). However, in recent years growing concerns have emerged about the deficiencies of geometric Brownian motion as an asset price model. A major problem is the fact that volatility is, in reality, stochastic, see Fig. 12.1.1. Many other parameterizations of asset price dynamics are possible. Ideally, there should be an economically based or plausible parametrization which may then explain a potential link between the changes in the WSI and its volatility. As explained in Sect. 12.1, the volatility of an index has, via the leverage effect, some qualitative link to the value of the underlying index. Still, the leverage effect should also be explained on the basis of formal economic reasoning and, ideally, should be specified quantitatively.

### Drift Parametrization

From an economic perspective it is clear that the WSI needs always to revert back to its underlying economic value even if this may take a long time. This property is a consequence of the conservation of value in an economy. We have observed that the drift of the discounted GOP can be interpreted as the change per unit of time of its underlying economic value. This drift provides an important link between the long term average evolution of the market index and the long term growth of the macro economy. By the law of conservation of value, the growth rate of the discounted index should in the long term, on average, match the growth rate of the total net wealth of the companies which comprise the market portfolio. Therefore, let us parameterize the discounted GOP dynamics, that is the SDE (13.1.1), by its trend. More precisely, we consider the *discounted GOP drift*

$$\alpha_t^{\delta_*} = \bar{S}_t^{\delta_*} \, |\theta_t|^2 \tag{13.1.4}$$

for $t \in [0, \infty)$, which is assumed to be a strictly positive, predictable parameter process, see (11.1.4). Using this parametrization we obtain from (13.1.4) the volatility $|\theta_t|$ of the GOP in the form

$$|\theta_t| = \sqrt{\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}}. \tag{13.1.5}$$

This structure provides a natural explanation for the leverage effect. When the index decreases, then the volatility increases and vice versa. This creates a feedback effect resulting from the structure of the SDE (13.1.1) for the discounted GOP.

By substituting (13.1.4) and (13.1.5) into (13.1.1), we obtain the following parametrization of the SDE of the discounted GOP:

$$d\bar{S}_t^{\delta_*} = \alpha_t^{\delta_*}\, dt + \sqrt{\bar{S}_t^{\delta_*}\, \alpha_t^{\delta_*}}\, dW_t \tag{13.1.6}$$

for $t \in [0, \infty)$. We emphasize that the square root of the discounted GOP appears in the diffusion coefficient. Note that the parameter process $\alpha^{\delta_*} = \{\alpha_t^{\delta_*},\, t \in [0, \infty)\}$ can be freely specified as a predictable stochastic process such that the SDE (13.1.6) has a unique strong solution.

With the quantity

$$A_t = A_0 + \int_0^t \alpha_s^{\delta_*}\, ds \tag{13.1.7}$$

we can rewrite (13.1.6) in the form

$$\bar{S}_t^{\delta_*} = \bar{S}_0^{\delta_*} + A_t - A_0 + \int_0^t \sqrt{\bar{S}_s^{\delta_*}\, \alpha_s^{\delta_*}}\, dW_s \tag{13.1.8}$$

for $t \in [0, \infty)$. Here $A_t$ can be interpreted as the *underlying value* at time $t$ of the discounted GOP, where $A_0$ needs to be appropriately chosen as the initial underlying value at time $t = 0$. One can say that the underlying value $A_t$ corresponds to the discounted wealth that underlies the discounted index $\bar{S}^{\delta_*}$. The drift parametrization above has, therefore, a formal economic meaning. If one expects the fluctuations of the increase per unit of time of the discounted underlying value to be reasonably independent of trading uncertainty, then the fitting of a model to market data is more likely to be effective and amenable to this drift parametrization rather than to the alternative formulation using volatility.

### Squared Bessel Process of Dimension Four

It is important to realize that the SDE (13.1.6) describes a very particular time transformed diffusion process. More precisely, it is the SDE of a time transformed squared Bessel process of dimension four, see Sect. 8.7 and Revuz & Yor (1999).

More precisely, with the specification of transformed time $\varphi(t)$ as

$$\varphi(t) = \frac{1}{4} \int_0^t \alpha_s^{\delta_*} \, ds \tag{13.1.9}$$

and with

$$X_{\varphi(t)} = \bar{S}_t^{\delta_*} \tag{13.1.10}$$

we obtain from (13.1.6) the SDE of a squared Bessel process of dimension four in the form

$$dX_{\varphi(t)} = 4\, d\varphi(t) + 2\sqrt{X_{\varphi(t)}}\, dW(\varphi(t)), \tag{13.1.11}$$

see (8.7.1), where

$$dW(\varphi(t)) = \sqrt{\frac{\alpha_t^{\delta_*}}{4}}\, dW_t \tag{13.1.12}$$

for $t \in [0, \infty)$. By (13.1.9) the increase of the transformed time equals a quarter of the underlying value $A_t$. This provides a simple economically founded parametrization of the discounted GOP dynamics. In the model the transformed time can be interpreted as business time or market time.

Note that we have still not specified the dynamics of the discounted GOP $\bar{S}_t^{\delta_*}$ because we have not fixed the dynamics of the discounted GOP drift process $\alpha_t^{\delta_*}$. So far almost any strictly positive, predictable process is possible here. The discounted GOP dynamics are in (13.1.6) and (13.1.11) only parameterized in an alternative way by using the drift instead of the volatility as parameter process.

### Time Transformed Bessel Process

By application of the Itô formula to the square root of the discounted GOP one obtains from (13.1.6) the SDE

$$d\sqrt{\bar{S}_t^{\delta_*}} = \frac{3\,\alpha_t^{\delta_*}}{8\sqrt{\bar{S}_t^{\delta_*}}}\, dt + \frac{1}{2}\sqrt{\alpha_t^{\delta_*}}\, dW_t, \tag{13.1.13}$$

see Exercise 13.2. This is the SDE of a time transformed Bessel process of dimension four, see (7.7.19). In a CFM the quadratic variation of $\sqrt{\bar{S}^{\delta_*}}$ equals

$$\left[\sqrt{\bar{S}^{\delta_*}}\right]_t = \frac{1}{4} \int_0^t \alpha_s^{\delta_*} \, ds \tag{13.1.14}$$

for $t \in [0, \infty)$, see Sect. 5.2. This means that by (13.1.9) in a CFM the increment of the transformed time $\varphi(t)$ equals the quadratic variation of the time transformed Bessel process $\sqrt{\bar{S}^{\delta_*}}$. That is

$$\varphi(t) - \varphi(0) = \left[\sqrt{\bar{S}^{\delta_*}}\right]_t \tag{13.1.15}$$

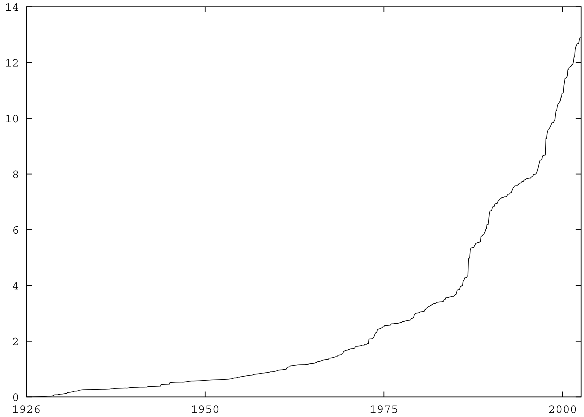



**Fig. 13.1.3.** Empirical quadratic variation of the square root of the discounted WSI

for $t \in [0, \infty)$. This is a surprisingly simple relationship. Hence, the transformed time process can be determined, in principle, from the quadratic variation of the square root of the discounted GOP, an observable quantity if we take the WSI as proxy for the GOP. For the discounted WSI shown in Fig. 13.1.1 we plot in Fig. 13.1.3 the empirical quadratic variation of its square root. One notes a reasonably smooth increase of this quadratic variation over the long time period.

We emphasize that so far we have not made any assumptions about the particular dynamics of the discounted GOP. The relationships revealed under the given drift parametrization hold generally for any CFM. In the next section we shall choose the discounted GOP drift as having a simple exponential function of time.

## 13.2 Stylized Minimal Market Model

Let us now apply the above results for the derivation of a parsimonious index model, the *minimal market model* (MMM), see Platen (2001, 2002, 2006c) and Sect. 7.5.

### Net Growth Rate

By conservation of value the long-term growth rate of the underlying value of the discounted GOP can be expected to correspond to the long-term net growth rate of the world economy. According to historical records we assume in the long term, as a first approximation, that the world economy has been growing exponentially, see Fig. 13.1.2. Such exponential growth will now be postulated for the discounted GOP drift. The following assumption leads us to the stylized version of the MMM.

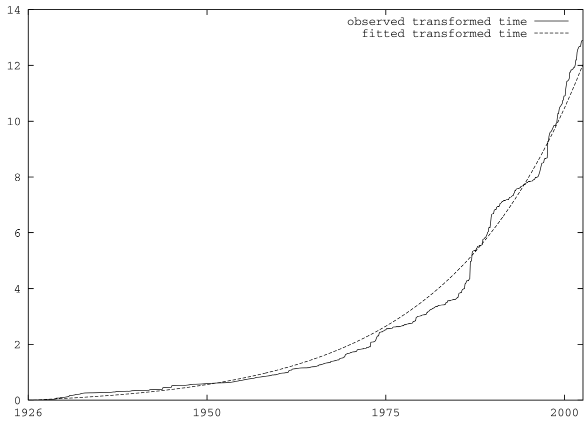



**Fig. 13.2.1.** Fitted and observed transformed time

**Assumption 13.2.1.** *The discounted GOP drift is an exponentially growing function of time.*

Note that this assumption can be considerably weakened and made more flexible, as will be shown later in Sect. 13.4, see also Heath & Platen (2005b). To satisfy Assumption 13.2.1 let us model the discounted GOP drift $\alpha_t^{\delta_*}$ as an exponential function of time of the form

$$\alpha_t^{\delta_*} = \alpha_0 \exp\{\eta\, t\} \tag{13.2.1}$$

for $t \in [0, \infty)$. In this equation we have as parameters, a nonnegative *initial value* $\alpha_0 > 0$ and a constant *net growth rate* $\eta > 0$. Note that the initial value parameter $\alpha_0$ depends on the initial date and also on the initial value of the discounted GOP. By equations (13.1.9) and (13.2.1) the underlying value at time $t$ satisfies under the given parametrization the equation

$$\varphi(t) = \frac{\alpha_0}{4} \int_0^t \exp\{\eta\, z\}\, dz \tag{13.2.2}$$

for $t \in [0, \infty)$. This demonstrates that the transformed time and the underlying value evolve asymptotically for long time periods in an exponential manner. More precisely, one obtains for the transformed time the explicit expression

$$\varphi(t) = \frac{\alpha_0}{4\,\eta} \left(\exp\{\eta\, t\} - 1\right). \tag{13.2.3}$$

By applying standard curve fitting methods, see Sect. 2.3, we fit the transformed time $\varphi(t)$, satisfying (13.2.3), to the observed quadratic variation of the square root of the discounted WSI shown in Fig. 13.1.3. In Fig. 13.2.1 we plot then the resulting fit for the parameter choice $\alpha_0 = 0.043$ and $\eta = 0.0528$. One notes that we achieve a reasonable fit of the theoretical transformed time when only using a constant net growth rate $\eta$ over the long time period.

The net growth rate for the market capitalization weighted world stock portfolio, when discounted by the US dollar savings account, has been estimated for the entire last century in Dimson et al. (2002) to be on average close to 0.049 under discrete annual compounding. This is reasonably close to what we have obtained as the annual net growth rate parameter $\eta = 0.0528$ under continuous compounding, as shown in Fig. 13.2.1.

## Normalized GOP

We now discuss the feedback effect in the dynamics of the market index that drives its value back to its long term exponentially growing average. The formulation (13.2.1) suggests that one should examine for this purpose the *normalized GOP*

$$Y_t = \frac{\bar{S}_t^{\delta_*}}{\alpha_t^{\delta_*}} \tag{13.2.4}$$

for $t \in [0, \infty)$.

By application of the Itô formula and using (13.1.5) and (13.1.6), we obtain for this case the SDE

$$dY_t = (1 - \eta\, Y_t)\, dt + \sqrt{Y_t}\, dW_t \tag{13.2.5}$$

for $t \in [0, \infty)$ with

$$Y_0 = \frac{\bar{S}_0^{\delta_*}}{\alpha_0}, \tag{13.2.6}$$

see Exercise 13.3. Note by (8.7.34) that $Y$ is a square root (SR) process of dimension four. The above stylized version of the MMM is an economically based, parsimonious model for the dynamics of the discounted GOP, and by extension for a WSI. We remark that we would still obtain the above type of SDE for $Y_t$ if $\eta$ were a stochastic process. This is rather important for extended versions of the MMM.

By using the SR process $Y = \{Y_t,\, t \in [0, \infty)\}$ and (13.2.5), the discounted GOP $\bar{S}_t^{\delta_*}$ can be expressed in the form

$$\bar{S}_t^{\delta_*} = Y_t\, \alpha_t^{\delta_*} \tag{13.2.7}$$

for $t \in [0, \infty)$. This leads us to a useful description of the GOP when expressed in units of the domestic currency given by

$$S_t^{\delta_*} = S_t^0\, \bar{S}_t^{\delta_*} = S_t^0\, Y_t\, \alpha_t^{\delta_*} \tag{13.2.8}$$

for $t \in [0, \infty)$. For the above model of the discounted GOP one needs only to specify the initial values $\bar{S}_0^{\delta_*}$ and $\alpha_0$ and the net growth rate process $\eta$. Note that $\alpha_0$ and $\bar{S}_0^{\delta_*}$ are linked through (13.2.6). Consequently, one can say that the stylized MMM assumes that the discounted GOP is the product of an SR process and an exponential function.

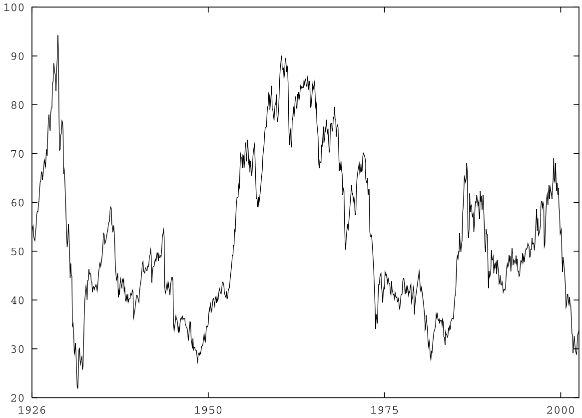



**Fig. 13.2.2.** Normalized GOP

One notes that the normalized GOP is an SR process of dimension four, see Sects. 4.4 and 7.5. The net growth rate $\eta$ is here the speed of adjustment parameter for the linear mean-reversion. Note that besides the scalar initial values $\alpha_0$ and $\bar{S}_0^{\delta_*}$, the net growth rate $\eta$ is the only parameter process needed for the characterization of the dynamics of the normalized GOP under the MMM.

According to our previous findings for the discounted WSI shown in Fig. 13.1.1 we set $\eta = 0.0528$ and choose $\alpha_0 = 0.043$. With this parameter choice we show in Fig. 13.2.2 the resulting normalized GOP $Y_t$, constructed according to (13.2.4) and (13.2.1).

For the above choice of $\eta$ the half life time of a major displacement of the normalized GOP would be about $\frac{\ln(2)}{\eta} \approx 13$ years. This rather long time period supports the view that it takes on average significant time to correct for major up- or downturns in the world financial market. One realizes that a look at the market performance over the last 10 or even 15 years may not be sufficient to judge its potential long term evolution. This is consistent with the impression that one obtains when studying in Fig. 13.1.2 the logarithm of the world stock index for the long period from 1926 until 2003. It seems to take about 25 years in this graph to go through a full "cycle" of random ups and downs for the market index. The MMM reflects well this type of long term mean reverting dynamics of the normalized GOP.

Since the normalized GOP $Y_t$ has for constant net growth rate a stationary density, so has $\ln(Y_t)$. This means, that $\ln(Y_t)$ has a uniformly bounded variance for all $t \in [0, \infty)$. In this sense the stylized MMM with constant parameters exhibits some kind of an equilibrium dynamic. By taking the logarithm on both sides of equation (13.2.7) we obtain the relation

$$\ln(\bar{S}_t^{\delta_*}) = \ln(Y_t) + \ln(\alpha_0) + \eta\, t \tag{13.2.9}$$

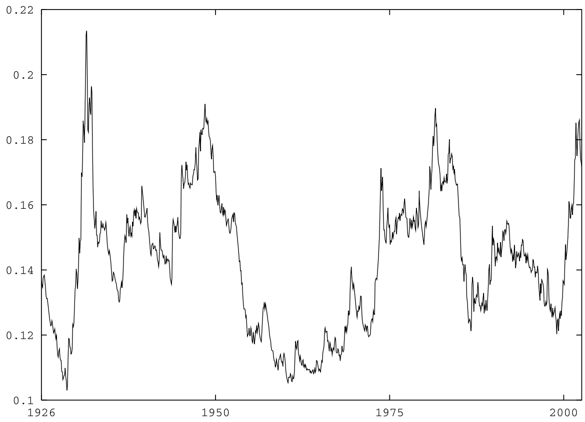



**Fig. 13.2.3.** Volatility of the WSI under the MMM

for $t \in [0, \infty)$, assuming the net growth rate to be a constant. The above stylized MMM suggests the logarithm of the discounted GOP will fluctuate around a straight line with slope $\eta$. The variance of $\ln(\bar{S}_t^{\delta_*})$ is therefore uniformly bounded under the MMM. This is also what one observes in Fig. 13.1.2.

The BS model and most of its extensions, with volatility processes independent of the trading noise, are not able to recover dynamics of the type shown in Fig. 13.1.2. In particular, exponential Lévy models, as mentioned in Sect. 3.6, share this problem. Stochastic volatility models, as discussed in Sect. 12.4, are better suited. However, they need an extra volatility process to generate some negative correlation between volatility and index as described earlier. The MMM is a parsimonious model that does not need any extra volatility process, but still generates a realistic long term dynamics for the GOP.

## Volatility under the MMM

The resulting model for the discounted GOP with constant net growth rate $\eta$ is called the stylized version of the MMM, which was originally proposed in Platen (2001). We now discuss the endogenous nature of volatility as it emerges under the MMM.

According to formula (13.2.4), under the MMM the discounted GOP has the volatility

$$|\theta_t| = \frac{1}{\sqrt{Y_t}} \tag{13.2.10}$$

for $t \in [0, \infty)$. We plot in Fig. 13.2.3 the path of the volatility for the discounted WSI, shown in Fig. 13.1.1, as it follows under the MMM for the default parameters $\eta = 0.0528$ and $\alpha_0 = 0.043$. It is interesting to note that according to this graph the volatility was, for instance, relatively high around 1975 and rather low during the period near the year 2000. The stochastic volatility process of the WSI under the MMM, as shown in Fig. 13.2.3, is negatively correlated to the normalized WSI and, therefore, also negatively correlated to the WSI.

From (13.2.10) and (13.2.5) it can be seen that the squared volatility $|\theta_t|^2$ satisfies the SDE

$$d|\theta_t|^2 = d\left(\frac{1}{Y_t}\right) = |\theta_t|^2\,\eta\,dt - \left(|\theta_t|^2\right)^{\frac{3}{2}}\,dW_t \tag{13.2.11}$$

for $t \in [0,\infty)$, see Exercise 13.4. This provides us with a stochastic volatility model in the sense as discussed in Sect. 12.4. Note that the diffusion coefficient of the squared volatility has in (13.2.11) the power $\frac{3}{2}$. In Platen (1997) such a 3/2 volatility model was suggested for the modeling of a market index, see also (12.4.24). This stochastic volatility model has been obtained under the benchmark approach by using economic arguments.

### Distribution of Log-Returns under the MMM

Let us now examine the distribution of log-returns of the GOP that can be expected to be estimated under the MMM. We show that under the MMM the estimated log-returns of the GOP, from sufficiently long observation periods, are Student $t$ distributed with four degrees of freedom. To see this, let us recall that under the MMM the squared volatility of the GOP is by (13.2.4) given as

$$|\theta_t|^2 = \frac{1}{Y_t}, \tag{13.2.12}$$

which is the inverse of an SR process. Note from (4.5.7) that the SR process $Y$ has as stationary density a gamma density with four degrees of freedom. Consequently, the squared volatility $\frac{1}{Y_t}$ has an *inverse gamma density* as its stationary density.

When estimating the density of, say, daily log-returns that are observed over a long time period, then the stationary density of the squared volatility acts as mixing density for the stochastic variance of the log-returns. This is similar to normal variance mixture models, as discussed in Sect. 2.5, and to stochastic volatility models, as described in Sect. 12.4. We refer also to Kessler (1997), Prakasa Rao (1999) and Kelly, Platen & Sørensen (2004) for more details on this issue. For log-returns of the GOP, the inverse gamma density acts under the MMM as a mixing density for their normal-mixture distribution. It follows from (1.2.16) and (1.2.28) that the resulting normal-mixture distribution is the Student $t$ distribution with four degrees of freedom. Therefore, this is the theoretically predicted log-return density that will be estimated under the stylized MMM dynamics. We emphasize that one needs a sufficiently long time period with log-return observations for this kind of estimation procedure to be reliable. Obviously, the path of the ergodic process $\frac{1}{Y}$ needs enough time to sufficiently act in its mixing role for the random variance of the conditionally Gaussian log-returns. Since we have seen that the half life time of shocks on the square root process $Y$ for the calibrated MMM is about 13 years, the available 33 years of daily data can be possibly considered to be just sufficient to confirm or reject the predicted Student $t$ feature of log-returns.

The above described distributional feature of the MMM is rather clear and testable. Most importantly, the Student $t$ log-return property has already been documented in the literature as an empirical stylized fact, as was pointed out in Sect. 2.6. Recall that Markowitz & Usmen (1996a) found that the Student $t$ distribution with about 4.5 degrees of freedom matches daily S&P500 log-return data well. Hurst & Platen (1997) found within the rich class of symmetric generalized hyperbolic distributions that for most stock market indices, daily log-returns are likely to be Student $t$ distributed with about four degrees of freedom. Fergusson & Platen (2006) confirmed with high accuracy this empirical stylized fact. Furthermore, in Breymann et al. (2003) the copula, see Sect. 1.5, of the joint distribution of log-returns of exchange rates has been identified as a Student $t$ copula with roughly four degrees of freedom. One can say that the MMM provides in its stylized version a rather accurate model for the probabilistic nature of the log-returns of a world stock index.

### Stylized Multi-Currency MMM (*)

We have examined under the MMM the properties of stochastic volatility for the discounted GOP in a currency denomination. By using the same arguments as above, we now show how to model exchange rates. This will result in a stylized multi-currency version of the MMM, similar to the one described in Platen (2001) and Heath & Platen (2005a).

Let us consider a market with $d+1$ currencies, $d \in \mathcal{N}$. We denote by $S_i^{\delta_*}(t)$ the GOP at time $t$ when denominated in units of the $i$th currency, $i \in \{0, 1, \ldots, d\}$. Furthermore, $r_t^i$ is the short rate for the $i$th currency and $\theta_i^k(t)$ the market price of risk for the $i$th currency denomination with respect to the $k$th Wiener process, $k \in \{1, 2, \ldots, d+1\}$, $t \in [0, \infty)$.

We derive now a stylized multi-currency version of the MMM. Assuming, for simplicity, constant net growth rates and constant short rates, we can describe at time $t$ the value of the GOP in the $i$th currency denomination according to (13.2.8) by the expression

$$S_i^{\delta_*}(t) = \alpha_t^i \, Y_t^i \, S_i^i(t). \tag{13.2.13}$$

Here we have

$$\alpha_t^i = \alpha_0^i \, \exp\{\eta^i \, t\}, \tag{13.2.14}$$

$$S_i^i(t) = \exp\{r^i \, t\}. \tag{13.2.15}$$

The $i$th normalized GOP $Y_t^i$ satisfies the SDE

$$dY_t^i = \left(1 - \eta^i\, Y_t^i\right) dt + \sqrt{Y_t^i}\, \sum_{k=1}^{d+1} q^{i,k}\, dW_t^k \tag{13.2.16}$$

for $t \in [0, \infty)$, with $Y_0^i > 0$ and $\eta^i$ the $i$th net growth rate, $i \in \{0, 1, \ldots, d\}$. Furthermore, we introduce constant *scaling levels* $q^{i,k}$, for $i \in \{0, 1, \ldots, d\}$ and $k \in \{1, 2, \ldots, d+1\}$ to model the covariations between normalized GOPs $Y^i$ and $Y^j$ for $i \neq j$. For the stylized multi-currency version of the MMM we set, for simplicity,

$$\sum_{k=1}^{d+1} (q^{i,k})^2 = 1 \tag{13.2.17}$$

for all $i \in \{0, 1, \ldots, d\}$. This constraint can be relaxed in extended versions of the MMM.

The $(i, j)$th exchange rate $X_t^{i,j}$ from the $j$th into the $i$th currency is given at time $t$ by the ratio

$$X_t^{i,j} = \frac{S_i^{\delta_*}(t)}{S_j^{\delta_*}(t)} = \frac{Y_t^i\, \alpha_t^i\, S_i^i(t)}{Y_t^j\, \alpha_t^j\, S_j^j(t)}. \tag{13.2.18}$$

This satisfies the SDE

$$dX_t^{i,j} = X_t^{i,j} \left( (r^i - r^j)\, dt + \sum_{k=1}^{d+1} \left( \frac{q^{i,k}}{\sqrt{Y_t^i}} - \frac{q^{j,k}}{\sqrt{Y_t^j}} \right) \left( \frac{q^{i,k}}{\sqrt{Y_t^i}}\, dt + dW_t^k \right) \right) \tag{13.2.19}$$

for $t \in [0, \infty)$ with $X_0^{i,j} > 0$, $i, j \in \{0, 1, \ldots, d\}$. Hence, this is the dynamics for an exchange rate, consistent with that of the GOP having the structure (13.2.13) and (13.2.16) in each currency denomination. Under the multi-currency MMM the exchange rate volatility depends on the volatilities of the GOPs in both currencies and, thus, on the fluctuations of the GOP in both denominations.

The $j$th savings account, when denominated in the $i$th currency, is given by the product

$$S_i^j(t) = X_t^{i,j}\, S_j^j(t). \tag{13.2.20}$$

Consequently, by the Itô formula it satisfies the SDE

$$dS_i^j(t) = S_i^j(t) \left( r^i\, dt + \sum_{k=1}^{d+1} \left( \frac{q^{i,k}}{\sqrt{Y_t^i}} - \frac{q^{j,k}}{\sqrt{Y_t^j}} \right) \left( \frac{q^{i,k}}{\sqrt{Y_t^i}}\, dt + dW_t^k \right) \right) \tag{13.2.21}$$

for all $t \in [0, \infty)$ with $S_i^j(0) > 0$ for $i, j \in \{0, 1, \ldots, d\}$.

One notes from (10.1.7) that the stochastic market price of risk with respect to the $k$th Wiener process under the $i$th currency denomination is of the form

$$\theta_i^k(t) = \frac{q^{i,k}}{\sqrt{Y_t^i}} \tag{13.2.22}$$

for all $t \in [0, \infty)$, $i \in \{0, 1, \ldots, d\}$ and $k \in \{1, 2, \ldots, d+1\}$. The $(j,k)$th volatility in the $i$th denomination is given by the expression

$$b_i^{j,k}(t) = \theta_i^k(t) - \theta_j^k(t) \tag{13.2.23}$$

for $t \in [0, \infty)$, $i, j \in \{0, 1, \ldots, d\}$ and $k \in \{1, 2, \ldots, d+1\}$ and is, therefore, stochastic. One notes that the volatility of an exchange rate is different and more complex than that of an index. The interplay between the volatilities of the denominations of the GOP in two different currencies under the MMM is visible in the above volatility structure of the corresponding exchange rate.

In fact, the above stylized multi-currency MMM, which characterizes a currency market, can be used to model an equity market. For equity markets the exdividend spot price of a stock is treated in the same manner as an exchange rate. The share savings account of a cum dividend stock is then similar to that of a foreign savings account. The dividend rate plays a similar role to that of the short rate for a foreign savings account. In Platen & Stahl (2003) it is shown that log-returns of many benchmarked US stocks are Student $t$ distributed with about four degrees of freedom. This suggests that, potentially, the above stylized multi-currency MMM can also be applied to a number of stocks.

In this context it is worth mentioning that the spot price of a commodity, like gold, copper, oil or electricity, can also be modeled like an exchange rate. Here, the, so-called, convenience yield, see Miltersen & Schwartz (1998), behaves in a similar manner as the foreign short rate. In this sense the above stylized multi-currency MMM can be used to model commodity prices. Forthcoming work will identify the dynamics of the GOP when denominated in units of equities or commodities.

## 13.3 Derivatives under the MMM

This section derives pricing formulas for standard derivatives under the stylized MMM. This includes zero coupon bonds, as well as, call and put options on an index. In this section we rely on the methodology presented in Chap. 12.

### Zero Coupon Bond under the MMM

First, we study the price of a zero coupon bond under the stylized MMM. For simplicity, we assume that the short rate $r_t$ is deterministic and the net growth rate $\eta$ is constant. The price $P(t,T)$ of a zero coupon bond that matures at time $T \in (0, \infty)$ is by the real world pricing formula (9.1.34) and (10.4.1) obtained from the conditional expectation

$$P(t,T) = S_t^{\delta_*}\, E\left(\frac{1}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) = \exp\left\{-\int_t^T r_s\, ds\right\} E\left(\frac{\bar{S}_t^{\delta_*}}{\bar{S}_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) \tag{13.3.1}$$

for $t \in [0,T]$. We recall that the discounted GOP $\bar{S}^{\delta_*}$ is a time transformed squared Bessel process of dimension $\delta = 4$. As in (13.2.1) we choose the discounted GOP drift

$$\alpha_t^{\delta_*} = \alpha_0\, \exp\{\eta\, t\} \tag{13.3.2}$$

with initial value $\alpha_0 > 0$ and constant net growth rate $\eta > 0$. The corresponding time transformation is given in (13.2.3) by

$$\varphi(t) = \frac{\alpha_0}{4\,\eta}\, (\exp\{\eta\, t\} - 1) \tag{13.3.3}$$

for $t \in [0,\infty)$, where we set $\varphi(0) = 0$. We know by the formula (8.7.17) the first negative moment of a squared Bessel process of dimension $\delta = 4$, which has the form

$$E\left(\left(\bar{S}_T^{\delta_*}\right)^{-1} \,\Big|\, \mathcal{A}_t\right) = \left(\bar{S}_t^{\delta_*}\right)^{-1} \left(1 - \exp\left\{-\frac{\bar{S}_t^{\delta_*}}{2\,(\varphi(T) - \varphi(t))}\right\}\right). \tag{13.3.4}$$

Therefore, we obtain by (13.3.1) and (13.3.4) the price for the fair zero coupon bond

$$P(t,T) = \exp\left\{-\int_t^T r_s\, ds\right\} \left(1 - \exp\left\{-\frac{\bar{S}_t^{\delta_*}}{2\,(\varphi(T) - \varphi(t))}\right\}\right) \tag{13.3.5}$$

for $t \in [0,T)$. Hence, for the stylized MMM an explicit formula exists for the price of a zero coupon bond, which was originally derived in Platen (2002).

### Forward Rates under the MMM

As introduced in Sect. 10.4, the forward rate $f(t,T)$ at time $t$ for the maturity date $T \in (0,\infty)$ is given by the formula

$$f(t,T) = -\frac{\partial}{\partial T}\, \ln(P(t,T)) \tag{13.3.6}$$

for $t \in [0,T)$. Using (13.3.5) for the stylized MMM with deterministic short rate, the forward rate follows in the form

$$f(t,T) = r_T + n(t,T), \tag{13.3.7}$$

where $n(t,T)$ describes the *market price of risk contribution*

$$n(t,T) = -\frac{\partial}{\partial T}\, \ln\left(1 - \exp\left\{-\frac{\bar{S}_t^{\delta_*}}{2(\varphi(T) - \varphi(t))}\right\}\right). \tag{13.3.8}$$

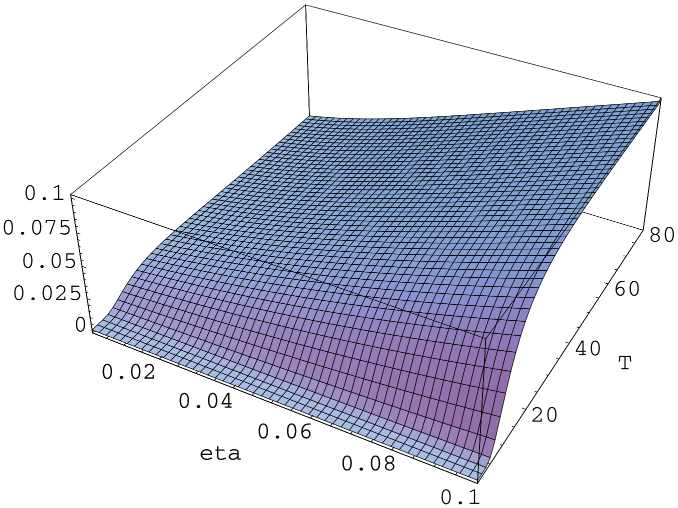



**Fig. 13.3.1.** Market price of risk contribution in dependence on $\eta$ and $T$

In our case with a deterministic short rate, the forward rate is the sum of the short rate at the maturity date and the market price of risk contribution. The existence of a nonzero market price of risk contribution is a consequence of the fact that the stylized MMM does not have an equivalent risk neutral probability measure. By performing the differentiation in (13.3.8) we obtain the equation

$$\begin{aligned} n(t,T) &= \frac{1}{\left(\exp\left\{\frac{\bar{S}_t^{\delta_*}}{2(\varphi(T)-\varphi(t))}\right\}-1\right)} \frac{\bar{S}_t^{\delta_*}}{(\varphi(T)-\varphi(t))^2} \frac{\alpha_T^{\delta_*}}{8} \\ &= \frac{2\,\eta^2\,Y_t}{\left(\exp\left\{\frac{2\,\eta\,Y_t}{(\exp\{\eta\,(T-t)\}-1)}\right\}-1\right)(\exp\{\eta\,(T-t)\}-1)} \\ &\quad \times \frac{1}{(1-\exp\{-\eta\,(T-t)\})} \end{aligned} \tag{13.3.9}$$

for $t \in [0,T)$, $T \in (0,\infty)$.

To illustrate the type of market price of risk contribution that the stylized MMM produces we plot in Fig. 13.3.1 this function for $t = 0$ and $Y_0 = 53$ as a function of the net growth rate $\eta \in [0.001, 0.1]$ and the maturity $T \in [0.001, 80.0]$. It can be seen that the market price of risk contribution is practically zero for short dated maturities of up to one or two years. Afterwards, one obtains an increase in the value of the market price of risk contribution. For larger net growth rates the market price of risk contribution is larger. For extremely large time to maturity it equals the net growth rate, that is,

$$\lim_{T\to\infty} n(t,T) = \eta. \tag{13.3.10}$$

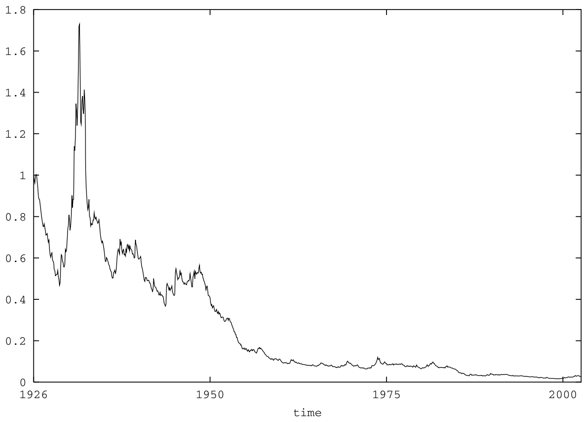



**Fig. 13.3.2.** Candidate Radon-Nikodym derivative for world market

In Platen (2005a) and Miller & Platen (2005) interest rate term structure models are discussed that are based on versions of the MMM.

### Absence of an Equivalent Risk Neutral Probability Measure

From the fair bond price (13.3.5) we note for $t \in [0, T)$ that

$$P(t,T) < P^*_T(t) = \exp\left\{-\int_t^T r_s\,ds\right\} = \frac{S^0_t}{S^0_T}, \tag{13.3.11}$$

which means for the stylized MMM that the fair zero coupon bond has a lower price than the savings bond $P^*_T(t)$. As discussed in the previous chapter, this demonstrates that the stylized MMM does not have an equivalent risk neutral probability measure. Indeed, the candidate Radon-Nikodym derivative process $\Lambda = \{\Lambda_t,\, t \in [0,\infty)\}$ for the hypothetical risk neutral measure, where

$$\Lambda_t = \frac{\hat{S}^0_t}{\hat{S}^0_0} = \frac{\bar{S}^{\delta_*}_0}{\bar{S}^{\delta_*}_t}, \tag{13.3.12}$$

is a strict $(\underline{\mathcal{A}}, P)$-supermartingale. This follows from our example in Sect. 8.7 for the inverse of a squared Bessel process of dimension four. Consequently, by Lemma 5.2.3 the process $\Lambda$ is a strict supermartingale. In Fig. 13.3.2 we show the candidate Radon-Nikodym derivative for the world market from 1926 until 2004, as it results when interpreting the discounted WSI in Fig. 13.1.1 as discounted GOP. We have for the hypothetical risk neutral measure $P_\theta$ on $[0, T]$ the inequality

$$P_{\theta,T}(\Omega) = E\left(\Lambda_T \,\middle|\, \mathcal{A}_0\right) = 1 - \exp\left\{-\frac{\bar{S}^{\delta_*}_0}{2\,\varphi(T)}\right\} < \Lambda_0 = 1. \tag{13.3.13}$$

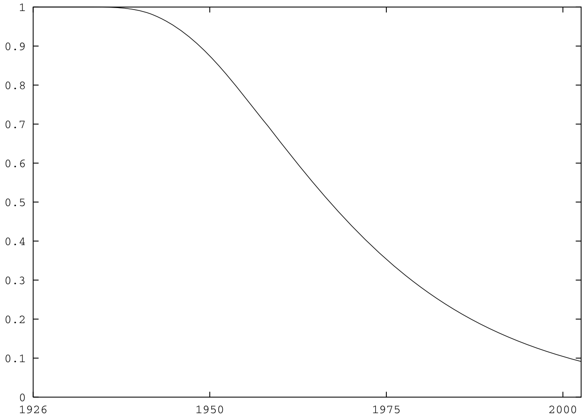



**Fig. 13.3.3.** Total mass of a hypothetical risk neutral measure

This shows that $P_\theta$ is not a probability measure because it does not give a total mass of one, see (1.1.4). This important fact does not create a problem, since we shall use the real world pricing formula to obtain derivative prices under the MMM and do not rely on risk neutral pricing. For illustration we show in Fig. 13.3.3 the total mass of the candidate risk neutral measure $P_\theta(\Omega)$ as a function of $T$. Here we use the default parameters $\eta = 0.0528$, $\alpha_0 = 0.043$ and $\bar{S}_0^{\delta_*} = 2.3$. One notes that the difference in total probability mass from the value one is very small for short time horizons $T$ of up to about ten years. In this range the hypothetical risk neutral measure is almost a probability measure. However, after ten years we observe the begin of a significant decline. After 40 years the total mass of the hypothetical risk neutral measure is only about 0.5. In these circumstances it is then not reasonable to expect a "risk neutral" price to be realistic for time horizons beyond ten years.

### Transition Density of the Stylized MMM

Before we price any particular European option we recall the transition density of the discounted GOP $\bar{S}^{\delta_*}$. According to (8.7.9) this transition density is of the form

$$p(s,x;t,y) = \frac{1}{2\,(\varphi(t)-\varphi(s))}\left(\frac{y}{x}\right)^{\frac{1}{2}} \exp\left\{-\frac{x+y}{2\,(\varphi(t)-\varphi(s))}\right\} I_1\left(\frac{\sqrt{x\,y}}{\varphi(t)-\varphi(s)}\right) \tag{13.3.14}$$

for $0 \le s < t < \infty$ and $x, y \in (0,\infty)$. Here $I_1(\cdot)$ is the modified Bessel function of the first kind with index $\nu = 1$, see (1.2.15). Note by (1.2.14) that (13.3.14) is the density of a non-central chi-square distributed random variable at time $t$ with value

$$\frac{y}{\varphi(t)-\varphi(s)} = \frac{\bar{S}_t^{\delta_*}}{\varphi(t)-\varphi(s)}$$

with $\delta = 4$ degrees of freedom and non-centrality parameter

$$\frac{x}{\varphi(t) - \varphi(s)} = \frac{\bar{S}_s^{\delta_*}}{\varphi(t) - \varphi(s)}.$$

Recall that Fig. 8.7.2 shows the transition density of a squared Bessel process of dimension four.

It is also interesting to consider the transition density for the normalized GOP

$$Y_t = \frac{\bar{S}_t^{\delta_*}}{\alpha_t^{\delta_*}},$$

see (13.2.4). This density is by (8.7.44) of the form

$$p(s,x;t,y) = \frac{1}{2\,\bar{s}_t\,\bar{\varphi}_t} \left(\frac{y}{x\,\bar{s}_t}\right)^{\frac{1}{2}} \exp\left\{-\frac{x + \frac{y}{\bar{s}_t}}{2\,\bar{\varphi}_t}\right\} I_1\left(\frac{\sqrt{x\,\frac{y}{\bar{s}_t}}}{\bar{\varphi}_t}\right) \tag{13.3.15}$$

for $0 \leq s < t < \infty$ and $x, y \in (0,\infty)$, where $\bar{s}_t = \exp\{-\eta(t-s)\}$ and $\bar{\varphi}_t = \frac{1}{4\eta}(\exp\{\eta(t-s)\} - 1)$. Recall that Fig. 4.4.1 shows the transition density of a square root process of dimension $\delta = 4$.

### European Call Options under the MMM

Since a diversified index is considered to be a proxy for the GOP, see Sect. 10.6, the MMM would appear to be a reasonable choice to model an index. We compute now the price $c_{T,K}(t, S_t^{\delta_*})$ of a fair European call option on the index with strike $K$ and maturity $T$ under the MMM. From the real world pricing formula (10.4.1) it follows that

$$\begin{aligned} c_{T,K}(t, S_t^{\delta_*}) &= S_t^{\delta_*}\,E\left(\frac{(S_T^{\delta_*} - K)^+}{S_T^{\delta_*}}\,\middle|\,\mathcal{A}_t\right) \\ &= E\left(\left(S_t^{\delta_*} - \frac{K\,S_t^{\delta_*}}{S_T^{\delta_*}}\right)^+\,\middle|\,\mathcal{A}_t\right) \end{aligned} \tag{13.3.16}$$

for $t \in [0,T]$. By applying the transition density of the time transformed squared Bessel process $\bar{S}^{\delta_*}$ of dimension four it has been shown in Hulley, Miller & Platen (2005) that the fair price of a European call option has the explicit formula

$$c_{T,K}(t, S_t^{\delta_*}) = S_t^{\delta_*}\,\left(1 - \chi^2(d_1; 4, \ell_2)\right) - K\,\exp\{-r\,(T-t)\}\,(1 - \chi^2(d_1; 0, \ell_2)) \tag{13.3.17}$$

with

$$d_1 = \frac{4\,\eta\,K\,\exp\{-r\,(T-t)\}}{S_t^0\,\alpha_t^{\delta_*}\,(\exp\{\eta\,(T-t)\} - 1)} \tag{13.3.18}$$

and

$$\ell_2 = \frac{2\,\eta\, S_t^{\delta_*}}{S_t^0\, \alpha_t^{\delta_*}\, (\exp\{\eta\,(T-t)\} - 1)} \tag{13.3.19}$$

for $t \in [0, T)$. This is an analytic pricing formula that involves the non-central chi-square distribution function, see (1.2.13). This formula has a similar level of complexity to that of the Black-Scholes formula, see (8.3.2). However, it provides more realistic European call option prices, as we shall see later.

We have shown in Fig. 12.3.1 an implied volatility surface for European call options on the GOP under the MMM. We noted a negatively skewed and slightly upwards sloping implied volatility surface.

### European Put Options under the MMM

For completeness, let us now determine the fair price of a European put option on the GOP when the underlying model is the stylized MMM. For this purpose it is appropriate to use the fair put-call parity relation (12.2.60) to calculate the put price $p_{T,K}(t, \bar{S}_t^{\delta_*})$ at time $t$ for maturity $T$ and strike $K$. This means that we apply the formula

$$p_{T,K}(t, \bar{S}_t^{\delta_*}) = c_{T,K}(t, \bar{S}_t^{\delta_*}) - S_t^{\delta_*} + K\, P(t,T) \tag{13.3.20}$$

for $t \in [0, T)$. This leads us by (13.3.17) to the explicit European put formula

$$\begin{aligned} p_{T,K}(t, S_t^{\delta_*}) = & -S_t^{\delta_*} \left(\chi^2(d_1; 4, \ell_2)\right) \\ & + K\, \exp\{-r\,(T-t)\} \left(\chi^2(d_1; 0, \ell_2) - \exp\{-\ell_2\}\right) \end{aligned} \tag{13.3.21}$$

for $t \in [0, T)$ when using the previous notation, see Hulley et al. (2005). One can calculate the implied volatilities for these put option prices as in Sect. 12.3. These are the same as those for the corresponding call prices.

As previously explained in Sect. 12.2, for the case of the modified CEV model, put-call parity breaks down if one uses the savings bond $P_T^*(t) = \frac{S_t^0}{S_T^0}$ instead of the fair bond $P(t,T)$ in relation (13.3.20).

It can be seen from (13.3.21) that when the GOP becomes very small, the put value also becomes small. As we have seen in (12.2.70), a put price derived under standard risk neutral pricing would be larger than the fair put price and would typically not become small when the GOP becomes small.

Note that one can explicitly calculate the forward price of a fair portfolio under the stylized MMM, as described at the end of Sect. 10.4. Furthermore, there are explicit formulas for fair European call and put options on primary security accounts, as will be discussed in Sect. 14.4.

## 13.4 MMM with Random Scaling (*)

### Model Formulation (*)

The version of the MMM described here, which generalizes the stylized version derived in Sect. 13.2, is governed by a particular choice of the discounted GOP drift $\alpha_t^{\delta_*}$. By (13.1.14) it can be seen that the $\alpha_t^{\delta_*}$, when integrated over time, yield the underlying value. One could argue that the underlying value is a non-decreasing, slowly varying stochastic process, where the randomness is caused by random trading activity potentially involving speculation.

If we take $\alpha_t^{\delta_*}$ to be a deterministic exponential function of time, as in (13.2.1), then we obtain the discounted GOP as a time transformed squared Bessel process of dimension $\delta = 4$. One could interpret this as an ideal or optimal market dynamics. Here $\alpha_t^{\delta_*}$ would express at time $t$ the discounted underlying value that is transferred per unit of time into the market. The discounted GOP evolves due to the conservation of underlying value according to a very specific probability law. Interestingly, the underlying value plays here the role of a transforming time, see (13.3.3).

In order to capture some possible delays or accelerations of this transfer of discounted underlying value into the market, we now employ a squared Bessel process with a more general dimension $\delta > 2$ and allow also for some randomness in its time transformation.

This is achieved by introducing the process $Z = \{Z_t,\, t \in [0,\infty)\}$ via the power transformation

$$Z_t = \left(\bar{S}_t^{\delta_*}\right)^{\frac{2}{\delta-2}} \tag{13.4.1}$$

for $t \in [0,\infty)$ and $\delta \in (2,\infty)$. The Itô formula applied to (13.1.8) and (13.4.1) yields

$$dZ_t = \frac{\delta}{4}\,\gamma_t\,dt + \sqrt{\gamma_t\,Z_t}\,dW_t. \tag{13.4.2}$$

The *scaling process* $\gamma = \{\gamma_t,\, t \in [0,\infty)\}$ with

$$\gamma_t = Z_t\,\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}\,\frac{4}{(\delta-2)^2} \tag{13.4.3}$$

will be specified later in an appropriate manner to reflect realistically the randomness of market activity or market time observed in the market. This means that $Z$ is a time transformed squared Bessel process of dimension $\delta > 2$, see Sect. 8.7. Note that for the standard choice $\delta = 4$ and $\gamma_t = 1$ we recover the stylized MMM, see Sect. 13.2. As we shall see, structuring the model equations in the above general form has the advantage that the model with deterministic $\gamma_t$ can generate different slopes of the implied volatility surface for European call and put options via the dimension $\delta$, see Heath & Platen (2005b).

Using the Itô formula together with (13.4.1) and (13.4.2), the GOP $S_t^{\delta_*}$ can be shown to satisfy the SDE

$$dS_t^{\delta_*} = S_t^{\delta_*}\left(\left[r + \left(\frac{\delta}{2}-1\right)^2\gamma_t\left(\frac{S_t^{\delta_*}}{S_t^0}\right)^{\frac{2}{2-\delta}}\right]dt + \left(\frac{\delta}{2}-1\right)\sqrt{\gamma_t}\left(\frac{S_t^{\delta_*}}{S_t^0}\right)^{\frac{1}{2-\delta}}dW_t\right) \tag{13.4.4}$$

for $t \in [0, \infty)$. For simplicity, we assume a constant short rate $r_t = r \geq 0$. The GOP volatility or total market price of risk is by (13.1.1) and (13.4.1) of the form

$$|\theta_t| = \left(\frac{\delta}{2} - 1\right) \sqrt{\frac{\gamma_t}{Z_t}} \tag{13.4.5}$$

for $t \in [0, \infty)$. This means that the volatility of the GOP is stochastic and depends at time $t$ on both the level of the discounted GOP with

$$\bar{S}_t^{\delta_*} = Z_t^{\frac{\delta-2}{2}} \tag{13.4.6}$$

and the random scaling quantity $\gamma_t$. Furthermore, the discounted GOP drift is by (13.1.4) and (13.4.5) given by

$$\alpha_t^{\delta_*} = \left(\frac{\delta}{2} - 1\right)^2 \gamma_t \, Z_t^{\frac{\delta-4}{2}} \tag{13.4.7}$$

for $t \in [0, \infty)$. By introducing a random scaling process, we model the discounted GOP drift in the form (13.4.7). Note that for the standard case with $\delta = 4$ the discounted GOP drift does not depend on $Z_t$. For $\delta > 4$ the discounted GOP drift increases when $\bar{S}_t^{\delta_*}$ increases. In the case $\delta \in (2, 4)$ the discounted GOP drift decreases when $\bar{S}_t^{\delta_*}$ increases.

### Random Scaling (*)

The random scaling process can be used to model the typical short term features of the market. For instance, it can model various random and seasonal features of trading activity. We assume here that the scaling process $\gamma = \{\gamma_t, \, t \in [0, \infty)\}$ is a nonnegative, adapted stochastic process that satisfies an SDE of the form

$$d\gamma_t = a(t, \gamma_t) \, dt + b(t, \gamma_t) \left(\varrho \, dW_t + \sqrt{1 - \varrho^2} \, d\tilde{W}_t\right) \tag{13.4.8}$$

for $t \in [0, \infty)$ with a random initial value $\gamma_0 > 0$. Here $\tilde{W}$ is a Wiener process that models some uncertainty in trading activity and is assumed to be independent of $W$. The scaling drift $a(\cdot, \cdot)$ and scaling diffusion coefficient $b(\cdot, \cdot)$ are given functions of time $t$ and scaling level $\gamma_t$. The scaling correlation $\varrho$ is, for simplicity, assumed to be constant. Under this formulation the dynamics of the diffusion process $\gamma$ can be chosen to match empirical evidence. Note that there seems to be no compelling reason to make the scaling correlation $\varrho$ different to zero, see Breymann, Kelly & Platen (2006). The main feedback effect for indices, is well captured under the MMM already. We keep $\varrho$ still flexible in the above model since it makes it similar to the stochastic volatility models presented in the previous section. For the preferred case $\varrho = 0$ we have independence between $\gamma_t$ and $W_t$, which simplifies the computation of derivative prices.

The above MMM with random scaling offers different choices for the dimension $\delta > 2$ and, thus, different volatility dynamics. It has some similarity with the CEV model, see Sect. 12.2, which also involves a squared Bessel process. In Heath & Platen (2003) the random scaling was chosen to be a geometric Brownian motion, whereas in Heath & Platen (2005a) the dynamics are similar to those outlined above. The results for short term and medium term options are similar to those that we are going to report in this section.

We provide now an example for the modeling of random scaling that is motivated by an intraday empirical analysis of trading activity, obtained in Breymann et al. (2006) for a diversified world stock index denominated in US dollars. The scaling is modeled as a product of the type

$$\gamma_t = \xi_t \, m_t \tag{13.4.9}$$

with

$$\xi_t = \xi_0 \, \exp\{\eta \, t\} \tag{13.4.10}$$

for $t \in [0, \infty)$ with $\xi_0 > 0$. As before, the parameter $\eta > 0$ is called the net growth rate. The *market activity* process $m = \{m_t, \, t \in [0, \infty)\}$ in (13.4.9) is designed to model normalized trading activity. Note that for constant market activity $m_t = 1$ and dimension $\delta = 4$ the stylized MMM of Sect. 13.2 is recovered. In Breymann et al. (2006) it was suggested that market activity appears to have multiplicative noise. Therefore, the market activity process is modeled as a nonnegative process that satisfies an SDE of the form

$$dm_t = k(m_t) \, \beta^2 \, dt + \beta \, m_t \left( \varrho \, dW_t + \sqrt{1 - \varrho^2} \, d\tilde{W}_t \right) \tag{13.4.11}$$

for $t \in [0, \infty)$ with random initial market activity $m_0 \geq 0$. In this SDE multiplicative noise is characterized by the constant *activity volatility* $\beta > 0$. The function $k(\cdot)$ controls the drift of the market activity. Let us choose this function to be of the form

$$k(m) = (p - g \, m) \, \frac{m}{2}, \tag{13.4.12}$$

with speed of adjustment parameter $g$ and reference level $p$. These constant, deterministic parameters are set so that the expected value of market activity is about one. The market activity process $m = \{m_t, \, t \in [0, \infty)\}$ has a stationary density, see (4.5.5), of the form

$$p_m(y) = \frac{g^{p-1}}{\Gamma(p-1)} \, y^{p-2} \exp\{-g \, y\} \tag{13.4.13}$$

for $y \in [0, \infty)$, where $\Gamma(\cdot)$ is the gamma function. This is a gamma density with mean $\frac{p-1}{g}$ and variance $\frac{1}{g}$ for parameters $p > 1$ and $g > 0$, see (1.2.9).

Figure 13.4.1 shows the stationary density (13.4.13) of $m_t$ for different levels of market activity $y$ and speed of adjustment parameter $g$, with $p = g+1$,

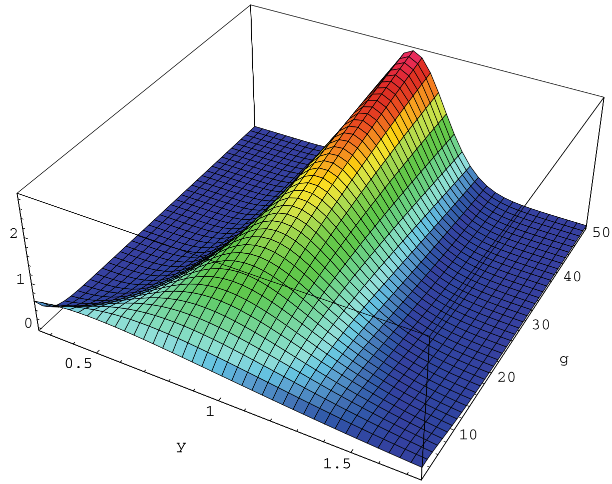



**Fig. 13.4.1.** Stationary density of market activity $m_t = y$ as function of $y$ and speed of adjustment parameter $g$

to ensure that the mean of the stationary density always equals one. Note that for a large speed of adjustment parameter the market activity remains close to one.

Applying the Itô formula and using (13.4.9)–(13.4.12), the drift and diffusion coefficients appearing in (13.4.8) take the form

$$a(t, \gamma) = \gamma \left( p - g \, \frac{\gamma}{\xi_t} \right) + \gamma \, \eta \tag{13.4.14}$$

and

$$b(t, \gamma) = \beta \, \gamma, \tag{13.4.15}$$

respectively, for $t \in [0, \infty)$.

Note that at any time $t \in [0, \infty)$ the actual value of the market activity $m_t$, and thus the random scaling $\gamma_t$, are not easily observable. These change very rapidly and can only be estimated after sufficient time has elapsed. Therefore, the initial value $m_0$ of the market activity itself may have to be modeled as a random variable. For instance, the stationary density (13.4.13) could be used as its probability density. In the following we shall discuss the impact of using different parameter choices on various derivatives.

## Zero Coupon Bond (*)

First, we consider a fair zero coupon bond that pays one unit of the domestic currency at the maturity date $T \in [0, \infty)$. An equivalent risk neutral probability measure does not exist for the above model. The benchmarked savings account $\hat{S}^0$ and, thus, the candidate Radon-Nikodym derivative process $\Lambda = \{\Lambda_t, \, t \in [0, \infty)\}$ with

$$\Lambda_t = \left( \frac{\bar{S}_t^{\delta_*}}{\bar{S}_0^{\delta_*}} \right)^{-1} = \left( \frac{Z_t}{Z_0} \right)^{1-\frac{\delta}{2}} \tag{13.4.16}$$

are by (8.7.24) strict local martingales when we assume no correlation, that is $\varrho = 0$. For this reason we shall use real world pricing to calculate derivative prices. By using (13.4.1) the benchmarked price $\hat{P}_T(t, Z_t, \gamma_t)$ for a zero coupon bond at time $t$ with maturity $T$ is then given by the conditional expectation

$$\hat{P}_T(t, Z_t, \gamma_t) = E\left( \frac{1}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t \right) = E\left( \frac{1}{S_T^0 \, Z_T^{\frac{\delta}{2}-1}} \,\bigg|\, \mathcal{A}_t \right) \tag{13.4.17}$$

for $t \in [0, T]$, see (10.4.8). Hence the corresponding zero coupon bond price $P_T(t, Z_t, \gamma_t)$ is given by

$$P_T(t, Z_t, \gamma_t) = S_t^{\delta_*} \, \hat{P}_T(t, Z_t, \gamma_t) = S_t^0 \, Z_t^{\frac{\delta}{2}-1} \, \hat{P}_T(t, Z_t, \gamma_t) \tag{13.4.18}$$

for $t \in [0, \infty)$.

In general, we do not have an explicit joint density of $(Z_T, \gamma_T)$, which we would need to calculate the conditional expectation in (13.4.17). Therefore, let us introduce the diffusion operator $\mathcal{L}^0$ for the Markovian factors $(Z_t, \gamma_t)$, which when applied to a sufficiently smooth function $f : (0, T) \times (0, \infty)^2 \to \Re$ is of the form

$$\mathcal{L}^0 f(t, Z, \gamma) = \left( \frac{\partial}{\partial t} + \frac{\delta \, \gamma}{4} \frac{\partial}{\partial Z} + a(t, \gamma) \frac{\partial}{\partial \gamma} + \frac{1}{2} \gamma \, Z \frac{\partial^2}{\partial Z^2} \right.$$
$$\left. + \varrho \, b(t, \gamma) \, \gamma^{\frac{1}{2}} \, Z^{\frac{1}{2}} \frac{\partial^2}{\partial Z \, \partial \gamma} + \frac{1}{2} b(t, \gamma)^2 \, \gamma \frac{\partial^2}{\partial \gamma^2} \right) f(t, Z, \gamma) \tag{13.4.19}$$

for $(t, Z, \gamma) \in (0, T) \times (0, \infty)^2$. Using (13.4.2) and (13.4.8) together with the Feynman-Kac formula, see Sect. 9.7, the benchmarked fair zero coupon bond pricing function $\hat{P}_T(\cdot, \cdot, \cdot)$ satisfies the Kolmogorov backward equation

$$\mathcal{L}^0 \, \hat{P}_T(t, Z, \gamma) = 0 \tag{13.4.20}$$

for $(t, Z, \gamma) \in (0, T) \times (0, \infty)^2$ with terminal condition

$$\hat{P}_T(T, Z, \gamma) = \frac{1}{S_T^0 \, Z^{\frac{\delta}{2}-1}} \tag{13.4.21}$$

for $(Z, \gamma) \in (0, \infty)^2$. By using numerical methods for solving partial differential equations (PDEs), as we shall describe in Sect. 15.7, one can numerically determine $\hat{P}_T(\cdot, \cdot, \cdot)$.

### Forward Rates (*)

For the above two-factor model we obtain by (10.4.12) the forward rate for the maturity date $T$ at time $t < T$ by the formula

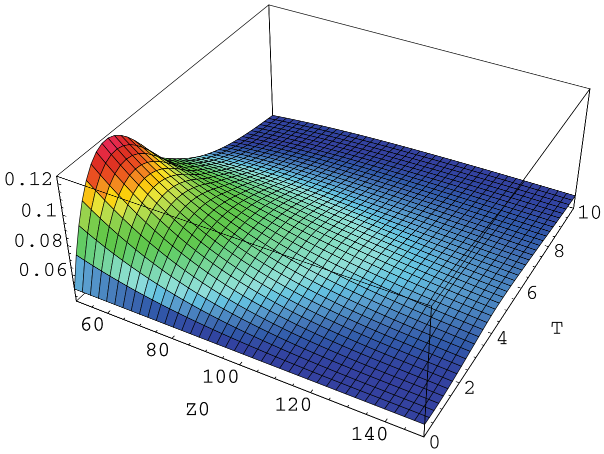



**Fig. 13.4.2.** Forward rates as a function of $Z_0$ and $T$

$$f_T(t, Z_t, \gamma_t) = -\frac{\partial}{\partial T} \ln\left(P_T(t, Z_t, \gamma_t)\right). \tag{13.4.22}$$

Figure 13.4.2 shows for different initial values of $Z_0 \in [50, 150]$ the forward rate curves at time $t = 0$ as functions of $T \in [0.25, 10]$. For this and subsequent plots the default parameters used are: $\delta = 4$, $r = 0.05$, $\varrho = 0$, $\eta_t = 0.048$, $\xi_0 = 10$, $p = 3$ and $g = 2$. Note that despite a constant short rate the forward rates are not constant and are always greater than the short rate. Furthermore, we observe a hump in the forward rate at about the time of two years to maturity. This is an important feature that has been observed in the market, see, for instance Bouchaud, Sagna, Cont, El Karoui & Potters (1999) and Matacz & Bouchaud (2000). These results together with those described below are numerically obtained by using the Crank-Nicolson finite difference method, which will be discussed in Sect. 15.7. The randomness of the initial value $m_0$ is generated by a two-point distributed random variable with mean $\frac{p-1}{g}$ and variance $\frac{1}{g}$. The fact that the realistic hump shaped forward rates, shown in Fig. 13.4.2, are greater than the constant short rate, demonstrates that the benchmarked savings account process $\hat{S}^0$, see (10.3.1), is a strict $(\underline{\mathcal{A}}, P)$-supermartingale, as was pointed out earlier.

## European Options on a Market Index (*)

As previously explained, the GOP is employed as proxy for a market index. Consider now a European put option on the index $S^{\delta_*}$ with strike $K$ and maturity date $T \in [0, \infty)$. Using the real world pricing formula (10.4.1), the put option price $p_{T,K}(t, Z_t, \gamma_t)$ is given by

$$p_{T,K}(t, Z_t, \gamma_t) = S_t^0 \, Z_t^{\frac{\delta}{2}-1} \, E\left(\left(\frac{K}{S_T^0 \, Z_T^{\frac{\delta}{2}-1}} - 1\right)^+ \,\middle|\, \mathcal{A}_t\right) \tag{13.4.23}$$

for $t \in [0, T]$.

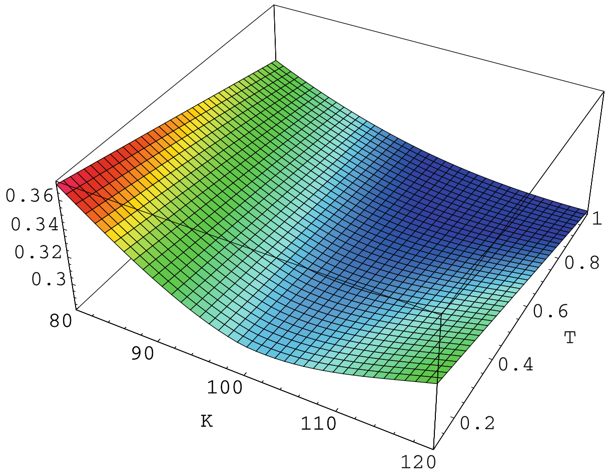



**Fig. 13.4.3.** Implied volatilities for put options on index as a function of strike $K$ and maturity $T$

To see the effect of random scaling on implied volatilities, Fig. 13.4.3 displays an implied volatility surface for European puts as a function of the maturity date $T$ and the strike $K$. These results were obtained using the zero coupon bond price (13.4.18) to infer the discount factor used in the Black-Scholes formula via (12.2.57). The implied volatilities shown in Fig. 13.4.3 are rather close to those observed for European index options in real markets, see Fig. 12.1.5. Note that the curvature of the implied volatility surface for short dated options results from the randomness of the scaling. One can show that this curvature is mainly generated by the randomness of the initial value $m_0$ of the market activity process. If a fixed initial value $m_0$ were used, then much of the curvature for the short dated implied volatility surface would disappear. This is important to notice since it tells us that the MMM with a random initial value already provides most of the stylized features observed in reality. The MMM with random initial scaling is also able to capture realistic implied volatility smiles for exchange rate and equity options, as observed in real markets, see Heath & Platen (2005a).

It is well-known that skew and smile patterns for implied volatility surfaces, as shown in Fig. 13.4.3, can be obtained by various stochastic volatility models, see Carr & Wu (2003) or Brigo, Mercurio & Rapisarda (2004) and our comments in Sect. 12.4. However, most of these models are difficult to calibrate to a range of standard and exotic derivatives. It has been demonstrated in Heath & Platen (2005a, 2005b) that the MMM avoids most of these problems.

To demonstrate the effect of making the scaling process $\gamma$ stochastic, Fig. 13.4.4 shows implied volatilities for European puts on the GOP as a

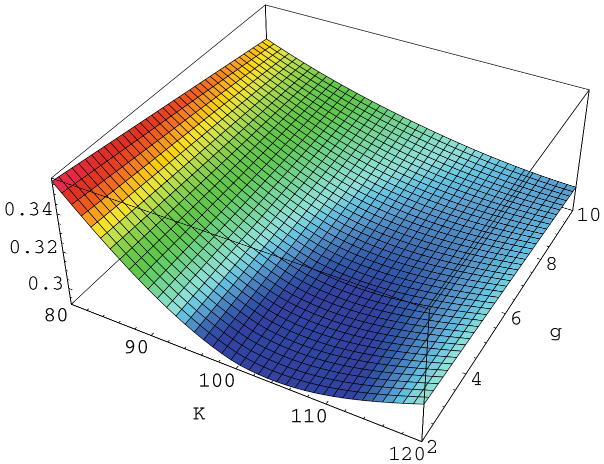



**Fig. 13.4.4.** Implied volatilities for put options as a function of strike $K$ and speed of adjustment $g$

function of the strike $K$ and the speed of adjustment parameter $g$ for a fixed maturity date $T = 0.25$ and with $p = g + 1$. The figure indicates that an increase in speed of adjustment $g$ decreases the curvature of the implied volatility curve, when viewed as a function of the strike $K$. For different values of $g$ the corresponding initial random market activity $m_0$ is also adjusted to match the mean and variance of the corresponding stationary distribution.

It should be noted that changing the dimension $\delta$ of the time transformed squared Bessel process $Z$ affects the slope of the implied volatility surface. That is, lowering the dimension $\delta$ produces a stronger negative skew for the implied volatility surface and vice versa.

For long dated European put or call options it can be seen that there is little curvature in the corresponding implied volatility curves for a given maturity date, see Fig. 13.4.5.

Note that a remarkably sustained increase in overall implied volatilities occurs for longer maturities. This is not usually obtained from a stochastic volatility model where an equivalent risk neutral probability measure exists. It can be observed that the impact of using random scaling, which is mainly reflected in the curvature of implied volatilities for short dated options, is not so prominent for longer dated maturities. This suggests that for long dated options deterministic scaling will suffice.

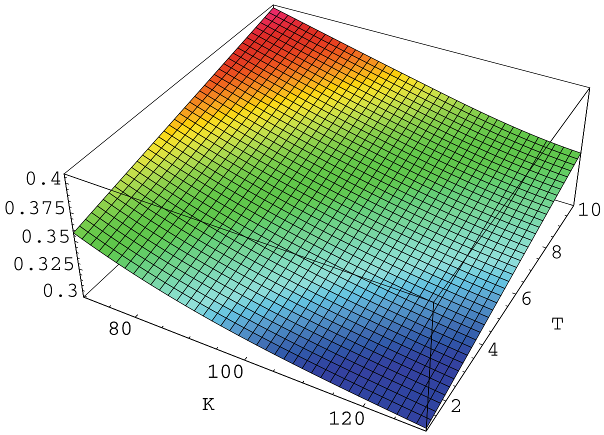



**Fig. 13.4.5.** Implied volatilities for long dated put options as a function of strike $K$ and maturity $T$

## 13.5 Exercises for Chapter 13

**13.1.** Calculate the SDE for the logarithm of the discounted GOP.

**13.2.** Derive the SDE of the square root for the discounted GOP.

**13.3.** Derive the SDE for the normalized GOP $Y_t = \frac{\bar{S}_t^{\delta_*}}{\alpha_t^{\delta_*}}$ if

$$\alpha_t^{\delta_*} = \alpha_0 \exp\{\int_0^t \eta_s \, ds\}.$$

**13.4.** Calculate the SDE for the squared volatility of the discounted GOP, given as in Exercise 13.3.

# 14

# Markets with Event Risk

After having studied continuous financial markets, this chapter applies the benchmark approach to markets that exhibit jumps due to event risk. It generalizes several results previously obtained to the case of *jump diffusion markets* (JDMs).

## 14.1 Jump Diffusion Markets

This section extends the results of Sect. 10.1. It provides a unified framework for financial modeling, portfolio optimization, derivative pricing and risk measurement when security price processes exhibit intensity based jumps. These jumps allow for the modeling of event risk in finance, insurance and other areas. Conditions are formulated under which such a market does not permit arbitrage. The natural numeraire for pricing is again shown to be the GOP, which relates to the concept of real world pricing as previously explained. Nonnegative portfolios, when expressed in units of the GOP, turn out to be supermartingales again. An equivalent risk neutral probability measure needs not to exist in the JDMs considered. The approach presented avoids the problem of dealing with risk neutral intensities and similar complications and restrictions that apply under the standard risk neutral approach.

### Continuous and Event Driven Uncertainty

We consider a market containing $d \in \mathcal{N}$ sources of trading uncertainty. *Continuous trading uncertainty* is represented by $m \in \{1, 2, \ldots, d\}$ independent standard Wiener processes $\tilde{W}^k = \{\tilde{W}^k_t,\, t \in [0, \infty)\},\ k \in \{1, 2, \ldots, m\}$. These are defined on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. We also model events of certain types, for instance, corporate defaults, credit rating changes, operational failures or specified insured events, when these are reflected in the prices of traded securities. Events of the $k$th type are counted by the $\underline{\mathcal{A}}$-adapted $k$th

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

*counting process* $p^k = \{p^k_t,\, t \in [0,\infty)\}$, whose *intensity* $h^k = \{h^k_t,\, t \in [0,\infty)\}$ is a given, predictable, strictly positive process with

$$h^k_t > 0 \tag{14.1.1}$$

and

$$\int_0^t h^k_s \, ds < \infty \tag{14.1.2}$$

almost surely for $t \in [0,\infty)$ and $k \in \{1,2,\dots,d-m\}$. The $k$th counting process $p^k$ leads to the $k$th *jump martingale* $q^k = \{q^k_t,\, t \in [0,\infty)\}$ with stochastic differential

$$dq^k_t = \left(dp^k_t - h^k_t \, dt\right) \left(h^k_t\right)^{-\frac{1}{2}} \tag{14.1.3}$$

for $k \in \{1,2,\dots,d-m\}$ and $t \in [0,\infty)$. It is assumed that the above jump martingales do not jump at the same time. They represent the compensated, normalized sources of *event driven trading uncertainty.*

The evolution of trading uncertainty is modeled by the vector process of independent $(\underline{\mathcal{A}}, P)$-martingales $\boldsymbol{W} = \{\boldsymbol{W}_t = (\tilde{W}^1_t, \dots, \tilde{W}^m_t, q^1_t, \dots, q^{d-m}_t)^\top,$ $t \in [0,\infty)\}$. Note that $W^1 = \tilde{W}^1$, $\dots$, $W^m = \tilde{W}^m$ are Wiener processes, while $W^{m+1} = q^1$, $\dots$, $W^d = q^{d-m}$ are compensated, normalized counting processes. The filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t \in [0,\infty)}$ satisfies the usual conditions and $\mathcal{A}_0$ is the trivial $\sigma$-algebra, see Sect. 5.1. Note that the conditional variance of the increment of the $k$th source of event driven trading uncertainty over a time interval of length $\Delta$ equals

$$E\left(\left(q^k_{t+\Delta} - q^k_t\right)^2 \,\Big|\, \mathcal{A}_t\right) = \Delta \tag{14.1.4}$$

for all $t \in [0,\infty)$, $k \in \{1,2,\dots,d-m\}$ and $\Delta \in [0,\infty)$. Note that in addition to trading uncertainties the market typically involves additional uncertainties that impact jump intensities, short rates, appreciation rates, volatilities and other financial quantities.

### Primary Security Accounts

As previously explained, a primary security account is an investment account, consisting of only one kind of security. The $j$th risky primary security account value at time $t$ is denoted by $S^j_t$, for $j \in \{1,2,\dots,d\}$ and $t \in [0,\infty)$. These primary security accounts model the evolution of wealth due to the ownership of primary securities, with all dividends and income reinvested. The 0th primary security account $S^0 = \{S^0_t,\ t \in [0,\infty)\}$ is again the domestic riskless savings account, which continuously accrues at the short term interest rate $r_t$. In the market considered, the denominating security is the domestic currency.

Without loss of generality, we assume that the nonnegative $j$th primary security account value $S^j_t$ satisfies the jump diffusion SDE

$$dS_t^j = S_{t-}^j \left( a_t^j \, dt + \sum_{k=1}^{d} b_t^{j,k} \, dW_t^k \right) \tag{14.1.5}$$

for $t \in [0, \infty)$ with initial value $S_0^j > 0$ and $j \in \{1, 2, \dots, d\}$, see Sect. 7.6. Recall that $S_{t-}^j$ denotes the value of the process $S^j$ just before time $t$, which is defined as the left hand limit at time $t$, see (5.2.17). This SDE formally looks similar to the SDE (10.1.2). However, we have here also the jump martingales $W_t^k = q_t^{k-m}$ for $k \in \{m+1, \dots, d\}$, $t \in [0, \infty)$.

We assume that the short rate process $r$, the appreciation rate processes $a^j$, the *generalized volatility processes* $b^{j,k}$ and the intensity processes $h^k$ are almost surely finite and predictable, $j \in \{1, 2, \dots, d\}$, $k \in \{1, 2, \dots, d-m\}$. They are assumed to be such that a unique strong solution of the system of SDEs (14.1.5) exists, see Sect. 7.7. To ensure nonnegativity for each primary security account we need to make the following assumption.

**Assumption 14.1.1.** *The condition*

$$b_t^{j,k} \geq -\sqrt{h_t^{k-m}} \tag{14.1.6}$$

*holds for all* $t \in [0, \infty)$, $j \in \{1, 2, \dots, d\}$ *and* $k \in \{m+1, \dots, d\}$.

Taking into account (14.1.3), it can be seen from the SDE (14.1.5) that this assumption excludes jumps that would lead to negative values for $S_t^j$, see Sect. 7.6. To securitize the sources of trading uncertainty properly, we introduce the *generalized volatility matrix* $\boldsymbol{b}_t = [b_t^{j,k}]_{j,k=1}^d$ for all $t \in [0, \infty)$ and make the following assumption.

**Assumption 14.1.2.** *The generalized volatility matrix* $\boldsymbol{b}_t$ *is* invertible *for Lebesgue-almost-every* $t \in [0, \infty)$.

Assumption 14.1.2 generalizes Assumption 10.1.1 and allows us to introduce the *market price of risk* vector

$$\boldsymbol{\theta}_t = (\theta_t^1, \dots, \theta_t^d)^\top = \boldsymbol{b}_t^{-1} \left[ \boldsymbol{a}_t - r_t \, \mathbf{1} \right] \tag{14.1.7}$$

for $t \in [0, \infty)$. Here $\boldsymbol{a}_t = (a_t^1, \dots, a_t^d)^\top$ is the *appreciation rate vector* and $\mathbf{1} = (1, \dots, 1)^\top$ the *unit vector*. Using (14.1.7), we can rewrite the SDE (14.1.5) similarly to (10.1.7) in the form

$$dS_t^j = S_{t-}^j \left( r_t \, dt + \sum_{k=1}^{d} b_t^{j,k} \left( \theta_t^k \, dt + dW_t^k \right) \right) \tag{14.1.8}$$

for $t \in [0, \infty)$ and $j \in \{0, 1, \dots, d\}$. For $k \in \{1, 2, \dots, m\}$, the quantity $\theta_t^k$ denotes the *market price of risk* with respect to the $k$th Wiener process $W^k$. If $k \in \{m+1, \dots, d\}$, then $\theta_t^k$ can be interpreted as the *market price of the* $(k-m)$*th event risk* with respect to the counting process $p^{k-m}$. As previously discussed, the market prices of risk play a central role, as they are invariants of the market and determine the risk premia that risky securities attract.

The vector process $\boldsymbol{S} = \{\boldsymbol{S}_t = (S_t^0, \ldots, S_t^d)^\top,\ t \in [0,\infty)\}$ characterizes the evolution of all primary security accounts. We say that a predictable stochastic process $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \ldots, \delta_t^d)^\top,\ t \in [0,\infty)\}$ is a strategy, see Sect. 10.1, if the Itô integral $I_{\delta,W}(t)$ of the corresponding gains from trade exists, see Sect. 5.3. As explained in Chap. 10, the $j$th component $\delta^j$ of $\boldsymbol{\delta}$ denotes the number of units of the $j$th primary security account held at time $t \in [0,\infty)$ in the portfolio $S^\delta$, $j \in \{0, 1, \ldots, d\}$. For a strategy $\boldsymbol{\delta}$ we denote by $S_t^\delta$ the value of the corresponding portfolio process at time $t$, when measured in units of the domestic currency. Thus, we set

$$S_t^\delta = \sum_{j=0}^{d} \delta_t^j \, S_t^j \tag{14.1.9}$$

for $t \in [0,\infty)$. As defined for a CFM, a strategy $\boldsymbol{\delta}$ and the corresponding portfolio process $S^\delta = \{S_t^\delta,\ t \in [0,\infty)\}$ are self-financing if

$$dS_t^\delta = \sum_{j=0}^{d} \delta_t^j \, dS_t^j \tag{14.1.10}$$

for all $t \in [0,\infty)$, see (10.1.10). We emphasize that $\boldsymbol{\delta}$ is assumed to be a predictable process and we consider only self-financing portfolios.

### Growth Optimal Portfolio

As before, let us denote by $\mathcal{V}^+$ the set of strictly positive portfolio processes. For a given strategy $\boldsymbol{\delta}$ with strictly positive portfolio process $S^\delta \in \mathcal{V}^+$ denote by $\pi_{\delta,t}^j$ the fraction of wealth that is invested in the $j$th primary security account at time $t$, that is,

$$\pi_{\delta,t}^j = \delta_t^j \, \frac{S_t^j}{S_t^\delta} \tag{14.1.11}$$

for $t \in [0,\infty)$ and $j \in \{0, 1, \ldots, d\}$, see (10.1.11). These fractions sum to one, see (10.1.13). In terms of the vector of fractions $\boldsymbol{\pi}_{\delta,t} = (\pi_{\delta,t}^1, \ldots, \pi_{\delta,t}^d)^\top$ we obtain from (14.1.10), (14.1.8) and (14.1.11) the SDE for $S_t^\delta$

$$dS_t^\delta = S_{t-}^\delta \left\{ r_t \, dt + \boldsymbol{\pi}_{\delta,t-}^\top \, \boldsymbol{b}_t \, (\boldsymbol{\theta}_t \, dt + d\boldsymbol{W}_t) \right\} \tag{14.1.12}$$

for $t \in [0,\infty)$, where $d\boldsymbol{W}_t = (dW_t^1, \ldots, dW_t^m, dq_t^1, \ldots, dq_t^{m-d})^\top$. Note by (14.1.3) that a portfolio process $S^\delta$ remains strictly positive if and only if

$$\sum_{j=1}^{d} \pi_{\delta,t}^j \, b_t^{j,k} > -\sqrt{h_t^{k-m}} \tag{14.1.13}$$

almost surely for all $k \in \{m+1, \ldots, d\}$ and $t \in [0,\infty)$.

For a strictly positive portfolio $S^\delta \in \mathcal{V}^+$ we obtain for its logarithm, by application of Itô's formula, the SDE

$$d\ln(S_t^\delta) = g_t^\delta\, dt + \sum_{k=1}^{m}\sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\, dW_t^k + \sum_{k=m+1}^{d} \ln\left(1 + \sum_{j=1}^{d} \pi_{\delta,t-}^j \frac{b_t^{j,k}}{\sqrt{h_t^{k-m}}}\right)\sqrt{h_t^{k-m}}\, dW_t^k \quad (14.1.14)$$

for $t \in [0,\infty)$. Similarly to (10.2.2), the *growth rate* in this expression is

$$g_t^\delta = r_t + \sum_{k=1}^{m}\left[\sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\, \theta_t^k - \frac{1}{2}\left(\sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\right)^2\right] + \sum_{k=m+1}^{d}\left[\sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\left(\theta_t^k - \sqrt{h_t^{k-m}}\right) + \ln\left(1 + \sum_{j=1}^{d} \pi_{\delta,t}^j \frac{b_t^{j,k}}{\sqrt{h_t^{k-m}}}\right) h_t^{k-m}\right] \quad (14.1.15)$$

for $t \in [0,\infty)$, see Exercise 14.1. Note that for the first sum on the right hand side of (14.1.15) a unique maximum exists, because it is a quadratic form with respect to the fractions. Careful inspection of the terms in the second sum reveals that, in general, a unique maximum growth rate only exists if the market prices of event risks are less than the square roots of the corresponding jump intensities. This leads to the following assumption.

**Assumption 14.1.3.** *The intensities and market price of event risk components satisfy*

$$\sqrt{h_t^{k-m}} > \theta_t^k \quad (14.1.16)$$

*for all* $t \in [0,\infty)$ *and* $k \in \{m+1,\ldots,d\}$.

We shall see that Assumption 14.1.3 guarantees that there are no portfolios that explode for the given market, which would otherwise lead to some form of arbitrage. Furthermore, this condition allows us to introduce the predictable vector process $\boldsymbol{c}_t = (c_t^1,\ldots,c_t^d)^\top$ with components

$$c_t^k = \begin{cases} \theta_t^k & \text{for} \quad k \in \{1,2,\ldots,m\} \\ \frac{\theta_t^k}{1-\theta_t^k\,(h_t^{k-m})^{-\frac{1}{2}}} & \text{for} \quad k \in \{m+1,\ldots,d\} \end{cases} \quad (14.1.17)$$

for $t \in [0,\infty)$. Note that a very large jump intensity with $h_t^{k-m} \gg 1$ or $\frac{\theta_t^k}{\sqrt{h_t^{k-m}}} \ll 1$ causes the corresponding component $c_t^k$ to approach the market price of jump risk $\theta_t^k$ asymptotically for given $t \in [0,\infty)$ and $k \in \{m+1,\dots,d\}$. In this case the structure of the $k$th component $c_t^k \approx \theta_t^k$ is similar to those obtained with respect to Wiener processes.

We now define the fractions

$$\boldsymbol{\pi}_{\delta_*,t} = (\pi^1_{\delta_*,t},\dots,\pi^d_{\delta_*,t})^\top = \left(\boldsymbol{c}_t^\top\, \boldsymbol{b}_t^{-1}\right)^\top \tag{14.1.18}$$

of a particular portfolio $S^{\delta_*} \in \mathcal{V}^+$, which will be later identified as a GOP, $t \in [0,\infty)$. By (14.1.12) and (14.1.17) it follows that $S_t^{\delta_*}$ satisfies the SDE

$$\begin{aligned} dS_t^{\delta_*} &= S_{t-}^{\delta_*}\Big(r_t\,dt + \boldsymbol{c}_t^\top\,(\boldsymbol{\theta}_t\,dt + d\boldsymbol{W}_t)\Big) \\ &= S_{t-}^{\delta_*}\left(r_t\,dt + \sum_{k=1}^m \theta_t^k\,(\theta_t^k\,dt + dW_t^k)\right. \\ &\qquad \left. + \sum_{k=m+1}^d \frac{\theta_t^k}{1-\theta_t^k\,(h_t^{k-m})^{-\frac{1}{2}}}\,(\theta_t^k\,dt + dW_t^k)\right) \end{aligned} \tag{14.1.19}$$

for $t \in [0,\infty)$, with $S_0^{\delta_*} > 0$. Inspection of (14.1.19) shows that Assumption 14.1.3 keeps the portfolio process $S^{\delta_*}$ strictly positive. Let us now define a GOP in the given market with intensity based jumps.

**Definition 14.1.4.** *In the given market a strictly positive portfolio process $S^{\underline{\delta}} \in \mathcal{V}^+$ that maximizes the growth rate $g_t^\delta$, see (14.1.15), of strictly positive portfolio processes is called a* GOP*, that is, $g_t^\delta \le g_t^{\underline{\delta}}$ almost surely for all $t \in [0,\infty)$ and $S^\delta \in \mathcal{V}^+$.*

This definition generalizes the Definition 10.2.1 of a GOP in a CFM. The proof of the following result is given at the end of this section, see also Platen (2004b).

**Corollary 14.1.5.** *Under Assumptions 14.1.1, 14.1.2 and 14.1.3 the portfolio process $S^{\delta_*} = \{S_t^{\delta_*},\ t \in [0,\infty)\}$, satisfying (14.1.19), is a GOP.*

By (14.1.15), (14.1.17) and (14.1.18) we obtain the optimal growth rate of the GOP in the form

$$g_t^{\delta_*} = r_t + \frac{1}{2}\sum_{k=1}^m (\theta_t^k)^2 - \sum_{k=m+1}^d h_t^{k-m}\left(\ln\left(1 + \frac{\theta_t^k}{\sqrt{h_t^{k-m}} - \theta_t^k}\right) + \frac{\theta_t^k}{\sqrt{h_t^{k-m}}}\right) \tag{14.1.20}$$

for $t \in [0,\infty)$. Note that the optimal growth rate is never less than the short rate. Furthermore, as long as $\frac{\theta_t^k}{\sqrt{h_t^{k-m}}} \ll 1$, that is, $\theta_t^k$ is significantly smaller than $\sqrt{h_t^{k-m}}$ for $k \in \{m+1,\dots,d\}$, we approximately obtain

$$g_t^{\delta_*} \approx r_t + \frac{1}{2}\sum_{k=1}^{d}(\theta_t^k)^2 = r_t + \frac{|\boldsymbol{\theta}_t|^2}{2} \tag{14.1.21}$$

and

$$d\ln(S_t^{\delta_*}) \approx g_t^{\delta_*}\,dt + \sum_{k=1}^{d}\theta_t^k\,dW_t^k. \tag{14.1.22}$$

This SDE is analogous to the SDE for the logarithm of the GOP of a CFM in Sect. 10.2. Also by (14.1.19) we can derive the approximation

$$dS_t^{\delta_*} \approx S_t^{\delta_*}\left(r_t + \sum_{k=1}^{d}\theta_t^k\left(\theta_t^k\,dt + dW_t^k\right)\right), \tag{14.1.23}$$

which is similar to (10.2.8). Now, let us formally characterize the given jump diffusion market.

**Definition 14.1.6.** *We denote the above financial market by* $\mathcal{S}_{(d)}^{JD} = \{\boldsymbol{S}, \boldsymbol{a}, \boldsymbol{b}, \boldsymbol{r}, \underline{\mathcal{A}}, P\}$ *and call it a* jump diffusion market *(JDM) when it has* $d \in \mathcal{N}$ *risky primary security accounts and satisfies Assumptions 14.1.1, 14.1.2 and 14.1.3.*

### Supermartingale Property

As is the case for a CFM, we call prices, when expressed in units of $S^{\delta_*}$ benchmarked prices. By the Itô formula and relations (14.1.12) and (14.1.19), a benchmarked portfolio process $\hat{S}^\delta = \{\hat{S}_t^\delta,\, t \in [0,\infty)\}$, with

$$\hat{S}_t^\delta = \frac{S_t^\delta}{S_t^{\delta_*}} \tag{14.1.24}$$

for $t \in [0,\infty)$, satisfies the SDE

$$d\hat{S}_t^\delta = \sum_{k=1}^{m}\left(\sum_{j=1}^{d}\delta_t^j\,\hat{S}_t^j\,b_t^{j,k} - \hat{S}_t^\delta\,\theta_t^k\right)dW_t^k$$
$$+ \sum_{k=m+1}^{d}\left(\left(\sum_{j=1}^{d}\delta_t^j\,\hat{S}_{t-}^j\,b_t^{j,k}\right)\left(1 - \frac{\theta_t^k}{\sqrt{h_t^{k-m}}}\right) - \hat{S}_{t-}^\delta\,\theta_t^k\right)dW_t^k \tag{14.1.25}$$

for $t \in [0,\infty)$.

The SDE (14.1.25) governs the dynamics of a benchmarked portfolio and generalizes the SDE (10.3.2). For example, by (14.2.6) and (14.2.5) the benchmarked savings account $\hat{S}_t^0$ satisfies the SDE

$$d\hat{S}_t^0 = -\hat{S}_{t-}^0\sum_{k=1}^{d}\theta_t^k\,dW_t^k \tag{14.1.26}$$

for $t \in [0,\infty)$.

Using previous notation, let us denote by $\mathcal{V}$ the set of all nonnegative portfolios in the given market. Note that the right hand side of (14.1.25) is driftless. Thus, for $S^\delta \in \mathcal{V}$ the nonnegative benchmarked portfolio $\hat{S}^\delta$ forms an $(\underline{\mathcal{A}}, P)$-local martingale when $\hat{S}^\delta$ is continuous, see Lemma 5.4.1. Also in the given JDM the driftless $\hat{S}^\delta$ is an $(\underline{\mathcal{A}}, P)$-local martingale, see Ansel & Stricker (1994). This provides by Lemma 5.2.3 for nonnegative $\hat{S}^\delta$ the important supermartingale property.

**Theorem 14.1.7.** *In a JDM any nonnegative benchmarked portfolio process $\hat{S}^\delta$ is an $(\underline{\mathcal{A}}, P)$-supermartingale, that is*

$$\hat{S}^\delta_t \geq E\left(\hat{S}^\delta_\tau \,\Big|\, \mathcal{A}_t\right) \tag{14.1.27}$$

*for all bounded $\tau \in [0, \infty)$ and $t \in [0, \tau]$.*

A proof of this theorem can be found for general semimartingale markets in Platen (2004a), or for jump diffusion markets driven by Poisson jump measures in Christensen & Platen (2005). We emphasize the fundamental fact that nonnegative benchmarked portfolios are supermartingales in general semimartingale markets as long as an almost surely finite GOP exists, see Platen (2004a). Based on the above supermartingale property of nonnegative benchmarked portfolios and the notion of arbitrage introduced in Definition 10.3.2, we can draw the following conclusion.

**Corollary 14.1.8.** *A JDM does not allow nonnegative portfolios that permit arbitrage.*

This result generalizes Corollary 10.3.3. Its proof is formally the same as the one given in Corollary 10.3.3. It is based on the fact that a nonnegative supermartingale that reaches zero remains afterwards at zero, see (10.5.4). This argument also can be used for a semimartingale market with a finite GOP to show that no nonnegative portfolio permits arbitrage, see Platen (2004a) and Christensen & Larsen (2007).

## Real World Pricing

Recall now the notion of a fair security, see Definition 9.1.2, where its benchmarked price is an $(\underline{\mathcal{A}}, P)$-martingale. Generalizing Corollary 10.4.2 yields by Lemma 10.4.1 the following result.

**Corollary 14.1.9.** *Consider a JDM with a bounded stopping time $\tau \in (0, \infty)$ and a given future $\mathcal{A}_\tau$-measurable payoff $H$ to be paid at $\tau$ with $E(\frac{H}{S^{\delta_*}_\tau}|\mathcal{A}_0) < \infty$. If there exists a fair nonnegative portfolio $S^\delta \in \mathcal{V}$ with $S^\delta_\tau = H$ almost surely, then this is the minimal nonnegative portfolio that replicates the payoff.*

This means that fair portfolios provide the best choice for an investor's tradable wealth. Otherwise, there exists a less expensive fair portfolio that achieves the same payoff $H$ at time $\tau$.

Let $H$ denote an $\mathcal{A}_\tau$-measurable payoff, with $E(\frac{H}{S_\tau^{\delta_*}}) < \infty$, to be paid at a stopping time $\tau \in [0, \infty)$. The real world pricing formula (9.1.30) can also be applied in a JDM context for pricing the payoff $H$. Its fair price $U_H(t)$ at time $t \in [0, \tau]$ is then given by the real world pricing formula

$$U_H(t) = S_t^{\delta_*} \, E\left( \frac{H}{S_\tau^{\delta_*}} \,\middle|\, \mathcal{A}_t \right), \tag{14.1.28}$$

see (10.4.1). This formula will be used when pricing derivatives in a JDM. In the same way, as discussed in Sect. 10.4, real world pricing is equivalent to risk neutral pricing as long as the candidate Radon-Nikodym derivative value $\Lambda_\theta(t) = \frac{S_t^0}{S_t^{\delta_*}} \frac{S_0^{\delta_*}}{S_0^0}$ for the hypothetical risk neutral probability measure forms an $(\underline{\mathcal{A}}, P)$-martingale.

We remark that the actuarial pricing formula in the form (9.2.6) follows from the real world pricing formula (14.1.28) also for a JDM, when the payoff $H$ paid at time $T$, is independent of the GOP value $S_T^{\delta_*}$. This is of particular importance in insurance, and also for measuring operational risk, as well as, for the pricing of weather derivatives and other payoffs that are not related to the fluctuations of the market index. We remark that even for semimartingale markets, the real world pricing formula is adequate for derivative pricing, as long as a finite GOP exists, see Platen (2004a) and Christensen & Platen (2005).

### Forward Rate Equation

As in Sect. 10.4, a simple example of a derivative is the fair zero coupon bond. It pays one unit of the domestic currency at the given maturity date $T \in [0, \infty)$. By the real world pricing formula (14.1.28) the price $P(t, T)$ at time $t$ for this derivative is given by the conditional expectation

$$P(t, T) = E\left( \frac{S_t^{\delta_*}}{S_T^{\delta_*}} \,\middle|\, \mathcal{A}_t \right) \tag{14.1.29}$$

for $t \in [0, T]$, $T \in [0, \infty)$. This leads to the benchmarked fair zero coupon bond value $\hat{P}(t, T) = \frac{P(t,T)}{S_t^{\delta_*}}$, where we can assume, similarly to (10.4.9), that it satisfies an SDE of the form

$$d\hat{P}(t, T) = -\hat{P}(t-, T) \sum_{k=1}^{d} \sigma^k(t, T) \, dW_t^k \tag{14.1.30}$$

for $t \in [0, T]$, with predictable generalized volatility process $\sigma^k(\cdot, T) = \{\sigma^k(t, T),\, t \in [0, T]\}$ for $k \in \{1, 2, \ldots, d\}$. By using a logarithmic transformation and an application of the Itô formula this becomes

$$\ln(\hat{P}(t,T)) = \ln(\hat{P}(0,T)) - \sum_{k=1}^{m} \left( \int_0^t \sigma^k(s,T)\, dW_s^k + \frac{1}{2} \int_0^t (\sigma^k(s,T))^2\, ds \right)$$

$$+ \sum_{k=m+1}^{d} \left( \int_0^t \sigma^k(s,T) \sqrt{h_s^{k-m}}\, ds + \int_0^t \ln\left(1 - \frac{\sigma^k(s,T)}{\sqrt{h_s^{k-m}}}\right) dp_s^k \right). \tag{14.1.31}$$

Hence, according to (10.4.12) the forward rate $f(t,T)$ at time $t \in [0,T]$ for the maturity $T \in [0,\infty)$ satisfies the equation

$$f(t,T) = -\frac{\partial}{\partial T} \ln(P(t,T)) = -\frac{\partial}{\partial T} \ln(\hat{P}(t,T)). \tag{14.1.32}$$

Consequently, by (14.1.31) we derive the *forward rate equation*

$$f(t,T) = f(0,T) + \sum_{k=1}^{m} \int_0^t \left( \frac{\partial}{\partial T} \sigma^k(s,T) \right) \left( \sigma^k(s,T)\, ds + dW_s^k \right)$$

$$+ \sum_{k=m+1}^{d} \int_0^t \frac{1}{1 - \frac{\sigma^k(s,T)}{\sqrt{h_s^{k-m}}}} \frac{\partial}{\partial T} \sigma^k(s,T) \left( \sigma^k(s,T)\, ds + dW_s^k \right) \tag{14.1.33}$$

for $t \in [0,T]$, see Exercise 14.2. This equation can also be found in Bruti-Liberati, Nikitopoulos-Sklibosios & Platen (2009). It is a generalization of (10.4.14) and the HJM equation (10.4.19). In the case when $\frac{\sigma^k(t,T)}{\sqrt{h_t^{k-m}}} \ll 1$ we obtain asymptotically the forward rate equation in the form of the CFM.

### GOP as Best Performing Portfolio

In Chap. 10 it was demonstrated by using various criteria that the GOP is the best performing portfolio for a CFM. Since the proofs of these results are based on the supermartingale property of nonnegative benchmarked portfolios, a similar set of proofs also applies for JDMs. Below, we generalize two of these results. First, let us formulate the property that in a JDM the GOP has the maximum long term growth rate, and, thus, almost surely, outperforms any other portfolio after a sufficiently long time.

**Theorem 14.1.10.** *In a JDM the GOP $S^{\delta_*}$ has almost surely the largest long term growth rate in comparison with that of any other strictly positive portfolio $S^\delta \in \mathcal{V}^+$, that is,*

$$\tilde{g}^{\delta_*} \stackrel{a.s.}{=} \limsup_{T\to\infty} \frac{1}{T} \ln\left(\frac{S_T^{\delta_*}}{S_0^{\delta_*}}\right) \geq \lim_{T\to\infty} \frac{1}{T} \ln\left(\frac{S_T^{\delta}}{S_0^{\delta}}\right) \stackrel{a.s.}{=} \tilde{g}^{\delta} \tag{14.1.34}$$

*almost surely.*

We now extend Corollary 10.5.3 by using the obvious extension of Definition 10.5.2 concerning the systematic outperformance of a portfolio.

**Corollary 14.1.11.** *In a JDM no strictly positive portfolio systematically outperforms the GOP in the sense of Definition 10.5.2.*

Thus, there also is no systematic way to beat the GOP in a JDM over any short or long term horizon. This is a fundamental property of the GOP and makes it very special for investment purposes. We emphasize that the proofs of the above theorem and corollary depend only on the supermartingale property of nonnegative benchmarked portfolios. As previously indicated, this supermartingale property holds for general semimartingale markets. Therefore, similar statements about the optimal performance of the GOP hold very generally, see Platen (2004a).

### Proof of Corollary 14.1.5 (*)

Under the Assumption 14.1.3 it follows from the first order conditions for identifying the maximum growth rate (14.1.15) that the optimal generalized portfolio volatilities are described by $\boldsymbol{c}_t$ as given in (14.1.17). Note from (14.1.12) that the generalized volatility of a portfolio $S^\delta \in \mathcal{V}^+$ has at time $t$ the form $\boldsymbol{\pi}_{\delta,t-}^\top \boldsymbol{b}_t$, which leads to the system of linear equations for the optimal fractions $\boldsymbol{\pi}_{\delta_*,t}$ for a GOP with

$$\boldsymbol{\pi}_{\delta_*,t-}^\top \boldsymbol{b}_t = \boldsymbol{c}_t. \tag{14.1.35}$$

By Assumption 14.1.2 the generalized volatility matrix $\boldsymbol{b}_t$ is invertible and the formula

$$\boldsymbol{\pi}_{\delta_*,t-}^\top = \boldsymbol{c}_t\, \boldsymbol{b}_t^{-1} \tag{14.1.36}$$

follows from (14.1.35) for the optimal fractions. This yields formula (14.1.18) for $t \in [0, \infty)$. These fractions are uniquely determined and so what is a GOP when its initial value is given. Consequently, the SDE (14.1.19) is, by (14.1.12), (14.1.18) and (14.1.17), the one that characterizes a GOP. □

## 14.2 Diversified Portfolios

This section considers diversified portfolios in a sequence of JDMs. It generalizes the Diversification Theorem of Sect. 10.6 to the case of markets with intensity based jumps.

### Sequence of JDMs

We rely again on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ with filtration $\underline{\mathcal{A}} = (\mathcal{A}_t)_{t\in[0,\infty)}$, satisfying the usual conditions. Continuous trading uncertainty is represented by independent standard Wiener processes $\tilde{W}^k = \{\tilde{W}_t^k, t \in$

$[0,\infty)\}$ for $k \in \mathcal{N}$. Event driven trading uncertainty is modeled by counting processes $p^k = \{p^k_t,\, t \in [0,\infty)\}$ characterized by corresponding predictable, strictly positive intensity processes $h^k = \{h^k_t,\, t \in [0,\infty)\}$ for $k \in \mathcal{N}$. We define the $k$th jump martingale $q^k = \{q^k_t,\, t \in [0,\infty)\}$ as in (14.1.3), for $k \in \mathcal{N}$.

In what follows, we consider a sequence $(\mathcal{S}^{\mathrm{JD}}_{(d)})_{d\in\mathcal{N}}$ of JDMs indexed by the number $d \in \mathcal{N}$ of risky primary security accounts. For a given integer $d$, the corresponding JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ comprises $d+1$ primary security accounts, denoted by $S^0_{(d)}, S^1_{(d)}, \ldots, S^d_{(d)}$. These include a savings account $S^0_{(d)} = \{S^0_{(d)}(t),\, t \in [0,\infty)\}$, which is a locally riskless primary security account, whose value at time $t$ is given by the exponential $S^0_{(d)}(t) = \exp\left\{\int_0^t r_s\, ds\right\}$ for $t \in [0,\infty)$. Here $r = \{r_t,\, t \in [0,\infty)\}$ denotes an adapted short rate process, which we assume, for simplicity, to be the same in each JDM. We include $d$ nonnegative, risky primary security account processes $S^j_{(d)} = \{S^j_{(d)}(t),\, t \in [0,\infty)\}$, $j \in \{1,2,\ldots,d\}$, each of which can be driven by the Wiener processes $\tilde{W}^1, \tilde{W}^2, \ldots, \tilde{W}^m$ and the jump martingales $q^1, q^2, \ldots, q^{d-m}$. Here $\mu \in [0,1]$ is a fixed real number and $m = [\mu d]$ denotes the largest integer not exceeding $\mu d$. In the $d$th JDM we have the trading uncertainty driven by the $d$-dimensional vector process $\boldsymbol{W} = \{\boldsymbol{W}_t = (\tilde{W}^1_t, \ldots, \tilde{W}^m_t, q^1_t, \ldots, q^{d-m})^\top$, $t \in [0,\infty)\}$. Obviously, if $\mu$ equals one, then we have no jumps. This covers the case of a CFM, as was discussed in Sect. 10.1.

As previously noted, for fixed $d \in \mathcal{N}$ we call a predictable stochastic process $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta^0_t, \delta^1_t, \ldots, \delta^d_t)^\top,\, t \in [0,\infty)\}$ a strategy if for each $j \in \{0,1,\ldots,d\}$ the Itô integral $\int_0^t \delta^j_s\, dS^j_{(d)}(s)$ exists. The corresponding portfolio value is then $S^\delta_{(d)}(t) = \sum_{j=0}^d \delta^j_t\, S^j_{(d)}(t)$ and satisfies the SDE

$$dS^\delta_{(d)}(t) = \sum_{j=0}^d \delta^j_t\, dS^j_{(d)}(t) \tag{14.2.1}$$

for $t \in [0,\infty)$. Note that in the $d$th JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ a given strategy $\boldsymbol{\delta}$ depends typically on $d$. However, for simplicity we shall initially suppress this dependence and shall only mention it when later required.

The corresponding $j$th fraction of a strictly positive portfolio $S^\delta_{(d)}$ is given by the expression $\pi^j_{\delta,t} = \delta^j_t \frac{S^j_{(d)}(t)}{S^\delta_{(d)}(t)}$ for $t \in [0,\infty)$ and $j \in \{0,1,\ldots,d\}$, as long as $S^\delta_{(d)}(t) > 0$.

As shown in Sect. 14.1, for each JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ there exists a unique GOP $S^{\delta_*}_{(d)} = \{S^{\delta_*}_{(d)}(t),\, t \in [0,\infty)\}$ satisfying the SDE (14.1.19) when we fix the initial value, which we set, for simplicity, to

$$S^{\delta_*}_{(d)}(0) = 1. \tag{14.2.2}$$

Any portfolio $S^\delta_{(d)}$ in the $d$th JDM, when expressed in units of $S^{\delta_*}_{(d)}$, yields a corresponding benchmarked portfolio $\hat{S}^\delta_{(d)} = \{\hat{S}^\delta_{(d)}(t),\, t \in [0,\infty)\}$, defined by

$$\hat{S}^{\delta}_{(d)}(t) = \frac{S^{\delta}_{(d)}(t)}{S^{\delta_*}_{(d)}(t)} \tag{14.2.3}$$

at time $t \in [0, \infty)$. It forms a driftless SDE, see (14.1.25).

To obtain a more compact formulation of the SDE (14.1.25), let us define the $(j,k)$th *specific generalized volatility* $\sigma^{j,k}_{(d)}(t)$, see (10.6.3)–(10.6.4), by setting

$$\sigma^{0,k}_{(d)}(t) = \theta^k_t \tag{14.2.4}$$

for $j = 0$ and $k \in \{1, 2, \ldots, d\}$ and

$$\sigma^{j,k}_{(d)}(t) = \begin{cases} \theta^k_t - b^{j,k}_t & \text{for} \quad k \in \{1, 2, \ldots, m\} \\ \theta^k_t - b^{j,k}_t \left(1 - \frac{\theta^k_t}{\sqrt{h^{k-m}_t}}\right) & \text{for} \quad k \in \{m+1, \ldots, d\} \end{cases} \tag{14.2.5}$$

for $t \in [0, \infty)$ and $j \in \{1, 2, \ldots, d\}$. By using (14.2.5) and (14.2.4) one can rewrite the SDE (14.1.25) in the form

$$d\hat{S}^{\delta}_{(d)}(t) = -\sum_{k=1}^{d} \sum_{j=0}^{d} \delta^j_t \, \hat{S}^j_{(d)}(t-) \, \sigma^{j,k}_{(d)}(t) \, dW^k_t, \tag{14.2.6}$$

and for strictly positive $S^{\delta}_{(d)}(t)$ as

$$d\hat{S}^{\delta}_{(d)}(t) = -\hat{S}^{\delta}_{(d)}(t-) \sum_{k=1}^{d} \sum_{j=0}^{d} \pi^j_{\delta,t-} \, \sigma^{j,k}_{(d)}(t) \, dW^k_t \tag{14.2.7}$$

for $t \in [0, \infty)$.

The following assumption asks for the property that the specific generalized volatilities are finite in a certain sense.

**Assumption 14.2.1.** *For all $d \in \mathcal{N}$, $T \in [0, \infty)$ and $j \in \{0, 1, \ldots, d\}$ suppose that*

$$\int_0^T \sum_{k=1}^{d} \left(\sigma^{j,k}_{(d)}(t)\right)^2 dt \le \bar{K}_T < \infty \tag{14.2.8}$$

*almost surely, where $\bar{K}_T < \infty$ denotes some finite $\mathcal{A}_T$-measurable random variable which does not depend on $d$. Furthermore, it is assumed that the inequality*

$$\sigma^{j,k}_{(d)}(t) < \sqrt{h^{k-m}_t} \tag{14.2.9}$$

*holds almost surely for all $t \in [0, \infty)$, $k \in \{m+1, m+2, \ldots, d\}$ and $j \in \{0, 1, \ldots, d\}$.*

### Sequences of Diversified Portfolios

Our aim is now to generalize the Diversification Theorem from Sect. 10.6 to the case of JDMs. Since for each $d \in \mathcal{N}$ the above model is a JDM, we can form a sequence of JDMs $(\mathcal{S}^{\mathrm{JD}}_{(d)})_{d\in\mathcal{N}}$, indexed by the number $d$ of risky primary security accounts. As in Sect. 10.6, for such a sequence of financial market models we identify a class of sequences of portfolios that approximate the corresponding sequence of GOPs.

Let us extend the Definition 10.6.2 for a sequence of diversified portfolios (DPs).

**Definition 14.2.2.** *For a sequence of JDMs* $(\mathcal{S}^{JD}_{(d)})_{d\in\mathcal{N}}$ *we call a corresponding sequence* $(S^{\delta}_{(d)})_{d\in\mathcal{N}}$ *of strictly positive portfolio processes* $S^{\delta}_{(d)}$ a sequence of DPs *if some constants* $K_1, K_2 \in (0,\infty)$ *and* $K_3 \in \mathcal{N}$ *exist, independently of* $d$*, such that for* $d \in \{K_3, K_3+1, \ldots\}$ *the inequality*

$$\left|\pi^j_{\delta,t}\right| \leq \frac{K_2}{d^{\frac{1}{2}+K_1}} \tag{14.2.10}$$

*holds almost surely for all* $j \in \{0,1,\ldots,d\}$ *and* $t \in [0,\infty)$.

Note that in (14.2.10) the strategy $\delta$ depends on $d$. Consider for fixed $d \in \mathcal{N}$ the $d$th JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ as an element of a given sequence of JDMs. By (14.2.7), when setting $\pi^j_{\delta,t} = 1$ and $\pi^i_{\delta,t} = 0$ for $i \neq j$, the $j$th benchmarked primary security account process $\hat{S}^j_{(d)} = \{\hat{S}^j_{(d)}(t),\ t \in [0,\infty)\}$, with

$$\hat{S}^j_{(d)}(t) = \frac{S^j_{(d)}(t)}{S^{\delta_*}_{(d)}(t)}, \tag{14.2.11}$$

satisfies the driftless SDE

$$d\hat{S}^j_{(d)}(t) = -\hat{S}^j_{(d)}(t-) \sum_{k=1}^{d} \sigma^{j,k}_{(d)}(t)\, dW^k_t \tag{14.2.12}$$

for $t \in [0,\infty)$ and $j \in \{0,1,\ldots,d\}$.

The $(j,k)$th specific generalized volatility $\sigma^{j,k}_{(d)}(t)$ of the benchmarked $j$th primary security account $\hat{S}^j_{(d)}(t)$ measures at time $t \in [0,\infty)$ the $j$th specific market risk with respect to the $k$th trading uncertainty $W^k$ for $k \in \{1,2,\ldots,d\}$, $j \in \{0,1,\ldots,d\}$, see Platen & Stahl (2003) and Sect. 10.6. Similarly as for CFMs, we introduce for all $t \in [0,\infty)$, $d \in \mathcal{N}$ and $k \in \{1,2,\ldots,d\}$ the $k$th *total specific volatility* for the $d$th JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ in the form

$$\hat{\sigma}^k_{(d)}(t) = \sum_{j=0}^{d} |\sigma^{j,k}_{(d)}(t)|. \tag{14.2.13}$$

Depending on $k$, the $k$th total specific volatility represents the sum of the absolute values of the specific generalized volatilities with respect to the $k$th trading uncertainty.

Similarly to Definition 10.6.3 the following regularity property of a sequence of markets ensures that each of the independent sources of trading uncertainty influences only a restricted range of benchmarked primary security accounts.

**Definition 14.2.3.** *A sequence of JDMs is called* regular *if there exists a constant* $K_5 \in (0, \infty)$*, independent of* $d$*, such that*

$$E\left(\left(\hat{\sigma}^k_{(d)}(t)\right)^2\right) \leq K_5 \tag{14.2.14}$$

*for all* $t \in [0, \infty)$*,* $d \in \mathcal{N}$ *and* $k \in \{1, 2, \ldots, d\}$*.*

### Sequence of Approximate GOPs

As in the case of a CFM, we consider for given $d \in \mathcal{N}$ in the $d$th JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ a strictly positive portfolio process $S^{\delta}_{(d)}$ with strategy $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta^0_t, \delta^1_t, \ldots, \delta^d_t)^\top,\ t \in [0, \infty)\}$. We introduce again the *tracking rate* $R^{\delta}_{(d)}(t)$ at time $t$ for the portfolio $S^{\delta}_{(d)}$ by setting

$$R^{\delta}_{(d)}(t) = \sum_{k=1}^{d} \left(\sum_{j=0}^{d} \pi^j_{\delta,t}\, \sigma^{j,k}_{(d)}(t)\right)^2 \tag{14.2.15}$$

for $t \in [0, \infty)$, see (10.6.22). By (14.2.7) one notes that the benchmarked portfolio $\hat{S}^{\delta}_{(d)}$ is constant with

$$\hat{S}^{\delta}_{(d)}(t) = \hat{S}^{\delta}_{(d)}(0) \tag{14.2.16}$$

almost surely, if and only if the tracking rate vanishes, that is,

$$R^{\delta}_{(d)}(t) = 0 \tag{14.2.17}$$

almost surely for all $t \in [0, \infty)$. Recall that by (14.2.2) $S^{\delta_*}_{(d)}(0) = 1$. In the case of a constant benchmarked portfolio $\hat{S}^{\delta}_{(d)}$, characterized by equation (14.2.16), the portfolio value $S^{\delta}_{(d)}(t)$ equals, by relation (14.2.3), a multiple of the GOP, that is,

$$S^{\delta}_{(d)}(t) = S^{\delta}_{(d)}(0)\, S^{\delta_*}_{(d)}(t) \tag{14.2.18}$$

almost surely for all $t \in [0, \infty)$. Therefore, a given portfolio process $S^{\delta}_{(d)}$ moves in step with the GOP if the tracking rate $R^{\delta}_{(d)}(t)$ remains small for all $t \in [0, \infty)$. Let us formalize this fact by extending Definition 10.6.4.

**Definition 14.2.4.** *For a sequence* $(\mathcal{S}^{JD}_{(d)})_{d\in\mathcal{N}}$ *of JDMs we call a sequence* $(S^{\delta}_{(d)})_{d\in\mathcal{N}}$ *of strictly positive portfolio processes a sequence of* approximate GOPs *if for all* $t \in [0,\infty)$ *the corresponding sequence of tracking rates vanishes in probability, see* (2.7.1)*. That is, we have*

$$\lim_{d\to\infty} R^{\delta}_{(d)}(t) \stackrel{P}{=} 0 \qquad (14.2.19)$$

*for all* $t \in [0,\infty)$*.*

To obtain a moment based sufficient condition for the identification of a sequence of approximate GOPs, we introduce, for any given $d \in \mathcal{N}$ and strictly positive portfolio process $S^{\delta}_{(d)}$, the *expected tracking rate*

$$e^{\delta}_{(d)}(t) = E\left(R^{\delta}_{(d)}(t)\right) \qquad (14.2.20)$$

at time $t \in [0,\infty)$. This leads to the following definition.

**Definition 14.2.5.** *For a sequence of JDMs* $(\mathcal{S}^{JD}_{(d)})_{d\in\mathcal{N}}$*, a sequence* $(S^{\delta}_{(d)})_{d\in\mathcal{N}}$ *of strictly positive portfolio processes is said to have a* vanishing expected tracking rate*, if their expected tracking rate converges to zero, that is,*

$$\lim_{d\to\infty} e^{\delta}_{(d)}(t) = 0 \qquad (14.2.21)$$

*for all* $t \in [0,\infty)$*.*

Using Definition 14.2.5 and the Markov inequality (1.3.57), we obtain for given $\varepsilon > 0$ and any sequence $(S^{\delta}_{(d)})_{d\in\mathcal{N}}$ of strictly positive portfolios with vanishing expected tracking rate the asymptotic inequality

$$\lim_{d\to\infty} P\left(R^{\delta}_{(d)}(t) > \varepsilon\right) \le \lim_{d\to\infty} \frac{1}{\varepsilon}\, e^{\delta}_{(d)}(t) = 0 \qquad (14.2.22)$$

for all $t \in [0,\infty)$. Therefore, by Definition 14.2.4 and inequality (14.2.22) we obtain the following result.

**Lemma 14.2.6.** *For a sequence of JDMs, any sequence of strictly positive portfolios with vanishing expected tracking rate is a sequence of approximate GOPs.*

### Diversification Theorem

Now, we state a crucial result of the benchmark approach. Using Definitions 14.2.2 and 14.2.3 the Lemma 14.2.6 allows us to extend the Diversification Theorem to the case of JDMs. Its proof is omitted since it is analogous to the one of Theorem 10.6.5 in Sect. 10.6 and can also be found in Platen (2005b).

**Theorem 14.2.7.** (Diversification Theorem for JDMs) *For a regular sequence of JDMs* $(\mathcal{S}^{JD}_{(d)})_{d\in\mathcal{N}}$*, each sequence* $(S^{\delta}_{(d)})_{d\in\mathcal{N}}$ *of DPs is a sequence of approximate GOPs. Moreover, for any* $d \in \{K_3, K_3+1, \ldots\}$ *and* $t \in [0,\infty)$*, the expected tracking rate of a given DP* $S^{\delta}_{(d)}$ *satisfies the inequality*

$$e^{\delta}_{(d)}(t) \leq \frac{(K_2)^2\, K_5}{d^{2K_1}}. \tag{14.2.23}$$

*Here the constants* $K_1, K_2, K_3$ *and* $K_5$ *are the same as in Definitions 14.2.2 and 14.2.3.*

The Diversification Theorem shows that for a regular sequence of JDMs any sequence of DPs approximates the GOP. This is highly relevant for the practical applicability of the benchmark approach, as previously discussed in Sect. 10.6. In particular, it allows one to approximate the GOP by a diversified market index without the need of an exact calculation of the fractions of the GOP. We emphasize that this result is model independent, which makes it very robust. The Diversification Theorem can be generalized under appropriate assumptions to the case of semimartingale markets, as will be shown in forthcoming work.

### Diversification in an MMM Setting

In Sect. 10.6 we provided some examples for diversified portfolios in a Black-Scholes type CFM. These examples demonstrate that the asymptotic properties of approximate GOPs do not need extremely large numbers of primary security accounts to be practically relevant. In a JDM with only a few rare events this is not so easy to demonstrate unless one generates an extremely large number of primary security accounts. However, in practice, the number of primary security accounts is indeed very large and the default of a single stock, even a large one, does not significantly change the value of the market portfolio. Let us now illustrate the fundamental phenomenon of diversification by simulating diversified portfolios in an MMM type setting, for simplicity, without jumps.

We consider the following multi-asset stylized MMM, which we discussed in Sect. 13.2. This example also demonstrates how to construct efficiently a market model under the benchmark approach. Firstly, we introduce the savings account in the form

$$S^0_{(d)}(t) = \exp\{r\, t\} \tag{14.2.24}$$

with constant short rate $r > 0$ for $t \in [0,T]$, $d \in \mathcal{N}$. The discounted GOP drift is set in all denominations to

$$\alpha^{\delta_*}_t = \alpha_0 \, \exp\{\eta\, t\} \tag{14.2.25}$$

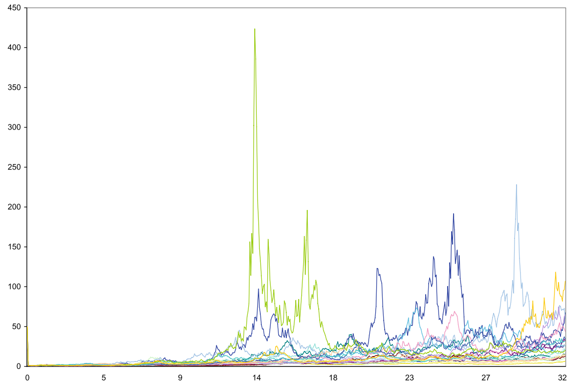



**Fig. 14.2.1.** Primary security accounts under the MMM

with net growth rate $\eta > 0$ and initial parameter $\alpha_0 > 0$. We model the $j$th benchmarked primary security account by the expression

$$\hat{S}^j_{(d)}(t) = \frac{1}{Y^j_t \, \alpha^{\delta_*}_t} \tag{14.2.26}$$

for all $j \in \{0, 1, \dots, d\}$. In this context $Y^j_t$ is the time $t$ value of the SR process $Y^j$, which satisfies the SDE

$$dY^j_t = \left(1 - \eta \, Y^j_t\right) dt + \sqrt{Y^j_t} \, dW^j_t \tag{14.2.27}$$

for $t \in [0, T]$, where we set $Y^j_0 = \frac{1}{\eta}$ for $j \in \{0, 1, \dots\}$. Also $W^0, W^1, \dots$ are independent standard Wiener processes.

Now, with (14.2.11) the GOP is obtained as the ratio

$$S^{\delta_*}_{(d)}(t) = \frac{S^0_{(d)}(t)}{\hat{S}^0_{(d)}(t)}. \tag{14.2.28}$$

Hence, by (14.2.11) the value of the $j$th primary security account is given by

$$S^j_{(d)}(t) = \hat{S}^j_{(d)}(t) \, S^{\delta_*}_{(d)}(t) \tag{14.2.29}$$

for $t \in [0, \infty)$, $j \in \{1, 2, \dots, d\}$ and $d \in \mathcal{N}$. By starting from the savings account and the benchmarked primary security accounts we have modeled all primary security accounts and the GOP in the denomination of the domestic currency.

We now simulate $d = 50$ primary security accounts $S^j_{(d)}$, $j \in \{0, 1, \dots, d\}$, for a period of $T = 32$ years, where we set $r = \eta = \alpha_0 = 0.05$. We show in Fig. 14.2.1 the trajectories of the first twenty risky primary security accounts. One notes their typical increase but also a decline of some of the securities. It is

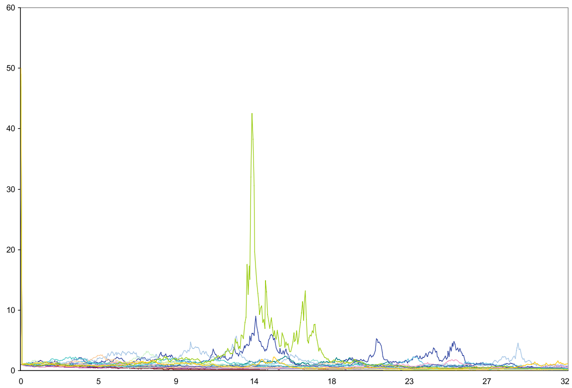



**Fig. 14.2.2.** Benchmarked primary security accounts

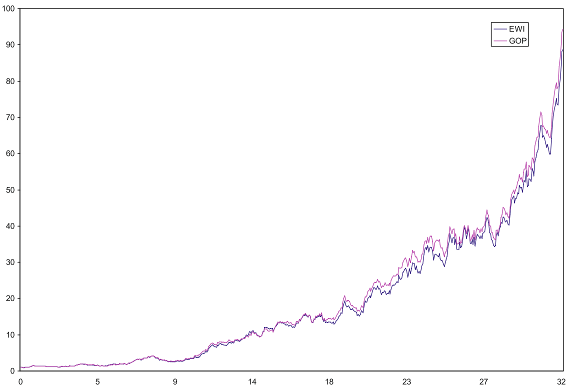



**Fig. 14.2.3.** GOP and EWI

noticeable that the primary security accounts have some common fluctuations. These are caused by the general market risk as captured by the GOP, which is shown in Fig. 14.2.3. In Fig. 14.2.2 we plot the corresponding benchmarked primary security accounts. These are strict supermartingales, as discussed previously in Chap. 13.

Figure 14.2.3 shows the equi-value weighted index (EWI) together with the GOP. One notes the closeness of the GOP and the EWI as predicted by the above Diversification Theorem. Figure 14.2.4 displays a market index, where its constituents represent simply one unit of each primary security account. Here one notes that the market index is initially a good proxy of the GOP. After an initial time period of about 13 years some extremely large stock values emerge, as can be seen in Fig. 14.2.1. The resulting large fractions of these stocks distort the performance of the market index. These fractions of the corresponding primary security accounts are simply too large to be acceptable as those of a DP and, thus, violate the conditions of the Diversification Theorem.

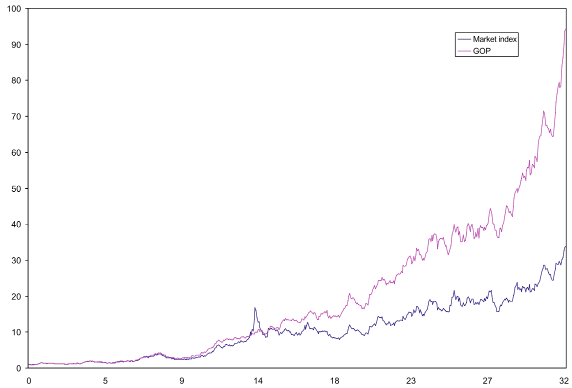



**Fig. 14.2.4.** GOP and market index

One can say that the market index is in our example, no longer interpretable as a DP after about 13 years because the fractions of a few excellent performing stocks are larger than the average fraction by magnitudes. The EWI does not suffer in this way and is in our example a good proxy for the GOP, as can be seen from Fig. 14.2.3. We emphasize that even for a market with only 50 risky primary security accounts a rather good approximation of the GOP by DPs like the EWI is obtained. Further experiments with other DPs reveal a similar behavior as shown in Fig. 14.2.3.

The Diversification Theorem identifies DPs as proxies for the GOP without any particular modeling assumptions on the market dynamics. This diversification phenomenon is, therefore, very robust. However, if the fractions of some primary security accounts become too large in a portfolio, then such a portfolio cannot be interpreted as a DP and it is unlikely to be a good proxy of the GOP.

## 14.3 Mean-Variance Portfolio Optimization

This section generalizes some of the results on mean-variance portfolio optimization that we presented in Chap. 11. It turns out that the kind of two fund separation of locally optimal portfolios into combinations of savings account and GOP, which we observed for a CFM, does not hold any longer in the same manner. Different classes of optimal portfolios arise in a JDM for different types of optimization objectives. For instance, Sharpe ratio maximization does not lead, in general, to portfolios that are a combination of the GOP and savings account.

### Locally Optimal Portfolios

Our objective here is to try to generalize the results of Sect. 11.1 on locally optimal portfolios. Given a strictly positive portfolio $S^\delta$, its discounted value $\bar{S}^\delta_t = \frac{S^\delta_t}{S^0_t}$ satisfies the SDE

$$d\bar{S}^\delta_t = \sum_{k=1}^{d} \psi^k_{\delta,t} \left(\theta^k_t \, dt + dW^k_t\right) \tag{14.3.1}$$

by (14.1.12) and an application of the Itô formula. Here

$$\psi^k_{\delta,t} = \sum_{j=1}^{d} \delta^j_t \, b^{j,k}_t \, \bar{S}^\delta_{t-} \tag{14.3.2}$$

is called the $k$th *generalized diffusion coefficient* at time $t$ for $k \in \{1, 2, \ldots, d\}$ and $t \in [0, \infty)$. Obviously, by (14.3.1) and (14.3.2), the discounted portfolio process $\bar{S}^\delta$ has *discounted drift*

$$\alpha^\delta_t = \sum_{k=1}^{d} \psi^k_{\delta,t} \, \theta^k_t \tag{14.3.3}$$

for $t \in [0, T]$. This drift measures the portfolio's trend at time $t$. The fluctuations of a discounted portfolio $\bar{S}^\delta$ can be measured at time $t$ by its *aggregate generalized diffusion coefficient*

$$\gamma^\delta_t = \sqrt{\sum_{k=1}^{d} \left(\psi^k_{\delta,t}\right)^2} \tag{14.3.4}$$

at time $t \in [0, \infty)$. Note that by relation (14.1.4) we have standardized the variances of the increments of the driving martingales $W^1, W^2, \ldots, W^d$ such that they equal the corresponding time increments, as is the case for standard Wiener processes.

For a given level of the aggregate generalized diffusion coefficient $\gamma^{\tilde{\delta}}_t > 0$, suppose that an investor aims to maximize the portfolio drift $\alpha^\delta_t$ of a discounted portfolio $\bar{S}^\delta$. This objective can be interpreted as a possible generalization of mean-variance portfolio optimization in the sense of Markowitz (1959) to the case of a JDM. More precisely, let us identify the class of SDEs for the portfolios of investors who prefer locally optimal portfolios, defined in the following sense:

**Definition 14.3.1.** *A strictly positive portfolio process $S^{\tilde{\delta}} \in \mathcal{V}^+$ that maximizes the portfolio drift (14.3.3) among all strictly positive portfolio processes $S^\delta \in \mathcal{V}^+$ with a given aggregate generalized diffusion coefficient level $\gamma^{\tilde{\delta}}_t$ is called* locally optimal*, that is,*

$$\gamma_t^\delta = \gamma_t^{\tilde{\delta}} \qquad \text{and} \qquad \alpha_t^\delta \le \alpha_t^{\tilde{\delta}} \tag{14.3.5}$$

*almost surely for all* $t \in [0, \infty)$.

This definition generalizes our Definition 11.1.1 to the case of JDMs.

## Mean-Variance Portfolio Selection Theorem

For the following analysis we use the *total market price of risk*

$$|\boldsymbol{\theta}_t| = \sqrt{\sum_{k=1}^{d} \left(\theta_t^k\right)^2} \tag{14.3.6}$$

and the weighting factor

$$G(t) = \sum_{k=1}^{d} \sum_{j=1}^{d} \theta_t^k \, b_t^{-1\,j,k} \tag{14.3.7}$$

for $t \in [0, \infty)$. The following condition generalizes Assumption 11.1.2. It excludes the trivial situation of having the savings account as GOP.

**Assumption 14.3.2.** *In a JDM suppose that*

$$0 < |\boldsymbol{\theta}_t| < \infty \tag{14.3.8}$$

*and*

$$G(t) \neq 0 \tag{14.3.9}$$

*almost surely for all* $t \in [0, \infty)$.

Now, we can formulate a mean-variance portfolio selection theorem which generalizes the results of Theorem 11.1.3. It identifies the structure of the drift and generalized diffusion coefficients of the SDE of a discounted locally optimal portfolio.

**Theorem 14.3.3.** *Under Assumption 14.3.2, any discounted locally optimal portfolio* $\bar{S}^\delta$ *satisfies in a JDM the SDE*

$$d\bar{S}_t^\delta = \bar{S}_t^\delta \, \frac{\left(1 - \pi_{\delta,t}^0\right)}{G(t)} \sum_{k=1}^{d} \theta_t^k \left(\theta_t^k \, dt + dW_t^k\right), \tag{14.3.10}$$

*with optimal fractions*

$$\pi_{\delta,t}^j = \frac{\left(1 - \pi_{\delta,t}^0\right)}{G(t)} \sum_{k=1}^{d} \theta_t^k \, b_t^{-1\,j,k} \tag{14.3.11}$$

*for all* $t \in [0, \infty)$ *and* $j \in \{1, 2, \ldots, d\}$.

The proof of this theorem is analogous to that of Theorem 11.1.3. It is, therefore, omitted, but it can be found in Platen (2006b). According to Theorem 14.3.3, the family of discounted locally optimal portfolios is characterized by a single parameter process, namely the fraction of wealth $\pi^0_{\delta,t}$ held in the savings account at time $t$. However, we shall see that, in general, it is not the GOP which arises as the mutual risky portfolio in the resulting two fund separation.

### Mutual Fund

Let us select a particular locally optimal portfolio $S^{\delta_{\mathrm{MF}}}$, which we call the *mutual fund* (MF), by choosing

$$\pi^0_{\delta_{\mathrm{MF}},t} = 1 - G(t) \tag{14.3.12}$$

for $t \in [0, \infty)$. By (14.3.10) the MF satisfies the SDE

$$dS^{\delta_{\mathrm{MF}}}_t = S^{\delta_{\mathrm{MF}}}_{t-} \left( r_t\, dt + \sum_{k=1}^{d} \theta^k_t \left( \theta^k_t\, dt + dW^k_t \right) \right) \tag{14.3.13}$$

for $t \in [0, \infty)$. Note that this SDE is very similar to that of a GOP in a CFM, see (10.2.8). However, in general, it is not the same SDE in the given JDM, as we shall see below.

By Theorem 14.3.3 it follows that any locally optimal portfolio $S^{\delta}$ can be obtained at any time by investing a fraction of wealth in the MF $S^{\delta_{\mathrm{MF}}}$ and holding the remaining fraction in the savings account. Therefore, Theorem 14.3.3 can be interpreted as a *mutual fund theorem*, see Merton (1973a). In this sense we have again two fund separation, see Corollary 11.1.4. The main difference here compared to the previous result obtained under a CFM is that the MF in a JDM, in general, does not coincide with the GOP. This can be seen when comparing the SDE (14.1.19) for the GOP and the SDE (14.3.13) for the MF. The MF coincides in a JDM with the GOP only if the market prices of event risk $\theta^{m+1}_t, \ldots, \theta^d_t$ are zero. Thus, mean-variance or Sharpe ratio maximization does, in general, not provide two fund separation into GOP and savings account. Further results in this direction can be found in Platen (2006b) and Christensen & Platen (2007).

For locally optimal portfolios the up and down movements of asset prices are weighted symmetrically by generalized diffusion coefficients. This is sufficient in a CFM for the purpose of identifying a superior asset allocation. For a practically useful portfolio selection in a JDM one needs to take into account the entire range of possible asset price jumps. Upward jumps are favorable for the investor, however, downward jumps can be disastrous. This asymmetric weighting of jumps can be conveniently modeled by utility functions.

The maximization of expected utility appears to be a useful objective in a JDM. In Sect. 11.3 we maximized expected utility from discounted terminal wealth for a CFM. The extension of this result to the case of a JDM is beyond the scope of this book.

## 14.4 Real World Pricing for Two Market Models

This section considers two examples of JDMs, a *Merton model* (MM) and a minimal market model with jumps (MMM). For both models real world pricing for some common payoffs is applied along the lines of results in Hulley, Miller & Platen (2005).

In the MM case, our aim is to illustrate how real world pricing retrieves the risk neutral prices for these instruments familiar from the literature. Of course, one could apply the standard risk neutral theory to obtain the pricing formulas under the MM, but this would defeat our purpose of illustrating real world pricing under the benchmark approach. In the case of the MMM, we wish to exhibit derivative pricing formulas where risk neutral pricing is not applicable and for what we believe is a more realistic market model.

### Specifying a Continuous GOP

In a JDM $\mathcal{S}^{\mathrm{JD}}_{(d)}$ let us interpret the GOP as a large diversified portfolio that is expressed in units of, say, US dollars, $d \in \mathcal{N}$. One may think of a diversified market portfolio or market index. Then aggregating all the jumps in the underlying primary security accounts is assumed to produce noise which is approximately continuous. In other words, we would expect the jumps to be invisible to an observer of the GOP. According to the SDE (14.1.19), the only way to eliminate jumps from the GOP dynamics is by setting the market prices of event risk equal to zero. This is a key assumption that has been used in Merton (1976) for the MM. Of course, small jumps can be asymptotically modeled by some Wiener processes. Henceforth, the following simplifying assumption will be used.

**Assumption 14.4.1.** *The market prices of event risks are zero, that is*

$$\theta_t^k = 0, \tag{14.4.1}$$

*for each $k \in \{m+1, \ldots, d\}$ and all $t \in [0, \infty)$.*

Note that there is technically no problem to extend the following examples to the case of nonzero market prices of event risk. Substitution of (14.4.1) into (14.1.19) produces the following SDE for the GOP

$$dS_t^{\delta_*} = S_t^{\delta_*} \left( r_t\, dt + \sum_{k=1}^{m} \theta_t^k \left( \theta_t^k\, dt + dW_t^k \right) \right), \tag{14.4.2}$$

for all $t \in [0, \infty)$, with

$$S_0^{\delta_*} = 1. \tag{14.4.3}$$

The solution to (14.4.2) is given by

$$S_t^{\delta_*} = \exp\left\{ \int_0^t \left( r_s + \frac{1}{2} \sum_{k=1}^{m} \left(\theta_s^k\right)^2 \right) ds + \sum_{k=1}^{m} \int_0^t \theta_s^k\, dW_s^k \right\}, \tag{14.4.4}$$

for all $t \in [0, T]$.

### Benchmarked Primary Security Accounts

The SDEs for the benchmarked primary security accounts are derived from (14.2.7) by setting $\pi^j_{\delta,t} = 1$ for $i = j$ and $\pi^i_{\delta,t} = 0$ otherwise, yielding

$$d\hat{S}^j_t = -\hat{S}^j_{t-} \sum_{k=1}^{d} \sigma^{j,k}_t \, dW^k_t, \tag{14.4.5}$$

for all $j \in \{0, 1, \ldots, d\}$ and $t \in [0, \infty)$, with $\hat{S}^j_0 = S^j_0$. Here in our JDM we have set $\sigma^{j,k}_t = \sigma^{j,k}_{(d)}(t)$ for all $j \in \{0, 1, \ldots, d\}$, $k \in \{1, 2, \ldots, d\}$ and $t \in [0, \infty)$. Recall that $W^{m+1}, \ldots, W^d$ are compensated, normalized jump martingales with corresponding intensity processes $h^1, \ldots, h^{d-m}$, respectively. From (14.4.5), via the Itô formula we obtain, see Sect. 6.4, the explicit expression

$$\begin{aligned}\hat{S}^j_t = S^j_0 \exp\left\{ -\frac{1}{2} \int_0^t \sum_{k=1}^{m} \left(\sigma^{j,k}_s\right)^2 ds - \sum_{k=1}^{m} \int_0^t \sigma^{j,k}_s \, dW^k_s \right\} \\ \times \exp\left\{ \int_0^t \sum_{k=m+1}^{d} \sigma^{j,k}_s \sqrt{h^{k-m}_s} \, ds \right\} \prod_{k=m+1}^{d} \prod_{l=1}^{p^{k-m}_t} \left( 1 - \frac{\sigma^{j,k}_{\tau^k_l -}}{\sqrt{h^{k-m}_{\tau^k_l -}}} \right)\end{aligned} \tag{14.4.6}$$

for each $j \in \{0, 1, \ldots, d\}$ and all $t \in [0, \infty)$. Here $(\tau^k_l)_{l \in \mathcal{N}}$ denotes the sequence of jump times of the counting process $p^k$ for the events of $k$th type, $k \in \{m+1, \ldots, d\}$.

Under the benchmark approach the benchmarked primary security accounts are the pivotal objects of study. The savings account together with the benchmarked primary security accounts are sufficient to specify the entire investment universe, see (14.2.28)–(14.2.29). For example, $S^{\delta_*}_t = \frac{S^0_t}{\hat{S}^0_t}$, for all $t \in [0, \infty)$, see (14.1.24), derives the GOP in terms of the savings account and the benchmarked savings account. Also, $S^j_t = \hat{S}^j_t S^{\delta_*}_t = \hat{S}^j_t \frac{S^0_t}{\hat{S}^0_t}$, for each $j \in \{1, \ldots, d\}$ and all $t \in [0, \infty)$, factors each primary security account in terms of the corresponding benchmarked primary security account, the savings account and the benchmarked savings account.

Before presenting the MM and the MMM we introduce some simplifying notation. Define the processes $|\sigma^j| = \{|\sigma^j_t|, t \in [0, \infty)\}$ for $j \in \{0, 1, \ldots, d\}$, by setting

$$|\sigma^j_t| = \sqrt{\sum_{k=1}^{m} \left(\sigma^{j,k}_t\right)^2}. \tag{14.4.7}$$

We also require the *aggregate continuous noise processes* $\hat{W}^j = \{\hat{W}^j_t, t \in [0, \infty)\}$ for $j \in \{0, 1, \ldots, d\}$, defined by

$$\hat{W}^j_t = \sum_{k=1}^{m} \int_0^t \frac{\sigma^{j,k}_s}{|\sigma^j_s|} \, dW^k_s. \tag{14.4.8}$$

By Lévy's Theorem for the characterization of the Wiener process, see Sect. 6.5, it follows that $\hat{W}^j$ is a Wiener process for each $j \in \{0, 1, \dots, d\}$. Note that these Wiener processes can be correlated. Furthermore, we require Assumption 14.1.2, such that the generalized volatility matrix $\boldsymbol{b}_t = [b_t^{j,k}]_{j,k=1}^d$ is for all $t \in [0, \infty)$ invertible. Recall by (14.2.5) that

$$b_t^{j,k} = \theta_t^k - \sigma_t^{j,k} \tag{14.4.9}$$

for $k \in \{1, 2, \dots, m\}$ and by (14.4.1) and (14.2.5) that

$$b_t^{j,k} = -\sigma_t^{j,k} \tag{14.4.10}$$

for $k \in \{m+1, \dots, d\}$, $j \in \{1, 2, \dots, d\}$ and $t \in [0, \infty)$.

In both models presented in this section we assume, for simplicity, that the parameters governing their jump behavior are constant. Thus, the counting processes $p^k$ are, in fact, time homogenous Poisson processes with constant intensities, such that

$$h_t^k = h^k > 0 \tag{14.4.11}$$

for each $k \in \{1, 2, \dots, d-m\}$ and all $t \in [0, \infty)$. Also, the jump ratios $\sigma_t^{j,k}$ for the benchmarked primary security accounts are assumed to be constant, and so that

$$\sigma_t^{j,k} = \sigma^{j,k} \leq \sqrt{h^{k-m}} \tag{14.4.12}$$

for all $j \in \{0, 1, \dots, d\}$, $k \in \{m+1, \dots, d\}$ and $t \in [0, \infty)$. Note that Assumption 14.4.1 on zero market prices of event risk ensures that (14.4.11) does not violate Assumption 14.1.3. Also, Assumption 14.4.1 and relation (14.2.5) ensure that (14.4.12) satisfies Assumption 14.1.1.

Using (14.4.7)–(14.4.12), we can rewrite the benchmarked $j$th primary security account in (14.4.6) as the product

$$\hat{S}_t^j = \hat{S}_t^{j,\mathrm{c}}\, S_t^{j,\mathrm{d}} \tag{14.4.13}$$

with continuous part

$$\hat{S}_t^{j,\mathrm{c}} = S_0^j \exp\left\{ -\frac{1}{2} \int_0^t |\sigma_s^j|^2\, ds - \int_0^t |\sigma_s^j|\, d\hat{W}_s^j \right\} \tag{14.4.14}$$

and compensated jump part

$$S_t^{j,\mathrm{d}} = \exp\left\{ \sum_{k=m+1}^d \sigma^{j,k} \sqrt{h^{k-m}}\, t \right\} \prod_{k=m+1}^d \left( 1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}} \right)^{p_t^{k-m}} \tag{14.4.15}$$

for each $j \in \{0, 1, \dots, d\}$ and all $t \in [0, \infty)$. The two specific models for the benchmarked primary security accounts, which we now present, differ in terms of how the continuous processes (14.4.14) are modeled. The jump processes (14.4.15) are, for simplicity, chosen to be the same in both cases. Forthcoming work will model stochastic intensities in natural extensions of the MMM.

### The Merton Model

The Merton model (MM) is the standard market model when including event risk with all parameters constant. We describe now a modification of the jump diffusion model introduced in Merton (1976), see Sect. 7.6. Each benchmarked primary security account can be expressed as the product of a driftless geometric Brownian motion and an independent jump martingale. Therefore, it is itself a martingale. The MM arises if one assumes that all parameter processes, that is, the short rate, the volatilities and the jump intensities, are constant. In addition to (14.4.11) and (14.4.12) we have then $r_t = r$ and $\sigma_t^{j,k} = \sigma^{j,k}$ for each $j \in \{0, 1, \ldots, d\}$, $k \in \{1, 2, \ldots, m\}$ and $t \in [0, \infty)$. In this case (14.4.14) can be written as

$$\hat{S}_t^{j,\mathrm{c}} = S_0^j \exp\left\{ -\frac{1}{2} \, |\sigma^j|^2 \, t - |\sigma^j| \, \hat{W}_t^j \right\} \tag{14.4.16}$$

for each $j \in \{0, 1, \ldots, d\}$ and all $t \in [0, \infty)$. In this special case, the benchmarked primary security accounts are the products of driftless geometric Brownian motions and compensated Poisson processes. The model is similar to that introduced in Samuelson (1965b), which was extended in Merton (1976) to include jumps. We refer to this model as the Merton model (MM). It is sometimes also called the Merton jump diffusion model.

By Assumption 14.4.1 and relations (14.4.13)–(14.4.15), the benchmarked savings account $\hat{S}^0$ exhibits no jumps. Furthermore, $\hat{S}^0$ satisfies Novikov's condition, see (9.5.12), and is, thus, a continuous martingale. Consequently, with this specification of the market, the benchmarked savings account is a Radon-Nikodym derivative process and an $(\underline{\mathcal{A}}, P)$-martingale. Therefore, Girsanov's theorem, see Sect. 9.5, is applicable, and so the standard risk neutral pricing approach can be used. While not advocating the MM as an accurate description of observed market behavior, its familiarity makes it useful for illustrating real world pricing under the benchmark approach.

### A Minimal Market Model with Jumps

The *minimal market model* (MMM) is generalized here to a case with jumps. For simplicity, we suppose the parameters associated with the jump parts of the benchmarked primary security accounts to be constant. Their continuous parts are modeled as inverted time transformed squared Bessel processes of dimension four. Consequently, each benchmarked primary security account is the product of an inverted, time transformed squared Bessel process of dimension four and an independent jump martingale. Since inverted squared Bessel processes of dimension four are strict local martingales, see (8.7.21), the benchmarked savings account is not a martingale in the MMM, and hence a viable equivalent risk neutral probability measure does not exist. We advocate real world pricing for derivatives using the GOP as numeraire and the real world probability measure as pricing measure.

Without imposing significant constraints on the parameter processes, and working within the full generality of Sect. 14.1, we have shown in Sect. 13.2 that the discounted GOP follows a time transformed squared Bessel process of dimension four. Since the discounted GOP is given by $\frac{S_t^{\delta_*}}{S_t^0} = \frac{1}{\hat{S}_t^0}$ for all $t \in [0, \infty)$, it follows that the benchmarked savings account is an inverted time transformed squared Bessel process of dimension four. A version of the MMM for the continuous part of the benchmarked primary security accounts, see Sect. 13.2, is obtained by modeling the resulting time transformations as exponential functions. We provide here an outline of this model in the context of this section. For further details we refer to Chap. 13 or Hulley et al. (2005).

For each $j \in \{0, 1, \ldots, d\}$, let $\eta^j \in \Re$ and define the function $\alpha^j : \Re^+ \to \Re^+$ by setting

$$\alpha^j(t) = \alpha_0^j \exp\{\eta^j t\} \tag{14.4.17}$$

for all $t \in [0, \infty)$ with $\alpha_0^j > 0$. We refer to $\eta^j$ again as the net growth rate of the $j$th primary security account, for $j \in \{0, 1, \ldots, d\}$. Next, we define the $j$th square root process $Y^j = \{Y_t^j, \, t \in [0, \infty)\}$ for $j \in \{0, 1, \ldots, d\}$, through the system of SDEs

$$dY_t^j = \left(1 - \eta^j Y_t^j\right) dt + \sqrt{Y_t^j}\, d\hat{W}_t^j \tag{14.4.18}$$

for each $j \in \{0, 1, \ldots, d\}$ and all $t \in [0, \infty)$, with $Y_0^j = \frac{1}{\alpha_0^j S_0^j}$. The continuous parts $\hat{S}_t^{j,\mathrm{c}}$ of the benchmarked primary security accounts (14.4.14) are modeled in terms of these square root processes by setting

$$\hat{S}_t^{j,\mathrm{c}} = \frac{1}{\alpha^j(t) Y_t^j} \tag{14.4.19}$$

for each $j \in \{0, 1, \ldots, d\}$ and all $t \in [0, \infty)$. Since (14.4.19) combined with (14.4.13) and (14.4.14) represents a version of the MMM for benchmarked primary security accounts we shall henceforth refer to it as such in this section.

As previously mentioned, between jumps the benchmarked primary security accounts are inverted time transformed squared Bessel processes of dimension four. The time transformations are deterministic in the given version of the MMM. More precisely, define the continuous strictly increasing functions $\varphi^j \colon \Re^+ \to \Re^+$ for $j \in \{0, 1, \ldots, d\}$ by setting

$$\varphi^j(t) = \varphi_0^j + \frac{1}{4} \int_0^t \alpha^j(s)\, ds \tag{14.4.20}$$

for each $j \in \{0, 1, \ldots, d\}$ and all $t \in [0, \infty)$ with $\varphi_0^j \in \Re^+$. Continuity and monotonicity imply that $\varphi^j$ possesses an inverse $(\varphi^j)^{-1} \colon [\varphi_0^j, \infty) \to \Re^+$ for each $j \in \{0, 1, \ldots, d\}$. Now define the processes $X^j = \{X_\varphi^j, \, \varphi \in [\varphi_0^j, \infty)\}$ for each $j \in \{0, 1, \ldots, d\}$ by setting

$$X^j_{\varphi^j(t)} = \alpha^j(t) Y^j_t = \frac{1}{\hat{S}^{j,c}_t} \tag{14.4.21}$$

for each $j \in \{0, 1, \ldots, d\}$ and all $t \in [0, \infty)$. It then follows, see Sect. 8.7, that $X^j$ is a squared Bessel process of dimension four, so that $\frac{1}{\hat{S}^{j,c}}$ is such time transformed squared Bessel process under the time transformation $(\varphi^j)^{-1}$ for each $j \in \{0, 1, \ldots, d\}$.

Under the MMM the benchmarked savings account is a strict local martingale, and hence a strict supermartingale, see Lemma 5.2.2 (i). This is also the candidate Radon-Nikodym derivative process employed by Girsanov's theorem to transform from the real world probability measure $P$ to a hypothetical equivalent risk neutral probability measure, see Sects. 9.4 and 13.3. However, the fact that the candidate Radon-Nikodym derivative is not an $(\underline{\mathcal{A}}, P)$-martingale rules out this measure transformation. Consequently, risk neutral derivative pricing is impossible within the MMM, and we shall resort to the more general real world pricing under the benchmark approach. Chapter 13 showed that the MMM is attractive for a number of reasons. In particular, it follows from economic reasoning when using the discounted GOP drift as the main parameter process. The modest number of parameters employed makes it a practical tool.

### Zero Coupon Bonds

We first consider a standard default-free zero coupon bond, paying one unit of the domestic currency at its maturity $T \in [0, \infty)$. According to the real world pricing formula (14.1.28), the value of the zero coupon bond at time $t$ is given by

$$P(t,T) = S^{\delta_*}_t \, E\left( \frac{1}{S^{\delta_*}_T} \,\middle|\, \mathcal{A}_t \right) = \frac{1}{\hat{S}^0_t} \, E\left( \exp\left\{ - \int_t^T r_s \, ds \right\} \hat{S}^0_T \,\middle|\, \mathcal{A}_t \right) \tag{14.4.22}$$

for all $t \in [0, T]$. We shall now examine (14.4.22) under the two market models outlined above.

In the MM case, since $\hat{S}^0$ is an $(\underline{\mathcal{A}}, P)$-martingale we obtain

$$P(t,T) = \exp\{-r(T-t)\} \frac{1}{\hat{S}^0_t} \, E\left( \hat{S}^0_T \,\middle|\, \mathcal{A}_t \right) = \exp\{-r(T-t)\} \tag{14.4.23}$$

for all $t \in [0, T]$. In other words, we obtain the usual bond pricing formula determined by discounting at the short rate. This is fully in line with the results under risk neutral pricing, see Sect. 9.4.

To simplify the notation let us set in the MMM case

$$\lambda^j_t = \frac{1}{\hat{S}^j_t (\varphi^j(t) - \varphi^j(T))} \tag{14.4.24}$$

for $t \in [0,T]$ and $j \in \{0,1,\ldots,d\}$, where $\lambda^j_T = \infty$. It is argued in Miller & Platen (2005), with some empirical support, that the interest rate process and the discounted GOP can be assumed to be independent. If we accept this, and apply it in the MMM case to (14.4.22), while remembering that $\hat{S}^0_T = \hat{S}^{0,\mathrm{c}}_T$, we obtain

$$\begin{aligned} P(t,T) &= E\left(\exp\left\{-\int_t^T r_s\,ds\right\}\,\Big|\,\mathcal{A}_t\right)\frac{1}{\hat{S}^0_t}\,E\left(\hat{S}^0_T\,\Big|\,\mathcal{A}_t\right) \\ &= E\left(\exp\left\{-\int_t^T r_s\,ds\right\}\,\Big|\,\mathcal{A}_t\right)\left(1-\exp\left\{-\frac{1}{2}\lambda^0_t\right\}\right) \end{aligned} \tag{14.4.25}$$

for all $t \in [0,T]$, from (8.7.23) and (14.4.24).

### Forward Contracts

In this subsection we fix $j \in \{0,1,\ldots,d\}$, $T \in [0,\infty)$ and $t \in [0,T]$. Consider now a *forward contract*, see (10.4.26), with the delivery of one unit of the $j$th primary security account at the maturity date $T$, which is written at time $t \in [0,T]$. The value of the forward contract at the writing time $t$ is defined to be zero. According to the real world pricing formula (14.1.28) the *forward price* $F^j(t,T)$ at time $t \in [0,T]$ for this contract is then determined by the relation

$$S^{\delta_*}_t\,E\left(\frac{F^j(t,T)-S^j_T}{S^{\delta_*}_T}\,\Big|\,\mathcal{A}_t\right) = 0. \tag{14.4.26}$$

By (14.4.22), solving this equation yields the forward price

$$F^j(t,T) = \frac{S^{\delta_*}_t\,E\left(\hat{S}^j_T\,\Big|\,\mathcal{A}_t\right)}{S^{\delta_*}_t\,E\left(\frac{1}{S^{\delta_*}_T}\,\Big|\,\mathcal{A}_t\right)} = \begin{cases} \frac{S^j_t}{P(t,T)}\,\frac{1}{\hat{S}^j_t}\,E\left(\hat{S}^j_T\,\big|\,\mathcal{A}_t\right) & \text{if } S^j_t > 0 \\ 0 & \text{if } S^j_t = 0 \end{cases} \tag{14.4.27}$$

for all $t \in [0,T]$.

In the MM case, with reference to (14.4.16), the same argument, which established that the benchmarked savings account is a continuous martingale, also applies to the driftless geometric Brownian motion $\hat{S}^{j,\mathrm{c}}$, while the compensated Poisson process $\hat{S}^{j,\mathrm{d}}$ is a jump martingale. Consequently, $\hat{S}^j$ is the product of independent martingales, and hence itself an $(\underline{\mathcal{A}},P)$-martingale. Together with (14.4.23) this enables us to write the forward price (14.4.27) as

$$F^j(t,T) = S^j_t\,\exp\{r(T-t)\} \tag{14.4.28}$$

for all $t \in [0,T]$. Thus, in the MM case we recover the standard expression for the forward price, see, for instance, Musiela & Rutkowski (2005).

In the MMM case, according to (14.4.21), $\hat{S}^{j,\mathrm{c}}$ is an inverted time transformed squared Bessel process of dimension four, while $S^{j,\mathrm{d}}$ is an independent jump martingale, as before. Thus, we obtain

$$\frac{1}{\hat{S}^j_t}\, E\left(\hat{S}^j_T \,\Big|\, \mathcal{A}_t\right) = \frac{1}{\hat{S}^{j,\mathrm{c}}_t}\, E\left(\hat{S}^{j,\mathrm{c}}_T \,\Big|\, \mathcal{A}_t\right) \frac{1}{S^{j,\mathrm{d}}_t}\, E\left(S^{j,\mathrm{d}}_T \,\Big|\, \mathcal{A}_t\right) = 1 - \exp\left\{-\frac{1}{2}\,\lambda^j_t\right\} \tag{14.4.29}$$

for all $t \in [0,T]$, by (8.7.23) and (14.4.24). Putting (14.4.27) together with (14.4.25) and (14.4.29) gives for the forward price the formula

$$F^j(t,T) = S^j_t\, \frac{1 - \exp\left\{-\frac{1}{2}\,\lambda^j_t\right\}}{1 - \exp\left\{-\frac{1}{2}\,\lambda^0_t\right\}} \left(E\left(\exp\left\{-\int_t^T r_s\, ds\right\} \,\Big|\, \mathcal{A}_t\right)\right)^{-1} \tag{14.4.30}$$

for all $t \in [0,T]$. This demonstrates that the forward price of a primary security account is a tractable quantity under the MMM.

### Asset-or-Nothing Binaries

Binary options may be regarded as basic building blocks for complex derivatives. This has been exploited in a recent approach to the valuation of exotic options, where a complex payoff is decomposed into a series of binaries, see Ingersoll (2000), Buchen (2004) and Buchen & Konstandatos (2005).

In this subsection we again fix $j \in \{0,1,\ldots,d\}$ and consider a derivative contract, with maturity $T$ and strike $K \in \Re^+$, on the $j$th primary security account. We also fix $k \in \{m+1,\ldots,d\}$ and assume that $\sigma^{j,k} \neq 0$ and $\sigma^{j,l} = 0$, for each $l \in \{m+1,\ldots,d\}$ with $l \neq k$. In other words, we assume that the $j$th primary security account responds only to the $(k-m)$th jump process. This does not affect the generality of our calculations below, but it does result in more manageable expressions. In addition, we shall assume a constant interest rate throughout the rest of this section, so that $r_t = r$, for all $t \in [0,T]$. Although this is already the case for the MM, we now require it to obtain also a compact pricing formula under the MMM.

The derivative contract under consideration is an *asset-or-nothing binary* on the $j$th primary security account. At its maturity $T$ it pays its holder one unit of the $j$th primary security account if this is greater than the strike $K$, and nothing otherwise. According to the real world pricing formula (14.1.28), its value is given by

$$\begin{aligned} A^{j,k}(t,T,K) &= S^{\delta_*}_t\, E\left(\mathbf{1}_{\{S^j_T \geq K\}}\, \frac{S^j_T}{S^{\delta_*}_T} \,\Bigg|\, \mathcal{A}_t\right) \\ &= \frac{S^j_t}{\hat{S}^j_t}\, E\left(\mathbf{1}_{\{\hat{S}^j_T \geq K (S^0_T)^{-1} \hat{S}^0_T\}}\, \hat{S}^j_T \,\Big|\, \mathcal{A}_t\right) \\ &= \frac{S^j_t}{\hat{S}^{j,\mathrm{c}}_t}\, E\Bigg(\mathbf{1}_{\{\hat{S}^{j,c}_T \geq g(p^{k-m}_T - p^{k-m}_t)\hat{S}^0_T\}} \\ &\qquad \times \exp\left\{\sigma^{j,k}\sqrt{h^{k-m}}\,(T-t)\right\}\left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^{p^{k-m}_T - p^{k-m}_t} \hat{S}^{j,\mathrm{c}}_T \,\Bigg|\, \mathcal{A}_t\Bigg) \end{aligned}$$

$$= \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^k(T-t))^n}{n!} \exp\left\{\sigma^{j,k}\sqrt{h^{k-m}}(T-t)\right\}$$

$$\times \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^n \frac{S_t^j}{\hat{S}_t^{j,c}} E\left(\mathbf{1}_{\{\hat{S}_T^{j,c} \geq g(n)\hat{S}_T^0\}} \hat{S}_T^{j,c} \,\middle|\, \mathcal{A}_t\right) \quad (14.4.31)$$

for all $t \in [0, T]$, where

$$g(n) = \frac{K}{S_t^0\, S_t^{j,d}} \exp\left\{-\left(r + \sigma^{j,k}\sqrt{h^{k-m}}\right)(T-t)\right\} \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^{-n} \quad (14.4.32)$$

for all $n \in \mathcal{N}$.

In the MM case, (14.4.31) yields the following explicit formula:

$$A^{j,k}(t, T, K) = \sum_{n=0}^{\infty} \exp\left\{-h^{k-m}(T-t)\right\} \frac{(h^{k-m}(T-t))^n}{n!}$$

$$\times \exp\left\{\sigma^{j,k}\sqrt{h^{k-m}}\,(T-t)\right\}\left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^n S_t^j\, N(d_1(n)) \quad (14.4.33)$$

for all $t \in [0, T]$, where

$$d_1(n) = \frac{\ln\left(\frac{S_t^j}{K}\right) + \left(r + \sigma^{j,k}\sqrt{h^{k-m}} + n\frac{\ln\left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)}{T-t} + \frac{1}{2}\left(\hat{\sigma}^{0,j}\right)^2\right)(T-t)}{\hat{\sigma}^{0,j}\sqrt{T-t}} \quad (14.4.34)$$

for each $n \in \mathcal{N}$. Here $N(\cdot)$ is the Gaussian distribution function. Deriving (14.4.33) is the subject of Exercise 14.3. In (14.4.34) we employ the following notation

$$\hat{\sigma}^{i,j} = \sqrt{|\sigma^i|^2 - 2\,\varrho^{i,j}\,|\sigma^i|\,|\sigma^j| + |\sigma^j|^2} \quad (14.4.35)$$

for $i, j \in \{0, 1, \ldots, d\}$, where $\varrho^{i,j}$ is the correlation between the Wiener processes $\hat{W}^i$ and $\hat{W}^j$.

For the MMM case, as we have just seen, calculating the price of a payoff written on a primary security account requires the evaluation of a double integral involving the transition density of a two-dimensional process. This is a consequence of choosing the GOP as numeraire. Closed form derivative pricing formulas can be obtained for the MM, but in the case of the MMM this is more difficult, because the joint transition densities of two squared Bessel processes are, in general, difficult to describe, see Bru (1991). A natural response to this is to solve the partial integro differential equation (PIDE) associated with the derivative price numerically by finite difference methods or Monte Carlo simulation as will be described in Chap. 15. However, to give the reader a feeling for the types of formulas that emerge from applying real world pricing in the MMM, we shall now assume, for simplicity, that the processes

$\hat{S}^0$ and $\hat{S}^{j,c}$ are independent, which is also a reasonable assumption in many practical situations. Combining (14.4.31) and (14.4.32), and remembering that $\hat{S}^0 = \hat{S}^{0,c}$, results in the formula

$$
\begin{aligned}
A^{j,k}(t,T,K) = \sum_{n=0}^{\infty} & \exp\left\{-h^{k-m}(T-t)\right\} \frac{(h^{k-m}(T-t))^n}{n!} \\
& \times \exp\left\{\sigma^{j,k}\sqrt{h^{k-m}}(T-t)\right\} \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^n \\
& \times S_t^j \left(G''_{0,4}\Big(\frac{\varphi^0(T)-\varphi^0(t)}{g(n)}; \lambda_t^j, \lambda_t^0\Big) - \exp\left\{-\frac{1}{2}\lambda_t^j\right\}\right)
\end{aligned}
\tag{14.4.36}
$$

for all $t \in [0,T]$, $k \in \{m+1,\dots,d\}$, see Exercise 14.5. Here $G''_{0,4}(x;\lambda,\lambda')$ equals the probability $P(\frac{Z}{Z'} \leq x)$ for the ratio $\frac{Z}{Z'}$ of a non-central chi-square distributed random variable $Z \sim \chi^2(0,\lambda)$ with degrees of freedom zero and non-centrality parameter $\lambda > 0$, and a non-central chi-square distributed random variable $Z' \sim \chi^2(4,\lambda')$ with four degrees of freedom and noncentrality parameter $\lambda'$. By implementing this special function one obtains the pricing formula given in (14.4.36), see Johnson et al. (1995) and Hulley et al. (2005).

### Bond-or-Nothing Binaries

In this subsection we price a *bond-or-nothing binary*, which pays the strike $K \in \Re^+$ at maturity $T$, when the $j$th primary security account at time $T$ is not less than $K$, where $j \in \{0,1,\dots,d\}$ is still fixed. As before, let us assume that the $j$th primary security account only responds to the $k$th jump martingale $W^k$, where $k \in \{m+1,\dots,d\}$ is fixed. We shall again require a constant interest rate for the MMM as well as the MM.

Since at its maturity the bond-or-nothing binary under consideration pays its holder the strike amount $K$ if the value of the $j$th primary security account is in excess of this, and nothing otherwise, the real world pricing formula (14.1.28), yields

$$
\begin{aligned}
B^{j,k}(t,T,K) &= S_t^{\delta_*} E\left(\mathbf{1}_{\{S_T^j \geq K\}} \frac{K}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t\right) \\
&= K\,P(t,T) - K S_t^{\delta_*} E\left(\mathbf{1}_{\{S_T^j < K\}} \frac{1}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_t\right) \\
&= K\,P(t,T) - K \frac{S_t^0}{\hat{S}_t^0} E\left(\mathbf{1}_{\left\{\hat{S}_T^0 > K^{-1} S_T^0 \hat{S}_T^j\right\}} \frac{\hat{S}_T^0}{S_T^0} \,\bigg|\, \mathcal{A}_t\right) \\
&= K\,P(t,T) - K \exp\{-r\,(T-t)\} \frac{1}{\hat{S}_t^0} E\left(\mathbf{1}_{\left\{\hat{S}_T^0 > g(p_T^{k-m} - p_t^{k-m})^{-1} \hat{S}_T^{j,c}\right\}} \hat{S}_T^0 \,\bigg|\, \mathcal{A}_t\right)
\end{aligned}
$$

$$= K\,P(t,T) - K\exp\{-r\,(T-t)\}\sum_{n=0}^{\infty}\exp\{-h^{k-m}\,(T-t)\}\frac{(h^{k-m}(T-t))^n}{n!}$$

$$\times\frac{1}{\hat{S}^0_t}\,E\left(\mathbf{1}_{\left\{\hat{S}^0_T > g(n)^{-1}\hat{S}^{j,c}_T\right\}}\hat{S}^0_T\,\Big|\,\mathcal{A}_t\right) \qquad (14.4.37)$$

for all $t\in[0,T]$, where $g(n)$ is given by (14.4.32), for each $n\in\mathcal{N}$.

In the MM case, (14.4.37) yields the following explicit formula:

$$B^{j,k}(t,T,K) = K\exp\{-r\,(T-t)\}$$

$$\times\left(1-\sum_{n=0}^{\infty}\exp\{-h^{k-m}\,(T-t)\}\,\frac{(h^{k-m}\,(T-t))^n}{n!}\,N(-d_2(n))\right)$$

$$= \sum_{n=0}^{\infty}\exp\{-h^{k-m}\,(T-t)\}\frac{(h^{k-m}\,(T-t))^n}{n!}\,K\exp\{-r(T-t)\}\,N(d_2(n))$$

$$(14.4.38)$$

for all $t\in[0,T]$, where

$$d_2(n) = \frac{\ln\left(\frac{S^j_t}{K}\right)+\left(r+\sigma^{j,k}\sqrt{h^{k-m}}+n\frac{\ln(1-\frac{\sigma^{j,k}}{\sqrt{h^{k-m}}})}{T-t}-\frac{1}{2}\left(\hat{\sigma}^{0,j}\right)^2\right)(T-t)}{\hat{\sigma}^{0,j}\sqrt{T-t}}$$

$$= d_1(n) - \hat{\sigma}^{0,j}\sqrt{T-t} \qquad (14.4.39)$$

for each $n\in\mathcal{N}$, see Hulley et al. (2005). Again $\hat{\sigma}^{0,j}$ is given by (14.4.35). Deriving (14.4.38) is the subject to Exercise 14.4.

For the MMM case, subject to the assumption that $\hat{S}^0_T$ and $\hat{S}^{j,c}_T$ are independent, we can combine (14.4.37), (14.4.32) and (14.4.25), to obtain

$$B^{j,k}(t,T,K) = K\exp\{-r(T-t)\}\left(1-\exp\left\{-\frac{1}{2}\lambda^0_t\right\}\right)$$

$$-\sum_{n=0}^{\infty}\exp\{-h^{k-m}\,(T-t)\}\frac{(h^{k-m}\,(T-t))^n}{n!}\,K\exp\{-r(T-t)\}$$

$$\times\left(G''_{0,4}\left((\varphi^j(T)-\varphi^j(t))\,g(n);\lambda^0_t,\lambda^j_t\right)-\exp\left\{-\frac{1}{2}\lambda^0_t\right\}\right)$$

$$= \sum_{n=0}^{\infty}\exp\{-h^{k-m}(T-t)\}\frac{(h^{k-m}\,(T-t))^n}{n!}$$

$$\times K\exp\{-r(T-t)\}\left(1-G''_{0,4}\Big((\varphi^j(T)-\varphi^j(t))g(n);\lambda^0_t,\lambda^j_t\Big)\right) \qquad (14.4.40)$$

for all $t \in [0, T]$, see Hulley et al. (2005). For the second equality in (14.4.40), we have once again used the fact that

$$\sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!}$$

is the total probability of a Poisson random variable with parameter $h^{k-m}(T-t)$. Deriving (14.4.40) is the subject of Exercise 14.6.

### European Call Options

In this subsection we fix $j \in \{0, 1, \ldots, d\}$ again and consider a European call option with maturity $T$ and strike $K \in \Re^+$ on the $j$th primary security account. As before, we make the simplifying assumption that the $j$th primary security account is only sensitive to the $(k-m)$th jump process, for some fixed $k \in \{m+1, \ldots, d\}$. We also continue to use a constant interest rate for both market models. According to the real world pricing formula (14.1.28) the European call option price is given by

$$\begin{aligned} c^{j,k}_{T,K}(t) &= S^{\delta_*}_t E\left( \frac{\left(S^j_T - K\right)^+}{S^{\delta_*}_T} \,\bigg|\, \mathcal{A}_t \right) = S^{\delta_*}_t E\left( \mathbf{1}_{\{S^j_T \geq K\}} \frac{S^j_T - K}{S^{\delta_*}_T} \,\bigg|\, \mathcal{A}_t \right) \\ &= A^{j,k}(t, T, K) - B^{j,k}(t, T, K) \end{aligned} \tag{14.4.41}$$

for all $t \in [0, T]$.

For the MM case, combining (14.4.33) and (14.4.38) gives

$$\begin{aligned} c^{j,k}_{T,K}(t) = &\sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \Bigg( \exp\left\{ \sigma^{j,k} \sqrt{h^{k-m}}\,(T-t) \right\} \\ &\times \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^n S^j_t\, N(d_1(n)) - K \exp\{-r(T-t)\} N(d_2(n)) \Bigg) \end{aligned} \tag{14.4.42}$$

for all $t \in [0, T]$, where $d_1(n)$ and $d_2(n)$ are given by (14.4.34) and (14.4.39), respectively, for each $n \in \mathcal{N}$.

It is easily seen that (14.4.42) corresponds to the original pricing formula for a European call on a stock whose price follows a jump diffusion, as given in Merton (1976). The only difference is that there the jump ratios are taken to be independent log-normally distributed, while in our case they are constant. Furthermore, this formula can be used to price an option to exchange the $j$th primary security account for the $i$th primary security account. In that case, the option pricing formula obtained instead of (14.4.42) is a generalization of that given in Margrabe (1978).

In the MMM case the European call option pricing formula is obtained by subtracting (14.4.40) from (14.4.36), according to (14.4.41), yielding

$$c_{T,K}^{j,k}(t) = \sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} \left[ \exp\left\{ \sigma^{j,k} \sqrt{h^{k-m}}\,(T-t) \right\} \right.$$
$$\times \left(1 - \frac{\sigma^{j,k}}{\sqrt{h^{k-m}}}\right)^n S_t^j \left( G''_{0,4}\left( \frac{\varphi^0(T) - \varphi^0(t)}{g(n)}; \lambda_t^j, \lambda_t^0 \right) - \exp\left\{ -\frac{1}{2}\lambda_t^j \right\} \right)$$
$$\left. - K \exp\{-r(T-t)\} \left( 1 - G''_{0,4}\left( \left(\varphi^j(T) - \varphi^j(t)\right) g(n); \lambda_t^0, \lambda_t^j \right) \right) \right] \quad (14.4.43)$$

for all $t \in [0, T]$, where $g(n)$ is given by (14.4.32), for each $n \in \mathcal{N}$ and $\lambda_t^j$ in (14.4.24).

### Defaultable Zero Coupon Bonds

We have incorporated default risk in our modeling. This allows us to study the pricing of credit derivatives. Here we consider the canonical example of such a contract, namely a defaultable zero coupon bond with maturity $T$. To keep the analysis simple, fix $k \in \{m+1, \ldots, d\}$ and assume that the bond under consideration defaults at the first jump time $\tau_1^{k-m}$ of $p^{k-m}$, provided that this time is not greater than $T$. In other words, default occurs if and only if $\tau_1^{k-m} \leq T$, in which case $\tau_1^{k-m}$ is the default time. As a further simplification, we assume zero recovery upon default. According to the real world pricing formula (14.1.28), the price of this instrument is given by

$$\tilde{P}^{k-m}(t,T) = S_t^{\delta_*} E\left( \frac{\mathbf{1}_{\{\tau_1^{k-m} > T\}}}{S_T^{\delta_*}} \,\middle|\, \mathcal{A}_t \right) = S_t^{\delta_*} E\left( \frac{1}{S_T^{\delta_*}} \,\middle|\, \mathcal{A}_t \right) E\left( \mathbf{1}_{\{\tau_1^{k-m} > T\}} \,\middle|\, \mathcal{A}_t \right)$$
$$= P(t,T) P(p_T^{k-m} = 0 \,|\, \mathcal{A}_t) \quad (14.4.44)$$

for all $t \in [0, T]$. Note that the second equality above follows from the independence of the GOP and the underlying Poisson process, see (14.4.2).

Equation (14.4.44) shows that the price of the defaultable bond can be expressed as the product of the price of the corresponding default-free bond and the conditional probability of survival. In our setup the latter may be further evaluated as

$$P\left(p_T^{k-m} = 0 \,\middle|\, \mathcal{A}_t\right) = E\left( \mathbf{1}_{\{p_t^{k-m}=0\}} \mathbf{1}_{\{p_T^{k-m} - p_t^{k-m} = 0\}} \,\middle|\, \mathcal{A}_t \right)$$
$$= \mathbf{1}_{\{p_t^{k-m}=0\}} P\left( p_T^{k-m} - p_t^{k-m} = 0 \,\middle|\, \mathcal{A}_t \right)$$
$$= \mathbf{1}_{\{p_t^{k-m}=0\}} E\left( \exp\left\{ -\int_t^T h_s^{k-m}\, ds \right\} \,\middle|\, \mathcal{A}_t \right) \quad (14.4.45)$$

for all $t \in [0, T]$.

One has to combine (14.4.44) and (14.4.45) with (14.4.23) to obtain an explicit pricing formula for the defaultable bond under consideration in the MM. Similarly, one can combine (14.4.44) and (14.4.45) with (14.4.25) to obtain the pricing formula for this instrument under the MMM.

Note that the expression obtained by combining (14.4.44) and (14.4.45) is similar to the familiar formula for the price of a defaultable zero coupon bond in a simple reduced form model for credit risk, see Schönbucher (2003). However, the difference is that for this standard case expectations are computed in the literature typically with respect to an equivalent risk neutral probability measure. In particular, the survival probability is usually a risk neutral probability. In (14.4.44) and (14.4.45), however, only the real world probability measure is in evidence. The crucial advantage of the benchmark approach in such a situation is that one avoids the undesirable dichotomy of distinguishing between real world default probabilities, as determined by historical data and credit rating agencies, and hypothetical risk neutral default probabilities, as determined by observed credit spreads. Note that substantial effort has been expended on the problem of trying to reconcile real world and risk neutral probabilities of default, see, for instance, Albanese & Chen (2005). This problem is, fortunately, avoided by using the benchmark approach with real world pricing since the real world probability measure is the pricing measure.

The above two market models highlight some aspects of the benchmark approach in derivative pricing for jump diffusion markets. This methodology can be applied generally and yields for many derivative and insurance instruments explicit formulas for the MMM and its extensions.

## 14.5 Exercises for Chapter 14

**14.1.** Calculate the growth rate of a strictly positive portfolio.

**14.2.** Derive the forward rate equation (14.1.33) from the benchmarked zero coupon bond SDE (14.1.30).

**14.3.** (*) Calculate for the Merton model, given in Sect. 14.4, the price of an asset-or-nothing binary from formula (14.4.31).

**14.4.** (*) Calculate for the Merton model, as in Sect. 14.4, the price of a bond-or-nothing binary from formula (14.4.37).

**14.5.** (*) Derive for the MMM, given in Sect. 14.4, the price of an asset-or-nothing binary from formula (14.4.31).

**14.6.** (*) Derive for the MMM, as in Sect. 14.4, the pricing formula of a bond-or-nothing binary from formula (14.4.37).

# 15

# Numerical Methods

This final chapter describes a range of numerical methods that have been used for the pricing of derivative contracts and other tasks in quantitative finance. First we describe random number generation and simulation methods for scenario and Monte Carlo simulation. Finally, we introduce tree methods and numerical schemes for the solution of partial differential equations.

## 15.1 Random Number Generation

The most flexible quantitative techniques that can be used for stochastic models in finance are simulation methods. In principle, one can simulate outcomes of any kind of random variable on a computer. For instance, we have used *scenario simulation* to produce many figures in this book.

When simulations are used to estimate a sample mean of some independent outcomes, this is called a *Monte Carlo simulation*, see Boyle (1977), Kloeden & Platen (1999) and Glasserman (2004). Monte Carlo methods provide important information about functionals of the underlying model, which in many cases cannot be easily obtained by other means. Monte Carlo simulation has been widely applied, for instance, in derivative pricing and also for the calculation of risk measures or expected utilities.

In a simulation, sample values of the random variables that appear in a model need to be generated. The outputs of many independent simulations are then analyzed statistically using the Law of Large Numbers, see Chap. 2. Monte Carlo methods typically require the generation of a large quantity of random numbers.

### Linear Congruential Random Number Generators

Before the introduction of computers, random numbers were often generated mechanically, for example, by tossing a die or turning a wheel of fortune.

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

In some cases they were listed in random number tables. This is clearly impractical for large-scale applications. Often a particular sequence of random numbers should be reproducible to allow for repetition of simulation studies. The implementation of simple deterministic algorithms to generate sequences of random numbers quickly and reproducibly is essential for efficient simulation methods. These numbers are usually not truly random, but can be made to resemble random numbers. They are called, *pseudo-random numbers* and have become increasingly important for advanced quantitative work with complex probabilistic models.

Typically, software products that produce random numbers do so by using *linear congruential pseudo-random number generators.* Let $x \pmod c$ denote the remainder when the real number $x$ is divided by a number $c \neq 0$, the *modulus.* These generators have the recursive form

$$X_{n+1} = aX_n + b \pmod c, \tag{15.1.1}$$

where $a$ and $c$ are positive integers and $b$ is a nonnegative integer. The formula (15.1.1) has the following interpretation: For a given integer initial value and *seed* $X_0$, the algorithm (15.1.1) generates a sequence of integer values in the range from 0 to $c - 1$. Numerical experiments show that when the coefficients $a, b$ and $c$ are chosen appropriately, the numbers

$$U_n = \frac{X_n}{c} \tag{15.1.2}$$

appear to be uniformly distributed on the unit interval $[0, 1)$.

Since only finitely many different numbers occur, the modulus $c$ is usually chosen to be rather large, and is often also selected as a power of 2, to take advantage of the binary arithmetic used in computers. To prevent cycling with a period less than $c$, the multiplier $a$ is usually also taken to be prime relative to $c$. Often $b$ is chosen equal to zero. Today most spreadsheets, symbolic manipulation packages and compilers use linear congruential generators. One should be aware of this fact in cases where the deterministic or cyclical nature of pseudo-random numbers creates apparently non-random numerical effects.

By interrupting the cycle through shuffling the random numbers in a random way, see Kloeden, Platen & Schurz (2003), much longer cycles can be achieved. There are also alternative ways of generating pseudo-random sequences of uniformly distributed random numbers. An important example is provided by *lagged Fibonacci generators*, studied, for instance, in Kahaner, Moler & Nash (1989) or Kloeden et al. (2003).

In what follows we assume that we have access to a routine that provides us with independent uniformly distributed $U(0, 1)$ random numbers. We are now going to show how we can use such a routine to produce random numbers with given distributions. Such random number generators are available in advanced software packages.

Some modern processors provide a hardware implementation of a natural uniform random number generator that does not generate any cycles, see

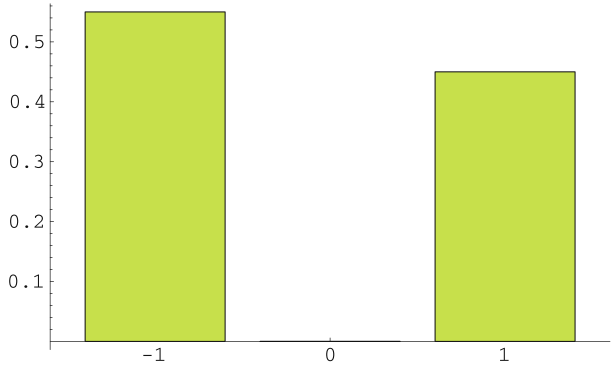



**Fig. 15.1.1.** Relative frequencies for three point distributed log-returns

Pivitt (1999), and provides outcomes close to true random numbers. The disadvantage of a natural random number generator is that one can never repeat a simulation unless one stores all the random numbers that were used in the experiment.

### *N*-Point Random Variable

A two-point random variable $X$, taking values $x_1 < x_2$ with probabilities $p_1$ and $p_2 = 1 - p_1$, respectively, can easily be generated from a $U(0,1)$ random variable $U$, by setting

$$X = \begin{cases} x_1 & \text{for} \quad 0 \leq U < p_1 \\ x_2 & \text{for} \quad p_1 \leq U < 1. \end{cases} \tag{15.1.3}$$

This idea readily extends to an $N$-state random variable $X$, taking values $x_1 < x_2 < \cdots < x_N$ with respective nonzero probabilities $p_1$, $p_2$, …, $p_N$, where

$$\sum_{i=1}^{N} p_i = 1.$$

With $s_0 = 0$ and

$$s_j = \sum_{i=1}^{j} p_i$$

we then set $X = x_j$ if $s_{j-1} \leq U < s_j$, for $j \in \{1, 2, \ldots, N\}$.

Figure 15.1.1 shows relative frequencies from 20 simulations of the three point distributed log-returns $X$ discussed in an example of the first section, with probabilities $p_1 = P(X = -1) = 0.465$, $p_2 = P(X = 0) = 0.072$, and $p_3 = P(X = 1) = 0.463$. We can compare the histogram with the corresponding probabilities that were given in Fig. 1.1.4 and note some differences between the true probabilities and the relative frequencies that were generated by the twenty observations. For this set of observations there were no outcomes with zero log-return. The relative frequencies in Fig. 15.1.1 somehow resemble the probabilities for the stock log-returns shown in Fig. 1.1.4.

Another efficient way of generating two-point distributed pseudo-random numbers is described in Bruti-Liberati & Platen (2004), where an implementation of a *random bit generator* is proposed. This hardware implementable fast generator is of particular use for large-scale Monte Carlo simulations.

### Inverse Transform Method

The following method can be applied for the generation of a continuous random variable $X$ with a probability distribution function $F_X$ that is invertible. For a number $0 < U < 1$, we define $x(U)$ by

$$U = F_X(x(U)),$$

so that

$$x(U) = F_X^{-1}(U)$$

if $F_X^{-1}$ exists, or in general,

$$x(U) = \inf\{x : U \leq F_X(x)\}. \tag{15.1.4}$$

Here $\inf\{x : U \leq F_X(x)\}$ is the greatest lower bound of the set $\{x : U \leq F_X(x)\}$. If $U$ is a $U(0,1)$ random variable, then $X(U)$ will be $F_X$-distributed. Consequently, this is called the *inverse transform method* and is conveniently used when the infimum (15.1.4) is easy to evaluate. For instance, the exponential random variable with parameter $\lambda > 0$, see (1.2.5), has an invertible distribution function with

$$x(U) = F_X^{-1}(U) = -\frac{\ln(1-U)}{\lambda}$$

for $0 \leq U < 1$.

Figure 15.1.2 shows the relative frequencies of 6000 simulated realizations of an exponentially distributed random variable with $\lambda = 2$. This histogram can be compared with Fig. 1.2.2 for the exponential density with the same intensity parameter $\lambda = 2$. In principle, the inverse transform method can be used for most continuous random variables. It may, however, require some computational effort to evaluate (15.1.4).

### Box-Muller Method

Since the distribution function of a standard Gaussian random variable can only be approximated, by numerical integration of its density, the inverse transform method is not very convenient in this case. The *Box-Muller method*,

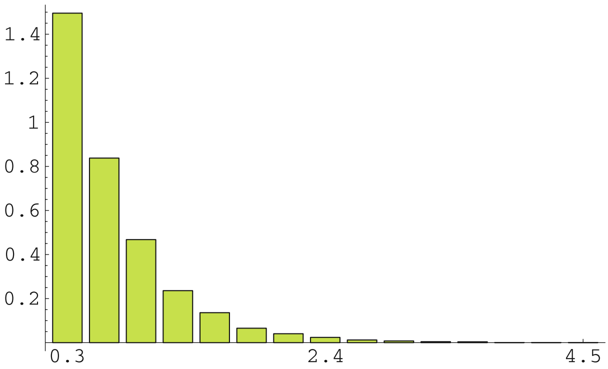



**Fig. 15.1.2.** Relative frequencies for an exponentially distributed random variable

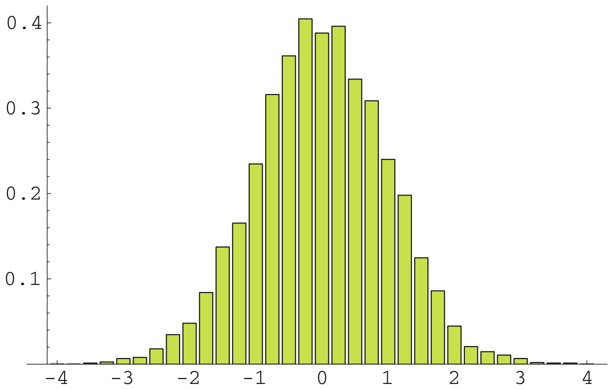



**Fig. 15.1.3.** Relative frequencies for a standard Gaussian random variable

which generates pairs of independent standard Gaussian random variables, avoids this problem.

The method is based on the observation that if $U_1$ and $U_2$ are two independent $U(0, 1)$ random variables, then $N_1$ and $N_2$ defined by

$$N_1 = \sqrt{-2\ln(U_1)}\,\cos(2\pi U_2)$$
$$N_2 = \sqrt{-2\ln(U_1)}\,\sin(2\pi U_2) \tag{15.1.5}$$

are two independent standard Gaussian random variables.

In Fig. 15.1.3 we show the frequency histogram for 6000 simulated realizations of a standard Gaussian random variable. This histogram should be compared with Fig. 1.2.3 for the standard Gaussian density.

### Marsaglia Method

There is a variant of the Box-Muller method, called the *Marsaglia method*, that avoids the time-consuming calculation of trigonometric functions. It starts by generating two independent $U(-1,1)$ random variables $V_1$ and $V_2$. One may interpret the pair $(V_1, V_2)$ as the coordinates of a point in the two-dimensional plane. We now accept only pairs $(V_1, V_2)$ that are in the interior of a circle centred at zero with radius one. The surviving pairs are such that the random variable

$$W = V_1^2 + V_2^2 \in [0,1) \tag{15.1.6}$$

is $U(0,1)$ distributed and

$$\theta = \arctan\left(\frac{V_1}{V_2}\right) \in [0, 2\,\pi) \tag{15.1.7}$$

is $U(0,2\pi)$ distributed. By using the trigonometric relations

$$\cos(\theta) = \frac{V_1}{\sqrt{W}} \quad \text{and} \quad \sin(\theta) = \frac{V_2}{\sqrt{W}} \tag{15.1.8}$$

together with (15.1.6), we can rewrite the equations (15.1.5) in the form

$$\begin{aligned} G_1 &= V_1 \sqrt{-2\,\frac{\ln(W)}{W}} \\ G_2 &= V_2 \sqrt{-2\,\frac{\ln(W)}{W}}. \end{aligned} \tag{15.1.9}$$

Therefore, those pairs $(V_1, V_2)$ which, with probability $\frac{\pi}{4} \approx 0.77$, fall in the interior of the circle yield a pair $(G_1, G_2)$ of independent standard Gaussian random variables. Even though roughly 23% of the original random numbers are discarded, this method is still more efficient than the Box-Muller method, because no trigonometric functions are involved. For extensive Monte Carlo simulations involving Gaussian random numbers the Marsaglia method can be quite efficient.

### Chi-Square Random Variables

We can generate independent realizations of a $\chi^2(n)$ random variable by constructing the sums of squares of sample values of $n$ independent standard Gaussian random variables. Figure 15.1.4 shows the corresponding frequency histogram for the range $[0, 30)$ with $n = 4$ based on 6000 samples. We can compare this plot with the probability density in Fig. 1.2.5. Note that we have a number of realizations in the range of 25 to 30, which is rather extreme. Recall that under the MMM the discounted GOP is chi-square distributed.

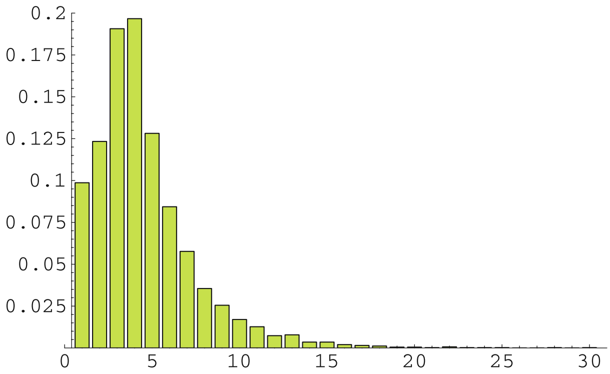



**Fig. 15.1.4.** Frequency histogram for $\chi^2(4)$ distributed random variables

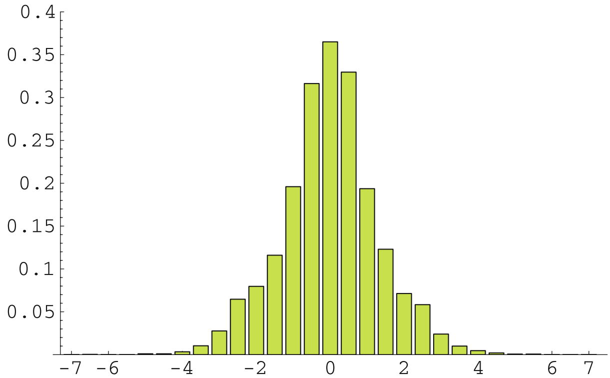



**Fig. 15.1.5.** Frequency histogram for $t(4)$ distributed random variables

## Student $t$ Random Variables

Given a $\chi^2(n)$ random variable $Z$ and an independent $N(0,1)$ random variable $Y$, (1.2.16) indicates that $X = Y(\frac{Z}{n})^{-\frac{1}{2}}$ is a central Student $t$ distributed random variable with $n$ degrees of freedom. A sequence of 6000 independent realizations of a $t(4)$ random variable yields the histogram shown in Fig. 15.1.5. We note that a few realizations are recorded far out in the tails. This means that extreme events occur from time to time. This histogram can be compared with the $t(4)$ density given in Fig. 1.2.6.

Another efficient way of generating Student $t$ distributed random variables for integer degrees of freedom uses the inverse transform method (15.1.4). If $U$ is a $U(0,1)$ random variable, then

$$x(U) = F_{t(1)}^{-1}(U) = \tan\left(\pi\left(U - \frac{1}{2}\right)\right) \qquad (15.1.10)$$

is Cauchy distributed, see Shaw (2005). Furthermore,

$$x(U) = F^{-1}_{t(2)}(U) = \frac{2\sqrt{2}\left(U - \frac{1}{2}\right)}{\sqrt{1 - 4\left(U - \frac{1}{2}\right)^2}} \tag{15.1.11}$$

is Student $t$ distributed with 2 degrees of freedom. To generate $t(4)$ distributed random variables one can apply the transform

$$x(U) = F^{-1}_{t(4)}(U) = 2\,\mathrm{sgn}\left(U - \frac{1}{2}\right)\sqrt{\frac{1}{\sqrt{z}}\cos\left(\frac{1}{3}\tan^{-1}\sqrt{\frac{1}{z} - 1}\right) - 1} \tag{15.1.12}$$

with

$$z = 1 - 4\left(U - \frac{1}{2}\right)^2. \tag{15.1.13}$$

## 15.2 Scenario Simulation

### Wagner-Platen Expansion

In quantitative finance it is essential to be able to approximate quantities that can be represented as functions of solutions of stochastic differential equations (SDEs). If a function is sufficiently smooth, then one can use the *Wagner-Platen expansion*, see Wagner & Platen (1978), Platen (1982b) and Kloeden & Platen (1999). This is a stochastic analogue of the classical deterministic Taylor expansion.

For illustration, consider the Wagner-Platen expansion for a process $X = \{X_t,\, t \in [t_0, T]\}$ which satisfies the SDE

$$X_t = X_{t_0} + \int_{t_0}^{t} a(X_s)\,ds + \int_{t_0}^{t} b(X_s)\,dW_s \tag{15.2.1}$$

for $t \in [t_0, T]$, $0 \leq t_0 < T < \infty$. Here $W$ is a standard Wiener process on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. The coefficients $a$ and $b$ are assumed to be sufficiently smooth real-valued functions, so that a unique strong solution of (15.2.1) exists. Then, for a twice continuously differentiable function $f : \Re \to \Re$, the Itô formula (6.1.12) provides the representation

$$\begin{aligned} f(X_t) &= f(X_{t_0}) + \int_{t_0}^{t}\left(a(X_s)\frac{\partial}{\partial x}f(X_s) + \frac{1}{2}\,b^2(X_s)\frac{\partial^2}{\partial x^2}f(X_s)\right)ds \\ &\quad + \int_{t_0}^{t} b(X_s)\frac{\partial}{\partial x}f(X_s)\,dW_s \\ &= f(X_{t_0}) + \int_{t_0}^{t} L^0 f(X_s)\,ds + \int_{t_0}^{t} L^1 f(X_s)\,dW_s \end{aligned} \tag{15.2.2}$$

for $t \in [t_0, T]$. Here we have used the operators

$$L^0 = a\,\frac{\partial}{\partial x} + \frac{1}{2} b^2 \frac{\partial^2}{\partial x^2} \tag{15.2.3}$$

and

$$L^1 = b\,\frac{\partial}{\partial x}. \tag{15.2.4}$$

Obviously, for the special case $f(x) \equiv x$, then $L^0 f = a$ and $L^1 f = b$ and so the representation (15.2.2) reduces to (15.2.1). If $a$ and $b$ are at least twice continuously differentiable, we can apply (15.2.2) to the functions $f = a$ and $f = b$ and substitute the resulting expressions into (15.2.1), to obtain

$$\begin{aligned} X_t &= X_{t_0} + \int_{t_0}^{t} \left( a(X_{t_0}) + \int_{t_0}^{s} L^0 a(X_z)\, dz + \int_{t_0}^{s} L^1 a(X_z)\, dW_z \right) ds \\ &\quad + \int_{t_0}^{t} \left( b(X_{t_0}) + \int_{t_0}^{s} L^0 b(X_z)\, dz + \int_{t_0}^{s} L^1 b(X_z)\, dW_z \right) dW_s \\ &= X_{t_0} + a(X_{t_0}) \int_{t_0}^{t} ds + b(X_{t_0}) \int_{t_0}^{t} dW_s + R_2 \end{aligned} \tag{15.2.5}$$

with remainder

$$\begin{aligned} R_2 &= \int_{t_0}^{t} \int_{t_0}^{s} L^0 a(X_z)\, dz\, ds + \int_{t_0}^{t} \int_{t_0}^{s} L^1 a(X_z)\, dW_z\, ds \\ &\quad + \int_{t_0}^{t} \int_{t_0}^{s} L^0 b(X_z)\, dz\, dW_s + \int_{t_0}^{t} \int_{t_0}^{s} L^1 b(X_z)\, dW_z\, dW_s. \end{aligned}$$

This is a simple example of a Wagner-Platen expansion. It can be extended through recursive applications of the Itô formula (15.2.2) to the integrands appearing in (15.2.5). For example, applying (15.2.2) to $f = L'b$ and substituting the resulting expression into (15.2.5) yields

$$X_t = X_{t_0} + a(X_{t_0}) \int_{t_0}^{t} ds + b(X_{t_0}) \int_{t_0}^{t} dW_s + L^1 b(X_{t_0}) \int_{t_0}^{t} \int_{t_0}^{s} dW_z\, dW_s + R_3, \tag{15.2.6}$$

where $R_3$ is another remainder. Notice that (15.2.6) represents $X_t$ as a weighted sum of functions of $X_{t_0}$ with multiple stochastic integrals as weights. The remainder term $R_3$ consists of the next group of multiple stochastic integrals with nonconstant integrands.

The Wagner-Platen expansion can be interpreted as a generalization of both the Itô formula and the classical deterministic Taylor formula. It is obtained via an iterated application of the Itô formula. It has many applications in quantitative finance, ranging from numerical applications, statistical and Value at Risk analysis to sensitivity analysis.

### Scenario Simulation for SDEs

In what follows we introduce methods for scenario simulation for SDEs. We consider pathwise converging discrete time approximations of solutions and list different numerical schemes.

Consider a discretization $0 = \tau_0 < \tau_1 < \cdots < \tau_n < \cdots < \tau_{n_T} = T$ of the time interval $[0, T]$ with $n_T \in \mathcal{N}$. We want to approximate a process $X = \{X_t,\, t \in [0, T]\}$ satisfying the one-dimensional SDE

$$dX_t = a(t, X_t)\, dt + \sum_{k=1}^{m} b^k(t, X_t)\, dW_t^k \tag{15.2.7}$$

for $t \in [0, T]$ with initial value $X_0 \in \Re$. Here $W^k = \{W^k,\, t \in [0, T]\}$, for $k \in \{1, 2, \ldots, m\}$, are possibly correlated Wiener processes.

One of the simplest discrete time approximations is the *Euler scheme*, see Maruyama (1955). Here a stochastic process $Y = \{Y_t, t \in [0, T]\}$ is constructed according to the iterative scheme

$$Y_{n+1} = Y_n + a(\tau_n, Y_n)\,(\tau_{n+1} - \tau_n) + \sum_{k=1}^{m} b^k(\tau_n, Y_n)\left(W^k_{\tau_{n+1}} - W^k_{\tau_n}\right), \tag{15.2.8}$$

for $n \in \{0, 1, \ldots, n_T - 1\}$, with initial value $Y_0 = X_0$ and $n_T \in \mathcal{N}$. In (15.2.8) $Y_n = Y_{\tau_n}$ denotes the value of the approximation at the discretization time $\tau_n$. Let

$$\Delta_n = \tau_{n+1} - \tau_n \tag{15.2.9}$$

for the $n$th increment of the time discretization and call

$$\Delta = \max_{n \in \{0, 1, \ldots, n_T - 1\}} \Delta_n \tag{15.2.10}$$

the *maximum step size*. We consider, for simplicity, *equidistant time discretizations* with

$$\tau_n = n\,\Delta, \tag{15.2.11}$$

where $\Delta = \Delta_n = \frac{T}{n_T} \in (0, 1)$ and some integer $n_T$.

The sequence $(Y_n)_{n \in \{0,1,\ldots,n_T\}}$ of values of the Euler approximation (15.2.8) at the discretization times $\tau_0, \tau_1, \ldots, \tau_{n_T}$ can be computed in an iterative manner. First we need to generate the random increments of the Wiener processes $W^k$, $k \in \{1, 2, \ldots, m\}$:

$$\Delta W_n^k = W^k_{\tau_{n+1}} - W^k_{\tau_n}, \tag{15.2.12}$$

for $n \in \{0, 1, \ldots, n_T - 1\}$. From (3.2.6) we know that these increments are independent Gaussian random variables with mean

$$E\left(\Delta W_n^k\right) = 0 \tag{15.2.13}$$

and variance

$$E\left(\left(\Delta W_n^k\right)^2\right) = \Delta. \tag{15.2.14}$$

To generate them we can use the Box-Muller method (15.1.5), for example.

To obtain a compact notation, we will henceforth resort to the short-hand

$$f = f(\tau_n, Y_n) \tag{15.2.15}$$

to indicate a function $f$ defined on $[0,T] \times \Re^d$, where $n \in \{0, 1, \ldots, n_T - 1\}$, when no misunderstanding is possible. Applying this abbreviation allows us to rewrite the Euler scheme (15.2.8) as

$$Y_{n+1} = Y_n + a\,\Delta_n + \sum_{k=1}^{m} b^k\,\Delta W_n^k, \tag{15.2.16}$$

for $n \in \{0, 1, \ldots, n_T - 1\}$. Usually we do not mention the initial condition, however, we typically set $Y_0 = X_0$.

The iterative structure of the Euler scheme, which generates approximate values of the diffusion process $X$ at the discretization times only, is one of the main features of the scheme. Note again that a discrete time approximation is considered to be a stochastic process defined on the whole interval $[0,T]$, although it will often be sufficient to consider its values at the discretization times only. If required, values at other times can be determined by interpolation. The simplest method is piecewise constant interpolation. Our figures often employ linear interpolation.

### Simulating Geometric Brownian Motion

To illustrate various aspects of a typical scenario simulation using a discrete time approximation of a diffusion process, we now examine a simple but important example in some detail. Let us consider the BS model, see (7.3.12). Here the diffusion is a geometric Brownian motion $X = \{X_t,\, t \in [0,T]\}$ satisfying the linear SDE

$$dX_t = a\,X_t\,dt + b\,X_t\,dW_t \tag{15.2.17}$$

for $t \in [0,\, T]$ with initial value $X_0 > 0$, where $W = \{W_t,\; t \in [0,T]\}$ is a Wiener process.

To simulate a trajectory of the Euler approximation for a given time discretization we start from the initial value $Y_0 = X_0$ and proceed by generating the sequence of values

$$Y_{n+1} = Y_n + a\,Y_n\,\Delta + b\,Y_n\,\Delta W_n \tag{15.2.18}$$

for $n \in \{0, 1, \ldots, n_T - 1\}$, according to (15.2.16). As mentioned before,

$$\Delta W_n = W_{\tau_{n+1}} - W_{\tau_n}, \tag{15.2.19}$$

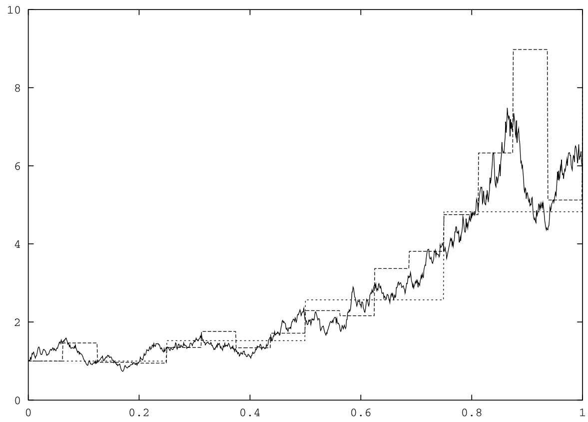



**Fig. 15.2.1.** Euler approximations with $\Delta = 0.25$ and $\Delta = 0.0625$ and the exact solution of the Black-Scholes SDE

see (15.2.12), is the $N(0, \Delta)$ distributed increment of $W$ over the subinterval $[\tau_n, \tau_{n+1}]$.

For comparison, note that the explicit solution of (15.2.17) at the discretization time is given by

$$X_{\tau_n} = X_0 \exp\left\{\left(a - \frac{1}{2}\, b^2\right) \tau_n + b \sum_{i=1}^{n} \Delta W_{i-1}\right\} \tag{15.2.20}$$

for $n \in \{0, 1, \ldots, n_T - 1\}$, see (7.3.3). We compare this with a piecewise constant Euler approximation over the time interval $[0, 1]$, using a constant step size $\Delta = 0.25$ and setting $X_0 = 1.0$, $a = 1.0$ and $b = 1.0$. Figure 15.2.1 plots the piecewise constant Euler approximation and the exact solution for the same sample path of the Wiener process. We see a substantial difference between the two paths due to the rather large step size. There is a considerable improvement in the Euler approximation when we use a smaller step size $\Delta = 0.0625$, as Fig. 15.2.1 also attests. Note that this improvement is not restricted to the terminal time, but takes the form of an overall better path.

## Strong Convergence

We shall now introduce the concept of strong order of convergence. This allows us to classify numerical schemes according to the rate at which their paths converge to those of the exact solution of an SDE, for asymptotically vanishing time step size.

One can estimate theoretically, and in some cases also calculate practically, the error of an approximation, using the following absolute error criterion. For a given maximum step size $\Delta$ and terminal time $T$, we define

$$\varepsilon(\Delta) = E\left(\left|X_T - Y_T^{\Delta}\right|\right). \tag{15.2.21}$$

Here $X_T$ is the exact solution of the SDE at time $T$ and $Y_T^\Delta$ is the discrete time approximation at time $T$. In order to classify different discrete time approximations, we introduce their *order of strong convergence*.

**Definition 15.2.1.** *We shall say that a discrete time approximation* $Y^\Delta$ converges strongly with order $\gamma > 0$ *at time* $T$ *if there exists a positive constant* $C$*, which does not depend on* $\Delta$*, and a* $\delta_0 > 0$*, such that*

$$\varepsilon(\Delta) = E\left(\left|X_T - Y_T^\Delta\right|\right) \leq C\,\Delta^\gamma \tag{15.2.22}$$

*for each* $\Delta \in (0, \delta_0)$.

We emphasize that this criterion has been constructed for the classification of, so-called, *strong approximations*. The literature contains many results establishing the strong order of convergence of particular schemes. We shall simply state results along these lines and refer the reader to Kloeden & Platen (1999) for more details.

In general, the *Euler scheme* takes the form

$$Y_{n+1} = Y_n + a\,\Delta + \sum_{k=1}^{m} b^k\,\Delta W^k, \tag{15.2.23}$$

where the Wiener process increments

$$\Delta W^k = \Delta W_n^k = W_{\tau_{n+1}}^k - W_{\tau_n}^k \tag{15.2.24}$$

for $k \in \{1, 2, \ldots, m\}$ and $n \in \{0, 1, \ldots, n_T - 1\}$ are independent of each other and $N(0, \Delta)$ distributed. For convenience, we will often suppress the dependence of random variables on the number $n$ and the length $\Delta$ of the time step. Assuming Lipschitz and linear growth conditions on the coefficients $a$ and $b$, the Euler scheme can be shown to exhibit strong order of convergence $\gamma = 0.5$. When the noise in the underlying SDE is additive, that is when the diffusion coefficient is a differentiable deterministic function of time, then it has strong order of convergence $\gamma = 1.0$.

The Euler scheme provides reasonable numerical results when the drift and diffusion coefficients are nearly constant and the time step size is sufficiently small. In general, however, it is not a very satisfactory method. Consequently, other more sophisticated schemes need to be considered. We next study a particular higher order scheme.

### Milstein Scheme

The Euler scheme uses the first two multiple stochastic integrals from the Wagner-Platen expansion (15.2.6). We now introduce the *Milstein scheme*, suggested in Milstein (1974), which exhibits strong order of convergence $\gamma = 1.0$ and is obtained by including one more term from the Wagner-Platen expansion (15.2.6). In the one-dimensional case, the Milstein scheme has the form

$$Y_{n+1} = Y_n + a\,\Delta + b\,\Delta W + \frac{1}{2}\,b\,\frac{\partial}{\partial x}\,b\left((\Delta W)^2 - \Delta\right). \tag{15.2.25}$$

In the presence of several Wiener processes it can be expressed as

$$Y_{n+1} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta W^j + \sum_{j_1,j_2=1}^{m} b^{j_1}\,\frac{\partial}{\partial x}\,b^{j_2}\,I_{(j_1,j_2)} \tag{15.2.26}$$

and involves the multiple stochastic integrals

$$I_{(j_1,j_2)} = \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s_2} dW_{s_1}^{j_1}\,dW_{s_2}^{j_2}. \tag{15.2.27}$$

These can be approximated, as demonstrated in Kloeden & Platen (1999) or Gaines & Lyons (1994).

## Commutativity

An important special case occurs when the diffusion coefficient matrix satisfies the *commutativity condition*

$$b^{j_2}\,\frac{\partial}{\partial x}\,b^{j_1} = b^{j_1}\,\frac{\partial}{\partial x}\,b^{j_2} \tag{15.2.28}$$

for all $j_1,\ j_2 \in \{1, 2, \ldots, m\}$ and $(t, x) \in [0, T] \times \Re^d$. The multi-asset Black-Scholes model is an example of a system of SDEs that satisfies this condition. Systems of SDEs with additive noise also satisfy it. Obviously, we have commutativity if the given SDE is driven by a single Wiener process.

When the commutativity condition applies, the Milstein scheme is simplified as follows:

$$Y_{n+1} = Y_n + a\Delta + \sum_{j=1}^{m} b^j \left(\Delta W^j - \frac{1}{2}\,\frac{\partial}{\partial x}\,b^j\,\Delta\right) + \frac{1}{2}\sum_{j_1,j_2=1}^{m} b^{j_1}\,\frac{\partial}{\partial x}\,b^{j_2}\,\Delta W^{j_1}\,\Delta W^{j_2}. \tag{15.2.29}$$

Note that the double Wiener integrals in (15.2.26) have been dispensed with.

## Strong Order 1.5 Taylor Scheme

By adding more terms from the Wagner-Platen expansion to the Milstein scheme in one dimension, one obtains the following *strong order 1.5 Taylor scheme*

$$Y_{n+1} = Y_n + a\,\Delta + b\,\Delta W + \frac{1}{2}\,b\,b'\left\{(\Delta W)^2 - \Delta\right\} + a'\,b\,\Delta Z$$
$$+ \frac{1}{2}\left(a\,a' + \frac{1}{2}\,b^2\,a''\right)\Delta^2 + \left(a\,b' + \frac{1}{2}\,b^2\,b''\right)\{\Delta W\,\Delta - \Delta Z\}$$
$$+ \frac{1}{2}\,b\left(b\,b'' + (b')^2\right)\left\{\frac{1}{3}\,(\Delta W)^2 - \Delta\right\}\Delta W, \qquad (15.2.30)$$

if the drift and diffusion coefficients are independent of time, see Platen & Wagner (1982). Here the random variable $\Delta Z$ is the following double stochastic integral

$$\Delta Z = \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s_2} dW_{s_1}\,ds_2. \qquad (15.2.31)$$

One can show that $\Delta Z$ is Gaussian, with mean $E(\Delta Z) = 0$, variance $E((\Delta Z)^2) = \frac{1}{3}\Delta^3$ and its covariance with the Wiener increment is $E(\Delta Z\,\Delta W) = \frac{1}{2}\Delta^2$.

Note that the random variables $\Delta W$ and $\Delta Z$, above, are easily sampled. One simply generates two independent $N(0,1)$ random variables $U_1$ and $U_2$ and then performs the following transformations

$$\Delta W = U_1\,\sqrt{\Delta}, \qquad \Delta Z = \frac{1}{2}\,\Delta^{\frac{3}{2}}\left(U_1 + \frac{1}{\sqrt{3}}\,U_2\right). \qquad (15.2.32)$$

### Explicit Strong Order 1.0 Schemes

Various first order derivative-free schemes can be obtained from the Milstein scheme (15.2.26), by replacing the derivatives in the Milstein scheme with difference approximations. The inclusion of these difference approximations requires the computation of, so-called, supporting values. An example is given by the following one-dimensional scheme, see Platen (1984),

$$Y_{n+1} = Y_n + a\,\Delta + b\,\Delta W + \frac{1}{2\sqrt{\Delta}}\left\{b(\tau_n, \bar{\Upsilon}_n) - b\right\}\left\{(\Delta W)^2 - \Delta\right\}, \quad (15.2.33)$$

with supporting value

$$\bar{\Upsilon}_n = Y_n + a\,\Delta + b\,\sqrt{\Delta}. \qquad (15.2.34)$$

The general *explicit strong order* 1.0 *scheme* is

$$Y_{n+1} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta W^j + \frac{1}{\sqrt{\Delta}} \sum_{j_1,j_2=1}^{m} \left\{b^{j_2}\left(\tau_n, \bar{\Upsilon}_n^{j_1}\right) - b^{j_2}\right\} I_{(j_1,j_2)}, \qquad (15.2.35)$$

with vector supporting values

$$\bar{\Upsilon}_n^j = Y_n + a\,\Delta + b^j\,\sqrt{\Delta} \qquad (15.2.36)$$

for $j \in \mathcal{N}$. The double Wiener integral $I_{(j_1,j_2)}$ is given by (15.2.27). If the SDE is commutative, as described in (15.2.28), then there exist explicit strong order 1.0 schemes that avoid multiple stochastic integrals, see (15.2.27). Further explicit higher strong order schemes are described in Kloeden & Platen (1999).

**Numerical Stability**

Numerical stability can be defined as the ability of a scheme to control the propagation of initial and roundoff errors. Such errors occur naturally in any simulation, but numerical methods differ in their ability to dampen them. Since numerical stability determines whether or not a scheme generates reasonable results at all, it is clearly more important than the order of convergence of the scheme.

Implicit schemes are characterized by the fact that the value of the scheme at any time is a function of itself. Therefore, it cannot be expressed as the subject of an equation involving only previous values. In general, implicit schemes offer better numerical stability than Taylor schemes or other explicit schemes. The simplest implicit strong scheme is the *drift implicit Euler scheme*

$$Y_{n+1} = Y_n + a\left(\tau_{n+1}, Y_{n+1}\right)\, \Delta + b\, \Delta W, \tag{15.2.37}$$

which has order of strong convergence $\gamma = 0.5$. As before, we employ the short-hand notation $b = b(\tau_n, Y_n)$. There is also a *family of drift implicit Euler schemes*

$$Y_{n+1} = Y_n + \left\{\theta\, a\left(\tau_{n+1}, Y_{n+1}\right) + (1-\theta)\, a\right\}\, \Delta + \sum_{j=1}^{m} b^j\, \Delta W^j, \tag{15.2.38}$$

where the parameter $\theta \in \Re$ is called the *degree of implicitness*. These exhibit an order of strong convergence $\gamma = 1.0$.

The *drift implicit Milstein scheme* is the simplest implicit counterpart of the Milstein scheme and has the form

$$Y_{n+1} = Y_n + a\left(\tau_{n+1}, Y_{n+1}\right)\, \Delta + b\, \Delta W + \frac{1}{2}\, b\, b' \left((\Delta W)^2 - \Delta\right). \tag{15.2.39}$$

In the commutative case, see (15.2.28), we obtain a *family of drift implicit Milstein schemes*

$$Y_{n+1} = Y_n + \left(\theta\, a\left(\tau_{n+1}, Y_{n+1}\right) + (1-\theta)\, a\right)\, \Delta + \sum_{j=1}^{m} b^j\, \Delta W^j$$

$$+ \frac{1}{2} \sum_{j_1, j_2 = 1}^{m} b^{j_1}\, \frac{\partial}{\partial x}\, b^{j_2} \left(\Delta W^{j_1} \Delta W^{j_2} - \mathbf{1}_{\{j_1 = j_2\}} \Delta\right). \tag{15.2.40}$$

Here the parameter $\theta \in \Re$ once again controls the degree of implicitness.

Finally, we mention a *family of drift implicit strong order* 1.0 *Runge-Kutta schemes*

$$Y_{n+1} = Y_n + \left(\theta\, a\left(\tau_{n+1}, Y_{n+1}\right) + (1-\theta)\, a^k\right) \Delta + \sum_{j=1}^{m} b^j\, \Delta W^j$$

$$+ \frac{1}{\sqrt{\Delta}} \sum_{j_1, j_2 = 1}^{m} \left(b^{j_2}\left(\tau_n, \bar{\Upsilon}_n^{j_1}\right) - b^{j_2}\right)\, I_{(j_1, j_2)} \tag{15.2.41}$$

with vector supporting values

$$\bar{\Upsilon}_n^j = Y_n + a\,\Delta + b^j\,\sqrt{\Delta}$$

for $j \in \{1, 2, \ldots, m\}$ and degree of implicitness parameter $\theta \in \Re$. Further implicit strong schemes can be found in Kloeden & Platen (1999).

### Balanced Implicit Method (*)

Note that the strong schemes discussed so far do not exhibit implicitness in their diffusion terms. Only the drift terms exhibit any implicitness. Introducing implicitness into the diffusion terms causes difficulties.

Milstein, Platen & Schurz (1998) propose a family of *balanced implicit methods* that demonstrate how to overcome this problem. In the simplest case, a balanced implicit method can be written in the form

$$Y_{n+1} = Y_n + a\,\Delta + b\,\Delta W + (Y_n - Y_{n+1})\,C_n, \tag{15.2.42}$$

where

$$C_n = c^0(Y_n)\,\Delta + c^1(Y_n)\,|\Delta W| \tag{15.2.43}$$

and $c^0$, $c^1$ represent positive real valued uniformly bounded functions. The freedom to choose $c^0$ and $c^1$ can be exploited to tailor a numerically stable scheme to the dynamics of any given SDE. However, the balanced implicit method is only of strong order $\gamma = 0.5$, since it is, in principle, a variation of the Euler scheme. It may be interpreted as a family of specific methods providing a kind of balance between approximating diffusion terms. In a number of applications, in particular, those involving SDEs with multiplicative noise, balanced implicit methods show better numerical stability than many other methods, see Fischer & Platen (1999). Such SDEs with multiplicative noise are typical in finance and filtering.

A balanced implicit method for the case with several Wiener processes can be written in the form

$$Y_{n+1} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta W_n^j + C_n(Y_n - Y_{n+1}), \tag{15.2.44}$$

where

$$C_n = c^0(\tau_n, Y_n)\,\Delta + \sum_{j=1}^{m} c^j(\tau_n, Y_n)\,|\Delta W_n^j|, \tag{15.2.45}$$

$\Delta W_n^j = W_{\tau_{n+1}}^j - W_{\tau_n}^j$ and $\Delta = \tau_{n+1} - \tau_n$ for $n \in \{0, 1, \ldots, n_T - 1\}$. Here $c^0, c^1, \ldots, c^m$ represent uniformly bounded functions.

**Strong Approximation of SDEs with Jumps (*)**

To complete this section, we briefly discuss numerical methods for scenario simulation of jump-diffusion processes with state dependent jump intensities. Such processes are important for credit, operational and insurance risk analysis. Standard models in this area are the Merton model, see (7.6.1), and the Crámer-Lundberg model, see (3.5.10) and (3.7.2).

Consider an asset price that exhibits jumps with random sizes. To incorporate this into an SDE one can use a *Poisson jump measure* $p_\varphi(\cdot,\cdot)$ with some intensity measure $\varphi(\cdot)$ on a *mark set* $\mathcal{E} = \Re\backslash\{0\}$, as described in Sect. 3.5. Here we assume that $\varphi(\mathcal{E}) < \infty$. Suppose that jumps arrive at times $\tau_1 < \tau_2 < \ldots$ with corresponding marks $v_1, v_2, \ldots$. Let $N_t$ count the number of jumps until time $t$ generated by $N_t = p_\varphi(\mathcal{E}, [0,t])$. Also let $c:\ \mathcal{E} \to \Re$ denote a real valued function. Within this framework one has

$$\sum_{k=1}^{N_t} c(v_k) = \int_0^t \int_{\mathcal{E}} c(v)\, p_\varphi(dv, dt)$$

for $t \in [0,T]$. The intensity $\frac{\varphi(dv)}{\varphi(\mathcal{E})}$ determines the probability that the magnitude of a jump lies in some infinitesimal interval $dv$ in the mark space.

A financial quantity $X_t$ that exhibits jumps can now be modeled by an SDE of the form

$$dX_t = a(t, X_t)\, dt + \sum_{k=1}^{m} b^k(t, X_t)\, dW_t^k + \int_{\mathcal{E}} c(v, t-, X_{t-})\, p_\varphi(dv, dt) \quad (15.2.46)$$

for $t \in [0,T]$ with $X_0 \in \Re$. Here the jump coefficient $c(\cdot,\cdot,\cdot)$ is dependent on the mark $v$, the time $t$ and the value $X_{t-}$ of the quantity just before time $t$. If at time $\tau$ the Poisson jump measure $p_\varphi$ generates a jump with mark $v$, then the process jumps by an amount $c(v, \tau-, X_{\tau-})$, that is,

$$X_\tau = X_{\tau-} + c(v, \tau-, X_{\tau-}). \quad (15.2.47)$$

The compensated jump measure

$$q_\varphi(dv, dt) = p_\varphi(dv, dt) - \varphi(dv)\, dt \quad (15.2.48)$$

is an $(\underline{\mathcal{A}}, P)$-martingale jump measure. Note that conditions must be imposed on $a$, $b$, $c$ and $\varphi$ to ensure the existence and uniqueness of solutions of the SDE (15.2.46), see Chap. 7.

The SDE above can be used to model stock prices with default, where the recovery rate is random, for example. Furthermore, it can model operational failures where the damage size varies randomly. Similarly, insurance claims with uncertain claim sizes can be modeled by (15.2.46) as well. Models driven by Lévy processes can also be expressed by SDEs of the type (15.2.46), see Sect. 3.6. Such models have been applied in finance by Madan & Seneta

(1990) and Barndorff-Nielsen (1998), among others. Affine jump-diffusions are considered in Duffie, Pan & Singleton (2000) and can also be characterized by (15.2.46). Furthermore, Björk et al. (1997) and Glasserman & Merener (2003) consider interest rate term structure models with jumps. The importance of SDEs with jumps in quantitative finance will increase in future years since the modeling of event risks becomes more important in applications.

Numerical methods for discrete time approximation of jump diffusions have been studied, for instance, in Platen (1982a, 1984), Platen & Rebolledo (1985), Maghsoodi & Harris (1987), Mikulevicius & Platen (1988), Maghsoodi (1996, 1998), Kubilius & Platen (2002), Glasserman & Merener (2003), Glasserman (2004), Bruti-Liberati, Nikitopoulos-Sklibosios & Platen (2006) and Bruti-Liberati & Platen (2007a, 2007b). We now follow Platen (1982a) by introducing a *jump adapted time discretization* $\{\tau_i\}_{i\in\{0,1,\ldots\}}$ with $\tau_0 < \tau_1 < \ldots$. This involves superposing the random jump times of the Poisson process $p_\varphi(\mathcal{E},[0,\cdot])$ on the discretization times of a deterministic grid. It can be constructed before the actual simulation of the solution of (15.2.46) begins. We need only to ensure that the maximum step size remains bounded by $\Delta > 0$. In other words, we require that consecutive time steps satisfy $\tau_{i+1}-\tau_i \le \Delta$ almost surely. This is achieved if one superposes the jump times on a regular time discretization with step size $\Delta$, for example.

The simplest jump adapted scheme is the *Euler scheme*. It is decoupled into a *diffusive part*

$$Y_{\tau_{i+1}-} = Y_{\tau_i} + a\,(\tau_{i+1}-\tau_i) + \sum_{j=1}^{m} b^j \left(W^j_{\tau_{i+1}} - W^j_{\tau_i}\right) \tag{15.2.49}$$

and a *jump part*

$$Y_{\tau_{i+1}} = Y_{\tau_{i+1}-} + \int_{\mathcal{E}} c(v,\tau_{i+1}-,Y_{\tau_{i+1}-})\, p_\varphi(dv,\{\tau_{i+1}\}). \tag{15.2.50}$$

The scheme (15.2.49) approximates the evolution of the diffusive part between jump times and represents a standard Euler scheme. If $\tau_{i+1}$ is a jump time of $p_\varphi$, then in (15.2.50) one typically has the value $Y_{\tau_{i+1}} \neq Y_{\tau_{i+1}-}$. Thus, (15.2.50) simulates the effect of jumps. Note that if the inter-jump diffusion of the process is captured exactly by (15.2.49), then the jump approximation (15.2.50) generates no errors and the scheme matches the exact solution of (15.2.49). The Euler scheme above can be shown to exhibit the same order of strong convergence, namely $\gamma = 0.5$, as in the pure diffusion case.

In principle, one can use all the strong schemes previously discussed for approximating the diffusive part of an SDE with jumps. Then the jumps can be handled in the same manner as shown in (15.2.49)–(15.2.50). The resulting strong order is, in general, that of the scheme that approximates the diffusion part, see Platen (1982a).

## 15.3 Classical Monte Carlo Method

In this section we describe the classical Monte Carlo method, which allows us to calculate expectations of given functions of random variables. In the next section we shall consider Monte Carlo methods for expectations of functions of solutions of SDEs.

### Monte Carlo Estimator

The classical *Monte Carlo method*, as described in Fishman (1996), for example, is probably the most flexible numerical method for evaluating an expression of the form

$$u = E(g(X)), \tag{15.3.1}$$

where $g : \Re^n \to \Re$ is some function and $X$ is an $n$-dimensional random variable on a probability space $(\Omega, \mathcal{A}, P)$. Such functionals appear frequently in finance, as derivative prices, risk measures or expected utilities, for example. The Monte Carlo method for computing (15.3.1) is simply a matter of generating $N$ independent sample paths of the random variable $g(X)$, say $g(X(\omega_1))$, $g(X(\omega_2)), \ldots, g(X(\omega_N))$, for $\omega_1, \omega_2, \ldots, \omega_N \in \Omega$. Then (15.3.1) is the sample average

$$\hat{u}_N = \frac{1}{N} \sum_{i=1}^{N} g(X(\omega_i)). \tag{15.3.2}$$

One can see that the *raw Monte Carlo estimator* $\hat{u}_N$ is *unbiased*, that is

$$E(\hat{u}_N) = u \tag{15.3.3}$$

for all $N \in \mathcal{N}$, see (2.3.6). If $\mathrm{Var}(g(X)) < \infty$, then by the strong Law of Large Numbers (LLN), see Theorem 2.1.2, it follows that $\hat{u}_N$ converges to $u$ almost surely, that is

$$\lim_{N \to \infty} \hat{u}_N \stackrel{\text{a.s.}}{=} u. \tag{15.3.4}$$

Note that $\hat{u}_N$ is a random variable for each $N \in \mathcal{N}$. Property (15.3.4) of the Monte Carlo estimator is called *strong consistency* and is a stronger condition than consistency, see (2.3.8). By a calculation used in the proof of (2.1.9), see also (1.4.49), it follows that

$$\mathrm{Var}(\hat{u}_N) = \frac{\mathrm{Var}(g(X))}{N}, \tag{15.3.5}$$

revealing that the variance of $\hat{u}_N$ decreases linearly as the number of simulated outcomes $N$ increases. This means that the deviation

$$\sqrt{\mathrm{Var}(\hat{u}_N)} = \frac{\sqrt{\mathrm{Var}(g(X))}}{\sqrt{N}}$$

only decreases like $N^{-\frac{1}{2}}$ as $N \to \infty$, see Sect. 2.2.

Although $\mathrm{Var}(g(X))$ is usually not known, an application of the weak LLN in Theorem 2.1.3 establishes that the estimate

$$\hat{V}_N = \frac{1}{N} \sum_{i=1}^{N} (g(X(\omega_i)))^2 - (\hat{u}_N)^2$$

converges in probability to $\mathrm{Var}(g(X))$, that is

$$\lim_{N \to \infty} \hat{V}_N \stackrel{P}{=} \mathrm{Var}(g(X)). \tag{15.3.6}$$

By the CLT from Theorem 2.1.4 we then know that

$$\hat{Z}_N = \frac{(\hat{u}_N - u)}{\sqrt{\mathrm{Var}(g(X))}} \sqrt{N} \tag{15.3.7}$$

converges in distribution to a standard Gaussian random variable $Z$. Therefore, we have the approximate relationship

$$\hat{u}_N \approx u + Z \sqrt{\frac{\mathrm{Var}(g(X))}{N}}, \tag{15.3.8}$$

which provides a proxy for the estimation error of a Monte Carlo simulation. It also tells us that the Monte Carlo estimator $\hat{u}_N$ is *asymptotically normal*.

### Classical Monte Carlo Simulation

As an illustrative example of classical Monte Carlo simulation, we choose a problem for which the exact answer is known. In particular, we shall compute $u = E(g(X))$, where $X \sim N(0,1)$ and $g(X)$ is defined by

$$g(X) = \left(\exp\left\{r\Delta + \sigma\sqrt{\Delta}\, X\right\}\right)^2 \tag{15.3.9}$$

with $r = 0.05$, $\sigma = 0.2$ and $\Delta = 1$. Of course, $u$ is then the second moment of a lognormally distributed random variable, which in turn could represent an asset price under the BS model. Since we know the Laplace transform of a Gaussian random variable we are able to determine the value of $u$ exactly.

$$u = E\left(\exp\left\{2\left(r\Delta + \sigma\sqrt{\Delta}\, X\right)\right\}\right) = \exp\left\{\left(r + \sigma^2\right) 2\,\Delta\right\}. \tag{15.3.10}$$

For the parameter values under consideration, we have $u \approx 1.197$. For sample sizes $N \in \{1, 2, \ldots, 2000\}$ we now calculate the raw Monte Carlo estimators

$$\hat{u}_N = \frac{1}{N} \sum_{i=1}^{N} \exp\left\{2\left(r\Delta + \sigma\sqrt{\Delta}\, X(\omega_i)\right)\right\}, \tag{15.3.11}$$

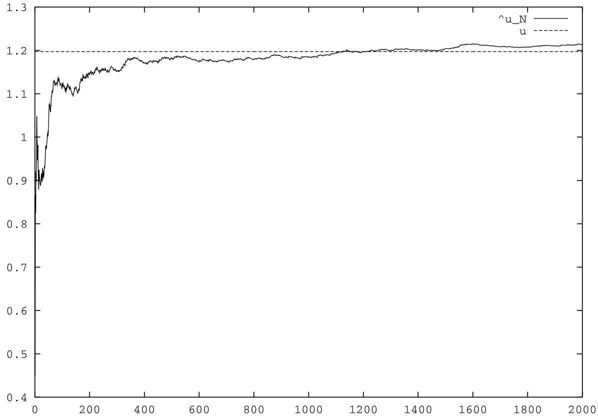



**Fig. 15.3.1.** Raw Monte Carlo estimates in dependence on the number of simulations

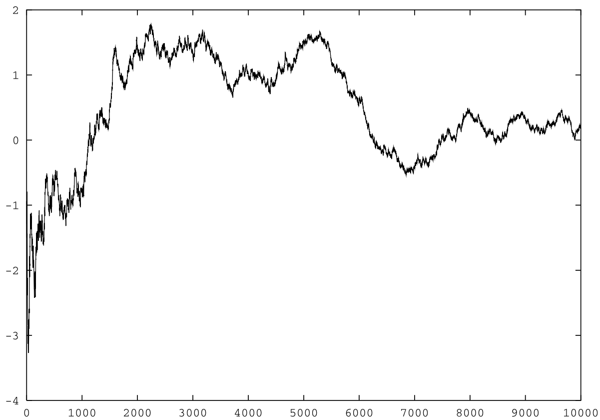



**Fig. 15.3.2.** Normalized raw Monte Carlo error

where $X(\omega_1), X(\omega_2), \ldots$ are independently sampled from the $N(0,1)$ distribution, as described in Sect. 15.1. Figure 15.3.1 plots the raw Monte Carlo estimates $\hat{u}_N$ for different sample sizes $N$. We note that after about $N \approx 1000$ simulations the estimates $\hat{u}_N$ stabilize and appear to converge towards the correct value $u \approx 1.2$. This illustrates the strong LLN in relation (15.3.4). We can also calculate the normalized quantity $\hat{Z}_N$ defined by (15.3.7), since

$$\mathrm{Var}(g(X)) = \exp\{4\,\Delta\,(r + 2\,\sigma^2)\}\,(1 - \exp\{-4\,\Delta\,\sigma^2\}) \approx 0.25.$$

Figure 15.3.2 illustrates the dependence of $\hat{Z}_N$ on $N$. It is interesting to see how $\hat{Z}_N$ converges towards some random level. In Fig. 15.3.3 we repeat the entire Monte Carlo estimation 50 times and plot the resulting normalized errors against sample size $N \in [9000, 10000]$. It is obvious that for each outcome a certain level is approached, which is clearly random. A statistical analy-

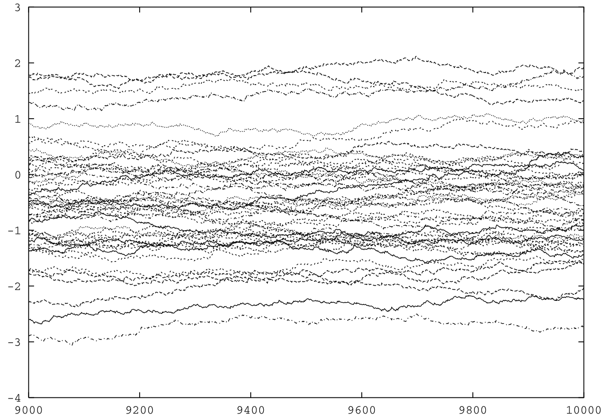



**Fig. 15.3.3.** Independent realizations of normalized Monte Carlo errors

sis would reveal that the distribution of this level is approximately standard Gaussian.

## Antithetic Variates

As we have seen in (15.3.7), the difference between the raw Monte Carlo estimator $\hat{u}_N$ and the exact value $u$ is proportional to $N^{-\frac{1}{2}}$. Unfortunately, we must accept this slow rate of convergence unless we can change the problem under consideration so that the variance of the unbiased random variable being simulated is reduced. There are many, so-called, *variance reduction* techniques that achieve this, see Fishman (1996) or Kloeden & Platen (1999). We now describe the antithetic variates technique, as an illustration of the general idea.

Assume that we are able to construct two random variables $X$ and $Y$ such that

$$E(g(X)) = E(g(Y)).$$

Then we can use the *antithetic variate Monte Carlo estimator*

$$\tilde{u}_N = \frac{1}{2}\left(\frac{1}{N}\sum_{i=1}^{N} g(X(\omega_i)) + \frac{1}{N}\sum_{i=1}^{N} g(Y(\omega_i))\right) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left(g(X)+g(Y)\right). \tag{15.3.12}$$

The variance of $\tilde{u}_N$ is obtained as

$$\begin{aligned}\mathrm{Var}(\tilde{u}_N) &= \frac{1}{4\,N}\,\mathrm{Var}(g(X)+g(Y))\\ &= \frac{1}{4\,N}\Big(\mathrm{Var}(g(X)) + 2\,\mathrm{Cov}(g(X),g(Y)) + \mathrm{Var}(g(Y))\Big).\end{aligned}$$

Obviously, if $\mathrm{Cov}(g(X),g(Y)) < 0$, then

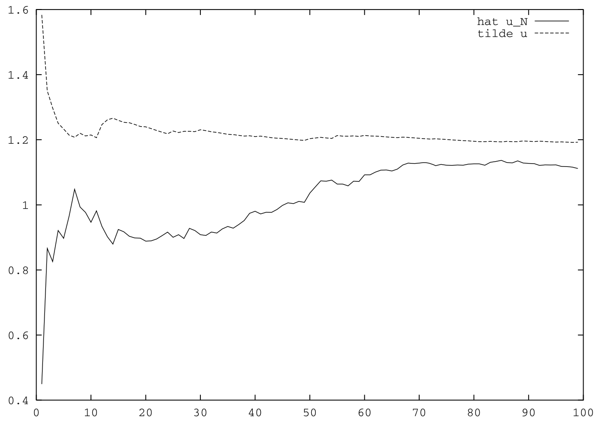



**Fig. 15.3.4.** Antithetic Monte Carlo estimator as a function of $N$

$$\mathrm{Var}(\tilde{u}_N) < \frac{1}{2\,N}\,\mathrm{Var}(g(X)) < \mathrm{Var}(\hat{u}_N), \tag{15.3.13}$$

which yields some variance reduction, when compared with (15.3.5). In particular, if the variates $g(X)$ and $g(Y)$ are perfectly negatively correlated, that is antithetic, then the variance is reduced to zero.

We illustrate the technique by applying it to our previous example. This is done by including the negative outcomes $-X(\omega_i)$ of each of the sample values $X(\omega_i)$ of the standard Gaussian random variable $X$, so that

$$\tilde{u}_N = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}\left(\exp\left\{2\left(r\Delta + \sigma\sqrt{\Delta}\,X(\omega_i)\right)\right\} + \exp\left\{2\left(r\Delta - \sigma\sqrt{\Delta}\,X(\omega_i)\right)\right\}\right)$$

is the antithetic variates estimator.

Figure 15.3.4 compares the results of using $\tilde{u}_N$ with those obtained by using the raw Monte Carlo estimator $\hat{u}_N$ for different sample sizes $N$. Note that the variance of $\tilde{u}_N$ is smaller than that of $\hat{u}_N$, indicating an improvement due to the use of antithetic variates.

## Control Variate Technique

We now discuss another variance reduction technique. Assume the existence of a *control variate function* $p : \Re \to \Re$, for which the value $\varphi = E(p(X))$ is known. We can then construct the *control variate estimator*

$$\bar{u}_N = \hat{u}_N - \alpha\,(\hat{p}_N - \varphi)\,, \tag{15.3.14}$$

where

$$\hat{p}_N = \frac{1}{N}\sum_{i=1}^{N} p(X(\omega_i))$$

and $\alpha \in \Re$ is some parameter. We aim to choose $\alpha$ so as to reduce the variance of $\bar{u}_N$. Its optimal value $\alpha^*$ is obtained by minimizing

$$\mathrm{Var}(\bar{u}_N) = \mathrm{Var}(\hat{u}_N) - 2\,\alpha\,\mathrm{Cov}(\hat{u}_N, \hat{p}_N) + \alpha^2\,\mathrm{Var}(\hat{p}_N).$$

This is easily done, since the above expression is a quadratic function of $\alpha$. The resulting optimal parameter value is

$$\alpha^* = \frac{\mathrm{Cov}(\hat{u}_N, \hat{p}_N)}{\mathrm{Var}(\hat{p}_N)} = \frac{\mathrm{Cov}(g(X), p(X))}{\mathrm{Var}(p(X))}. \tag{15.3.15}$$

For this parameter choice we obtain

$$\mathrm{Var}(\bar{u}_N) = \mathrm{Var}(\hat{u}_N) - \frac{\mathrm{Cov}(\hat{u}_N, \hat{p}_N)}{\mathrm{Var}(p_N)} = \mathrm{Var}(\hat{u}_N)\left(1 - (\varrho_{\hat{u}_N, \hat{p}_N})^2\right), \tag{15.3.16}$$

indicating a substantial reduction in variance if the control variate is highly correlated with the raw Monte Carlo estimator $\hat{u}_N$.

We now illustrate the control variate technique by applying it to our previous problem. The following control variate function will be used

$$p(X) = \sqrt{g(X)} = \exp\left\{r\,\Delta + \sigma\,\sqrt{\Delta}\,X\right\}.$$

We know that $\varphi = E(p(X)) \approx 1.1$. The control variate estimator (15.3.14) is then

$$\begin{aligned}\bar{u}_N &= \frac{1}{N}\sum_{i=1}^{N} \exp\left\{2\left(r\Delta + \sigma\,\sqrt{\Delta}\,X(\omega_i)\right)\right\} \\ &\quad - \alpha\left(\frac{1}{N}\sum_{i=1}^{N}\left(\exp\left\{r\Delta + \sigma\,\sqrt{\Delta}\,X(\omega_i)\right\} - \varphi\right)\right),\end{aligned}$$

where $\alpha$ is chosen appropriately. Typically, its optimal value $\alpha^*$ is not known. However, one can often find a good value by reasoning or experimentation. Here, (15.3.15) can be evaluated, to give

$$\alpha^* = \frac{\exp\{3\,\Delta\left(r + \frac{3}{2}\,\sigma^2\right)\}\,(1 - \exp\{-2\,\Delta\,\sigma^2\})}{\exp\{2\,\Delta\,(r + \sigma^2)\}\,(1 - \exp\{-\Delta\,\sigma^2\})} \approx 2.28.$$

With the choice $\alpha = 2.3$, Fig. 15.3.5 shows that we have an excellent control variate Monte Carlo estimator. The graph shows quick convergence to the correct value.

Note that in this example we have extracted considerable advantage from our knowledge of the properties of the functional $g(X)$. The general rule is, the more one knows about the underlying random variables, the greater is the advantage one can extract from a variance reduction technique. This will also apply to the application of variance reduction techniques to Monte Carlo methods for SDEs, as will be described in the next section.

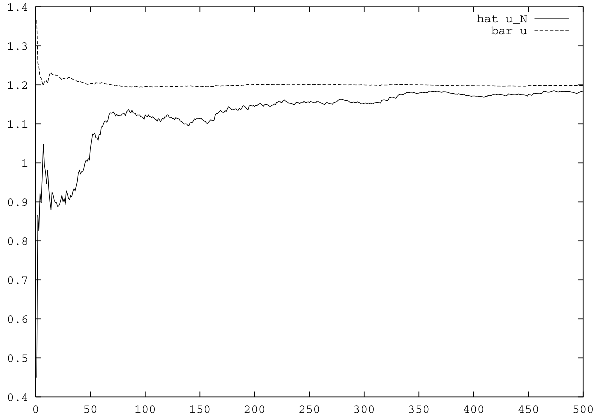



**Fig. 15.3.5.** Control variate Monte Carlo estimator as a function of $N$

## Stratified Sampling (*)

*Stratified sampling* is a variance reduction technique of long standing, which has been widely used in classical Monte Carlo simulation, see, for example, Ross (1990) and Curran (1994). Let $A_i \in \mathcal{A},\ i \in \{1, 2, \ldots, N\}$, be a partition of $\Omega$, so that

$$\bigcup_{i=1}^{N} A_i = \Omega, \qquad A_i \cap A_j = \emptyset$$

if $i \neq j$. We assume that $P(A_i) = \frac{1}{N}$ for all $i \in \{1, 2, \ldots, N\}$. Since $(A_i)_{1 \leq i \leq N}$ partitions $\Omega$, it follows that $\sigma\{A_i \,\big|\, 1 \leq i \leq N\}$ is nothing other than the collection of all finite unions, including the empty union, of the $A_i$'s. This makes the sigma-algebras $\mathcal{A}_i$ trivial. It also means that one only has piecewise constant random variables. Now let $Z : \Omega \to \Re$ be any random variable, and define the random variables $Z_{A_i}$ for $i \in \{1, 2, \ldots, N\}$ by setting $Z_{A_i} = \mathbf{1}_{A_i} Z$. Next define the random variable $\bar{Z}$ as follows

$$\bar{Z} = \frac{1}{N} \sum_{i=1}^{N} \bar{Z}_{A_i},$$

where $\bar{Z}_{A_1}, \bar{Z}_{A_2}, \ldots, \bar{Z}_{A_N}$ are independent random variables with $E(\bar{Z}_{A_i}) = E(Z_{A_i})$ and $\mathrm{Var}(\bar{Z}_{A_i}) = \mathrm{Var}(Z_{A_i})$ for $i \in \{1, 2, \ldots, N\}$. Since

$$E(\bar{Z}) = \frac{1}{N} \sum_{i=1}^{N} E\left(\bar{Z}_{A_i}\right) = \frac{1}{N} \sum_{i=1}^{N} E(Z_{A_i}) = \sum_{i=1}^{N} \int_{A_i} Z\, dP = \int_{\Omega} Z\, dP = E(Z), \tag{15.3.17}$$

$\bar{Z}$ is an unbiased estimator for $E(Z)$. Furthermore, using the independence property of $\bar{Z}_{A_i},\ i \in \{1, 2, \ldots, N\}$, we obtain the following from (1.4.35) and (1.4.43):

$$\operatorname{Var}(\bar{Z}) = \sum_{i=1}^{N} \frac{\operatorname{Var}(\bar{Z}_{A_i})}{N^2} = \sum_{i=1}^{N} \frac{\operatorname{Var}(Z_{A_i})}{N^2}$$

$$= \frac{1}{N} E(\operatorname{Var}(Z \,|\, \mathcal{A})) \leq \frac{1}{N} \operatorname{Var}(Z). \tag{15.3.18}$$

This inequality will be strict if $\operatorname{Var}(E(Z \,|\, \mathcal{A})) > 0$. Consequently, if we set

$$Z = H(X_T),$$

then we obtain an unbiased estimator for $E(H(X_T))$ from (15.3.17), with reduced variance, according to (15.3.18).

### Quasi Monte Carlo Method (*)

We now shift our attention to Monte Carlo methods that do not use pseudo-random numbers or true random numbers. There is an extensive literature on *quasi Monte Carlo methods* with overviews provided by Ripley (1983) and Niederreiter (1992). Applications to financial modeling problems have been considered, for instance, in Barraquand (1993), Paskov & Traub (1995) and Joy, Boyle & Tan (1996).

To illustrate a commonly used procedure, consider a $d$-dimensional asset price vector $\boldsymbol{X}$ and a payoff functional $H(\boldsymbol{X})$. We assume that the joint density function $p_X : \Re^d \to \Re$ of $\boldsymbol{X}$ is known, so that

$$u = E\left(H\left(\boldsymbol{X}\right)\right) = \int_{\Re^d} H(y)\, p_X(y)\, dy \tag{15.3.19}$$

can be computed if the integral on the right can be evaluated. Consequently, estimation of $u$ can be considered as a numerical integration problem over $\Re^d$. For a one-dimensional asset price $X$, if $F_X : \Re \to [0,1]$ is the distribution function for $X$, when $p_X(y) = F'_X(y)$, for all $y \in \Re$, and (15.3.19) can be expressed as

$$u = \int_0^1 H(F_X^{-1}(z))\, dz. \tag{15.3.20}$$

The valuation problem has now been transformed into the computation of a Riemann integral over the unit interval $[0,1]$. For a $d$-dimensional asset price vector $\boldsymbol{X}$, subject to certain conditions on the joint distribution function $F_X$, the right hand side of (15.3.20) can be written as a standard Riemann integral over the $d$-dimensional unit cube. Note that this technique requires the density or the distribution function of $\boldsymbol{X}$ to be known.

A Monte Carlo estimate of (15.3.20) would usually take the form

$$\frac{1}{N} \sum_{i=1}^{N} H(F_X^{-1}(Z_i)),$$

where $Z_i$ for $i \in \{1, 2, \ldots, N\}$, are independent, uniformly distributed random variables. In practice, pseudo-random numbers would usually be used. However, with the use of, so-called, *low discrepancy* numbers, see Niederreiter (1992), the above estimator is referred to as a *quasi Monte Carlo estimator*. Low discrepancy point sets, such as *Sobol* or *Halton sequences*, see Niederreiter (1992), exhibit more regularity than pseudo-random point sets, especially for small sample sizes. This often leads to faster rates of convergence.

## 15.4 Monte Carlo Simulation for SDEs

Monographs on Monte Carlo methods for SDEs include, for instance, Kloeden & Platen (1999), Kloeden et al. (2003), Milstein (1995), Jäckel (2002) and Glasserman (2004). As previously explained, when other numerical methods for functionals of SDEs fail or are difficult to implement, Monte Carlo methods can still be applied. The main advantage of Monte Carlo simulation is that it can also be applied to high dimensional systems of SDEs.

### Weak Convergence

As was the case for scenario simulation, we may introduce a criterion that allows us to classify different discrete time approximations that can be used in raw Monte Carlo simulation. Recall that in Monte Carlo simulation one wishes to estimate the expected value of a certain payoff function. To obtain a suitable class of potential payoff functions, let us denote by $\tilde{\mathcal{C}}_P(\Re^d, \Re)$ the family of all polynomials $g : \Re^d \to \Re$. Consider the process $X = \{X_t,\, t \in [0, T]\}$, which is assumed to be the unique strong solution of the SDE

$$dX_t = a(t, X_t)\, dt + \sum_{j=1}^{m} b^j(t, X_t)\, dW_t^j \tag{15.4.1}$$

for $t \in [0, T]$, with $X_0 \in \Re^d$.

**Definition 15.4.1.** *Let $Y^\Delta$ be a discrete time approximation of $X$. Then $Y_T^\Delta$* converges with weak order *$\beta > 0$ to $X_T$ if for each $g \in \tilde{\mathcal{C}}_P(\Re^d, \Re)$ there exists a $\Delta_0 \in [0, 1]$ and a constant $C_g$, independent of $\Delta$, such that*

$$\mu(\Delta) = |E(g(X_T)) - E(g(Y_T^\Delta))| \le C_g\, \Delta^\beta \tag{15.4.2}$$

*for each $\Delta \in (0, \Delta_0)$.*

### Systematic and Statistical Error

Under the weak convergence criterion (15.4.1), functionals of the form

$$u = E(g(X_T)) \tag{15.4.3}$$

are approximated using *weak approximations* $Y^\Delta$ of the solution of the SDE (15.4.1). One can construct a raw Monte Carlo estimate, using the sample average

$$u_{N,\Delta} = \frac{1}{N} \sum_{k=1}^{N} g(Y_T^\Delta(\omega_k)), \tag{15.4.4}$$

where $Y_T^\Delta(\omega_1), Y_T^\Delta(\omega_2), \ldots, Y_T^\Delta(\omega_N)$ are $N$ independent simulated realizations of $Y_T^\Delta$, with $\omega_k \in \Omega$, for $k \in \{1, 2, \ldots, N\}$. The *weak error* $\hat{\mu}_{N,\Delta}$ has the form

$$\hat{\mu}_{N,\Delta} = u_{N,\Delta} - E(g(X_T)) \tag{15.4.5}$$

and can be decomposed into a *systematic error* $\mu_{\text{sys}}$ and a *statistical error* $\mu_{\text{stat}}$ with mean zero, so that

$$\hat{\mu}_{N,\Delta} = \mu_{\text{sys}} + \mu_{\text{stat}}. \tag{15.4.6}$$

Thus, the systematic error is given by

$$\begin{aligned} \mu_{\text{sys}} &= E(\hat{\mu}_{N,\Delta}) \\ &= E\left(\frac{1}{N} \sum_{k=1}^{N} g(Y_T^\Delta(\omega_k))\right) - E(g(X_T)) \\ &= E(g(Y_T^\Delta)) - E(g(X_T)). \end{aligned} \tag{15.4.7}$$

From (15.4.2) it then follows that

$$\mu(\Delta) = |\mu_{\text{sys}}|. \tag{15.4.8}$$

For a large number $N$ of independent simulated realizations of $Y^\Delta$, we can conclude from the Central Limit Theorem, see (2.1.27), that the statistical error $\mu_{\text{stat}}$ becomes asymptotically Gaussian with mean zero and variance

$$\text{Var}(\mu_{\text{stat}}) = \text{Var}(\hat{\mu}_{N,\Delta}) = \frac{1}{N} \text{Var}(g(Y_T^\Delta)). \tag{15.4.9}$$

This reveals the main disadvantage of raw Monte Carlo methods, namely that the variance of the statistical error only decreases like $\frac{1}{N}$, see (15.3.5).

### Simplified Weak Taylor Scheme

In Sect. 15.2 we studied the Euler scheme (15.2.23) for strong approximations. For weak convergence we only need to approximate the probability measure induced by the process $X$. Therefore, we can replace the Gaussian increments $\Delta W_n^j$ in (15.2.23) with simpler random variables $\Delta \hat{W}_n^j$, as long as their moments are similar to those of $\Delta W_n^j$. This leads to the *simplified weak Euler scheme*

$$Y_{n+1} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta\hat{W}_n^j, \tag{15.4.10}$$

where the $\Delta\hat{W}_n^j$ are independent $\mathcal{A}_{\tau_{n+1}}$-measurable random variables whose moments satisfy the following condition

$$\left|E\left(\Delta\hat{W}_n^j\right)\right| + \left|E\left(\left(\Delta\hat{W}_n^j\right)^3\right)\right| + \left|E\left(\left(\Delta\hat{W}_n^j\right)^2\right) - \Delta\right| \leq K\,\Delta^2 \tag{15.4.11}$$

for some constant $K$ and $j \in \{1, 2, \ldots, m\}$. The simplest choice $\Delta\hat{W}_n^j$, which satisfies (15.4.11), is a two-point distributed random variable, with

$$P\left(\Delta\hat{W}_n^j = \pm\sqrt{\Delta}\right) = \frac{1}{2}. \tag{15.4.12}$$

More accurate weak Taylor schemes can be derived by including additional multiple stochastic integrals from a Wagner-Platen expansion of the type shown in (15.2.6). Naturally, the objective is to achieve a desired weak order of convergence $\beta$, as defined in (15.4.2), with the minimal number of terms from the expansion. The choice of terms is typically different from the case where one tries to obtain the same strong order of convergence.

The *weak order* 2.0 *Taylor scheme*

$$\begin{aligned} Y_{n+1} = Y_n + a\,\Delta + b\,\Delta W_n + \frac{1}{2}\,b\,b'\left((\Delta W_n)^2 - \Delta\right) + a'\,b\,\Delta Z_n \\ + \frac{1}{2}\left(a\,a' + \frac{1}{2}\,a''b^2\right)\Delta^2 + \left(a\,b' + \frac{1}{2}\,b''b^2\right)(\Delta W_n\,\Delta - \Delta Z_n) \end{aligned} \tag{15.4.13}$$

is obtained by adding all double stochastic integrals from a Wagner-Platen expansion to the Euler scheme. Here the random variable $\Delta Z_n$ represents the double integral given by (15.2.31). As shown in Platen (1984), for $\beta \in \mathcal{N}$ to construct a weak order $\beta$ Taylor scheme, one needs to include all multiple Itô integrals with multiplicity less or equal than $\beta$ from a Wagner-Platen expansion.

Under the weak convergence criterion one has more freedom to construct appropriate schemes than under the strong convergence criterion. For example, for a weak order 2.0 scheme it is possible to avoid the random variable $\Delta Z_n$ appearing in (15.4.13). One can show that it is sufficient to include only a single random variable $\Delta\tilde{W}_n$ with similar moments to those of $\Delta W_n$. The random variable $\Delta Z_n$ is then replaced by $\frac{1}{2}\Delta\tilde{W}_n\Delta$. For simplicity, we again suppose the dependence of the random variables on the number $n$ of the time step. The *simplified weak order* 2.0 *Taylor scheme* can then be expressed as

$$\begin{aligned} Y_{n+1} = Y_n + a\,\Delta + b\,\Delta\tilde{W} + \frac{1}{2}\,b\,b'\left(\left(\Delta\tilde{W}\right)^2 - \Delta\right) \\ + \frac{1}{2}\left(a'\,b + a\,b' + \frac{1}{2}\,b''b^2\right)\Delta\tilde{W}\,\Delta + \frac{1}{2}\left(a\,a' + \frac{1}{2}\,a''b^2\right)\Delta^2, \end{aligned} \tag{15.4.14}$$

where $\Delta\tilde{W}$ has to satisfy the moment condition

$$\left|E\left(\Delta\tilde{W}\right)\right| + \left|E\left(\left(\Delta\tilde{W}\right)^3\right)\right| + \left|E\left(\left(\Delta\tilde{W}\right)^5\right)\right|$$
$$+ \left|E\left(\left(\Delta\tilde{W}\right)^2\right) - \Delta\right| + \left|E\left(\left(\Delta\tilde{W}\right)^4\right) - 3\Delta^2\right| \leq K\,\Delta^3 \qquad (15.4.15)$$

for some constant $K$.

Obviously, an $N(0, \Delta)$ random variable satisfies this condition. A three-point distributed random variable $\Delta\tilde{W}$ with

$$P\left(\Delta\tilde{W} = \pm\sqrt{3\Delta}\right) = \frac{1}{6} \quad \text{and} \quad P\left(\Delta\tilde{W} = 0\right) = \frac{2}{3} \qquad (15.4.16)$$

also satisfies (15.4.15).

## Derivative Free Weak Approximations

Higher order weak Taylor schemes require the evaluation of derivatives of various orders of the drift and diffusion coefficients. As with strong schemes, we can construct derivative free weak approximations, which avoid such derivatives. For the simple case, with time independent coefficients, the following *explicit weak order* 2.0 *scheme* is proposed in Platen (1984):

$$Y_{n+1} = Y_n + \frac{1}{2}\left(a\left(\bar{\Upsilon}\right) + a\right)\Delta + \frac{1}{4}\left(b\left(\bar{\Upsilon}^+\right) + b\left(\bar{\Upsilon}^-\right) + 2b\right)\Delta\tilde{W}$$
$$+ \frac{1}{4}\left(b\left(\bar{\Upsilon}^+\right) - b\left(\bar{\Upsilon}^-\right)\right)\left(\left(\Delta\tilde{W}\right)^2 - \Delta\right)\Delta^{-\frac{1}{2}} \qquad (15.4.17)$$

with supporting values

$$\bar{\Upsilon} = Y_n + a\,\Delta + b\,\Delta\tilde{W}$$

and

$$\bar{\Upsilon}^{\pm} = Y_n + a\,\Delta \pm b\sqrt{\Delta}.$$

Here $\Delta\tilde{W}$ is required to satisfy the moment condition (15.4.15). For example, $\Delta\tilde{W}$ can be the three-point distributed random variable in (15.4.16).

The above scheme can be generalized to yield the general *explicit weak order* 2.0 *scheme*

$$Y_{n+1} = Y_n + \frac{1}{2}\left(a\left(\bar{\Upsilon}\right) + a\right)\Delta + \frac{1}{4}\sum_{j=1}^{m}\Bigg[\left(b^j\left(\bar{R}_+^j\right) + b^j\left(\bar{R}_-^j\right) + 2b^j\right)\Delta\tilde{W}^j$$

$$+ \sum_{\substack{r=1\\r\neq j}}^{m}\left(b^j\left(\bar{U}_+^r\right) + b^j\left(\bar{U}_-^r\right) - 2b^j\right)\Delta\tilde{W}^j\,\Delta^{-\frac{1}{2}}\Bigg]$$

$$+\frac{1}{4}\sum_{j=1}^{m}\Bigg[\left(b^j\left(\bar{R}_+^j\right) - b^j\left(\bar{R}_-^j\right)\right)\left(\left(\Delta\tilde{W}^j\right)^2 - \Delta\right)$$

$$+ \sum_{\substack{r=1\\r\neq j}}^{m}\left(b^j\left(\bar{U}_+^r\right) - b^j\left(\bar{U}_-^r\right)\right)\left(\Delta\tilde{W}^j\Delta\tilde{W}^r + V_{r,j}\right)\Bigg]\Delta^{-\frac{1}{2}} \quad (15.4.18)$$

with supporting values

$$\bar{\Upsilon} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta\tilde{W}^j, \qquad \bar{R}_\pm^j = Y_n + a\,\Delta \pm b^j\sqrt{\Delta}$$

and

$$\bar{U}_\pm^j = Y_n \pm b^j\sqrt{\Delta}.$$

Here the random variables $\Delta\tilde{W}^j$ are defined as in (15.4.16), while $V_{j_1,j_2}$ are independent two-point distributed random variables satisfying

$$P\left(V_{j_1,j_2} = \pm\Delta\right) = \frac{1}{2} \quad (15.4.19)$$

for $j_2 \in \{1, \ldots, j_1 - 1\}$,

$$V_{j_1,j_1} = -\Delta \quad (15.4.20)$$

and

$$V_{j_1,j_2} = -V_{j_2,j_1} \quad (15.4.21)$$

for $j_2 \in \{j_1+1, \ldots, m\}$ and $j_1 \in \{1, 2, \ldots, m\}$.

### Extrapolation

Extrapolation provides an efficient yet simple way of obtaining a higher order weak approximation, while using only lower order weak schemes. Only equidistant time discretizations of the time interval $[0, T]$ with $\tau_{n_T} = T$ will be used in what follows. As usual, we shall denote a discrete time approximation with time step size $\Delta > 0$ by $Y^\Delta$. Its values at discretization times $\tau_n$ will be denoted by $Y_n^\Delta$ and the corresponding approximation with twice this step size will be written as $Y^{2\Delta}$, etc. Suppose that we have estimated the functional

$$E\left(g\left(Y_T^\Delta\right)\right),$$

using a weak order 1.0 approximation, such as the simplified Euler scheme (15.4.10), with step size $\Delta$. Let us repeat this simulation with double the step size $2\Delta$, to obtain the following estimate

$$E\left(g\left(Y_T^{2\Delta}\right)\right).$$

We can now combine the above two estimates in the *weak order* 2.0 *extrapolation*

$$V_{g,2}^{\Delta}(T) = 2E\left(g\left(Y_T^{\Delta}\right)\right) - E\left(g\left(Y_T^{2\Delta}\right)\right). \qquad (15.4.22)$$

This method, which was proposed in Talay & Tubaro (1990), is a stochastic generalization of the well-known *Richardson extrapolation* technique.

As is shown in Kloeden & Platen (1999), if a weak method exhibits an appropriate leading error term representation, then a corresponding extrapolation method can be constructed. For instance, one can take a weak order $\beta = 2.0$ approximation $Y^{\Delta}$ and extrapolate it to obtain a *weak order* 4.0 *extrapolation* as follows

$$V_{g,4}^{\Delta}(T) = \frac{1}{21}\left(32\,E\left(g\left(Y_T^{\Delta}\right)\right) - 12\,E\left(g\left(Y_T^{2\Delta}\right)\right) + E\left(g\left(Y_T^{4\Delta}\right)\right)\right). \quad (15.4.23)$$

Suitable weak order 2.0 approximations include the simplified weak order 2.0 Taylor scheme (15.4.14) and the explicit weak order 2.0 scheme (15.4.17).

The practical applicability of extrapolations depends strongly on the numerical stability of the underlying weak schemes over a range of time step sizes.

### Implicit Methods (*)

As is the case for scenario simulation, numerical stability is the most important criterion for Monte Carlo simulation. The following implicit and predictor corrector methods are efficient weak schemes offering reasonable numerical stability.

The simplest implicit weak scheme is the *drift implicit simplified Euler scheme*, which has the form

$$Y_{n+1} = Y_n + a\left(\tau_{n+1}, Y_{n+1}\right)\Delta + \sum_{j=1}^{m} b^j\, \Delta\hat{W}^j, \qquad (15.4.24)$$

where the random variables $\Delta\hat{W}^j$ for $j \in \{1, 2, \ldots, m\}$ and $n \in \mathcal{N}$ are independent and two-point distributed, as in (15.4.12).

One can also consider the *family of drift implicit simplified Euler schemes*

$$Y_{n+1} = Y_n + \big((1-\theta)\, a\left(\tau_{n+1}, Y_{n+1}\right) + \theta\, a\big)\Delta + \sum_{j=1}^{m} b^j\, \Delta\hat{W}^j \qquad (15.4.25)$$

with $\Delta\hat{W}^j$ as in (15.4.12) for $j \in \{1, 2, \ldots, m\}$. The parameter $\theta$ is again the degree of drift implicitness. For $\theta = 0$, the scheme reduces to the simplified Euler scheme (15.4.10), while for $\theta = 0.5$ it represents a stochastic generalization of the trapezoidal method. For all values of $\theta$, it can be shown to converge with weak order $\beta = 1.0$, and with a degree of implicitness $\theta \geq \frac{1}{2}$ it displays good numerical stability, see Kloeden & Platen (1999).

To achieve better numerical stability for a weak scheme, one can also make the diffusion coefficient implicit. An example is the *family of implicit weak Euler schemes*

$$\begin{aligned} Y_{n+1} = Y_n &+ \big(\theta\, \bar{a}_\eta\, (\tau_{n+1}, Y_{n+1}) + (1-\theta)\, \bar{a}_\eta\big)\, \Delta \\ &+ \sum_{j=1}^{m} \big(\eta\, b^j\, (\tau_{n+1}, Y_{n+1}) + (1-\eta)\, b^j\, (\tau_n, Y_n)\,\big)\, \Delta\hat{W}^j, \end{aligned} \tag{15.4.26}$$

where the random variables $\Delta\hat{W}^j$ are as in (15.4.12) and $\theta, \eta \in [0, 1]$ are degrees of implicitness parameters. Here $\bar{a}_\eta$ is a corrected drift coefficient, defined by

$$\bar{a}_\eta(\cdot,\cdot) = a(\cdot,\cdot) - \eta \sum_{j_1,j_2=1}^{m} b^{j_1}(\cdot,\cdot)\frac{\partial b^{j_2}(\cdot,\cdot)}{\partial x}. \tag{15.4.27}$$

In Kloeden & Platen (1999) it is shown that such weak Euler schemes exhibit weak order $\beta = 1.0$, under certain conditions.

## Predictor-Corrector Methods (*)

To achieve a higher weak order without compromising numerical stability, predictor-corrector methods have been developed in Platen (1995). These are similar to implicit methods, but do not require the solution of an algebraic equation at each time step.

One can form the following *family of weak order* 1.0 *predictor-corrector methods* with corrector

$$\begin{aligned} Y_{n+1} = Y_n &+ \big(\theta\, \bar{a}_\eta\, \big(\tau_{n+1}, \bar{Y}_{n+1}\big) + (1-\theta)\, \bar{a}_\eta\big)\, \Delta \\ &+ \sum_{j=1}^{m} \big(\eta\, b^j\, \big(\tau_{n+1}, \bar{Y}_{n+1}\big) + (1-\eta)\, b^j\big)\, \Delta\hat{W}^j \end{aligned} \tag{15.4.28}$$

for $\theta$, $\eta \in [0, 1]$, where

$$\bar{a}_\eta(\cdot,\cdot) = a(\cdot,\cdot) - \eta \sum_{j_1,j_2=1}^{m} \sum_{k=1}^{d} b^{k,j_1}(\cdot,\cdot)\frac{\partial b^{j_2}(\cdot,\cdot)}{\partial x^k}. \tag{15.4.29}$$

The corresponding predictor is given by

$$\bar{Y}_{n+1} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta\hat{W}^j. \tag{15.4.30}$$

Here the random variables $\Delta\hat{W}^j$ are as in (15.4.12). Note that for $\eta > 0$ the corrector (15.4.28) allows some implicitness in the diffusion terms. This scheme is in many cases numerically stable.

It is very helpful when a scheme supports different degrees of implicitness. By varying the degree of implicitness in such a scheme, the corresponding simulation results provide a feeling for the numerical stability that can be achieved.

For SDEs with time independent coefficients, Platen (1995) examines a derivative free *weak order 2.0 predictor-corrector method* with corrector

$$Y_{n+1} = Y_n + \frac{1}{2}\left(a\left(\bar{Y}_{n+1}\right) + a\right)\Delta + \phi_n, \tag{15.4.31}$$

where

$$\phi_n = \frac{1}{4}\sum_{j=1}^{m}\left[b^j\left(\bar{R}^j_+\right) + b^j\left(\bar{R}^j_-\right) + 2\,b^j + \sum_{\substack{r=1\\r\neq j}}^{m}\left(b^j\left(\bar{U}^r_+\right) + b^j\left(\bar{U}^r_-\right) - 2\,b^j\right)\Delta^{-\frac{1}{2}}\right]\Delta\tilde{W}^j$$
$$+\frac{1}{4}\sum_{j=1}^{m}\left[\left(b^j\left(\bar{R}^j_+\right) - b^j\left(\bar{R}^j_-\right)\right)\left(\left(\Delta\tilde{W}^j\right)^2 - \Delta\right)\right.$$
$$\left. + \sum_{\substack{r=1\\r\neq j}}^{m}\left(b^j\left(\bar{U}^r_+\right) - b^j\left(\bar{U}^r_-\right)\right)\left(\Delta\tilde{W}^j\,\Delta\tilde{W}^r + V_{r,j}\right)\right]\Delta^{-\frac{1}{2}}$$

and

$$\bar{R}^j_\pm = Y_n + a\,\Delta \pm b^j\sqrt{\Delta} \quad \text{and} \quad \bar{U}^j_\pm = Y_n \pm b^j\sqrt{\Delta}$$

are supporting values. The predictor

$$\bar{Y}_{n+1} = Y_n + \frac{1}{2}\left\{a\left(\bar{\Upsilon}\right) + a\right\}\Delta + \phi_n \tag{15.4.32}$$

employs a supporting value

$$\bar{\Upsilon} = Y_n + a\,\Delta + \sum_{j=1}^{m} b^j\,\Delta\tilde{W}^j.$$

The independent random variables $\Delta\tilde{W}^j$ and $V_{r,j}$ can be chosen as in (15.4.18).

In all the weak schemes above, one first computes the predicted approximate value $\bar{Y}_{n+1}$ and then the corrected value $Y_{n+1}$ at each time step. The difference

$$Z_{n+1} = \bar{Y}_{n+1} - Y_{n+1}$$

provides information about the, so-called, local error at each time step. This information can be used to control the size of the time step during the simulation. So, for example, if the mean of $Z_{n+1}$ is too large, then one can switch to a smaller time step size. For further references on predictor-corrector methods we refer again to Kloeden & Platen (1999).

### Weak Approximations with Jumps (*)

To complete this section, we now present some weak schemes suitable for the Monte Carlo simulation of solutions of SDEs with jumps. Results in this direction were obtained in Platen (1982a), Mikulevicius & Platen (1988) and Kubilius & Platen (2002) and Bruti-Liberati & Platen (2007a).

We consider an SDE with jumps, as described in (15.2.46)–(15.2.48). Since we specify a finite intensity $\varphi(\mathcal{E}) < \infty$, the Poisson process $p_\varphi = \{p_\varphi(\mathcal{E}, [0,t]), t \in [0,T]\}$ generates a sequence of jump times. As in Sect. 15.2, we assume that the discretization of the interval $[0,T]$ includes all jump times of $p_\varphi$ not greater than $T$. As before, we use a *jump adapted time discretization* $\{\tau_i\}_{i\in\{0,1,\ldots\}}$ with maximum step size $\Delta > 0$. This is a sequence of $\underline{\mathcal{A}}$-stopping times, including every jump time of $p_\varphi$ not greater than $T$, satisfying

$$0 = \tau_0 < \tau_1 < \ldots < \tau_{n_T} = T$$

and $n_T < \infty$, almost surely.

The following jump adapted simplified Euler scheme is a simple discrete time weak approximation of the jump diffusion process (15.2.46)–(15.2.48)

$$Y_{\tau_{n+1}-} = Y_{\tau_n} + a\,(\tau_{n+1} - \tau_n) + \sum_{j=1}^{m} b^j\, \Delta\hat{W}_n^j$$

$$Y_{\tau_{n+1}} = Y_{\tau_{n+1}-} + \int_{\mathcal{E}} c(v, Y_{\tau_{n+1}-})\, p_\varphi(dv, \{\tau_{n+1}\}) \qquad (15.4.33)$$

for $n \in \{0, 1, \ldots, n_T - 1\}$ with $Y_0 = x$. Here the random variables $\Delta\hat{W}_n^j$ can be chosen as in (15.4.12). It has been shown that under sufficient conditions, the above approximation converges with weak order $\beta = 1$, see Mikulevicius & Platen (1988).

Similarly, a family of implicit simplified weak Euler schemes, or weak order 1.0 predictor-corrector schemes, as described previously, can be employed for approximating the diffusion part, thereby enhancing numerical stability. In principle, any of the previously mentioned weak schemes can be employed to approximate the diffusion part. The weak order of the resulting jump adapted scheme is then determined by the weak order of convergence of that scheme.

## 15.5 Variance Reduction of Functionals of SDEs

As previously explained, a raw Monte Carlo estimate of the form (15.4.4) for the expectation appearing in (15.4.3) can be very expensive in terms of computational time when a certain accuracy is needed. In Sect. 15.3 we introduced several classical variance reduction techniques. In this section we describe efficient variance reduction for functionals of the type (15.4.3) when solutions of SDEs are involved.

### Control Variate Method for SDEs

The classical control variate technique has already been introduced in Sect. 15.3. It is a very flexible and powerful technique, which we now explore in the context of SDEs. Let $\boldsymbol{X} = \{\boldsymbol{X}_t = (X_t^1, \ldots, X_t^d)^\top, t \in [0,T]\}$ be the solution of a $d$-dimensional SDE with initial value $\boldsymbol{x} = (x^1, \ldots, x^d)^\top \in \Re^d$. We wish to compute the following expected value

$$u = E\left(H(\boldsymbol{X}_T)\right), \tag{15.5.1}$$

where $H : \Re^d \to \Re$ is some payoff function. Our aim will be to find a control variate, similar to that in Sect. 15.3, by exploiting the particular structure of the underlying SDE. This will enable us to obtain an accurate and fast estimate of $E(H(\boldsymbol{X}_T))$.

As described in Sect. 15.3, the control variate technique is based on finding a suitable random variable $Y$ with known mean $E(Y)$. We then estimate $E(H(\boldsymbol{X}_T))$ by computing the expected value of

$$Z = H(\boldsymbol{X}_T) - \alpha(Y - E(Y))$$

for some suitable choice of $\alpha \in \Re$, rather than attempting to compute the expected value of $H(\boldsymbol{X}_T)$ directly. The parameter $\alpha$ is chosen to minimize the variance of $Z$. Because

$$E(Z) = E(H(\boldsymbol{X}_T)),$$

$\bar{u}_N = \frac{1}{N}\sum_{i=1}^N Z(\omega_i)$ is an unbiased estimator for $E(H(\boldsymbol{X}_T))$. We assume that both $H(\boldsymbol{X}_T)$ and $Y$ can be evaluated for any realization $\omega \in \Omega$. With this type of formulation the random variable $Y$ is a control variate for the estimation of $E(H(\boldsymbol{X}_T))$.

As an example, we consider a stochastic volatility model with a vector process $\boldsymbol{X} = \{\boldsymbol{X}_t = (S_t, \sigma_t)^\top, t \in [0,T]\}$ that satisfies the two-dimensional SDE

$$\begin{aligned} dS_t &= \sigma_t\, S_t\, dW_t^1 \\ d\sigma_t &= \gamma\,(\kappa - \sigma_t)\, dt + \xi\, \sigma_t\, dW_t^2 \end{aligned} \tag{15.5.2}$$

for $t \in [0,T]$ with initial values $S_0 = s > 0$, $\sigma_0 = \sigma > 0$ and $\gamma, \kappa, \xi > 0$. Here the short rate is set to zero and the pricing is performed under an equivalent risk neutral probability measure $P$. We take $W^1$ and $W^2$ to be two independent standard Wiener processes defined on the probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. This type of stochastic volatility model has been examined in Chap. 12. We shall consider the payoff of a European call option

$$H(\boldsymbol{X}_T) = (S_T - K)^+,$$

where $K > 0$ is the strike and $T$ is the maturity.

Suppose now that $\tilde{\boldsymbol{X}} = \{\tilde{\boldsymbol{X}}_t = (\tilde{S}_t, \tilde{\sigma}_t),\ t \in [0,T]\}$ is an adjusted asset price process with deterministic volatility, which evolves according to the equations

$$\begin{aligned} d\tilde{S}_t &= \tilde{\sigma}_t \tilde{S}_t \, dW_t^1 \\ d\tilde{\sigma}_t &= \gamma\,(\kappa - \tilde{\sigma}_t)\, dt \end{aligned} \tag{15.5.3}$$

for $t \in [0,T]$ with the same initial values $\tilde{S}_0 = s$ and $\tilde{\sigma}_0 = \sigma$. The adjusted price $\tilde{u}$ for the European call payoff $Y = (\tilde{S}_T - K)^+$ is then

$$\tilde{u} = E(Y) = E\left((\tilde{S}_T - K)^+\right). \tag{15.5.4}$$

Since the volatility process is deterministic, this payoff can be evaluated using the Black-Scholes formula (8.3.2). Consequently, the random variable

$$\begin{aligned} Z &= (S_T - K)^+ - \alpha((\tilde{S}_T - K)^+ - E(\tilde{S}_T - K)^+) \\ &= (S_T - K)^+ - \alpha(Y - \tilde{u}) \end{aligned}$$

is easily computed. To price the option under consideration, we now estimate $E(Z)$, which is the control variate estimator, with $Y$ playing the role of the control variate. Note that this means simulating $\sigma_t$, $S_t$ and $\tilde{S}_t$ for $t \in [0,T]$. To do this, we could use a weak order two predictor-corrector scheme, for example. The above method can be very powerful since it takes advantage of the structure of the underlying SDE.

### Measure Transformation Method (*)

We now describe another important variance reduction method that is applicable to the problem of estimating expected values of functionals of SDEs. It is based on deep results from stochastic calculus.

Consider a $d$-dimensional diffusion process $X^{s,x} = \{X_t^{s,x},\ t \in [s,T]\}$, starting at $x \in \Re^d$ at time $s \in [0,T)$, which satisfies the SDE

$$dX_t^{s,x} = a(t, X_t^{s,x})\, dt + \sum_{j=1}^{m} b^j(t, X_t^{s,x})\, dW_t^j \tag{15.5.5}$$

for $t \in [s,T]$ with $X_s^{s,x} = x$. Here $W$ is a standard $m$-dimensional Wiener process on the filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$. Our aim is to approximate the functional

$$u(s,x) = E\left(H(X_T^{s,x}) \,\middle|\, \mathcal{A}_s\right), \tag{15.5.6}$$

where $H(\cdot)$ is a given real-valued payoff and $(s,x) \in [0,T] \times \Re^+$.

If we assume that $H$ and the drift and diffusion coefficients $a$ and $b$, respectively, are sufficiently smooth, then $u(\cdot,\cdot)$ satisfies a Kolmogorov backward equation, according to the Feynman-Kac formula (9.7.3)–(9.7.5). That is,

$$L^0 \, u(s,x) = 0 \tag{15.5.7}$$

for all $(s,x) \in (0,T) \times \Re^d$ with boundary condition

$$u(T,y) = H(y) \tag{15.5.8}$$

for all $y \in \Re^d$. In this case $L^0$ is the differential operator

$$L^0 = \frac{\partial}{\partial s} + \sum_{k=1}^{d} a^k \frac{\partial}{\partial x^k} + \frac{1}{2} \sum_{k,\ell=1}^{d} \sum_{j=1}^{m} b^{k,j} \, b^{\ell,j} \, \frac{\partial^2}{\partial x^k \, \partial x^\ell}.$$

Milstein (1995) applies the Girsanov transformation (9.5.11) to produce an equivalent probability measure $\tilde{P}$ under which the process $\tilde{W}$, defined by

$$\tilde{W}_t^j = W_t^j - \int_0^t d^j\left(z, \tilde{X}_z^{0,x}\right) dz \tag{15.5.9}$$

for $t \in [0,T]$ and $j \in \{1,2,\ldots,m\}$, is a Wiener process. The transformed probability measure $\tilde{P}$ is determined by the Radon-Nikodym derivative process

$$\Lambda_t = \frac{d\tilde{P}}{dP}\bigg|_{\mathcal{A}_t} = \frac{\Theta_t}{\Theta_0}, \tag{15.5.10}$$

where $\Lambda = \{\Lambda_t, \, t \in [0,T]\}$ is assumed to be an $(\underline{\mathcal{A}}, P)$-martingale. The process $\tilde{X}^{0,x}$ in (15.5.9) is given by the SDE

$$\begin{aligned} d\tilde{X}_t^{0,x} &= a\left(t, \tilde{X}_t^{0,x}\right) dt + \sum_{j=1}^{m} b^j\left(t, \tilde{X}_t^{0,x}\right) d\tilde{W}_t^j \\ &= \left( a\left(t, \tilde{X}_t^{0,x}\right) - \sum_{j=1}^{m} b^j\left(t, \tilde{X}_t^{0,x}\right) d^j\left(t, \tilde{X}_t^{0,x}\right)\right) dt + \sum_{j=1}^{m} b^j\left(t, \tilde{X}_t^{0,x}\right) dW_t^j. \end{aligned} \tag{15.5.11}$$

The *adjustment process* $\Theta = \{\Theta_t, \, t \in [0,T]\}$ satisfies the equation

$$\Theta_t = \Theta_0 + \sum_{j=1}^{m} \int_0^t \Theta_z \, d^j\left(z, \tilde{X}_z^{0,x}\right) dW_z^j \tag{15.5.12}$$

with $\Theta_0 > 0$. In all the expressions above, the functions $d^j$ for $j \in \{1, 2, \ldots, m\}$ can be chosen quite freely. Note that $\tilde{X}_t^{0,x}$ is $d$-dimensional, while $\Theta_t$ is only one-dimensional.

Obviously, the process $\tilde{X}^{0,x}$ in (15.5.11) is a diffusion process with respect to $\tilde{P}$ with the same drift and diffusion coefficients as the diffusion process $X^{0,x}$ in (15.5.5). It then follows from (15.5.10) that

$$\begin{aligned} E\left(H\left(X_T^{0,x}\right)\right) &= \int_\Omega H\left(X_T^{0,x}\right) dP \;=\; \int_\Omega H\left(\tilde{X}_T^{0,x}\right) d\tilde{P} \\ &= \int_\Omega H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}\, dP \;=\; E\left(H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}\right). \end{aligned} \tag{15.5.13}$$

Hence, we can estimate (15.5.6) by estimating the expected value of the following expression

$$H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}. \tag{15.5.14}$$

So far, our analysis does not depend on the particular choice of the functions $d^j$, $j \in \{1, 2, \ldots, m\}$. Therefore, the adjustment process $\Theta$ can be chosen quite freely to reduce the variance of the random variable (15.5.14).

Now, let us study the following idealized situation from a theoretical viewpoint. Assume that we know that $u(\cdot, \cdot) > 0$ and that the solutions of (15.5.11) and (15.5.12) exist. Furthermore, let us choose the parameter functions $d^j$ as

$$d^j(t, x) = -\frac{1}{u(t, x)} \sum_{k=1}^{d} b^{k,j}(t, x)\, \frac{\partial u(t, x)}{\partial x^k} \tag{15.5.15}$$

for all $(t, x) \in [0, T] \times \Re^d$ and $j \in \{1, 2, \ldots, m\}$. Then it follows from an application of the Itô formula and (15.5.11), (15.5.12), (15.5.15) and (15.5.7) that

$$u\left(t, \tilde{X}_t^{0,x}\right) \Theta_t = u(0, x)\, \Theta_0. \tag{15.5.16}$$

Combining (15.5.8) and (15.5.16), we conclude that

$$u(0, x) = H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}. \tag{15.5.17}$$

Consequently, with the choice of the parameter functions (15.5.15), the variable

$$H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0} \tag{15.5.18}$$

is *not* random and so its variance is zero, in this idealized case.

Unfortunately, the construction of the parameter functions in (15.5.15) only works if one already knows the solution $u(\cdot, \cdot)$ of the Kolmogorov backward equation, which is exactly what we are trying to determine by means of Monte Carlo simulation. However, the above discussion does show that a substantial variance reduction may be achieved by an application of a measure transformation, if one exploits information about the pricing function.

In practice, when implementing a measure transformation method, one needs to find or guess a function $\bar{u}$ which is sufficiently close to the solution $u$ of the Kolmogorov backward equation (15.5.7) and (15.5.8). One can then use $\bar{u}$ instead of $u$ to define the parameter functions in (15.5.15):

$$d^j(t,x) = -\frac{1}{\bar{u}(t,x)} \sum_{k=1}^{d} b^{k,j}(t,x) \frac{\partial \bar{u}(t,x)}{\partial x^k} \tag{15.5.19}$$

for all $(t,x) \in [0,T] \times \Re^d$ and $j \in \{1,2,\ldots,m\}$. Then the quantity

$$H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}$$

is still random and

$$E\left(H\left(\tilde{X}_T^{0,x}\right) \frac{\Theta_T}{\Theta_0}\right) = E\left(H\left(X_T^{0,x}\right)\right).$$

In general, the variance is substantially reduced if $\bar{u}$ is chosen close to $u$ in some sense.

There exist other powerful variance reduction techniques, such as, the *integral representation variance reduction*, developed in Heath & Platen (2002c). In principle, these methods can usually be combined. It is an art to tailor an efficient variance reduction method to the problem at hand.

## 15.6 Tree Methods

Simulation involves the generation of many random numbers and can therefore be very time consuming. Deterministic numerical methods, when applicable, are often desirable. In what follows, we consider deterministic, discrete time approximations of the paths of an asset price process. We study binomial tree models, which are widely used for risk neutral option pricing, see Cox, Ross & Rubinstein (1979) and van der Hoek & Elliott (2006). The main aim of this section is to illustrate how the benchmark approach enables the numerical approximation of option prices using trees, without the requirement of a risk neutral probability measure.

### Single-Period Binomial Model

Consider a regular discretization $\{t_0, t_1, \ldots, t_{n_T}\}$ of $[0,T]$, with $t_0 = 0$ and $t_{n_T} = T$. The step size $\Delta$ is then here $\frac{T}{n_T}$. Suppose that $S^{\delta_*} = \{S_t^{\delta_*},\, t \in [0,T]\}$ approximates the GOP process, where $S^{\delta_*}$ may be interpreted as a diversified index. We also introduce a primary security account process $S = \{S_t,\, t \in$

$[0, T]\}$, where $S_t$ denotes the cum dividend price of a stock at time $t$. $S^{\delta_*}$ and $S$ are taken to be piecewise constant between discretization points and are assumed to be right continuous. The benchmarked primary security account price $\hat{S}_t$ is then

$$\hat{S}_t = \frac{S_t}{S_t^{\delta_*}} \tag{15.6.1}$$

for $t \in [0, T]$. The process $\hat{S} = \{\hat{S}_t,\, t \in [0, T]\}$ and the GOP are modeled on a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$, which satisfies the usual conditions, see Sect. 5.1.

The value of a benchmarked portfolio $\hat{S}^\delta = \{\hat{S}_t^\delta,\, t \in [0, T]\}$ with predictable strategy $\boldsymbol{\delta} = \{\boldsymbol{\delta}_t = (\delta_t^0, \delta_t^1)^\top,\, t \in [0, T]\}$, which at time $t$ holds $\delta_t^0$ units of the GOP and $\delta_t^1$ units of the stock, is given by

$$\hat{S}_t^\delta = \delta_t^0 + \delta_t^1\, \hat{S}_t \tag{15.6.2}$$

for $t \in \{t_0, t_1, \ldots, t_{n_T}\}$.

In the following we consider a very simple single period *binomial model*, see Sect. 3.3. This is one of the simplest ways of modeling the randomness of a security. We model the uncertain value of the benchmarked security $\hat{S}_\Delta$ at time $t_1 = \Delta > 0$, by only allowing the two possible values $(1+u)\hat{S}_0$ and $(1+d)\hat{S}_0$, such that

$$P\left(\hat{S}_\Delta = (1+u)\,\hat{S}_0\right) = p \tag{15.6.3}$$

and

$$P\left(\hat{S}_\Delta = (1+d)\,\hat{S}_0\right) = 1 - p$$

for $p \in (0, 1)$ and $-d, u \in (0, \infty)$. Note that $P$ denotes the real world probability measure. An upward move $u = \frac{\hat{S}_\Delta - \hat{S}_0}{\hat{S}_0}$ can be interpreted as a positive return and a downward move $d = \frac{\hat{S}_\Delta - \hat{S}_0}{\hat{S}_0}$ as a negative return. Note that since we aim to apply the real world pricing concept, we consider here benchmarked securities. In Cox et al. (1979) one can find similar derivations, as presented below. However, these are performed under some equivalent risk neutral probability measure, whereas we shall work entirely under the real world probability measure.

## Pricing and Hedging

Now, consider a European call option on the benchmarked stock with benchmarked payoff

$$\hat{H} = \frac{H}{S_T^{\delta_*}} = \frac{(S_T - K)^+}{S_T^{\delta_*}} = (\hat{S}_T - \hat{K})^+ \tag{15.6.4}$$

at maturity $T$. The deterministic benchmarked strike price $\hat{K} = \frac{K}{S_\Delta^{\delta_*}}$ is chosen from the interval $((1+d)\,\hat{S}_0, (1+u)\,\hat{S}_0)$ and the maturity is set to $T = \Delta$. At maturity $\varDelta$ this instrument yields a benchmarked payoff $((1+u)\,\hat{S}_0 - \hat{K})^+$ in the event of an upward move and is otherwise zero. Assuming that the GOP has initial value $S_0^{\delta_*} = 1$, the real world pricing formula then yields the following price for the call option

$$S_0^{\delta_H} = S_0^{\delta_*}\, E\left(\hat{H} \,\middle|\, \mathcal{A}_0\right) = \left((1+u)\,\hat{S}_0 - \hat{K}\right) p. \tag{15.6.5}$$

If we additionally assume that $S$ is a fair price process, then

$$\begin{aligned} S_0 = \hat{S}_0 = E\left(\hat{S}_\varDelta \,\middle|\, \mathcal{A}_0\right) &= p\,(1+u)\,\hat{S}_0 + (1-p)\,(1+d)\,\hat{S}_0 \\ &= \left(p\,(1+u) + (1-p)\,(1+d)\right)\hat{S}_0. \end{aligned} \tag{15.6.6}$$

Consequently, we obtain the relation

$$1 = p\,(1+u) + (1-p)(1+d),$$

from which it follows that

$$p = \frac{-d}{u-d}. \tag{15.6.7}$$

Putting (15.6.5) and (15.6.7) together, the price of the *European call* at time $t = 0$ can be expressed as

$$S_0^{\delta_H} = \hat{S}_0^{\delta_H} = \left((1+u)\,\hat{S}_0 - \hat{K}\right)\frac{-d}{u-d}. \tag{15.6.8}$$

Note that the option price increases with $u$ and decreases as $d$ increases.

We now calculate the *hedge ratio* for the above option. Since we assume a self-financing strategy, we must have

$$\hat{S}_0^{\delta_H} = \delta_0^0 + \delta_0^1\,\hat{S}_0 \tag{15.6.9}$$

and

$$\hat{S}_\varDelta^{\delta_H} = \delta_0^0 + \delta_0^1\,\hat{S}_\varDelta. \tag{15.6.10}$$

This provides a system of equations, where $\hat{S}_0^{\delta_H}$ is given by (15.6.9) and

$$\hat{S}_\varDelta^{\delta_H} = \hat{H} = \left(\hat{S}_\varDelta - \hat{K}\right)^+. \tag{15.6.11}$$

It then follows from (15.6.9)–(15.6.11) that

$$\delta_0^1 = \frac{\hat{S}_\varDelta^{\delta_H} - \hat{S}_0^{\delta_H}}{\hat{S}_\varDelta - \hat{S}_0}.$$

In the case of an upward move, after which $\hat{S}_\varDelta = (1+u)\hat{S}_0$, this gives

$$\delta_0^1 = \frac{(1+u)\,\hat{S}_0 - \hat{K} - \left((1+u)\,\hat{S}_0 - \hat{K}\right)\frac{-d}{u-d}}{u\,\hat{S}_0} = \frac{(1+u)\,\hat{S}_0 - \hat{K}}{u-d}. \tag{15.6.12}$$

While in the case of a downward move, after which $\hat{S}_\Delta = (1+d)\,\hat{S}_0$, we have

$$\delta_0^1 = \frac{-\left((1+u)\,\hat{S}_0 - \hat{K}\right)\,\frac{-d}{u-d}}{d\,\hat{S}_0} = \frac{(1+u)\,\hat{S}_0 - \hat{K}}{u-d}. \tag{15.6.13}$$

Note that the hedge ratio is the same in (15.6.12) and (15.6.13). This shows that the price (15.6.5) can be hedged. We conclude that the above real world pricing allows perfect replication of the payoff.

We emphasize that $p$ is a real world probability and may be estimated directly from historical data, see Sect. 1.1. An interesting point to emphasize from this simple example is that an option can be hedged with a stock and the index. This is different to the standard approach, as described in Chap. 8, where hedge portfolios are constructed from the underlying security and the savings account. We emphasize that the benchmarked strike price $\hat{K}$ is deterministic, which means that it describes how many units of the index are exchanged for one unit of the stock at maturity. Therefore, the above option may be interpreted as an option to exchange the index and the stock.

### Binomial Volatility

The variability of the benchmarked stock price $\hat{S}_\Delta$ can be measured by the variance of the ratio

$$\frac{\hat{S}_\Delta}{\hat{S}_0} = 1 + \eta, \tag{15.6.14}$$

which involves the two-point distributed random variable $\eta \in \{d, u\}$, where

$$P(\eta = u) = p \tag{15.6.15}$$

and

$$P(\eta = d) = 1 - p. \tag{15.6.16}$$

It then follows by (1.3.15) that the variance of $\eta$ and thus of $\frac{\hat{S}_\Delta}{\hat{S}_0}$ is

$$E\left(\left(\frac{\hat{S}_\Delta}{\hat{S}_0} - 1\right)^2 \,\Big|\, \mathcal{A}_0\right) = (u-d)^2\,p\,(1-p).$$

Therefore, from (15.6.7) we obtain

$$E\left(\left(\frac{\hat{S}_\Delta}{\hat{S}_0} - 1\right)^2 \,\Big|\, \mathcal{A}_0\right) = -u\,d. \tag{15.6.17}$$

Obviously, this variance increases when $u$ or $-d$ increase. We call

$$\sigma_\Delta = \sqrt{\frac{1}{\Delta} E\left(\left(\frac{\hat{S}_\Delta}{\hat{S}_0} - 1\right)^2 \,\bigg|\, \mathcal{A}_0\right)} = \sqrt{\frac{-u\,d}{\Delta}} \tag{15.6.18}$$

the *binomial volatility* of the benchmarked security price. We know that the Black-Scholes option price (8.3.2) increases with increasing squared volatility. This appears natural from a risk management perspective. However, according to (15.6.8) and (15.6.18), it is not necessarily the case for a binomial model. In the formula

$$S_0^{\delta_H} = \left((1+u)\,\hat{S}_0 - \hat{K}\right) \frac{(\sigma_\Delta)^2\,\Delta}{u\,(u-d)}$$

we have the freedom to choose $u$ and $d$ eventually so that the option price does not increase with $\sigma_\Delta$. This economically unsatisfactory feature is a result of the simplicity of the binomial model.

### Multi-Period Binomial Model

To make the binomial model more realistic, we now consider a multi-period binomial tree, see Sect. 3.3. Here the benchmarked price of the stock at time $t_i = i\Delta,\ i \in \{1, 2, \ldots, n_T\}$ is given by

$$\hat{S}_{i\Delta} = \hat{S}_{(i-1)\Delta}\,(1+u) \tag{15.6.19}$$

with probability

$$p = \frac{-d}{u-d} \tag{15.6.20}$$

in the case of an upward move. In the case of a downward move, the benchmarked stock price becomes

$$\hat{S}_{i\Delta} = \hat{S}_{(i-1)\Delta}\,(1+d) \tag{15.6.21}$$

with probability $1-p$.

Consider a European call option on the benchmarked stock price with maturity $T = t_2 = 2\Delta$ and a deterministic benchmarked strike price $\hat{K}$. It is easily seen that $\hat{S}_{2\Delta}$ can take three possible values: $\hat{S}_{uu} = (1+u)^2\,\hat{S}_0$, $\hat{S}_{ud} = \hat{S}_{du} = (1+u)(1+d)\hat{S}_0$ and $\hat{S}_{dd} = (1+d)^2\hat{S}_0$. These events can be schematically represented as a binomial tree, as in Fig. 3.1.5. The benchmarked value $\hat{S}^\delta_{2\Delta}$ of a portfolio at time $T = t_2 = 2\Delta$ is given by

$$\hat{S}^\delta_{2\Delta} = \delta^0_{2\Delta} + \delta^1_{2\Delta}\,\hat{S}_{2\Delta}. \tag{15.6.22}$$

Using our previous result (15.6.5), we can compute the option price at time $t_1 = \Delta$ for the case when $\hat{S}_\Delta = (1+u)\,\hat{S}_0$, giving

$$\hat{S}^{\delta_H}_\Delta = \left((1+u)^2\,\hat{S}_0 - \hat{K}\right)^+ \frac{-d}{u-d} + \left((1+u)\,(1+d)\,\hat{S}_0 - \hat{K}\right)^+ \frac{u}{u-d}. \tag{15.6.23}$$

Similarly, if $\hat{S}_\Delta = (1+d)\hat{S}_0$, then the option price is

$$\hat{S}_\Delta^{\delta_H} = \left((1+u)\,(1+d)\,\hat{S}_0 - \hat{K}\right)^+ \frac{-d}{u-d} + \left((1+d)^2\,\hat{S}_0 - \hat{K}\right)^+ \frac{-u}{u-d}. \tag{15.6.24}$$

If we interpret (15.6.23) and (15.6.24) as benchmarked payoffs at time $t_1 = \Delta$, then repeating the procedure we obtain, the benchmarked price at time $t_0 = 0$ as

$$\begin{aligned}\hat{S}_0^{\delta_H} &= p\left\{\left((1+u)^2\,\hat{S}_0 - \hat{K}\right)^+ \frac{-d}{u-d} + \left((1+u)\,(1+d)\,\hat{S}_0 - \hat{K}\right)^+ \frac{u}{u-d}\right\}\\ &+(1-p)\left\{\left((1+u)\,(1+d)\,\hat{S}_0 - \hat{K}\right)^+ \frac{-d}{u-d} + \left((1+d)^2\,\hat{S}_0 - \hat{K}\right)^+ \frac{-u}{u-d}\right\}.\end{aligned} \tag{15.6.25}$$

A repeat of the earlier analysis gives us a perfect hedging strategy, as before.

In general, by using backward recursion and the well-known binomial probabilities, see (2.1.32) and (3.3.12), one can derive the following Cox-Ross-Rubinstein (CRR) type *binomial option pricing formula*, see Cox et al. (1979), for a benchmarked European call option

$$\begin{aligned}\hat{S}_0^{\delta_H} &= E\left(\left(\hat{S}_T - \hat{K}\right)^+ \,\Big|\, \mathcal{A}_0\right)\\ &= \sum_{k=0}^{n_T} \frac{n_T\,!}{k\,!\,(n_T-k)\,!}\, p^k\,(1-p)^{n_T-k}\left((1+u)^k(1+d)^{n_T-k}\,\hat{S}_0 - \hat{K}\right)^+\\ &= \hat{S}_0 \sum_{k=t_k}^{n_T} \frac{n_T\,!}{k\,!\,(n_T-k)\,!}\, p^k\,(1-p)^{n_T-k}\,(1+u)^k(1+d)^{n_T-k}\\ &\quad - \hat{K} \sum_{k=t_k}^{n_T} \frac{n_T\,!}{k\,!\,(n_T-k)\,!}\, p^k\,(1-p)^{n_T-k},\end{aligned} \tag{15.6.26}$$

where $t_k$ denotes the first integer $k$ for which $(1+u)^k(1+d)^{n_T-k}\hat{S}_0 > \hat{K}$. The maturity of this instrument is $T = n_T\Delta$.

Note that this is not exactly the CRR binomial option pricing formula in the literature, see Pliska (1997), Elliott & Kopp (2005) or van der Hoek & Elliott (2006). The formula from the literature considers discounted securities under a risk neutral probability measure in an exchange of stock for cash. Here we have an option pricing formula that refers to benchmarked securities under the real world probability measure, where the stock is exchanged for a market index. The main difference is that the strike price $\hat{K}$ describes how many units of the index are exchanged for one unit of the stock.

Let

$$t_N(t_k, n_T, p) = \sum_{k=t_k}^{n_T} \frac{n_T\,!}{k\,!\,(n_T - k)\,!}\, p^k\,(1-p)^{n_T - k} \tag{15.6.27}$$

denote the, so-called, *complementary binomial distribution*. Then we can express the binomial option pricing formula (15.6.26) as

$$S_0^{\delta_H} = S_0\, t_N(t_k, n_T, 1{-}p) - \hat{K}\, t_N(t_k, n_T, p). \tag{15.6.28}$$

As already mentioned, this is similar to the Cox-Ross-Rubinstein binomial option pricing formula.

We can also deduce the hedging strategy $\boldsymbol{\delta}_H = \{\boldsymbol{\delta}_{k\Delta} = (\delta^0_{k\Delta}, \delta^1_{k\Delta})^\top,\ k \in \{0, 1, \ldots, n_T{-}1\}\}$ that replicates $\hat{H}$. The benchmarked hedge portfolio value is given by

$$\hat{S}^{\delta_H}_{k\Delta} = \delta^0_{(k-1)\Delta} + \delta^1_{(k-1)\Delta}\, \hat{S}_{k\Delta} = \delta^0_{k\Delta} + \delta^1_{k\Delta}\, \hat{S}_{k\Delta} \tag{15.6.29}$$

for $k \in \{1, 2, \ldots, n_T\}$. As in (15.6.13) the hedge ratio is

$$\delta^1_{k\Delta} = \sum_{\ell = t_{k\,k}}^{n_T - k} \frac{(n_T - k)\,!}{\ell\,!\,(n_T - k - \ell)\,!}\, p^{(n_T - k - \ell)}\,(1-p)^\ell, \tag{15.6.30}$$

where $t_{k\,k}$ denotes the smallest integer $\ell$ for which $(1{+}u)^\ell\,(1{+}d)^{n_T - k - \ell}\hat{S}_{k\Delta} > \hat{K}$.

From (15.6.29) it finally follows that

$$\delta^0_{k\Delta} = \hat{S}^{\delta_H}_{k\Delta} - \delta^1_{k\Delta}\, \hat{S}_{k\Delta}. \tag{15.6.31}$$

The binomial model is convenient but also rather crude. It allows for easy calculation of option prices and is useful as a first approximation. We emphasize that due to its simplicity it has severe short comings.

### Approximating the Black-Scholes Price

It is interesting to note that the binomial option pricing formula (15.6.28) approaches the Black-Scholes formula (8.3.2) in the limit as $\Delta \to 0$. This is due to the Central Limit Theorem, see Sect. 2.1, which ensures that the binomial distribution approaches a Gaussian one as its number of states increases, see Fig. 2.1.5. We will now show how the binomial model approximates the BS model asymptotically as $\Delta \to 0$. A proof for the weak convergence of tree methods when interpreted as Markov chains is contained in Platen (1992). Intuitively, the random walk represented by the binomial tree converges weakly to a limiting process as the time step size $\Delta$ decreases to zero. This limiting process turns out to be a geometric Brownian motion.

To describe this asymptotic behavior formally, let us assume a particular form for $u$ and $d$, by setting

$$\ln(1+u) = \sigma\sqrt{\Delta} \tag{15.6.32}$$

and

$$\ln(1+d) = -\sigma\sqrt{\Delta}. \tag{15.6.33}$$

Here the *volatility parameter* $\sigma > 0$ has a given fixed value. Thus, we obtain the Cox-Ross-Rubinstein (CRR) spatial steps

$$u = \exp\left\{\sigma\sqrt{\Delta}\right\} - 1 \approx \sigma\sqrt{\Delta} \tag{15.6.34}$$

and

$$d = \exp\left\{-\sigma\sqrt{\Delta}\right\} - 1 \approx -\sigma\sqrt{\Delta}. \tag{15.6.35}$$

Recall the binomial volatility $\sigma_\Delta$ introduced in (15.6.18). If $\Delta \ll 1$, we can see from (15.6.34)–(15.6.35) that

$$\sigma_\Delta = \frac{1}{\sqrt{\Delta}}\sqrt{-u\,d} \approx \sigma. \tag{15.6.36}$$

We now introduce a continuous time stochastic process $Y^\Delta = \{Y_t^\Delta,\ t \in [0,T]\}$, defined by

$$Y_t^\Delta = \hat{S}_{i\Delta} \tag{15.6.37}$$

for $t \in [i\Delta, (i+1)\Delta)$, $i \in \{0,1,\ldots,n_T\}$. Then by (15.6.19)–(15.6.21) it follows that

$$Y_{i\Delta+\Delta}^\Delta = Y_{i\Delta}^\Delta + Y_{i\Delta}^\Delta\, q_i \tag{15.6.38}$$

for some independent random variable $q_i$ satisfying

$$P(q_i = u) = p = \frac{-d}{u-d} \tag{15.6.39}$$

and

$$P(q_i = d) = 1 - p = \frac{u}{u-d}, \tag{15.6.40}$$

see (15.6.14)–(15.6.16). By looking at (15.6.32)–(15.6.33) we realize that (15.6.38) can be rewritten approximately as

$$Y_{i\Delta+\Delta}^\Delta \approx Y_{i\Delta}^\Delta + Y_{i\Delta}^\Delta\,\sigma\sqrt{\Delta}\,\xi_i. \tag{15.6.41}$$

Here $\xi_i$ is an independent random variable $X$ with

$$P(\xi_i = 1) = p = \frac{\exp\{-\sigma\sqrt{\Delta}\} - 1}{\exp\{\sigma\sqrt{\Delta}\} - \exp\{-\sigma\sqrt{\Delta}\}} \tag{15.6.42}$$

and

$$P(\xi_i = -1) = 1 - p.$$

Note that

$$E\left(Y^{\Delta}_{i\Delta+\Delta} - Y^{\Delta}_{i\Delta} \,\big|\, \mathcal{A}_{i\Delta}\right) = 0 \tag{15.6.43}$$

and

$$E\left(\left(Y^{\Delta}_{i\Delta+\Delta} - Y^{\Delta}_{i\Delta}\right)^2 \,\Big|\, \mathcal{A}_{i\Delta}\right) = \left(Y^{\Delta}_{i\Delta}\right)^2 \sigma^2_{\Delta}\, \Delta \approx \left(Y^{\Delta}_{i\Delta}\right)^2 \sigma^2\, \Delta. \tag{15.6.44}$$

More precisely, it can be shown that $Y^{\Delta}$ converges weakly to the following diffusion process $X = \{X_t,\, t \in [0, T]\}$ as $\Delta \to 0$:

$$dX_t = X_t\, \sigma\, dW_t \tag{15.6.45}$$

for $t \in [0, T]$ with $X_0 = \hat{S}_0$, see Platen (1992). Thus, for a suitable payoff function $H$,

$$\lim_{\Delta \to 0} E\left(H\left(Y^{\Delta}_T\right) \,\Big|\, \mathcal{A}_0\right) = E\left(H(X_T) \,\big|\, \mathcal{A}_0\right). \tag{15.6.46}$$

Consequently, the binomial option pricing formula (15.6.28) approximates the following Black-Scholes formula as $\Delta \to 0$

$$S^{\delta_H}_0 = S_0\, N(\hat{d}_1) - \hat{K}\, N(\hat{d}_2). \tag{15.6.47}$$

Here

$$\hat{d}_1 = \frac{\ln(\frac{S_0}{\hat{K}}) + \frac{1}{2}\sigma^2\, T}{\sigma\sqrt{T}}, \quad \hat{d}_2 = \hat{d}_1 - \sigma\sqrt{T} \tag{15.6.48}$$

and $N(\cdot)$ denotes the standard Gaussian distribution function (1.2.7). The difference between the Black-Scholes formulas (8.3.2) and (15.6.46) is that in the latter, the strike price $\hat{K}$ is expressed in units of the index.

*American options* can be priced efficiently with binomial trees. This is a significant advantage of tree methods. First, one builds a tree as described above. Then one uses the same backward algorithm as before to pull back the maturity values of the option to one time step before maturity. These values are compared with the payoffs resulting from immediate exercise, and the larger values are chosen as the option values at the corresponding nodes. The whole process is then repeated, until the initial node is reached. Similarly one also obtains the prices of barrier options and a range of other exotic derivatives.

Finally, we emphasize that tree methods are explicit. They are very similar to simplified explicit weak schemes. For this reason they inherit many of the same numerical stability problems. Making the calculation of an option price implicit using a tree seems impossible. However, the following important class of numerical methods allows some implicitness.

## 15.7 Finite Difference Methods

In practice, derivatives are often priced by solving numerically the underlying partial differential equation (PDE) for the pricing function. The most common methods are *finite difference methods*. There is an extensive literature on them that has evolved over several decades. We refer to standard textbooks, such as Richtmeyer & Morton (1967) or Smith (1985). Monographs that deal directly with applications of finite differences to problems in finance include Wilmott, Dewynne & Howison (1993), Shaw (1998) and Tavella & Randall (2000).

### Derivative Pricing

Let us consider an appropriate real valued factor process $X = \{X_t,\, t \in [0,T]\}$ that drives the market dynamics and satisfies the SDE

$$dX_t = a(t, X_t)\, dt + b(t, X_t)\, d\tilde{W}_t \tag{15.7.1}$$

for $t \in [0,T]$ with $X_0 \in \Re$. Consider now a European option with payoff

$$u(T, X_T) = h(X_T) \tag{15.7.2}$$

at maturity $T$, expressed in units of an appropriate numeraire. Assume that the corresponding pricing formula

$$u(t, X_t) = \tilde{E}\left(h(X_T) \,\middle|\, \mathcal{A}_t\right), \tag{15.7.3}$$

yields a sufficiently smooth pricing function $u(\cdot,\cdot)$. By application of the Itô formula, this pricing function satisfies the PDE

$$\frac{\partial u(t,x)}{\partial t} + a(t,x)\,\frac{\partial u(t,x)}{\partial x} + \frac{1}{2}\,(b(t,x))^2\,\frac{\partial^2 u(t,x)}{\partial x^2} = 0 \tag{15.7.4}$$

for $t \in (0,T)$ and $x \in [0,\infty)$ with terminal condition

$$u(T,x) = h(x) \tag{15.7.5}$$

for $x \in [0,\infty)$. Recall that this is the Kolmogorov backward equation, which arises from the Feynman-Kac formula, see Chap. 9, when calculating the conditional expectation (15.7.3). Here $\tilde{E}$ denotes expectation under an appropriate equivalent probability measure $\tilde{P}$, while $\tilde{W}$ is a standard Wiener process under $\tilde{P}$. Under real world pricing $u(t,X_t)$ is the benchmarked option price, $h(X_T)$ is the benchmarked payoff at time $T$ and $\tilde{P}$ is the real world probability measure. Under risk neutral pricing $u(t,X_t)$ is the discounted option price, $h(X_T)$ is the discounted payoff at time $T$ and $\tilde{P}$ is an equivalent risk neutral probability measure.

### Spatial Discretization

We now solve numerically the PDE (15.7.4)–(15.7.5). The key idea of the finite difference method is to replace the partial derivatives in (15.7.4) with finite difference approximations. When applying the deterministic Taylor formula, we see that the first spatial partial derivative $\frac{\partial u(t,x)}{\partial x}$ can be written as

$$\frac{\partial u(t,x)}{\partial x} = \frac{u(t,x+\Delta x) - u(t,x-\Delta x)}{2\,\Delta x} + R_1(t,x) \tag{15.7.6}$$

and the second partial derivative as

$$\begin{aligned}\frac{\partial^2 u(t,x)}{\partial x^2} &= \frac{\frac{u(t,x+\Delta x)-u(t,x)}{\Delta x} - \frac{(u(t,x)-u(t,x-\Delta x))}{\Delta x}}{2\,\Delta x} + R_2(t,x)\\ &= \frac{u(t,x+\Delta x) - 2\,u(t,x) + u(t,x-\Delta x)}{2\,(\Delta x)^2} + R_2(t,x).\end{aligned} \tag{15.7.7}$$

Here $R_1(t,x)$ and $R_2(t,x)$ are the remainder terms of the Taylor expansions and $\Delta x > 0$ is a small increment in the direction of the spatial coordinate $x$.

We next construct an equally spaced grid

$$\mathcal{X}^1_{\Delta x} = \{x_k = k\,\Delta x :\ k \in \{0,1,\ldots,N\}\} \tag{15.7.8}$$

for $N \in \{2,3,\ldots\}$, called a *spatial discretization*. The spatial derivative (15.7.6) and (15.7.7) at a point $(t,x_k)$ can then be expressed as

$$\frac{\partial u(t,x_k)}{\partial x} = \frac{u_{k+1}(t) - u_{k-1}(t)}{2\,\Delta x} + R_1(t,x_k) \tag{15.7.9}$$

and

$$\frac{\partial^2 u(t,x_k)}{\partial x^2} = \frac{u_{k+1}(t) - 2\,u_k(t) + u_{k-1}(t)}{2\,(\Delta x)^2} + R_2(t,x_k) \tag{15.7.10}$$

for $k \in \{1,2,\ldots,N-1\}$. For each $k \in \{1,2,\ldots,N-1\}$ and for $t \in (0,T)$, we call $u_k(t)$ an interior value of the function, while $u_0(t)$ and $u_N(t)$ are called boundary values. Hence by using the above spatial discretization we can replace the spatial partial derivatives appearing in (15.7.4) by finite differences.

### System of Coupled ODEs

Let $\boldsymbol{u}(t) = (u_0(t), u_1(t), \ldots, u_N(t))^\top$ describe the evolution of the solution to the PDE (15.7.4) at the grid points (15.7.8), over time. Using (15.7.9) and (15.7.10), we can show that $\boldsymbol{u}$ is determined by the following coupled system of ODEs

$$\frac{d\boldsymbol{u}(t)}{dt} + \frac{\boldsymbol{A}(t)}{(\Delta x)^2}\,\boldsymbol{u}(t) + \boldsymbol{R}(t) = 0 \tag{15.7.11}$$

for $t \in [0,T]$. Here $\boldsymbol{R}(t)$ is a vector of remainder terms, while $\boldsymbol{A}(t) = [A^{i,j}(t)]_{i,j=0}^N$ is the following matrix

$$\begin{aligned}
A^{0,0}(t) &= A^{0,2}(t) = A^{N,N-2}(t) = A^{N,N}(t) = (\Delta x)^2 \\
A^{0,1}(t) &= A^{N,N-1}(t) = -2\,(\Delta x)^2 \\
A^{k,k}(t) &= -(b(t,x_k))^2 \\
A^{k,k-1}(t) &= -\frac{1}{2}\,(A^{k,k}(t) + a(t,x_k)\,\Delta x) \\
A^{k,k+1}(t) &= -\frac{1}{2}\,(A^{k,k}(t) - a(t,x_k)\,\Delta x) \\
A^{k,j}(t) &= 0
\end{aligned} \tag{15.7.12}$$

for $k \in \{1,2,\ldots,N-1\}$ and $|k-j| > 1$. Neglecting time dependence, $\boldsymbol{A}(t) = \boldsymbol{A}$ can be expressed as

$$\boldsymbol{A} = \begin{pmatrix}
(\Delta x)^2 & -2(\Delta x)^2 & (\Delta x)^2 & 0 & \ldots & 0 & 0 & 0 \\
A^{1,0} & A^{1,1} & A^{1,2} & 0 & \ldots & 0 & 0 & 0 \\
0 & A^{2,1} & A^{2,2} & A^{2,3} & \ldots & 0 & 0 & 0 \\
\vdots & \vdots & \vdots & \vdots & \ldots & \vdots & \vdots & \vdots \\
0 & 0 & 0 & 0 & \ldots & A^{N-2,N-2} & A^{N-2,N-1} & 0 \\
0 & 0 & 0 & 0 & \ldots & A^{N-1,N-2} & A^{N-1,N-1} & A^{N-1,N} \\
0 & 0 & 0 & 0 & \ldots & (\Delta x)^2 & -2(\Delta x)^2 & (\Delta x)^2
\end{pmatrix}.$$

By ignoring the remainder terms in (15.7.11), we obtain the following approximate vector ODE

$$d\boldsymbol{u}(t) \approx \frac{\boldsymbol{A}(t)}{(\Delta x)^2}\,\boldsymbol{u}(t)\,dt \tag{15.7.13}$$

for $t \in (0,T)$ with terminal condition

$$\boldsymbol{u}(T) = (h(x_0),\ldots,h(x_N))^\top.$$

This provides us with approximate values at the grid points.

There are alternatives to (15.7.9) for approximating first order spatial derivatives. We list some of these below, suppressing the dependence on $t$ on the right hand side

$$
\begin{aligned}
\frac{\partial u(t,x_k)}{\partial x} &= \frac{u_{k+1}-u_k}{\Delta x} + O(\Delta x) \\
\frac{\partial u(t,x_k)}{\partial x} &= \frac{u_k-u_{k-1}}{\Delta x} + O(\Delta x) \\
\frac{\partial u(t,x_k)}{\partial x} &= \frac{u_{k+1}-u_{k-1}}{2\,\Delta x} + O((\Delta x)^2) \\
\frac{\partial u(t,x_k)}{\partial x} &= \frac{3\,u_k-4\,u_{k+1}+u_{k-2}}{2\,\Delta x} + O((\Delta x)^2) \\
\frac{\partial u(t,x_k)}{\partial x} &= \frac{-3\,u_k+4\,u_{k+1}-u_{k-2}}{2\,\Delta x} + O((\Delta x)^2). \qquad (15.7.14)
\end{aligned}
$$

Recall that an expression of the form $O((\Delta x)^q)$ denotes a function of $\Delta x$ that satisfies $\lim_{\Delta x\to 0}\frac{O((\Delta x)^q)}{(\Delta x)^q} < \infty$. In the context of (15.7.14) it describes the convergence behavior of the error term associated with a given finite difference approximation. The higher the value of $q$, the quicker the approximation converges.

For second order spatial derivatives, the following finite difference approximations can be used instead of (15.7.10)

$$
\begin{aligned}
\frac{\partial^2 u(t,x_k)}{\partial x^2} &= \frac{u_k-2\,u_{k-1}+u_{k-2}}{(\Delta x)^2} + O(\Delta x) \\
\frac{\partial^2 u(t,x_k)}{\partial x^2} &= \frac{u_{k+2}-2\,u_{k+1}+u_k}{(\Delta x)^2} + O(\Delta x) \\
\frac{\partial^2 u(t,x_k)}{\partial x^2} &= \frac{u_{k+1}-2\,u_k+u_{k-1}}{(\Delta x)^2} + O((\Delta x)^2) \\
\frac{\partial^2 u(t,x_k)}{\partial x^2} &= \frac{2\,u_k-5\,u_{k-1}+4\,u_{k-2}-u_{k-3}}{(\Delta x)^2} + O((\Delta x)^2) \\
\frac{\partial^2 u(t,x_k)}{\partial x^2} &= \frac{-u_{k+3}+4\,u_{k+2}-5\,u_{k+1}+2\,u_k}{(\Delta x)^2} + O((\Delta x)^2). \qquad (15.7.15)
\end{aligned}
$$

**Time Discretization**

The finite difference solution of (15.7.4) is constructed in two steps. First one applies any of the above spatial discretizations to approximate the spatial derivatives. This produces a coupled system of ODEs, see (15.7.13). The second step involves solving the above system of ODEs. Fortunately, efficient discrete time numerical methods are available for this. For example, one can use discrete time approximations of ODEs that are in fact just special cases of the simulation schemes presented in Sect. 15.4, such as the Euler scheme and the drift implicit Euler scheme.

We shall solve the system of ODEs (15.7.13) numerically, using an equidistant time discretization. We denote the time step size by $\Delta \ll 1$, so that the points in the discretization are given by $\tau_n = n\Delta$, $n \in \{0, 1, \ldots, n_T\}$. Here $n_T$ is the largest integer $n$ for which $\tau_n$ is not greater than $T$. If we neglect higher order terms and apply the Euler scheme to (15.7.13), then we obtain the algorithm

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \boldsymbol{A}(\tau_n)\, \boldsymbol{u}(\tau_n)\, \frac{\Delta}{(\Delta\, x)^2} \tag{15.7.16}$$

for $n \in \{0, 1, \ldots, n_T\}$. To simplify notation, we do not discriminate between the exact solution $\boldsymbol{u}(\cdot)$ and the approximate solution resulting from (15.7.16). Obviously, we know the value of $\boldsymbol{u}(T) = \boldsymbol{u}(\tau_{n_T})$, since the payoff at maturity is known. This is

$$u_k(T) = u(T, x_k) = h(x_k) \tag{15.7.17}$$

for all $k \in \{0, 1, \ldots, N\}$. Next, the system (15.7.16) of discrete time difference equations with terminal condition (15.7.17) can be solved in backward steps, starting from the maturity date. Under certain conditions, one obtains in this way an approximate solution of the system (15.7.16) and, therefore, a numerical solution of the PDE (15.7.4).

The above algorithm is called a *finite difference method*. Such methods are widely used in derivative pricing. They are suitable for a wide range of option pricing problems. Their key limitation is that one can apply these only for low dimensional problems. In general, it is difficult to obtain a reasonable PDE solution for three or higher dimensional models.

Note that the finite difference method described above has some features in common with tree methods. For certain choices of finite difference approximations, some finite difference methods can actually be shown to be equivalent to a tree method. Understanding the parallels between tree methods, simplified weak schemes and finite difference methods is useful for analyzing their numerical properties.

There are two major sources of error in a finite difference method. These are the truncation errors resulting from the spatial discretization, and the time discretization errors, respectively. We have seen in Sects. 15.5 and 15.6 that weak schemes and tree approximations require certain moment relationships for their increments to be satisfied. These conditions involve the spatial and the time discretization step sizes. In particular, these step sizes cannot be chosen with complete freedom. Instead, their values must be chosen to ensure that negative probabilities are avoided. Due to the above mentioned similarities, these constraints have analogues for explicit finite-difference methods, such as (15.7.16). For an explicit method we require $\frac{\Delta}{(\Delta x)^2} \in (0, \frac{1}{2}]$, otherwise the Markov chain approximation to the underlying diffusion represented by the finite difference scheme may feature negative probabilities, see Platen (1992). Only when weak convergence and reasonable levels of numerical stability can be ensured will a finite difference method produce accurate prices.

When using a finite difference method one should be aware that the underlying diffusion process is often defined beyond the boundaries of the state space used in the approximation. Fortunately, for most models the truncation of the state space by a finite difference method does not seriously harm the computational results, as we shall see later.

We emphasize that the approximate solution generated by a finite difference method differs from the exact solution of a PDE. A weak convergence result establishes the link. In the case of weak simulation schemes, we have already seen that numerical stability is crucially important. In the rich literature on finite difference methods, such as Richtmeyer & Morton (1967), one can find conditions and convergence results that make numerical stability properties precise.

### The Theta Method

The following method allows one to obtain better numerical stability than is typically achieved by the finite difference method (15.7.16), which was based on the Euler method. We have seen in Sect. 15.4 that implicit methods are extremely important for achieving appropriate levels of numerical stability. So far we have only considered the simple explicit Euler scheme (15.7.16) to solve the coupled system of ODEs (15.7.13). To improve numerical stability we can instead use the family of implicit Euler schemes, see (15.2.38). These yield approximate solutions of the ODE (15.7.13) for which implicitness can be introduced and controlled. If we use the ODE version of the family of implicit Euler schemes (15.2.38), then we obtain the, so-called, *theta method*

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \Big(\theta\, \boldsymbol{A}(\tau_{n+1})\, \boldsymbol{u}(\tau_{n+1}) + (1-\theta)\, \boldsymbol{A}(\tau_n)\, \boldsymbol{u}(\tau_n)\Big) \frac{\Delta}{(\Delta x)^2} \tag{15.7.18}$$

with *degree of implicitness* $\theta \in \Re$. As in (15.2.38), the $\theta = 0$ case recovers the explicit finite difference method (15.7.16). If we choose $\theta = 1$, then we obtain the *fully implicit finite difference method*

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \boldsymbol{A}(\tau_{n+1})\, \boldsymbol{u}(\tau_{n+1}) \frac{\Delta}{(\Delta x)^2}. \tag{15.7.19}$$

Some literature refers to this method as being unconditionally stable, meaning that errors are not propagated or amplified. This is similar to what we discussed in the case of SDEs, see Shaw (1998).

### Crank-Nicolson Method

On the boundary of unconditionally stable methods lies the popular and important Crank-Nicolson method, obtained by choosing $\theta = \frac{1}{2}$ such that

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \frac{1}{2}\Big(\boldsymbol{A}(\tau_{n+1})\, \boldsymbol{u}(\tau_{n+1}) + \boldsymbol{A}(\tau_n)\, \boldsymbol{u}(\tau_n)\Big) \frac{\Delta}{(\Delta x)^2}. \tag{15.7.20}$$

The Crank-Nicolson scheme is often recommended for pricing derivatives. Because of its time symmetry, it achieves a higher order of convergence than either the explicit or fully implicit methods.

We have seen in previous sections that numerical stability must be given priority before attempting to use higher order methods. Despite its popularity, the Crank-Nicolson method sometimes exhibits surprising numerical instability and any application should be monitored carefully, see Shaw (1998). If more numerical stability is required, then the fully implicit finite difference method (15.7.19) is a better choice.

Finite difference methods with a degree of implicitness $\theta > 0$ introduce a different numerical problem, namely that of solving a coupled system of algebraic equations. The reason is that $\boldsymbol{u}(\tau_{n+1})$ also appears on the right hand side of the difference equation (15.7.18). At each time step one thus has to solve a large linear system of equations. For instance, in the case where the matrix $\boldsymbol{A}(\tau_n) = \boldsymbol{A}$ is time independent, the Crank-Nicolson method (15.7.20) becomes

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \frac{1}{2}\,\boldsymbol{A}\left(\boldsymbol{u}(\tau_{n+1}) + \boldsymbol{u}(\tau_n)\right)\frac{\Delta}{(\Delta x)^2}. \tag{15.7.21}$$

This may be reformulated as follows:

$$\left(\boldsymbol{I} - \frac{1}{2}\,\boldsymbol{A}\,\frac{\Delta}{(\Delta x)^2}\right)\boldsymbol{u}(\tau_{n+1}) = \left(\boldsymbol{I} + \frac{1}{2}\,\boldsymbol{A}\,\frac{\Delta}{(\Delta x)^2}\right)\boldsymbol{u}(\tau_n), \tag{15.7.22}$$

where $\boldsymbol{I}$ is the unit matrix. Setting

$$\boldsymbol{M} = \boldsymbol{I} - \frac{1}{2}\,\boldsymbol{A}\,\frac{\Delta}{(\Delta x)^2}$$

and

$$\boldsymbol{B}_n = \left(\boldsymbol{I} + \frac{1}{2}\,\boldsymbol{A}\,\frac{\Delta}{(\Delta x)^2}\right)\boldsymbol{u}(\tau_n)$$

we can rewrite (15.7.22) as

$$\boldsymbol{M}\,\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{B}_n. \tag{15.7.23}$$

To obtain $\boldsymbol{u}(\tau_{n+1})$ from (15.7.22) one must invert $\boldsymbol{M}$. In general, this is a difficult problem because $\boldsymbol{M}$ is usually a large matrix. Fortunately, most of its elements are zero. Additionally, the nonzero elements form diagonal bands. Such matrices are called *sparse* and there is an area of scientific computing devoted to solving sparse linear systems.

### Sparse Matrix Solvers (*)

For the solution of sparse linear systems one typically uses either *direct solvers* or *iterative solvers*. A direct solver computes the solution up to the limit of precision of the computer. A popular direct solver is the tridiagonal solver, also known as Gauss elimination method. It can only be applied in certain situations with reasonable effort. An iterative solver on the other hand, produces a solution only as accurate as specified by the user, but provides more flexibility and efficiency. This type of solver is usually sufficient for a finite difference method because there is no need to obtain exact solutions for the difference equation, since it is only an approximation of the PDE.

Iterative solvers, see Barrett, Berry, Chan, Demmel, Dongarra, Eijkhout, Pozo, Romine & van der Vorst (1994), start with an initial guess, which is iteratively improved. An example is the *Jacobi method*, which for a system

$$\boldsymbol{M}\,\boldsymbol{u} = \boldsymbol{B},$$

see (15.7.23), with $\boldsymbol{M} = [M^{i,j}]_{i,j=1}^{N}$, $\boldsymbol{u} = (u^1, \dots, u^N)^\top$ and $\boldsymbol{B} = (B^1, \dots, B^N)^\top$, performs the following calculation in the $k$th iteration

$$u_k^i = \frac{1}{M^{i,i}} \left( B^i - \sum_{j \neq i} M^{i,j}\, u_{k-1}^j \right). \tag{15.7.24}$$

Here $\boldsymbol{u}_k = (u_k^1, \dots, u_k^N)^\top$ denotes the approximate solution after the $k$th iteration.

A simple generalization of the above method is the *Gauss-Seidel method*, whose $k$th iteration is

$$u_k^i = \frac{1}{M^{i,i}} \left( B^i - \sum_{j<i} M^{i,j}\, u_k^j - \sum_{j>i} M^{i,j}\, u_{k-1}^j \right). \tag{15.7.25}$$

Here the improvements already obtained are worked into the solution at each step.

Finally, we mention the *successive overrelaxation method*, which at each step averages its result from the previous iteration with a value determined by the Gauss-Seidel procedure. In detail, one obtains

$$u_k^i = \alpha\, \bar{u}_k^i + (1-\alpha)\, u_{k-1}^i$$

with

$$\bar{u}_k^i = \frac{1}{M^{i,i}} \left( B^i - \sum_{j<i} M^{i,j}\, u_k^j - \sum_{j>i} M^{i,j}\, u_{k-1}^j \right). \tag{15.7.26}$$

The parameter $\alpha \in (0, 1]$ is called *overrelaxation parameter*. If $\alpha = 1$ we simply retrieve the Gauss-Seidel method.

### Predictor-Corrector Methods (*)

It is quite possible that a predictor-corrector method for ODEs of the type (15.4.28) may work well for obtaining numerical solutions of the coupled system of ODEs (15.7.13). With such a method one does not need to solve a large system of algebraic equations or invert a matrix in each time step. For the approximation of the ODE (15.7.13) the *modified trapezoidal method*, see also (15.4.31), uses the corrector

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \frac{1}{2}\Big(\boldsymbol{A}(\tau_{n+1})\,\bar{\boldsymbol{u}}(\tau_{n+1}) + \boldsymbol{A}(\tau_n)\,\boldsymbol{u}(\tau_n)\Big)\,\frac{\Delta}{(\Delta x)^2} \tag{15.7.27}$$

and the predictor

$$\bar{\boldsymbol{u}}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \boldsymbol{A}(\tau_n)\,\boldsymbol{u}(\tau_n)\,\frac{\Delta}{(\Delta x)^2}. \tag{15.7.28}$$

This method inherits its order of convergence from the Crank-Nicolson scheme. However, as with the Crank-Nicolson method, it lies on the border of unconditional stability. Consequently, it can have still problems with its numerical stability. Better numerical stability can be expected from the following *predictor-corrector method*, with corrector

$$\boldsymbol{u}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \boldsymbol{A}(\tau_{n+1})\,\bar{\boldsymbol{u}}(\tau_{n+1})\,\frac{\Delta}{(\Delta x)^2} \tag{15.7.29}$$

and predictor

$$\bar{\boldsymbol{u}}(\tau_{n+1}) = \boldsymbol{u}(\tau_n) + \boldsymbol{A}(\tau_n)\,\boldsymbol{u}(\tau_n)\,\frac{\Delta}{(\Delta x)^2}.$$

This method is similar to the implicit Euler method (15.4.24) and constrains error propagation, even if the ratio $\frac{\Delta}{(\Delta x)^2}$ is not very small.

### Boundary Conditions (*)

Most PDE solutions are required to satisfy certain boundary conditions. For example, at maturity $T$ the price of a European call option $c_{T,K}(t,S)$ has to match its payoff. Also, in the case of a European call option under the Black-Scholes model, boundary conditions are required to capture the following limiting behavior:

$$\lim_{S\to 0} c_{T,K}(t,S) = 0 \tag{15.7.30}$$

and

$$\lim_{S\to\infty} c_{T,K}(t,S) = S - K\,\exp\{-r\,(T-t)\}, \tag{15.7.31}$$

assuming a constant interest rate $r \geq 0$. These have to be translated into the boundary conditions for the finite difference method. Similar observations apply to other Markovian asset price models.

There is typically some flexibility in constructing the spatial grid, and the decision of where to truncate is often quite subjective. In practice, for a finite difference method the domain of the underlying asset $S$ cannot be $(0,\infty)$, so we must reduce the problem to a finite interval

$$x_{\min} = x_0 < \ldots < x_N = x_{\max}.$$

For a European call option we can set $x_{\min} = x_0 = 0$ and specify the following boundary condition there as

$$u_0(t) = 0. \tag{15.7.32}$$

At the other end of the grid we take $x_{\max} = x_N < \infty$ large enough so that the boundary condition, where one could then set

$$u_N(t) \approx x_N - K \exp\{-r\,(T-t)\}, \tag{15.7.33}$$

is reasonable. Other choices are also possible, as we shall discuss below. In particular, for barrier options and other exotic options the design of the boundary can become quite important.

### Truncation at the Boundaries (*)

As mentioned, one disadvantage of finite difference methods is that they can only handle bounded state variables. So the state space must be truncated if it is naturally infinite. This introduces a truncation error, which is avoided when one uses another method, such as Monte Carlo simulation with variance reduction, see Sect. 15.5. To highlight the truncation problem, let us consider an underlying security $S$ subject to the Black-Scholes model. Under the risk neutral measure the process $X = \{X_t = \ln(S_t),\ t \in [0,T]\}$ satisfies the SDE

$$dX_t = \left(r - \frac{\sigma^2}{2}\right) dt + \sigma\, dW_t \tag{15.7.34}$$

for $t \in [0,T]$ with $X_0 = \ln(S_0)$. Let $u(t,X_t)$ denote the price of a European option on the above security, with payoff $H(X_T) \geq 0$ at maturity $T$. Define the following operator

$$\hat{L}\, f(t,x) = \frac{1}{2}\,\sigma^2\, \frac{\partial^2 f(t,x)}{\partial x^2} + \left(r - \frac{\sigma^2}{2}\right) \frac{\partial f(t,x)}{\partial x} - r\, f(t,x) = 0.$$

By the Feynman-Kac formula, see Sect. 9.7, $u(\cdot,\cdot)$ satisfies the PDE

$$\hat{L}\, u(t,x) = 0 \tag{15.7.35}$$

for all $(t,x) \in (0,T) \times \Re$, subject to the terminal condition

$$u(T,x) = H(x) \tag{15.7.36}$$

for all $x \in \Re$. Notice that the terminal value problem (15.7.35)–(15.7.36) is defined over the semi-infinite domain $[0,T] \times \Re$.

Let us now consider a truncated region $(0,T) \times (-\ell,\ell)$, where $\ell \gg 1$ is large. The finite difference method can be applied over this bounded domain.

However, the boundaries $-\ell$ and $\ell$ have been introduced artificially and boundary conditions must be specified there. One possibility is to impose *Dirichlet conditions*, which means that one specifies the value of $u(t,\ell)$ and $u(t,-\ell)$. Alternatively, one may introduce *von Neumann conditions*, where one specifies the first derivatives $\frac{\partial u(t,\ell)}{\partial x}$ and $\frac{\partial u(t,-\ell)}{\partial x}$.

For simplicity, we consider the application of Dirichlet boundary conditions at the newly introduced boundaries. These take the form of lower and upper functions $\underline{u}_\ell(t)$ and $\bar{u}_\ell(t)$ for $t \in (0,T)$. We have now transformed (15.7.35)–(15.7.36) into the following boundary value problem

$$\hat{L}\, u_\ell(t,x) = 0 \tag{15.7.37}$$

for all $(t,x) \in (0,T) \times (-\ell,\ell)$, with Dirichlet boundary conditions

$$u_\ell(t,-\ell) = \underline{u}_\ell(t) \quad \text{and} \quad u_\ell(t,\ell) = \bar{u}_\ell(t) \tag{15.7.38}$$

for all $t \in [0,T)$ and terminal condition

$$u_\ell(T,x) = H(x) \tag{15.7.39}$$

for all $x \in [-\ell,\ell]$.

The important question is whether the truncated pricing function $u_\ell(t,x)$, which solves (15.7.37)–(15.7.39), converges to $u(t,x)$, which solves (15.7.35)–(15.7.36), as $\ell \to \infty$, for all $(t,x) \in (0,T) \times \Re$. If the payoff $H(\cdot)$ is bounded by a constant $\bar{F} < \infty$, then Lamberton & Lapeyre (1996) answer this question affirmatively, with the following result.

**Lemma 15.7.1.** *For all* $(t,x) \in (0,T) \times \Re$

$$\lim_{\ell\to\infty} u_\ell(t,x) = u(t,x). \tag{15.7.40}$$

### Convergence of Finite Difference Method (*)

Once the domain of the PDE has been truncated, one can discretize it in the spatial direction, using a spatial step size $\Delta x$ and in the time direction, using a time step size $\Delta$. In the time direction it is possible to use, for instance, the theta method to solve the system of ODEs (15.7.18), which we can now express as

$$\boldsymbol{u}^\Delta(\tau_{n+1}) = \boldsymbol{u}^\Delta(\tau_n) + \Big(\theta\, \boldsymbol{A}(\tau_{n+1})\, \boldsymbol{u}^\Delta(\tau_{n+1}) + (1-\theta)\, \boldsymbol{A}(\tau_n)\, \boldsymbol{u}^\Delta(\tau_n)\Big) \frac{\Delta}{(\Delta x)^2} \tag{15.7.41}$$

for $n \in \{0,1,\ldots,n_T\}$ and degree of implicitness $\theta \in [0,1]$. Here we have used $u^\Delta$ instead of $u$ to indicate the dependence on $\Delta$. Let

$$u_k^\Delta(\tau_n) = u^\Delta(\tau_n, x_k),$$

again show the dependence of the approximate solution on $\Delta$. By applying this method, we thus obtain an approximate value $u^{\Delta}(t,x)$. Assuming uniform boundedness of $u$ and $\frac{\partial u}{\partial x}$, the following result from Raviart & Thomas (1983) is applicable.

**Lemma 15.7.2.** *For a degree of implicitness $\theta \in [0, \frac{1}{2})$, if $\Delta \to 0$ and $\Delta x \to 0$ in such a way that $\frac{\Delta}{(\Delta x)^2} \to 0$, then*

$$\lim_{\Delta \to 0} u^{\Delta}(t,x) = u(t,x) \tag{15.7.42}$$

*for all $(t,x) \in (0,T) \times (-\ell, \ell)$.*

This means that for low degrees of implicitness extremely small time step sizes are required in order to achieve a certain level of accuracy. This problem can be quite severe if the spatial step size is already small. However, Raviart & Thomas (1983) also provide the following result.

**Lemma 15.7.3.** *For any degree of implicitness $\theta \in [\frac{1}{2}, 1]$, if $\Delta \to 0$ and $\Delta x \to 0$, then*

$$\lim_{\Delta \to 0} u^{\Delta}(t,x) = u_{\ell}(t,x) \tag{15.7.43}$$

*for all $(t,x) \in (0,T) \times (-\ell, \ell)$.*

In other words, for methods with high levels of implicitness one does not need to be too concerned about the ratio $\frac{\Delta}{(\Delta x)^2}$ in order to ensure convergence. This provides more freedom in designing a finite difference method, since it allows large time step sizes to be combined with small spatial step sizes. It also signals important numerical stability and robustness properties, which are essential for the implementation of reliable derivative pricing tools.

When applying the benchmark approach with real world pricing, we have seen that different expectations and PDEs arise, compared with the standard risk neutral approach. However, the available solution methods are very similar in both cases.

## 15.8 Exercises for Chapter 15

**15.1.** Consider a Monte Carlo simulation that generates a sequence of independent, identically distributed nonnegative random variables $X_1, X_2, \ldots$, each having mean $\mu \in \Re$, variance $\sigma^2 \in (0,\infty)$, skewness zero and kurtosis $\kappa \in [3,\infty)$. How does the variance of the estimator

$$\hat{\varrho}_N = \frac{1}{N} \sum_{i=1}^{N} (X_i)^{\frac{1}{2}}$$

decrease with $N$? Is this estimator strongly consistent in the sense of Chap. 2?

**15.2.** Use a Wagner-Platen expansion, with time increment $h > 0$ and Wiener process increment $W_{t_0+h} - W_{t_0}$ in the expansion part, to expand the increment $X_{t_0+h} - X_{t_0}$ of a geometric Brownian motion at time $t_0$, where

$$dX_t = a\, X_t\, dt + b\, X_t\, dW_t$$

for $t \in [t_0, \infty)$ and $X_{t_0} > 0$.

**15.3.** Write down the Euler scheme and the Milstein scheme for the linear SDE

$$dX_t = (\mu\, X_t + \eta)\, dt + \gamma\, X_t\, dW_t$$

for $t \in [0, \infty)$ with $X_0 = 1$.

**15.4.** Determine the Euler and Milstein schemes for the Vasicek short rate model

$$dr_t = \gamma(\bar{r} - r_t)\, dt + \beta\, dW_t$$

for $t \in [0, \infty)$ with $r_0 > 0$. What are the differences between the two schemes?

**15.5.** Derive the explicit order 1.0 strong scheme for the Black-Scholes SDE

$$dX_t = \mu\, X_t\, dt + \sigma\, X_t\, dW_t$$

for $t \in [0, \infty)$ with $X_0 = 1$.

**15.6.** Consider the two-dimensional SDE

$$\begin{aligned} dX_t^1 &= dW_t^1 \\ dX_t^2 &= X_t^1\, dW_t^2 \end{aligned}$$

for $t \in [0, \infty)$, with $X_0^1 = X_0^2 = 1$. Here $W^1$ and $W^2$ are independent Wiener processes. Does the above system of SDEs have commutative noise?

**15.7.** Apply the Milstein scheme to the system of SDEs in Exercise 15.6.

**15.8.** Verify that the two point distributed random variable $\Delta\hat{W}$ with

$$P(\Delta\hat{W} = \pm\sqrt{\Delta}) = \frac{1}{2}$$

satisfies the condition

$$\left|E\left(\Delta\hat{W}\right)\right| + \left|E\left(\left(\Delta\hat{W}\right)^3\right)\right| + \left|E\left(\left(\Delta\hat{W}\right)^2\right) - \Delta\right| \leq K\, \Delta^2.$$

**15.9.** Construct an antithetic variance reduction method to estimate $E(1 + Z + \frac{1}{2}(Z)^2)$, where $Z \sim N(0,1)$. For this purpose combine the raw Monte Carlo estimate

$$V_N^+ = \frac{1}{N} \sum_{k=1}^{N} \left(1 + Z(\omega_k) + \frac{1}{2}(Z(\omega_k))\right)^2,$$

which uses outcomes of $Z$, with the antithetic estimate

$$V_N^- = \frac{1}{N} \sum_{k=1}^{N} \left(1 - Z(\omega_k) + \frac{1}{2}(-Z(\omega_k))\right)^2,$$

which uses the same outcomes, but with a negative sign. Find the degree of variance reduction that can be achieved for the estimator

$$\hat{V}_N = \frac{1}{2}\,(V_N^+ + V_N^-).$$

Is $\hat{V}_N$ an unbiased estimator?

**15.10.** For estimating $E((1 + Z + \frac{1}{2}\,(Z)^2))$, use the control variate

$$V_N^* = \frac{1}{N} \sum_{k=1}^{N} (1 + Z(\omega_k))\,.$$

Analyze the variance reduction that can be achieved by the estimate

$$\tilde{V}_N = V_N^+ + \alpha(\gamma - V_N^*)$$

for different $\alpha \in \Re$. For which choice of $\gamma \in \Re$ does one obtain an unbiased estimator? Which $\alpha \in \Re$ achieves the minimum variance ?

**15.11.** For a European put option on a stock with benchmarked strike $\hat{K} > 0$ and maturity $T = n_T \Delta,\ n_T \in \mathcal{N}$, and time step size $\Delta \in (0, 1)$, as described in Sect. 15.6, write down a multi-period binomial tree option pricing formula at time $t = 0$. Here a positive benchmarked return of the underlying security is $u = \exp\{\sigma\sqrt{\Delta}\}$ and it is assumed that the growth optimal portfolio has the value one at time $t = 0$.

**15.12.** Show that the Box-Muller method generates a pair of independent standard Gaussian random variables.

# 16

# Solutions for Exercises

### Solutions for Exercises of Chapter 1

**1.1** For a random variable $X$ with second moment we have

$$\begin{aligned}\mathrm{Var}(X) &= E((X - E(X))^2) = E(X^2 - 2\,X\,E(X) + (E(X))^2) \\ &= E(X^2) - 2(E(X))^2 + (E(X))^2 = E(X^2) - (E(X))^2.\end{aligned}$$

**1.2** For a Poisson distributed random variable $X \sim P(\lambda)$ with intensity $\lambda$ we have the mean

$$E(X) = \sum_{i=0}^{\infty} i\,p_i = \lambda\,e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda\,e^{-\lambda}\,e^{\lambda} = \lambda$$

and the second moment

$$\begin{aligned}E(X^2) &= \sum_{i=0}^{\infty} i^2\,p_i = \lambda\,e^{-\lambda} \sum_{i=1}^{\infty} i\,\frac{\lambda^{i-1}}{(i-1)!} \\ &= \lambda\,e^{-\lambda} \left( \sum_{i=1}^{\infty} \left( \frac{\lambda^{i-1}}{(i-1)!} + (i-1)\,\frac{\lambda^{i-1}}{(i-1)!} \right) \right) \\ &= \lambda\,e^{-\lambda} \left( \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} + \lambda \sum_{i=1}^{\infty} \frac{\lambda^{i-2}}{(i-2)!} \right) \\ &= \lambda\,e^{-\lambda}\,(e^{\lambda} + \lambda\,e^{\lambda}) = \lambda\,(1+\lambda).\end{aligned}$$

Thus, by the result from Exercise 1.1 we obtain the variance

$$\mathrm{Var}(X) = E(X^2) - (E(X))^2 = \lambda + \lambda^2 - \lambda^2 = \lambda.$$

E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

**1.3** We have for a uniformly distributed random variable $X \sim U(a,b)$ the mean

$$E(X) = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{2(b-a)}\,(b^2 - a^2) = \frac{a+b}{2}.$$

the second moment

$$E(X^2) = \int_a^b \frac{x^2}{b-a}\,dx = \frac{1}{3(b-a)}\,(b^3 - a^3) = \frac{1}{3}\,(b^2 + ab + a^2)$$

and, thus, the variance

$$\mathrm{Var}(X) = \frac{1}{3}\,(b^2 + ab + b^2) - \frac{1}{4}\,(b+a)^2 = \frac{(b-a)^2}{12}.$$

**1.4** For an exponentially distributed $X \sim Exp(\lambda)$ with intensity $\lambda$ it follows

$$E(X) = \int_0^\infty x\,\lambda\,e^{-\lambda x}\,dx = \lim_{x\to\infty} \frac{1}{\lambda}\,(1 - (\lambda\,x + 1)\,e^{-\lambda x}) = \frac{1}{\lambda},$$

$$E(X^2) = \int_0^\infty x^2\,\lambda\,e^{-\lambda x}\,dx = \lim_{x\to\infty} \frac{1}{\lambda^2}\,(2 - (\lambda^2\,x^2\,2\,\lambda\,x + 2)\,e^{-\lambda x}) = \frac{2}{\lambda^2},$$

and therefore

$$\mathrm{Var}(X) = 2\,\lambda^{-2} - (\lambda^{-1})^2 = \lambda^{-2}.$$

**1.5** For standard Gaussian $X \sim N(0,1)$ we have the mean

$$\begin{aligned}
E(X) &= \int_{-\infty}^0 \frac{x}{\sqrt{2\,\pi}}\,e^{-\frac{1}{2}x^2}\,dx + \int_0^\infty \frac{x}{\sqrt{2\,\pi}}\,e^{-\frac{1}{2}x^2}\,dx \\
&= \sqrt{2\,\pi}\left(\lim_{h\to-\infty}\left(e^{-\frac{1}{2}h^2} - 1\right) + \lim_{h\to\infty}\left(1 - e^{-\frac{1}{2}h^2}\right)\right) \\
&= \sqrt{2\,\pi}\,(-1+1) = 0
\end{aligned}$$

and the variance

$$\begin{aligned}
\mathrm{Var}(X) = E(X^2) &= \int_{-\infty}^0 \frac{x^2}{\sqrt{2\,\pi}}\,e^{-\frac{1}{2}x^2}\,dx + \int_0^\infty \frac{x^2}{\sqrt{2\,\pi}}\,e^{-\frac{1}{2}x^2}\,dx \\
&= \frac{1}{\sqrt{2\,\pi}}\lim_{h\to\infty}\left(-h\,e^{-\frac{1}{2}\,h^2}\right) + \frac{1}{\sqrt{2\,\pi}}\int_0^\infty e^{-\frac{1}{2}\,x^2}\,dx \\
&\quad + \frac{1}{\sqrt{2\,\pi}}\lim_{h\to\infty}\left(-h\,e^{-\frac{1}{2}\,h^2}\right) + \frac{1}{\sqrt{2\,\pi}}\int_{-\infty}^0 e^{-\frac{1}{2}\,x^2}\,dx \\
&= 0 + \frac{1}{2} + 0 + \frac{1}{2} = 1.
\end{aligned}$$

**1.6** For $X \sim N(0,1)$ standard Gaussian distributed and $k \in \mathcal{N}$ we have

$$E(X^{2k}) = \frac{1}{\sqrt{2\,\pi}} \int_{-\infty}^{\infty} x^{2k}\, e^{-\frac{1}{2}\,x^2}\, dx = \sqrt{\frac{2}{\sqrt{\pi}}} \int_{0}^{\infty} z^{2k}\, e^{-\frac{1}{2}\,z^2}\, dz$$

$$= 2^{\frac{2k-1}{2}} \sqrt{\frac{2}{\pi}} \int_{0}^{\infty} t^{\frac{2k-1}{2}}\, e^{-t}\, dt$$

$$= 2^{\frac{2k-1}{2}} \sqrt{\frac{2}{\pi}}\; \Gamma\left(k + \frac{1}{2}\right),$$

where $\Gamma(\cdot)$ is the gamma function, see (1.2.10). Thus we obtain

$$E(X^{2k}) = 1 \cdot 3 \cdot 5 \cdot \ldots \cdot (2k-1).$$

**1.7** We show that

$$E(X) = \int_{-\infty}^{\infty} \frac{y-\mu}{\sigma}\, \frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}\, dy = 0$$

and

$$E\left((X)^2\right) = \int_{-\infty}^{\infty} \left(\frac{y-\mu}{\sigma}\right)^2 \frac{1}{\sqrt{2\,\pi}\,\sigma}\, e^{-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}}\, dy = 1$$

and notice that a linear transform of a Gaussian random variable is Gaussian.

**1.8** The square $Y^2$ of a standard Gaussian random variable $Y \sim N(0,1)$ is $\chi^2(1)$, that is chi-square distributed with $n = 1$ degree of freedom. This means, it is $G(\frac{1}{2}, \frac{1}{2})$ gamma distributed.

**1.9** We obtain by using the Gaussian density and the definition of an expectation that

$$E(Y) = E\left(\exp\{X\}\right) = \int_{-\infty}^{\infty} \exp\{x\}\, \frac{1}{\sqrt{2\,\pi}\,\sigma}\, \exp\left\{-\frac{(x-\mu)^2}{2\,\sigma^2}\right\} dx$$

$$= \exp\left\{\mu + \frac{\sigma^2}{2}\right\},$$

see also (1.3.76).

**1.10** (*) We rely on the following property of the standard Gaussian density, which can be verified by completing the square in its exponent:

$$N'(x-\theta) = \exp\left\{-\frac{1}{2}\,\theta^2 + \theta\,x\right\} N'(x)$$

for all $x, \theta \in \Re$. It then follows by change of variable that

$$\begin{aligned} E(H(X+\theta)) &= \int_{-\infty}^{\infty} H(x+\theta)\, N'(x)\, dx \\ &= \int_{-\infty}^{\infty} H(\bar{x})\, N'(\bar{x}-\theta)\, d\bar{x} \\ &= \int_{-\infty}^{\infty} \exp\left\{-\frac{1}{2}\,\theta^2 + \theta\,\bar{x}\right\} H(\bar{x})\, N'(\bar{x})\, d\bar{x} \\ &= E\left(\exp\left\{-\frac{1}{2}\,\theta^2 + \theta\, X\right\} H(X)\right). \end{aligned}$$

**1.11** (*) Assume that the inverse $C$ of the covariance matrix $D = C^{-1}$ has the form

$$C = [c^{i,j}] = \begin{bmatrix} 4 & -6 \\ -6 & 12 \end{bmatrix}.$$

Its determinant is then $\det(C) = 12$. Furthermore, assume that $X_1 \sim N(0,1)$ and $X_2 \sim N(0, \frac{1}{3})$. We have then from (1.4.16) the joint density

$$\begin{aligned} p(x_1, x_2) &= \frac{\sqrt{\det(C)}}{2\,\pi} \exp\left\{-\frac{1}{2} \sum_{i,j=1}^{2} C^{i,j}\,(x_i-\mu_i)\,(x_j-\mu_j)\right\} \\ &= \frac{\sqrt{12}}{2\,\pi} \exp\left\{-\frac{1}{2}\left(4\,x_1^2 - 12\,x_1\,x_2 + 12\,x_2^2\right)\right\} \\ &\neq \frac{1}{\sqrt{2\,\pi}} \exp\left\{-\frac{x_1^2}{2}\right\} \frac{1}{\sqrt{\frac{2\,\pi}{3}}} \exp\left\{-\frac{1}{2}\,\frac{x_2^2}{3}\right\} \\ &= p(x_1)\cdot p(x_2), \end{aligned}$$

which shows that $X_1$ and $X_2$ are not independent. The random variables would be independent if $C$ would be a diagonal matrix.

**1.12** (*) Differentiating the function

$$F(x) = \frac{1}{\pi}\,\ln(\sqrt{1+x^2})$$

yields

$$F'(x) = x\,p(x) = x\,[\pi\,(1+x^2)]^{-1}.$$

We observe that both one sided improper integrals

$$\int_{-\infty}^{0} x\,p(x)\,dx \quad \text{and} \quad \int_{0}^{\infty} x\,p(x)\,dx$$

diverge. Therefore, the two sided improper integral $\int_{-\infty}^{\infty} x\, p(x)\, dx$ diverges.

**1.13** (*) The conditional density for $X$ with $f_X(x) = x$ with respect to the event $A = \{\omega \in [0, 0.5]\}$ is

$$f_X\left(x \,|\, A\right) = \begin{cases} 0 & \text{for} \quad x \notin [0, 0.5] \\ 8\, x & \text{for} \quad x \in [0, 0.5]. \end{cases}$$

Therefore, the conditional expectation amounts to

$$E\left(X \,|\, A\right) = \int_{-\infty}^{\infty} x\, f_X\left(x \,|\, A\right) dx = \int_0^{0.5} 8\, x^2\, dx = \frac{1}{3}.$$

## Solutions for Exercises of Chapter 2

**2.1** We can apply the strong Law of Large Numbers since

$$\sum_{i=1}^{\infty} \frac{Var(X_i)}{i^2} = K \sum_{i=1}^{\infty} (i)^{-2} = K\, \frac{\pi^2}{6} < \infty.$$

Therefore, it holds that

$$\mu \overset{\text{a.s.}}{=} \lim_{n\to\infty} \hat{\mu}_n.$$

**2.2** By the Central Limit Theorem it follows that the sequence $\hat{Y}_n$ converges in distribution for $n \to \infty$ to a Gaussian random variable with mean zero and variance $\sigma^2$.

**2.3** The $100(1-\alpha)\%$ confidence interval for $2Z$ uses its mean $2E(Z)$ and variance $4\text{Var}(Z)$ and is given in the form

$$\left(2\left(E(Z) - \sqrt{\text{Var}(Z)}\right) p_{1-\alpha}, 2\left(E(Z) + \sqrt{\text{Var}(Z)}\, p_{1-\alpha}\right)\right)$$

with $p_{1-\alpha} \approx 2.58$ for $\alpha = 99\%$.

**2.4** We need to satisfy the relation

$$P\left(\frac{Z - E(Z)}{\sqrt{\text{Var}(Z)}}\right) < -z_\alpha$$

with

$$z_\alpha = \frac{\text{VaR}((1-\alpha)\%) + E(Z)}{\sqrt{\text{Var}(Z)}}.$$

The one sided confidence interval is of the form

$$(-\infty, z_\alpha),$$

where $z_\alpha = z_{0.01} \approx 2.35$ for $\alpha = 99\%$.

## Solutions for Exercises of Chapter 3

**3.1** Let $W$ be a standard Wiener process and let $s \in [0, t]$. Then

$$\begin{aligned} C(s,t) &= E((W_t - E(W_t))(W_s - E(W_s))) = E(W_t\, W_s) \\ &= E((W_t - W_s + W_s)W_s) \\ &= E((W_t - W_s)W_s) + E(W_s^2) \\ &= E(W_t - W_s)\, E(W_s) + E(W_s^2) = 0 \cdot 0 + s = s \end{aligned}$$

since $W_s$ and $W_t - W_s$ are independent for $s < t$. Analogously, $C(s,t) = t$ for $t < s$. Hence

$$C(s,t) = \min(s,t) = \frac{1}{2}\,(|s+t| - |s-t|).$$

**3.2** The covariance of the Wiener process is not a function of $(t - s)$ only, so the Wiener process is not stationary. A similar argument applies for a random walk.

**3.3** Relation (3.3.12) relates to a Bernoulli trial with $n$ independent outcomes and $\frac{j-(k-n)}{2}$ successes (here upward moves) that occur with probability 0.5. The probability for such an event is given by the binomial distribution with

$$p_j(n) = \frac{n!}{\left(\frac{j-(k-n)}{2}\right)!\,\left(n - \frac{j-(k-n)}{2}\right)!}\,\left(\frac{1}{2}\right)^n.$$

**3.4** The probability $q_j(n)$ for having $j$ upwards moves in a non-symmetric random walk in $n$ time steps is related to the binomial distribution with probability $p$ for an upward move. Therefore it is

$$q_j(n) = \frac{n!}{j!\,(n-j)!}\,(p)^j\,(1-p)^{n-j}.$$

**3.5** The stationary probability vector is $(0.5, 0.5)$ and, therefore, we have the mean $\mu = 0.5 \cdot 0.05 + 0.5 \cdot 0.06 = 0.055$, and the variance $v = 0.5(0.05 - 0.055)^2 + 0.5(0.06 - 0.055)^2 = 0.000025$.

**3.6** The long term expected squared interest rate is computed by using the ergodicity and, thus, the stationary probability vector $(0.5, 0.5)$. It then follows

$$E\left((X_t)^2\right) = 0.5(0.05)^2 + 0.5(0.06)^2 = 0.00305.$$

**3.7** We obtain by the formula (3.5.1) for the Poisson probabilities that

$$E(N_t) = \sum_{k=1}^{\infty} \frac{k^2}{k!} e^{-\lambda t} (\lambda t)^k$$
$$= \lambda t \left( \sum_{k=1}^{\infty} (k-1) e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} + \sum_{k=1}^{\infty} e^{-\lambda t} \frac{(\lambda t)^{k-1}}{(k-1)!} \right)$$
$$= \lambda t (\lambda t + 1).$$

**3.8** By the independence of the marks from the Poisson process it follows that

$$E(Y_t) = E\left( \sum_{k=1}^{N_t} \xi_k \right) = E(N_t) E(\xi_k) = \frac{\lambda t}{2}.$$

**3.9** The probability for a compound Poisson process with intensity $\lambda > 0$ of having no jumps until time $t > 0$ equals the probability of the Poisson process $N$ of having no jumps until that time. Thus, by (3.5.1) we obtain

$$P(N_t = 0) = e^{-\lambda t}.$$

**3.10** (*) The given Lévy process is by (3.6.2) at time $t \in [0, T]$ of the form

$$X_t = \alpha t + \beta W_t + \frac{1}{2} \left( p_\varphi \left( \left\{ \frac{1}{2} \right\}, [0, t] \right) - \lambda t \right).$$

Therefore it follows by the formulas for the means of the Wiener and Poisson process that $E(X_t) = \alpha t$.

Similarly, we obtain from the formulas for the variance of the Wiener and Poisson process the variance of $X_t$ as

$$E\left((X_t - \alpha t)^2\right) = E\left( \left( \beta W_t + \frac{1}{2} \left( p_\varphi \left( \left\{ \frac{1}{2} \right\}, [0, t] \right) - \lambda t \right) \right)^2 \right)$$
$$= E\left((\beta W_t)^2\right) + \frac{1}{4} E\left( \left( p_\varphi \left( \left\{ \frac{1}{2} \right\}, [0, t] \right) - \lambda t \right)^2 \right)$$
$$= \beta^2 (\varphi - \varphi(0)) + \frac{\lambda t}{4}.$$

### Solutions for Exercises of Chapter 4

**4.1** The transition density of the standard Ornstein-Uhlenbeck process is a Gaussian one and of the form (4.2.3). For $t \to \infty$ it converges towards the standard Gaussian density. Thus, the process is stationary with mean 0 and variance 1.

**4.2** According to (4.2.1) the transition density $p(s, x; t, y)$ of the standard Wiener process is Gaussian with mean $x$ and variance $(t - s)$. Therefore, we obtain from (4.3.4) $a(s, x) = 0$ and from (4.3.5) $b(s, x) = 1$.

**4.3** The standard Ornstein-Uhlenbeck process has the Gaussian transition density $p(s, x; t, y)$ given in (4.2.3) with mean $x \exp\{-(t - s)\}$ and variance $(1 - e^{-2(t-s)})$. Thus by (4.3.4) we have

$$a(s, x) = \lim_{t \downarrow s} \frac{1}{(t - s)} (x \exp\{-(t - s)\} - x) = -x$$

and by (4.3.5) it follows

$$\begin{aligned}
b^2(s, x) &= \lim_{t \downarrow s} \frac{1}{t - s} E\left((X_t - X_s)^2 \,\middle|\, X_s = x\right) \\
&= \lim_{t \downarrow s} \frac{1}{t - s} \left[E\left(\left(X_t - X_s - E\left(X_t - X_s \,\middle|\, X_s = x\right)\right)^2\right.\right. \\
&\qquad \left.\left. \middle| X_s = x\right) + E\left(\left(X_t - X_s \,\middle|\, X_s = x\right)\right)^2\right] \\
&= \lim_{t \downarrow s} \frac{1}{t - s} \left[\left(1 - e^{-2(t-s)}\right) + x^2 \left(\exp\{-(t - s)\} - 1\right)^2\right] \\
&= 2.
\end{aligned}$$

Therefore we have $b(s, x) = \sqrt{2}$.

**4.4** For the Gaussian transition density (4.2.1) of the standard Wiener process it holds

$$\frac{\partial}{\partial y} p(s, x; t, y) = p(s, x; t, y) \left(-\frac{(y - x)}{(t - s)}\right)$$

$$\frac{\partial^2}{\partial y^2} p(s, x; t, y) = p(s, x; t, y) \frac{(y - x)^2}{(t - s)^2} - \frac{p(s, x; t, y)}{(t - s)}$$

and

$$\frac{\partial}{\partial t} p(s, x; t, y) = -\frac{1}{2(t - s)} p(s, x; t, y) + \frac{(y - x)^2}{2(t - s)^2} p(s, x; t, y).$$

Therefore

$$\frac{\partial}{\partial t} p(s, x; t, y) - \frac{1}{2} \frac{\partial^2 p(s, x; t, y)}{\partial y^2} = 0,$$

for $(s,x)$ fixed, which provides the Kolmogorov forward equation (4.4.1) with boundary condition (4.4.3).

Similarly we have

$$\frac{\partial p(s,x;t,y)}{\partial s} = \frac{1}{2}\,\frac{p(s,x;t,y)}{(t-s)} - p(s,x;t,y)\left(\frac{(y-x)^2}{2(t-s)^2}\right)$$

and

$$\frac{\partial p(s,x;t,y)}{\partial x} = p(s,x;t,y)\,\frac{(y-x)}{(t-s)}$$

$$\frac{\partial^2 p(s,x;t,y)}{\partial x^2} = p(s,x;t,y)\,\frac{(y-x)^2}{(t-s)^2} - p(s,x;t,y)\,\frac{1}{(t-s)}.$$

Thus

$$\frac{\partial p(s,x;t,y)}{\partial s} + \frac{1}{2}\,\frac{\partial^2 p(s,x;t,y)}{\partial x^2} = 0$$

for $(t,y)$ fixed, which represents the Kolmogorov backward equation (4.4.2) with boundary condition (4.4.3).

**4.5** For the standard Ornstein-Uhlenbeck process we have the Kolmogorov forward equation, see (4.4.1),

$$\frac{\partial p(s,x;t,y)}{\partial t} - \frac{\partial}{\partial y}\,(y\,p(s,x;t,y)) - \frac{1}{2}\,\frac{\partial^2}{\partial y^2}\,(2\,p(s,x;t,y)) = 0$$

that is

$$\frac{\partial p(s,x;t,y)}{\partial t} - p(s,x;t,y) - y\,\frac{\partial}{\partial y}\,p(s,x;t,y) - \frac{\partial^2}{\partial y^2}\,p(s,x;t,y) = 0$$

with boundary condition (4.4.3).

**4.6** The Kolmogorov backward equation for the standard Ornstein-Uhlenbeck process is, see (4.4.2),

$$\frac{\partial p(s,x;t,y)}{\partial s} - x\,\frac{\partial p(s,x;t,y)}{\partial x} + \frac{\partial^2 p(s,x;t,y)}{\partial x^2} = 0$$

with boundary condition (4.4.3). Taking the partial derivatives of the transition density (4.2.3) it follows that

$$\frac{\partial p(s,x;t,y)}{\partial s} = -\frac{1}{2}\, p(s,x;t,y)\, \frac{-2e^{-2(t-s)}}{1-e^{-2(t-s)}} + p(s,x;t,y) \cdot \left( \frac{2\left(y - x\, e^{-(t-s)}\right)\, x\, e^{-(t-s)}}{2\left(1-e^{-2(t-s)}\right)} - \frac{\left(y - x\, e^{-(t-s)}\right)^2\, 2e^{-2(t-s)}}{2\left(1-e^{-2(t-s)}\right)^2} \right)$$

$$\frac{\partial p(s,x;t,y)}{\partial x} = p(s,x;t,y)\, \frac{\left(y - x\, e^{-(t-s)}\right)\, e^{-(t-s)}}{1-e^{-2(t-s)}}$$

$$\frac{\partial^2 p(s,x;t,y)}{\partial x^2} = p(s,x;t,y) \left[ \frac{\left(y - x\, e^{-(t-s)}\right)^2\, e^{-2(t-s)}}{\left(1-e^{-2(t-s)}\right)^2} - \frac{e^{-2(t-s)}}{1-e^{-2(t-s)}} \right].$$

Then we obtain by substituting these partial derivatives into the left hand side of the above Kolmogorov backward equation that

$$\begin{aligned} p(s,x;t,y) \Bigg[ & \frac{e^{-2(t-s)}}{(1-e^{-2(t-s)})} + \frac{y\, x\, e^{-(t-s)}}{(1-e^{-2(t-s)})} - \frac{x^2\, e^{-2(t-s)}}{(1-e^{-2(t-s)})} \\ & - \frac{y^2\, e^{-2(t-s)}}{(1-e^{-2(t-s)})^2} + \frac{2x\, y\, e^{-3(t-s)}}{(1-e^{-2(t-s)})^2} - \frac{x^2\, e^{-4(t-s)}}{(1-e^{-2(t-s)})^2} \\ & - \frac{x\, y\, e^{-(t-s)}}{1-e^{-2(t-s)}} + \frac{x^2\, e^{-2(t-s)}}{1-e^{-2(t-s)}} + \frac{y^2\, e^{-2(t-s)}}{(1-e^{-2(t-s)})^2} \\ & - \frac{2\, y\, x\, e^{-3(t-s)}}{(1-e^{-2(t-s)})^2} + \frac{x^2\, e^{-4(t-s)}}{(1-e^{-2(t-s)})^2} - \frac{e^{-2(t-s)}}{(1-e^{-2(t-s)})} \Bigg] \\ = 0. & \end{aligned}$$

Obviously, for $t = s$ the transition density (4.2.3) equals the Dirac delta function (4.4.3).

**4.7** The stationary density for the standard Ornstein-Uhlenbeck can be taken from formula (4.5.5) or for $(t-s) \to \infty$ from equation (4.2.3). It is with

$$\bar{p}(y) = \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{y^2}{2} \right\}$$

the density of a standard Gaussian random variable.

**4.8** Geometric Brownian motion is not a stationary process because its transition density given in (4.2.2) does not converge for $(t-s) \to \infty$ to a stationary density.

**4.9** The geometric Ornstein-Uhlenbeck process is a stationary process. Its stationary density is the log-normal probability density

$$\bar{p}(y) = \frac{1}{y\sqrt{2\pi}} \exp\left\{-\frac{(\ln(y))^2}{2}\right\}.$$

**4.10** (*) Geometric Brownian motion is not an ergodic process because it does not have a stationary density.

**4.11** (*) With the transition densities (4.2.1) of a standard Wiener process we can write

$$\begin{aligned}
&\int_{-\infty}^{\infty} p(s,x;r,z)\, p(r,z;t,y)\, dz \\
&= \int_{-\infty}^{\infty} \frac{1}{2\pi\sqrt{(r-s)(t-r)}} \exp\left\{-\frac{1}{2}\left(\frac{(z-x)^2}{r-s} + \frac{(y-z)^2}{t-r}\right)\right\} dz \\
&= \frac{1}{\sqrt{2\pi(t-s)}} \exp\left\{-\frac{(y-x)^2}{2(t-s)}\right\} \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{1}{2}u^2\right\} du \\
&= p(s,x;t,y)\cdot 1 = p(s,x;t,y),
\end{aligned}$$

where we used the substitution

$$u = u(z) = \left(z - \frac{x(t-r)+y(r-s)}{t-s}\right)\sqrt{\frac{t-s}{(r-s)(t-r)}}.$$

**4.12** (*) The Ornstein-Uhlenbeck process is an ergodic process, because we have according to (4.5.11) the scale measure

$$s(x) = \exp\left\{\int_0^x y\, dy\right\} = \exp\left\{\frac{x^2}{2}\right\}$$

with the properties

$$\int_0^{\infty} s(x)\, dx = \int_{-\infty}^{0} s(x)\, dx = \int_0^{\infty} \exp\left\{\frac{x^2}{2}\right\} dx = \infty$$

and

$$\int_{-\infty}^{\infty} \frac{1}{2s(x)}\, dx = \int_{-\infty}^{\infty} \frac{1}{2} \exp\left\{-\frac{x^2}{2}\right\} dx = \sqrt{\frac{\pi}{2}} < \infty$$

that prove the conditions for ergodicity (4.5.12) and (4.5.13).

**4.13** (*) Using (4.5.5) we have

$$\frac{d\bar{p}(y)}{dy} = 2\bar{p}(y)\,\frac{a(y)}{b^2(y)} - \bar{p}(y)\,\frac{1}{b^2(y)}\,\frac{db^2(y)}{dy} = \frac{\bar{p}(y)}{b^2(y)}\left(2a(y) - \frac{db^2(y)}{dy}\right).$$

Then it holds

$$\begin{aligned} Q(y) &= a(y)\,\bar{p}(y) - \frac{1}{2}\,\frac{d}{dy}\left(b^2(y)\,\bar{p}(y)\right) \\ &= a(y)\,\bar{p}(y) - \frac{1}{2}\,\bar{p}(y)\,\frac{db^2(y)}{dy} - \frac{1}{2}\,b^2(y)\,\frac{d\bar{p}(y)}{dy} \\ &= \bar{p}(y)\left[a(y) - \frac{1}{2}\,\frac{db^2(y)}{dy} - a(y) + \frac{1}{2}\,\frac{db^2(y)}{dy}\right] = 0 \end{aligned}$$

and it follows

$$\frac{dQ(y)}{dy} = 0$$

which proves (4.5.1).

**4.14** (*) By (4.3.4) and (4.1.2) we obtain by using the Taylor expansion for the exponential

$$\begin{aligned} a(s,x) &= \lim_{t\downarrow s} \frac{1}{t-s}\,E\left(X_t - X_s \,\middle|\, X_s = x\right) \\ &= \lim_{t\downarrow s} E\left(\frac{X_s\left[\exp\{g(t-s) + b(W_t - W_s)\} - 1\right]}{t-s} \,\middle|\, X_s = x\right) \\ &= x\,\lim_{t\downarrow s} E\left(\frac{g(t-s) + b(W_t - W_s)}{t-s}\right. \\ &\qquad \left. + \frac{1}{2}\,\frac{(g(t-s) + b(W_t - W_s))^2}{t-s} \,\middle|\, X_s = x\right) \\ &= x\left(g + \frac{1}{2}\,b^2\right). \end{aligned}$$

## Solutions for Exercises of Chapter 5

**5.1** Assuming a filtered probability space $(\Omega, \mathcal{A}, \underline{\mathcal{A}}, P)$ we have for $0 \le s \le t \le T$ by the martingale property for Wiener processes that

$$\begin{aligned} E\left(Y_t \,\middle|\, \mathcal{A}_s\right) &= E\left(\alpha_1 W_t^1 + \alpha_2 W_t^2 \,\middle|\, \mathcal{A}_s\right) \\ &= \alpha_1\,E\left(W_t^1 \,\middle|\, \mathcal{A}_s\right) + \alpha_2\,E\left(W_t^2 \,\middle|\, \mathcal{A}_s\right) \\ &= \alpha_1\,W_s^1 + \alpha_2\,W_s^2 \\ &= Y_s \end{aligned}$$

which proves the martingale property (5.1.2).

**5.2** We compute for $0 \le s \le t \le T < \infty$ the conditional expectation

$$\begin{aligned} E\left(Y_t \,\middle|\, \mathcal{A}_s\right) &= E\left(W_t^2 \,\middle|\, \mathcal{A}_s\right) \\ &= E\left((W_t - W_s)^2 + W_s^2 \,\middle|\, \mathcal{A}_s\right) \\ &= (t-s) + Y_s \\ &\ge Y_s, \end{aligned}$$

which shows that $Y$ is a submartingale as defined in (5.1.7).

**5.3** We obtain by the properties of the Wiener process for $0 \le t \le s \le T$

$$\begin{aligned} E\left(M_s \,\middle|\, \mathcal{A}_t\right) &= E\left((W_s - W_0)^2 - s \,\middle|\, \mathcal{A}_t\right) \\ &= E\left(((W_s - W_t) + (W_t - W_0))^2 - s \,\middle|\, \mathcal{A}_t\right) \\ &= E\left((W_s - W_t)^2 \,\middle|\, \mathcal{A}_t\right) + W_t^2 - s = W_t^2 - t = M_t, \end{aligned}$$

which shows that $M$ is a martingale.

**5.4** We consider for $0 \le s \le t \le T$ the conditional expectation

$$\begin{aligned} E\left(\bar{X}_t \,\middle|\, \mathcal{A}_s\right) &= E\left(\exp\left\{-\frac{1}{2}\,\sigma^2\,(\varphi - \varphi(0)) + \sigma\,W_t\right\} \,\middle|\, \mathcal{A}_s\right) \\ &= \exp\left\{-\frac{1}{2}\,\sigma^2\,s + \sigma\,W_s\right\} \\ &\quad \times E\left(\exp\left\{-\frac{1}{2}\,\sigma^2\,(t-s) + \sigma\,(W_t - W_s)\right\} \,\middle|\, \mathcal{A}_s\right) \\ &= \bar{X}_s, \end{aligned}$$

where we used the Laplace transform of the Gaussian increment $W_t - W_s$ of the Wiener process $W$ in the form

$$E\left(\exp\{\sigma(W_t - W_s)\} \,\middle|\, \mathcal{A}_s\right) = \exp\left\{\frac{1}{2}\,\sigma^2(t-s)\right\}.$$

$\bar{X}$ is an $(\underline{\mathcal{A}}, P)$-martingale.

**5.5** We have from the covariation property (5.4.5) of Itô integrals that

$$\left[\int_0 a\,du + \int_0 b\,dW_u\right]_t = \int_0^t b^2\,du = b^2\,(\varphi - \varphi(0)).$$

**5.6** Similarly as in Exercise 5.5 we obtain

$$\left[\int_0 a\,du + \int_0 b\,dW_u, \int_0 1\,dW_u\right]_t = \int_0^t b\,du = b\,t.$$

**5.7** By using Jensen's inequality, see (1.3.52), it follows for $0 \leq t \leq s \leq T$ that

$$E\left(g(X_s)\,\big|\,\mathcal{A}_t\right) \geq g\left(E\left(X_s\,\big|\,\mathcal{A}_t\right)\right) = g(X_t),$$

which shows that $g(x)$ is a submartingale.

**5.8** (*) For $f$ being a deterministic step function corresponding to the partition $0 = t_1 < t_2 < \ldots < t_{n+1} = T$ with $f_t = f_j$ for $t \in [t_j, t_{j+1})$ for $j \in \{1, 2, \ldots, n\}$ we have for $0 \leq s \leq t \leq T$ that

$$\begin{aligned}
E\left(I_{f,W}(t)\,\big|\,\mathcal{A}_s\right) &= E\left(\int_0^t f_u\, dW_u\,\Big|\,\mathcal{A}_s\right)\\
&= E\left(\sum_{j=1}^{i_s-1} f_j(W_{t_{j+1}} - W_{t_j}) + f_{i_s}\,(W_s - W_{t_{i_s}})\right.\\
&\qquad + f_{i_s}\,\left(W_{t_{i_s+1}} - W_s\right) + \sum_{j=i_s+1}^{i_t-1} f_j\,(W_{t_{j+1}} - W_{t_j})\\
&\qquad \left. + f_{i_t}\,(W_t - W_{t_{i_t}})\,\Big|\,\mathcal{A}_s\right),
\end{aligned}$$

where

$$i_t = \max\{k \in \{1, 2, \ldots\} : \, t_k \leq t\}.$$

Thus, we obtain by the zero mean property of Wiener process increments that only the first two terms in the above expectation survive so that

$$E\left(I_{f,W}(t)\,\big|\,\mathcal{A}_s\right) = I_{f,W}(s),$$

which proves the martingale property for $I_{f,W}(s)$.

**5.9** (*) Using the notation and representation of the Itô integral of Exercise 5.8 we have for $0 \leq s \leq t \leq T < \infty$ and deterministic step functions $f$ and $\bar{f}$

$$E\left((I_{f,W}(t) - I_{f,W}(s))(I_{\bar{f},W}(t) - I_{\bar{f},W}(s))\,\big|\,\mathcal{A}_s\right)$$

$$\begin{aligned}
&= E\left(\left[f_{i_s}(W_{t_{i_s+1}} - W_s) + \sum_{j=i_s+1}^{i_t+1} f_j\,(W_{t_{j+1}} - W_{t_j}) + f_{i_t}\,(W_t - W_{t_{i_t}})\right]\right.\\
&\quad \left.\times\left[\bar{f}_{i_s}\,(W_{t_{i_s+1}} - W_s) + \sum_{j=i_s+1}^{i_t+1} \bar{f}_j\,(W_{t_{j+1}} - W_{t_j}) + \bar{f}_{i_t}\,(W_t - W_{t_{i_t}})\right]\,\Bigg|\,\mathcal{A}_s\right).
\end{aligned}$$

Thus, it follows by the expectation properties for products of Wiener process increments that

$$E\left((I_{f,W}(t) - I_{f,W}(s))(I_{\bar{f},W}(t) - I_{\bar{f},W}(s)) \,\middle|\, \mathcal{A}_s\right) = \int_s^t E\left(f_u\, \bar{f}_u \,\middle|\, \mathcal{A}_s\right) du.$$

**5.10** (*) Using the notation and representation of the Itô integrals of the Exercises 5.8 and 5.9 we have for $0 \leq s \leq t \leq T < \infty$, and deterministic step functions $f$ and $\bar{f}$ and $\mathcal{A}_s$-measurable constants $\alpha$ and $\bar{\alpha}$ the equation

$$\begin{aligned}\int_s^t \left(\alpha\, f_u + \bar{\alpha}\, \bar{f}_u\right) dW_u =& \left(\alpha\, f_{t_{i_s}} + \bar{\alpha}\, \bar{f}_{t_{i_s}}\right) (W_{t_{i_s+1}} - W_s) \\ &+ \sum_{j=i_s+1}^{i_t-1} \left(\alpha\, f_{t_j} + \bar{\alpha}\, \bar{f}_{t_j}\right) (W_{t_{j+1}} - W_{t_j}) \\ &+ \left(\alpha\, f_{t_{i_t}} + \bar{\alpha}\, \bar{f}_{t_{i_t}}\right) (W_t - W_{t_{i_t}})\end{aligned}$$

$$\begin{aligned}=&\alpha \left[ f_{t_{i_s}} (W_{t_{i_s+1}} - W_s) + \sum_{j=i_s+1}^{i_t-1} f_{t_j} (W_{t_{j+1}} - W_{t_j}) + f_{t_{i_t}} (W_t - W_{t_{i_t}}) \right] \\ &+ \bar{\alpha} \left[ \bar{f}_{t_{i_s}} (W_{t_{i_s+1}} - W_s) + \sum_{j=i_s+1}^{i_t-1} \bar{f}_{t_j} (W_{t_{j+1}} - W_{t_j}) + \bar{f}_{t_{i_t}} (W_t - W_{t_{i_t}}) \right] \\ =&\alpha \int_s^t f_u\, dW_u + \bar{\alpha} \int_s^t \bar{f}_u\, dW_u,\end{aligned}$$

which proves the linearity property (5.4.2) for deterministic step functions.

**5.11** (*) Obviously, we have $E(X_t - X_0 \,|\, \mathcal{A}_0)$ for all $t \in [0, T]$. Since the Lévy process $X$ has stationary independent increments it follows for $0 \leq s \leq t \leq T$ that

$$E\left(X_s \,\middle|\, \mathcal{A}_t\right) = E\left(X_s - X_t \,\middle|\, \mathcal{A}_t\right) + X_t = X_t.$$

This proves that $X$ is a martingale.

## Solutions for Exercises of Chapter 6

**6.1** By the Itô formula it follows that

$$d(Y_t)^2 = (2Y_t\, a + b^2)\, dt + 2\, Y_t\, b\, dW_t.$$

**6.2** By application of the Itô formula we obtain

$$dZ_t = Z_t \left( \mu + \frac{1}{2}\,\sigma^2 \right) dt + Z_t\,\sigma\,dW_t.$$

Applying again the Itô formula we obtain

$$\begin{aligned} d\ln(Z_t) &= \left( \mu + \frac{1}{2}\,\sigma^2 - \frac{1}{2}\,\sigma^2 \right) dt + \sigma\,dW_t \\ &= \mu\,dt + \sigma\,dW_t. \end{aligned}$$

**6.3** It follows by the Itô formula that

$$\begin{aligned} d(Z_t)^2 &= 2\,Z_t^2 \left( \mu + \frac{1}{2}\,\sigma^2 \right) dt + Z_t^2\,\sigma^2\,dt + 2\,Z_t^2\,\sigma\,dW_t \\ &= 2\,Z_t^2 \left( \mu + \sigma^2 \right) dt + 2\,Z_t^2\,\sigma\,dW_t. \end{aligned}$$

**6.4** We have by the Itô formula that

$$\begin{aligned} dZ_t^{-1} &= Z_t^{-1} \left( -\mu - \frac{1}{2}\,\sigma^2 + \sigma^2 \right) dt - Z_t^{-1}\,\sigma\,dW_t \\ &= Z_t^{-1} \left( -\mu + \frac{1}{2}\,\sigma^2 \right) dt - Z_t^{-1}\,\sigma\,dW_t. \end{aligned}$$

**6.5** We obtain by the Itô formula

$$\begin{aligned} d(Y_t\,Z_t) &= \left( Z_t\,a + Y_t\,Z_t \left( \mu + \frac{1}{2}\,\sigma^2 \right) + b\,Z_t\,\sigma \right) dt \\ &\quad + Z_t\,b\,dW_t + Y_t\,Z_t\,\sigma\,dW_t \\ &= Z_t \left( a + Y_t \left( \mu + \frac{1}{2}\,\sigma^2 \right) + b\,\sigma \right) dt \\ &\quad + Z_t\,(b + Y_t\,\sigma)\,dW_t. \end{aligned}$$

**6.6** We have by the Itô formula the stochastic differential

$$d(Y_t^1\,Y_t^2) = (Y_t^2\,a_1 + Y_t^1\,a_2)\,dt + Y_t^2\,b_1\,dW_t^1 + Y_t^1\,b_2\,dW_t^2.$$

**6.7** The stochastic differential is obtained by the Itô formula and we obtain

$$
\begin{aligned}
dZ_t &= d\left(\exp\{Y_t^1\}\,\exp\{Y_t^2\}\right)\\
&= d\left(\exp\{Y_t^1 + Y_t^2\}\right)\\
&= Z_t\left(a_1 + a_2 + \frac{1}{2}\,(b_1^2 + b_2^2)\right) dt + Z_t\, b_1\, dW_t^1 + Z_t\, b_2\, dW_t^2.
\end{aligned}
$$

**6.8** Applying the Itô formula we obtain

$$d(W_t)^2 = 2\, W_t\, dW_t + dt.$$

Now by the covariation property (5.4.5) of Itô integrals we have

$$
\begin{aligned}
[W, (W)^2]_t &= \left[\int_0 dW_s, 2\int_0 W_s\, dW_s + \int_0 ds\right]_t\\
&= \int_0^t 2\, W_s\, ds.
\end{aligned}
$$

**6.9** (*) By the Itô formula we have

$$d(X_t)^2 = 2\, X_t\, \xi_t\, dW_t + (\xi_t)^2\, dt$$

and by the covariation property (5.4.5) of Itô integrals it follows

$$d[X]_t = (\xi_t)^2\, dt.$$

Therefore, it holds

$$
\begin{aligned}
dY_t &= d\left((X_t)^2 - [X]_t\right)\\
&= 2\, X_t\, \xi_t\, dW_t
\end{aligned}
$$

and $Y_t$ is represented by an Itô integral. Thus, by the martingale property (5.4.3) of Itô integrals $Y$ is a martingale.

**6.10** (*) The stochastic differential of $X$ is

$$dX_t = \sigma\, dW_t + \xi\, dp(t)$$

for $t \in [0, T]$. By the Itô formula (6.4.11) it follows that

$$d\,\exp\{X_t\} = \exp\{X_t\}\left(\sigma\, dW_t + \frac{1}{2}\,\sigma^2\, dt\right) + \exp\{X_{t-}\}\,(\exp\{\xi\} - 1)\, dp(t)$$

for $t \in [0, T]$.

**6.11** (*) The stochastic differential of $X$ is

$$dX_t = a\,dp^1(t) + b\,dp^2(t)$$

for $t \in [0, T]$. Using the Itô formula (6.4.11) we obtain

$$\begin{aligned} d\,\exp\{X_t\} = \exp\{X_{t-}\}\,(\exp\{a\} - 1)\,dp^1(t) \\ + \exp\{X_{t-}\}\,(\exp\{b\} - 1)\,dp^2(t) \end{aligned}$$

for $t \in [0, T]$.

## Solutions for Exercises of Chapter 7

**7.1** We obtain from (7.2.6) or (7.3.5) the mean

$$\mu(t) = E(X_t) = \exp\{-(t - t_0)\}.$$

Furthermore we obtain from (7.2.6) or (7.3.9) the variance

$$\begin{aligned} v(t) &= E\left((X_t - E(X_t))^2\right) \\ &= E\left(\left(\int_{t_0}^{t} \sqrt{2}\,\exp\{-(t-s)\}\,dW_s\right)^2\right) \\ &= 2\int_{t_0}^{t} \exp\{-2(t-s)\}\,ds \\ &= 1 - \exp\{-2(t - t_0)\}. \end{aligned}$$

**7.2** According to (7.3.5) we obtain for the mean

$$\mu(t) = \exp\{0.05\,t\}.$$

It follows for the variance

$$\begin{aligned} v(t) &= E\left((X_t - \mu(t))^2\right) \\ &= E(X_t^2) - (\mu(t))^2 \\ &= P(t) - (\mu(t))^2, \end{aligned}$$

where with (7.3.8) we obtain

$$dP(t) = (0.1 + 0.04)\,P(t)\,dt$$

with $P(0) = 1$ such that

$$v(t) = \exp\{0.14\, t\} - \exp\{0.1\, t\}$$

for $t \geq 0$.

**7.3** We apply formula (7.4.5), where

$$\begin{aligned} X_t &= X_0\, \Psi_{t,0} \\ &= X_0 \exp\left\{ \left( -\frac{1}{2} - \frac{1}{2} - \frac{1}{2} \right) t + \sum_{l=1}^{2} \left( W_t^l - W_0^l \right) \right\} \\ &= X_0 \exp\left\{ -\frac{3}{2}\, t + W_t^1 + W_t^2 \right\}. \end{aligned}$$

**7.4** (*) By the Itô formula one obtains

$$dX_t = X_{t-} \left[ \left( k\, a + \frac{k^2\, b^2}{2} \right) dt + k\, b\, dW_t + (\exp\{k\, c\} - 1)\, dN_t \right]$$

for $t \in [0, T]$ with $X_0 = 1$.

**7.5** (*) In the linear SDE for $X_t$ we take the expectation and obtain

$$d\mu(t) = \mu(t-) \left[ k\, a + \frac{k^2\, b^2}{2} + \lambda\, (\exp\{k\, c\} - 1) \right] dt$$

for $t \in [0, T]$ with $\mu(0) = 1$.

**7.6** (*) The explicit solution is of the form

$$X_t = \Psi_{t,0} \left( X_0 + \int_0^t (a_2 - b_1\, b_2)\, \Psi_{s,0}^{-1}\, ds + \int_0^t b_2\, \Psi_{s,0}^{-1}\, dW_s \right)$$

with

$$\Psi_{t,0} = \exp\left\{ \int_0^t \left( a_1 - \frac{1}{2}\, b_1^2 \right) ds + \int_0^t b_1\, dW_s \right\}$$

for $t \in [0, T]$. By application of the Itô formula we obtain

$$\begin{aligned} dX_t &= \Psi_{t,0} \left[ (a_2 - b_1\, b_2)\, \Psi_{t,0}^{-1}\, dt + b_2\, \Psi_{t,0}^{-1}\, dW_t \right] \\ &\quad + \frac{X_t}{\Psi_{t,0}}\, d\Psi_{t,0} + d \left[ \Psi_{\cdot,0}, \int_0^{\cdot} b_2\, \Psi_{s,0}^{-1}\, dW_s \right]_t . \end{aligned}$$

Noting by the Itô formula that

$$d\Psi_{t,0} = \Psi_{t,0}(a_1\, dt + b_1\, dW_t)$$

it follows

$$\begin{aligned} dX_t &= (a_2 - b_1\, b_2)\, dt + b_2\, dW_t \\ &\quad + X_t\, (a_1\, dt + b_1\, dW_t) + b_1\, b_2\, dt \\ &= (X_t\, a_1 + a_2)\, dt + (X_t\, b_1 + b_2)\, dW_t \end{aligned}$$

for $t \in [0, T]$.

## Solutions for Exercises of Chapter 8

**8.1** From (8.3.2) we obtain the discounted option price

$$\bar{V}(t, \bar{S}_t) = \bar{S}_t\, N(d_1(t)) - K\, \exp\left\{ - \int_0^T r_s\, ds \right\} N(d_2(t))$$

with

$$d_1(t) = \left( \ln\left( \frac{\bar{S}_t}{K\, (B_T)^{-1}} \right) + \int_t^T \frac{1}{2}\, \sigma_s^2\, ds \right) \left( \int_t^T \sigma_s^2\, ds \right)^{-\frac{1}{2}}$$

and

$$d_2(t) = d_1(t) - \left( \int_t^T \sigma_s^2\, ds \right)^{\frac{1}{2}}.$$

Then it is

$$\frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}} = N(d_1(t)) + Q_t,$$

where

$$\begin{aligned} Q_t &= \bar{S}_t\, N'(d_1(t))\, \frac{\partial d_1(t)}{\partial \bar{S}} - \frac{K}{B_T}\, N'(d_2(t))\, \frac{\partial d_2(t)}{\partial \bar{S}} \\ &= \frac{\partial d_1(t)}{\partial \bar{S}\, \sqrt{2\pi}} \left[ \bar{S}_t\, \exp\left\{ -\frac{(d_1(t))^2}{2} \right\} - \frac{K}{B_T}\, \exp\left\{ -\frac{(d_2(t))^2}{2} \right\} \right] \\ &= \frac{\partial d_1(t)}{\partial \bar{S}\, \sqrt{2\pi}}\, \exp\left\{ -\frac{(d_1(t))^2}{2} \right\} \left[ \bar{S}_t - \frac{K}{B_T}\, \exp\left\{ \ln\left( \frac{\bar{S}_t}{\frac{K}{B_T}} \right) \right.\right. \\ &\qquad \left.\left. - \frac{1}{2} \int_t^T \sigma_s^2\, ds + \frac{1}{2} \int_t^T \sigma_s^2\, ds \right\} \right] \\ &= 0. \end{aligned}$$

Furthermore, we have

$$\begin{aligned} \frac{\partial^2 \bar{V}(t, \bar{S}_t)}{\partial \bar{S}^2} &= N'(d_1(t))\, \frac{\partial d_1(t)}{\partial \bar{S}} \\ &= N'(d_1(t))\, \bar{S}_t^{-1} \left( \int_t^T \sigma_s^2\, ds \right)^{-\frac{1}{2}}. \end{aligned}$$

We also obtain the time derivative

$$\frac{\partial \bar{V}(t, \bar{S}_t)}{\partial t} = \frac{1}{2} \sigma_t^2 \left( \int_t^T \sigma_s^2 \, ds \right)^{-\frac{1}{2}} \left( \bar{S}_t \, N'(d_1(t)) \right.$$
$$\times \left[ \ln \left( \frac{\bar{S}_t \, B_T}{K} \right) \left( \int_t^T \sigma_s^2 \, ds \right)^{-1} - \frac{1}{2} \right]$$
$$\left. - \frac{K}{B_T} \, N'(d_2(t)) \left[ \ln \left( \frac{\bar{S}_t \, B_T}{K} \right) \left( \int_t^T \sigma_s^2 \, ds \right)^{-1} + \frac{1}{2} \right] \right)$$

$$= \frac{1}{2} \sigma_t^2 \left( \int_t^T \sigma_s^2 \, ds \right)^{-\frac{1}{2}} N'(d_1(t)) \left( \ln \left( \frac{\bar{S}_t \, B_T}{K} \right) \left( \int_t^T \sigma_s^2 \, ds \right)^{-1} \right.$$
$$\times \left[ \bar{S}_t - \frac{K}{B_T} \exp \left\{ d_1(t) \left( \int_t^T \sigma_s^2 \, ds \right)^{\frac{1}{2}} - \frac{1}{2} \int_t^T \sigma_s^2 \, ds \right\} \right]$$
$$\left. - \frac{1}{2} \left[ \bar{S}_t - \frac{K}{B_T} \exp \left\{ d_1(t) \left( \int_t^T \sigma_s^2 \, ds \right)^{\frac{1}{2}} - \frac{1}{2} \int_t^T \sigma_s^2 \, ds \right\} \right] \right)$$
$$= -\frac{1}{2} \sigma_t^2 \left( \int_t^T \sigma_s^2 \, ds \right)^{-\frac{1}{2}} N'(d_1(t)) \, \bar{S}_t.$$

We note that

$$\frac{\partial \bar{V}(t, \bar{S}_t)}{\partial t} + \frac{1}{2} \sigma_t^2 \, \bar{S}_t^2 \, \frac{\partial^2 \bar{V}(t, \bar{S}_t)}{\partial \bar{S}^2} = 0$$

and also that

$$\bar{V}(T, \bar{S}_T) = \left( \bar{S}_T - \frac{K}{B_T} \right)^+ .$$

This shows that the discounted European call option price (8.3.2) satisfies the discounted BS-PDE (8.2.21) with terminal condition (8.2.22).

**8.2** The hedge ratio is given by the expression

$$\frac{\partial V(t, S_t)}{\partial S} = \left( \frac{\partial}{\partial \bar{S}_t} \, V(t, S_t) \right) \frac{\partial \bar{S}_t}{\partial S_t}$$
$$= \left( \frac{\partial}{\partial \bar{S}_t} \left( \bar{V}(t, \bar{S}_t) \, B_t \right) \right) \frac{1}{B_t}$$
$$= \frac{\partial}{\partial \bar{S}_t} \bar{V}(t, \bar{S}_t).$$

Then it follows from our calculations in Exercise 8.1 that

$$\frac{\partial V(t,S_t)}{\partial S} = N(d_1(t))$$

which corresponds to (8.4.3).

**8.3** The gamma for the European put option is given by the expression

$$\begin{aligned}\frac{\partial^2 V(t,S_t)}{\partial S^2} &= \frac{\partial}{\partial S} N(d_1(t)) \\ &= N'(d_1(t))\, S_t^{-1} \left(\int_t^T \sigma_s^2\, ds\right)^{-\frac{1}{2}},\end{aligned}$$

see (8.4.5) and (8.5.5), which is the same gamma as for the European call option.

**8.4** We obtain from (8.2.4) for the European put option the number of units $\delta_t^0$ to be held at time $t$ in the savings account in the form

$$\delta_t^0 = \frac{V(t,S_t)}{B_t} - \delta_t^1 \frac{S_t}{B_t}.$$

Therefore, it follows from (8.5.3) and (8.5.4) that

$$\begin{aligned}\delta_t^0 &= \bar{S}_t\,(N(d_1(t)) - 1) - \frac{K}{B_T}\,(N(d_2(t)) - 1) - (N(d_1(t)) - 1)\,\bar{S}_t \\ &= \frac{K}{B_T}\,(1 - N(d_2(t))).\end{aligned}$$

**8.5** By using the notation $\bar{S} = \frac{S}{B_t}$ and $V(t,S) = \bar{V}(t,\bar{S})\,B_t$ we obtain with the partial derivatives $\frac{\partial S}{\partial \bar{S}} = B_t$

$$\frac{\partial \bar{V}(t,\bar{S})}{\partial \bar{S}} = \frac{1}{B_t}\,\frac{\partial V(t,S)}{\partial S}\,\frac{\partial S}{\partial \bar{S}} = \frac{\partial V(t,S)}{\partial S}$$

$$\frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} = \frac{\partial V^2(t,S)}{\partial S^2}\,\frac{\partial S}{\partial \bar{S}} = \frac{\partial^2 V(t,S)}{\partial S^2}\,B_t$$

and

$$\begin{aligned}\frac{\partial \bar{V}(t,\bar{S})}{\partial t} &= \frac{\partial}{\partial t}\left(\frac{1}{B_t}\,V(t,\bar{S}\,B_t)\right) \\ &= -r\,\frac{1}{B_t}\,V(t,S) + \frac{1}{B_t}\,\frac{\partial V(t,S)}{\partial t} + \frac{1}{B_t}\,\frac{\partial V(t,\bar{S}\,B_t)}{\partial S}\,\bar{S}\,B_t\end{aligned}$$

by (8.2.21) the PDE

$$\begin{aligned}0 &= \frac{\partial \bar{V}(t,\bar{S})}{\partial t} + \frac{1}{2}\,\sigma^2\,\bar{S}^2\,\frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} \\ &= \frac{1}{B_t}\left(-r\,V(t,S) + \frac{\partial V(t,S)}{\partial t} + \frac{\partial V(t,S)}{\partial S}\,S\,r + \frac{1}{2}\,\sigma^2\,\bar{S}^2\,B_t^2\,\frac{\partial^2 V(t,S)}{\partial S^2}\right),\end{aligned}$$

which proves (8.2.23).

**8.6** The discounted P&L process $\bar{C}$ for a European put option has according to (8.2.13) and (8.2.20) the form

$$\bar{C}_t = \bar{V}(t,\bar{S}_t) - \bar{V}(0,\bar{S}_0) - \int_0^t \frac{\partial \bar{V}(s,\bar{S}_s)}{\partial \bar{S}}\,d\bar{S}_s.$$

On the other hand, we have by the discounted BS-PDE for a European put option

$$d\bar{V}(t,\bar{S}_t) = \frac{\partial \bar{V}(t,\bar{S}_t)}{\partial \bar{S}}\,d\bar{S}_t$$

and it follows

$$\begin{aligned}d\bar{C}_t &= d\bar{V}(t,\bar{S}_t) - \frac{\partial \bar{V}(t,\bar{S}_t)}{\partial \bar{S}}\,d\bar{S}_t \\ &= 0.\end{aligned}$$

This means, the discounted P&L

$$\bar{C}_t = \bar{C}_0 = 0$$

equals the constant zero. Consequently, by (8.2.12) the P&L

$$C_t = \bar{C}_t\,B_t = 0$$

is zero for all $t \in [0,T]$.

**8.7** We consider the square root process $Y = \{Y_t,\, t \in [0,\infty)\}$ of dimension $\delta > 2$ satisfying the SDE

$$dY_t = \left(\frac{\delta}{4}\,c^2 + b\,Y_t\right)dt + c\,\sqrt{Y_t}\,dW_t$$

for $t \in [0,\infty)$, $Y_0 > 0$, $c > 0$ and $b < 0$. The Itô integral

$$M_t = c\int_0^t \sqrt{Y_s}\,dW_s$$

forms a martingale due to Lemma 5.2.2 (iii), since the square root process, as a transformed time changed squared Bessel process, has moments of any positive order. Consequently,

$$dE(Y_t) = \left(\frac{\delta}{4}\,c^2 + b\,E(Y_t)\right) dt$$

for $t \in [0,\infty)$ with $E(Y_0) > 0$. Therefore, we obtain

$$E(Y_t) = E(Y_0)\,\exp\{b\,t\} + \frac{\delta\,c^2}{4\,b}\,(\exp\{b\,t\} - 1).$$

**8.8** (*) Using the notation of Sect. 8.7 we have for $\delta > 2$, $\alpha \geq -\frac{\delta}{2}$ and $\varphi > \varphi(0)$ by (8.7.7) and (8.7.9) the $\alpha$th moment in the form

$$\begin{aligned}
E(X_\varphi^\alpha) &= \int_0^\infty y^\alpha\, p_\delta(\varphi(0), x; \varphi, y)\, dy \\
&= \int_0^\infty y^\alpha\, \frac{1}{2\,(\varphi - \varphi(0))} \left(\frac{y}{x}\right)^{\frac{\delta}{4}-\frac{1}{2}} \exp\left\{-\frac{x+y}{2\,(\varphi - \varphi(0))}\right\} \\
&\qquad \times \sum_{k=0}^\infty \frac{\left(\frac{\sqrt{x\,y}}{2\,(\varphi-\varphi(0))}\right)^{2k+\frac{\delta}{2}-1}}{k!\,\Gamma\left(k+\frac{\delta}{2}\right)}\, dy \\
&= \sum_{k=0}^\infty \frac{\exp\left\{-\frac{x}{2\,(\varphi-\varphi(0))}\right\} x^k \left(\frac{1}{2\,(\varphi-\varphi(0))}\right)^{2k+\frac{\delta}{2}}}{k!\,\Gamma\left(k+\frac{\delta}{2}\right)} \int_0^\infty y^{\alpha+k+\frac{\delta}{2}-1} \\
&\qquad \times \exp\left\{-\frac{y}{2\,(\varphi-\varphi(0))}\right\} dy.
\end{aligned}$$

According to the gamma function (1.2.10) it holds for $\beta = \alpha + k + \frac{\delta}{2} > 0$ that

$$\int_0^\infty y^{\beta-1}\, \exp\left\{-\frac{y}{q}\right\} dy = \Gamma(\beta)\,(q)^\beta$$

and thus

$$\begin{aligned}
E(X_\varphi^\alpha) &= \sum_{k=0}^\infty \frac{\exp\left\{-\frac{x}{q}\right\} x^k \left(\frac{1}{q}\right)^{2k+\frac{\delta}{2}}}{k!\,\Gamma\left(k+\frac{\delta}{2}\right)}\,(q)^{\alpha+k+\frac{\delta}{2}}\,\Gamma\left(\alpha+k+\frac{\delta}{2}\right) \\
&= (q)^\alpha\, \exp\left\{-\frac{x}{q}\right\} \sum_{k=0}^\infty \left(\frac{x}{q}\right)^k \frac{\Gamma\left(\alpha+k+\frac{\delta}{2}\right)}{k!\,\Gamma\left(k+\frac{\delta}{2}\right)}
\end{aligned}$$

with $q = 2\,(\varphi - \varphi(0))$, which shows the first equation in (8.7.16). For $k \geq 1$ and $\alpha \leq 0$ we have from the properties of the gamma function that $\Gamma\left(\alpha+k+\frac{\delta}{2}\right) \leq \Gamma\left(k+\frac{\delta}{2}\right)$ and thus the estimate

$$E(X_\varphi^\alpha) = (2\,(\varphi-\varphi(0)))^\alpha\, \exp\left\{-\frac{x}{q}\right\}\left(\frac{\Gamma\left(\alpha+\frac{\delta}{2}\right)}{\Gamma\left(\frac{\delta}{2}\right)} + \exp\left\{\frac{x}{q}\right\}\right) < \infty,$$

which provides also the second part of (8.7.16).

**8.9** (*) Using (8.7.9) it follows

$$\int_0^\infty y^{1-\frac{n}{2}}\, p_n(\varphi(0),x;\varphi,y)\,dy$$

$$= \int_0^\infty \frac{y^{1-\frac{n}{2}}}{2(\varphi-\varphi(0))} \left(\frac{y}{x}\right)^{\frac{n}{4}-\frac{1}{2}} \exp\left\{-\frac{x+y}{2(\varphi-\varphi(0))}\right\} I_{\frac{n}{2}-1}\left(\frac{\sqrt{x\,y}}{\varphi-\varphi(0)}\right) dy$$

$$= \int_0^\infty \frac{x^{1-\frac{n}{2}}}{2(\varphi-\varphi(0))} \left(\frac{y}{x}\right)^{\frac{1}{2}-\frac{n}{4}} \exp\left\{-\frac{x+y}{2(\varphi-\varphi(0))}\right\} I_{\frac{n}{2}-1}\left(\frac{\sqrt{x\,y}}{\varphi-\varphi(0)}\right) dy$$

$$= x^{1-\frac{n}{2}} \int_0^\infty p_{4-n}(\varphi(0),y;\varphi,x)\,dy.$$

**8.10** (*) Combining (8.7.9) and (8.7.19) we obtain

$$\int_0^\infty p_{4-n}(\varphi(0),y;\varphi,x)\,dy$$

$$= \int_0^\infty \frac{1}{2(\varphi-\varphi(0))} \left(\frac{y}{x}\right)^{\frac{1}{2}-\frac{n}{4}} \exp\left\{-\frac{x+y}{2(\varphi-\varphi(0))}\right\} I_{\frac{n}{2}-1}\left(\frac{\sqrt{x\,y}}{\varphi-\varphi(0)}\right) dy$$

$$= \int_0^\infty \frac{1}{2(\varphi-\varphi(0))} \left(\frac{y}{x}\right)^{\frac{1}{2}-\frac{n}{4}} \exp\left\{-\frac{x+y}{2(\varphi-\varphi(0))}\right\} \sum_{k=0}^\infty \frac{\left(\frac{\sqrt{x\,y}}{2(\varphi-\varphi(0))}\right)^{2k+\frac{n}{2}-1}}{k\,!\,\Gamma\left(\frac{n}{2}+k\right)}\, dy$$

$$= \sum_{k=0}^\infty \frac{x^{\frac{n}{2}-1+k} \exp\left\{-\frac{x}{2(\varphi-\varphi(0))}\right\}}{(2(\varphi-\varphi(0)))^{\frac{n}{2}+2k}\, k\,!\,\Gamma\left(\frac{n}{2}+k\right)} \int_0^\infty y^k \exp\left\{-\frac{y}{2(\varphi-\varphi(0))}\right\} dy$$

$$= \sum_{k=0}^\infty \left(\frac{x}{2(\varphi-\varphi(0))}\right)^{\frac{n}{2}-1+k} \frac{\exp\left\{-\frac{x}{2(\varphi-\varphi(0))}\right\}}{\Gamma\left(\frac{n}{2}+k\right)}$$

$$= \left(\frac{x}{2(\varphi-\varphi(0))}\right)^{\frac{n}{2}-1} \exp\left\{-\frac{x}{2(\varphi-\varphi(0))}\right\} \sum_{k=0}^\infty \left(\frac{x}{2(\varphi-\varphi(0))}\right)^k \frac{1}{\Gamma\left(\frac{n}{2}+k\right)}.$$

Using the series expansion

$$\Gamma(a) - \Gamma(a,z) = e^{-z}\, z^a \sum_{k=0}^\infty \frac{\Gamma(a)}{\Gamma(a+1+k)}\, z^k,$$

see Abramowitz & Stegun (1972), with $a = \frac{n}{2} - 1$ and $z = \frac{x}{2(\varphi-\varphi(0))}$ the above equation becomes

$$\int_0^\infty p_{4-n}(\varphi(0),y;\varphi,x)\,dy = 1 - \frac{\Gamma\left(\frac{n}{2}-1, \frac{x}{2(\varphi-\varphi(0))}\right)}{\Gamma\left(\frac{n}{2}-1\right)}.$$

## Solutions for Exercises of Chapter 9

**9.1** By the SDE (9.4.14) it follows that the discounted stock price $\bar{S}_t$ satisfies the SDE

$$d\bar{S}_t = \sigma\, \bar{S}_t\, dW_{\theta\, t},$$

where $W_\theta$ is a standard Wiener process under the risk neutral measure $P_\theta$. Since this SDE is driftless $\bar{S}$ is an $(\underline{\mathcal{A}}, P_\theta)$-local martingale. Furthermore, because $\bar{S}$ is a geometric Brownian motion with bounded second moment, see (7.3.13)–(7.3.14), it follows that the diffusion coefficient $\sigma\bar{S}_t$ is square integrable for all $t \in [0, T]$. Consequently, by (5.4.1) $\sigma\bar{S}$ is from $\mathcal{L}_T^2$ and by the martingale property (5.4.3) of Itô integrals an $(\underline{\mathcal{A}}, P_\theta)$-martingale.

**9.2** By application of the Itô formula it follows by (8.3.2) and the discounted BS-PDE as in Exercise 9.1 that

$$d\bar{V}(t, \bar{S}_t) = \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}}\, \sigma\, \bar{S}_t\, (\theta_t\, dt + dW_t),$$

where the hedge ratio

$$\frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}} = \frac{\partial V(t, S_t)}{\partial S}$$

is by (8.4.3) bounded. With

$$dW_{\theta\, t} = \theta_t\, dt + dW_t$$

it follows by the Girsanov Theorem that $W_\theta$ is a $P_\theta$-Wiener process. Since $\bar{S} \in \mathcal{L}_T^2$ it follows by the martingale property (5.4.3) of Itô integrals from the above SDE that the $(\underline{\mathcal{A}}, P_\theta)$-local martingale $\bar{V}$ is an $(\underline{\mathcal{A}}, P_\theta)$-martingale.

**9.3** We obtain by the Itô formula (6.1.12) and the discounted BS-PDE (8.2.21) for the discounted put option price the SDE

$$\begin{aligned} d\left(\frac{p_{T,K}(t, S_t)}{B_t}\right) &= d\bar{V}(t, \bar{S}_t) \\ &= \left(\frac{\partial}{\partial t}\, \bar{V}(t, \bar{S}_t) + \frac{1}{2}\, \sigma^2\, \bar{S}_t^2\, \frac{\partial^2 \bar{V}(t, \bar{S}_t)}{\partial \bar{S}^2}\right) dt + \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}}\, d\bar{S}_t \\ &= \frac{\partial \bar{V}(t, \bar{S}_t)}{\partial \bar{S}}\, d\bar{S}_t \end{aligned}$$

for $t \in [0, T]$. Therefore, by the Itô formula and (8.2.1)–(8.2.2) we obtain

$$\begin{aligned}
dp_{T,K}(t,S_t) &= d(\bar{V}(t,\bar{S}_t)\,B_t)\\
&= p_{T,K}(t,S_t)\,r\,dt + B_t\,d\bar{V}(t,\bar{S}_t)\\
&= p_{T,K}(t,S_t)\,r\,dt + \left(\frac{\partial \bar{V}(t,\bar{S}_t)}{\partial \bar{S}}\right) B_t\,d\bar{S}_t\\
&= p_{T,K}(t,S_t)\,r\,dt + \left(\frac{\partial p_{T,K}(t,S_t)/B_t}{\partial S}\right)\frac{\partial S_t}{\partial \bar{S}}\,B_t\,d\bar{S}_t\\
&= p_{T,K}(t,S_t)\,r\,dt + \frac{\partial p_{T,K}(t,S_t)}{\partial S}\,B_t\,[(a-r)\,\bar{S}_t\,dt + \sigma\,\bar{S}_t\,dW_t]\\
&= \left(p_{T,K}(t,S_t)\,r + \frac{\partial p_{T,K}(t,S_t)}{\partial S}\,(a-r)\,S_t\right) dt\\
&\quad + \frac{\partial p_{T,K}(t,S_t)}{\partial S}\,\sigma\,S_t\,dW_t.
\end{aligned}$$

We obtain with (9.4.1) and (9.1.16) for the real world dynamics of $p_{T,K}(t,S_t)$ the SDE

$$\begin{aligned}
dp_{T,K}(t,S_t) &= p_{T,K}(t,S_t)\,r dt + \frac{\partial p_{T,K}(t,S_t)}{\partial S}\,\sigma\,S_t\left(dW_t + \frac{a-r}{\sigma}\,dt\right)\\
&= r\,p_{T,K}(t,S_t)\,dt + \frac{\partial p_{T,K}(t,S_t)}{\partial S}\,\sigma\,S_t\,(dW_t + \theta\,dt).
\end{aligned}$$

Here $W$ is a standard Wiener process under $P$. Since under the risk neutral measure $P_\theta$ the value

$$W_\theta(t) = W_t + \theta\,t$$

forms an $(\underline{\mathcal{A}}, P)$-Wiener process we obtain directly the risk neutral SDE

$$dp_{T,K}(t,S_t) = r\,p_{T,K}(t,S_t)\,dt + \frac{\partial p_{T,K}(t,S_t)}{\partial S}\,\sigma\,S_t\,dW_\theta(t)$$

for $t \in [0,T]$.

**9.4** Using (9.4.3) we have

$$dS_t = r\,S_t\,dt + \sigma\,S_t\,(dW_t + \theta\,dt),$$

where $W$ is a Wiener process under the real world probability measure $P$. By Itô's formula combined with (9.1.15) the SDE for the benchmarked security

$$\hat{S}_t = \frac{S_t}{D_t}$$

is given by

$$d\hat{S}_t = (\sigma - \theta)\,\hat{S}_t\,dW_t.$$

This SDE is that of a driftless geometric Brownian motion, which has by (6.3.2) the explicit solution

$$\hat{S}_t = \hat{S}_0 \exp\left\{-\frac{1}{2}(\sigma-\theta)^2 t + (\sigma-\theta) W_t\right\}.$$

By the mean (7.3.13) and variance (7.3.14) of a geometric Brownian motion it follows that

$$E(((\sigma-\theta)\hat{S}_t)^2) < \infty$$

for $t \in [0,T]$ so that by (5.4.1) $(\sigma-\theta)\hat{S} \in \mathcal{L}_T^2$. Consequently, $\hat{S}$ is an $(\underline{\mathcal{A}}, P_\theta)$-martingale by the martingale property (5.4.3) of Itô integrals.

**9.5** From (9.4.13), (8.3.4) and (8.1.1) we obtain

$$\begin{aligned}
c_{T,K}(t,S) &= \int_{-\infty}^{\infty} \exp\{-r(T-t)\} \\
&\quad \left(S \exp\left\{\left(r - \frac{1}{2}\sigma^2\right)(T-t) + \sigma\sqrt{T-t}\,x\right\} - K\right)^+ N'(x)\,dx \\
&= \int_{-\infty}^{\infty} \left(S \exp\left\{-\frac{1}{2}\sigma^2 (T-t) + \sigma\sqrt{T-t}\,x\right\}\right. \\
&\qquad \left. - K \exp\{-r(T-t)\}\right)^+ N'(x)\,dx \\
&= \int_{-d_2(t)}^{\infty} \left(S \exp\left\{-\frac{1}{2}\sigma^2 (T-t) + \sigma\sqrt{T-t}\,x\right\}\right. \\
&\qquad \left. - K \exp\{-r(T-t)\}\right) N'(x)\,dx \\
&= \int_{-\infty}^{d_2(t)} \left(S \exp\left\{-\frac{1}{2}\sigma^2 (T-t) - \sigma\sqrt{T-t}\,x\right\}\right. \\
&\qquad \left. - K \exp\{-r(T-t)\}\right) N'(x)\,dx \\
&= S \int_{-\infty}^{d_2(t)} \exp\left\{-\frac{1}{2}\sigma^2 (T-t) - \sigma\sqrt{T-t}\,x\right\} \\
&\qquad \times \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{x^2}{2}\right\} dx - K \exp\{-r(T-t)\}\, N(d_2(t)).
\end{aligned}$$

With the change of variables

$$z = x + \sigma \sqrt{T-t}$$

and (8.3.3)–(8.3.4) we finally obtain

$$\begin{aligned} c_{T,K}(t,S) = S \int_{-\infty}^{d_1(t)} \frac{1}{\sqrt{2\pi}} \exp\left\{ -\frac{1}{2}\sigma^2 (T-t) - z\,\sigma \sqrt{T-t} + \sigma^2 (T-t) \right. \\ \left. - \frac{z^2}{2} + z\,\sigma \sqrt{T-t} - \frac{\sigma^2}{2}(T-t) \right\} dz \\ - K \exp\{-r(T-t)\}\, N(d_2(t)) \\ = S\, N(d_1(t)) - K \exp\{-r(T-t)\}\, N(d_2(t)), \end{aligned}$$

which proves the Black-Scholes European call option pricing formula.

**9.6** (*) By (9.4.8) the Radon-Nikodym derivative at time $t$ for the standard BS model equals the expression

$$\Lambda_\theta(t) = \frac{\hat{S}^0_t}{\hat{S}^0_0}$$

for $t \in [0,T]$. By (9.1.21) and (9.1.20) we obtain

$$\begin{aligned} d\Lambda_\theta(t) &= -\theta \frac{\hat{S}^0_t}{\hat{S}^0_0}\, dW_t \\ &= -\theta\, \Lambda_\theta(t)\, dW_t \end{aligned}$$

for $t \in [0,T]$, where $\Lambda_\theta(0) = 1$.

**9.7** (*) Under the BS model with savings account $B_t = \exp\{r\,t\}$ and risky security

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

we have the GOP in the form

$$S^{\delta_*}_t = S^{\delta_*}_0 \exp\left\{ r\,t + \frac{1}{2}\theta^2\, t + \theta\, W_t \right\}.$$

The fair zero coupon bond price $P(t,T)$ at time $t$, when $T$ is the maturity date, is obtained by the real world pricing formula and the Laplace transform for Gaussian random variables

$$
\begin{aligned}
P(t,T) &= S_t^{\delta_*}\, E\left(\frac{1}{S_T^{\delta_*}} \,\Big|\, \mathcal{A}_t\right) \\
&= E\left(\exp\left\{r\,(t-T) + \frac{1}{2}\,\theta^2\,(t-T) + \theta\,(W_t - W_T)\right\} \,\Big|\, \mathcal{A}_t\right) \\
&= \exp\{-r\,(T-t)\}\, E\left(\exp\left\{-\frac{1}{2}\,\theta^2\,(T-t) - \theta\,(W_T - W_t)\right\} \,\Big|\, \mathcal{A}_t\right) \\
&= \exp\{-r\,(T-t)\} = \exp\{-r\,T\}\, B_t.
\end{aligned}
$$

The benchmarked zero coupon bond price, when normalized to one at time zero, has the form

$$
\Lambda_{\theta_{P(\cdot,T)}}(t) = \frac{\hat{P}(t,T)}{\hat{P}(0,T)} = \frac{\exp\{-r\,T\}\,B_t}{S_t^{\delta_*}}\,\frac{S_0^{\delta_*}}{\exp\{-r\,T\}} = S_0^{\delta_*}\,\frac{B_t}{S_t^{\delta_*}}
$$

and satisfies the SDE

$$
d\Lambda_{\theta_{P(\cdot,T)}}(t) = \Lambda_{\theta_{P(\cdot,T)}}(t)\,(-\theta)\,dW_t
$$

with market price of risk $\theta = \frac{a-r}{\sigma}$. The Radon-Nikodym derivative for the zero coupon bond $P(\cdot,T)$ as numeraire is by (9.6.21)–(9.6.23) of the form

$$
\frac{dP_{\theta_{P(\cdot,T)}}}{dP} = \Lambda_{\theta_{P(\cdot,T)}}(T).
$$

The drifted Wiener process $W_{\theta_{P(\cdot,T)}}$ with

$$
dW_{\theta_{P(\cdot,T)}}(t) = dW_t + \theta_{P(\cdot,T)}(t)\,dt
$$

and

$$
\theta_{P(\cdot,T)}(t) = \theta
$$

is a Wiener process under the probability measure $P_{\theta_{P(\cdot,T)}}$. Therefore, the corresponding numeraire pair is $(P(\cdot,T), P_{\theta_{P(\cdot,T)}}) = (P(\cdot,T), P_\theta)$. The risk neutral measure $P_\theta$ equals here the, so-called, *T-forward measure* $P_{\theta_{P(\cdot,T)}}$ is an important observation under the BS model.

**9.8** (*) We obtain under the BS model the discounted price $\bar{V}(t,\bar{S}_t) = \frac{V(t,S_t)}{B_t}$ at time $t$ of the payoff $f(S_T) = (S_T)^2$ by the risk neutral pricing formula (9.6.10) in the form

$$
\bar{V}(t,\bar{S}_t) = E_\theta\left(\frac{(\bar{S}_T\,B_T)^2}{B_T} \,\Big|\, \mathcal{A}_t\right),
$$

where

$$
d\bar{S}_t = \bar{S}_t\,\sigma\,dW_{\theta\,t}
$$

under the risk neutral probability measure $P_\theta$ for $t \in [0,T]$. We have here only a terminal payoff at time $T$. When using the Feynman-Kac formula (9.7.3), then we obtain by (9.7.4) the PDE

$$\frac{\partial \bar{V}(t,\bar{S})}{\partial t} + \frac{1}{2}\,\sigma^2\,\bar{S}^2\,\frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} = 0$$

for $(t,\bar{S}) \in [0,T) \times (0,\infty)$ with terminal condition

$$\bar{V}(T,\bar{S}) = \frac{(\bar{S}_T)^2}{B_T}.$$

This PDE has, by using the explicit solution for geometric Brownian motion, the solution

$$\begin{aligned}\bar{V}(t,\bar{S}_t) &= \frac{1}{B_T}\,E_\theta\left((\bar{S}_t)^2\,\exp\left\{2\left[-\frac{1}{2}\,\sigma^2(T-t) + \sigma\,(W_{\theta T} - W_{\theta t})\right]\right\}\,\Big|\,\mathcal{A}_t\right)\\ &= \frac{(\bar{S}_t)^2}{B_T}\,\exp\{\sigma^2(T-t)\},\end{aligned}$$

since

$$\begin{aligned}\frac{\partial \bar{V}(t,\bar{S})}{\partial t} &= -\bar{V}(t,\bar{S})\,\sigma^2,\\ \frac{\partial \bar{V}(t,\bar{S})}{\partial \bar{S}} &= 2\,\frac{\bar{V}(t,\bar{S})}{\bar{S}},\\ \frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} &= -2\,\frac{\bar{V}(t,\bar{S})}{\bar{S}^2} + \frac{2}{\bar{S}}\,\frac{\partial \bar{V}(t,\bar{S})}{\partial \bar{S}} = \frac{2}{\bar{S}^2}\,\bar{V}(t,\bar{S}).\end{aligned}$$

Therefore we have

$$\begin{aligned}\frac{\partial \bar{V}(t,\bar{S})}{\partial t} + \frac{1}{2}\,\sigma^2\,\bar{S}^2\,\frac{\partial^2 \bar{V}(t,\bar{S})}{\partial \bar{S}^2} &= -\bar{V}(t,\bar{S})\,\sigma^2 + \frac{1}{2}\,\sigma^2\,\bar{S}^2\,\frac{2}{\bar{S}^2}\,\bar{V}(t,\bar{S})\\ &= 0.\end{aligned}$$

## Solutions for Exercises of Chapter 10

**10.1** The growth rate $g_t^\delta$ of a portfolio is defined in (10.2.1) as the drift of the SDE of the logarithm of the portfolio $S^\delta$. By application of the Itô formula to $\ln(S_t^\delta)$ one obtains the SDE

$$d\ln(S^\delta_t) = \frac{1}{S^\delta_t}\,dS^\delta_t - \frac{1}{2(S^\delta_t)^2}\,d[S^\delta_t]$$

$$= \left( r_t + \sum_{k=1}^d \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t\, \theta^k_t - \frac{1}{2} \sum_{k=1}^d \left( \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t\, \theta^k_t \right)^2 \right) dt$$

$$+ \sum_{k=1}^d \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t\, dW^k_t.$$

The drift of this SDE, which is the growth rate of $S^\delta$, is then

$$g^\delta_t = r_t + \sum_{k=1}^d \left( \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t\, \theta^k_t - \frac{1}{2} \left( \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t\, \theta^k_t \right)^2 \right).$$

**10.2** We apply for the benchmarked value $\hat{S}^\delta_t = \frac{S^\delta_t}{S^{\delta_*}_t}$ the integration by parts formula (6.3.1) and obtain the SDE

$$d\hat{S}^\delta_t = S^\delta_t\, d\left(\frac{1}{S^{\delta_*}_t}\right) + \frac{1}{S^{\delta_*}_t}\, dS^\delta_t + d\left[\frac{1}{S^{\delta_*}}, S^\delta\right]_t$$

$$= \frac{S^\delta_t}{S^{\delta_*}_t}\left(-r_t\,dt - \sum_{k=1}^d \theta^k_t\, dW^k_t\right) + \frac{S^\delta_t}{S^{\delta_*}_t}\left(r_t\,dt + \sum_{k=1}^d \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t \left(\theta^k_t\,dt + dW^k_t\right)\right)$$

$$+ \frac{S^\delta_t}{S^{\delta_*}_t} \sum_{k=1}^d \left( -\theta^k_t \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t \right) dt$$

$$= \hat{S}^\delta_t \sum_{k=1}^d \left( \sum_{j=1}^d \pi^j_{\delta,t}\, b^{j,k}_t - \theta^k_t \right) dW^k_t.$$

This SDE is driftless. Therefore, by Lemma 5.4.1 a square integrable $\hat{S}^\delta$ ensures that $\hat{S}^\delta$ is an $(\underline{\mathcal{A}}, P)$-local martingale. However, this is not sufficient to guarantee that $\hat{S}^\delta$ is, in general, an $(\underline{\mathcal{A}}, P)$-martingale. A counter example is the unfair portfolio (9.1.42) in Sect. 9.1. Since $\hat{S}^\delta$ is a nonnegative local martingale it is by Lemma 5.2.3 an $(\underline{\mathcal{A}}, P)$-supermartingale.

**10.3** According to Definition 10.6.3 and (10.6.19) we need to show that

$$E\left(\left(\hat{\sigma}^k_{(d)}(t)\right)^2\right) = E\left(\left(\sum_{j=0}^d \left|\sigma^{j,k}_{(d)}(t)\right|\right)^2\right)$$

is bounded by a constant for all $k \in \mathcal{N}$. Due to (10.6.27) we have for $d \in \mathcal{N}$ and $k \in \{1, 2, \dots, d\}$

$$\sum_{j=0}^{d} \left| \sigma_{(d)}^{j,k}(t) \right| \leq \sigma \left( 1 + \frac{1}{\sqrt{d}} \right) \leq 2\,\sigma$$

and, therefore, $E\left( \left( \hat{\sigma}_{(d)}^{k}(t) \right)^2 \right) \leq 4\,\sigma^2$. This demonstrates that the corresponding sequence of CFMs is regular.

**Solutions for Exercises of Chapter 11**

**11.1** The expected log-utility $v^{\tilde{\delta}}$ follows by (11.3.3), (11.3.8), (11.3.11) and (10.2.8), as

$$\begin{aligned} v^{\tilde{\delta}} &= E\left( U\left( U'^{-1}\left( \frac{\lambda}{\bar{S}_T^{\delta_*}} \right) \right) \,\middle|\, \mathcal{A}_0 \right) = E\left( \ln\left( \bar{S}_T^{\delta_*} \right) - \ln(\lambda) \,\middle|\, \mathcal{A}_0 \right) \\ &= E\left( \ln\left( \bar{S}_T^{\delta_*} \right) - \ln\left( \bar{S}_0^{\delta_*} \right) + \ln(S_0) \,\middle|\, \mathcal{A}_0 \right) \\ &= \frac{1}{2}\, E\left( \int_0^T |\theta(s, \bar{S}_s^{\delta_*})|^2 \, ds \,\middle|\, \mathcal{A}_0 \right) + \ln(S_0) \\ &= \frac{1}{2} \int_0^T E\left( |\theta(s, \bar{S}_s^{\delta_*})|^2 \,\middle|\, \mathcal{A}_0 \right) ds + \ln(S_0). \end{aligned} \tag{S.1}$$

For the BS model with $\theta(s, \bar{S}_s^{\delta_*}) = \theta$ we obtain, therefore, $v^{\tilde{\delta}} = \frac{\theta^2}{2} T + \ln(S_0)$. In the case of other discounted GOP dynamics one has simply to calculate the conditional expectation in (S.1), which is possible for certain models.

**11.2** Similarly as in the above exercise the expected power utility for $\gamma < 0$ under the BS model is obtained by (11.3.3), (11.3.8), (11.3.16) and (10.2.8) as

$$v^{\tilde{\delta}} = E\left( \frac{1}{\gamma} \left( \left( \lambda\, \hat{S}_T^0 \right)^{\frac{1}{\gamma-1}} \hat{S}_T^0 \right)^{\gamma} \,\middle|\, \mathcal{A}_0 \right) = \frac{1}{\gamma} \left( S^0 \right)^{\gamma} \left( S_0^{\delta_*} \right)^{-\gamma} \exp\left\{ \frac{\theta^2}{2}\, T \, \frac{\gamma}{1-\gamma} \right\}.$$

**11.3** (*) The benchmarked fair price $\hat{V}_t$ at time $t \in [0, T]$ of the payoff $H$ paid at time $T \in (0, \infty)$ satisfies according to (11.5.9) the SDE

$$d\hat{V}_t = \sum_{k=1}^{m} x_H^k(t)\, dW_t^k \tag{S.2}$$

with

$$\hat{V}_0 = E\left(\frac{H}{S_T^{\delta_*}} \,\bigg|\, \mathcal{A}_0\right). \tag{S.3}$$

On the other hand, according to (10.2.8) the discounted GOP is characterized by the SDE

$$d\bar{S}_t^{\delta_*} = \bar{S}_t^{\delta_*} \sum_{k=1}^{d} \theta_t^k \,(\theta_t^k\, dt + dW_t^k) \tag{S.4}$$

for $t \in [0, T]$ with $\bar{S}_0^{\delta_*} > 0$.

The discounted payoff can now be expressed as

$$\bar{H} = \frac{H}{S_T^0} = \hat{H}\, \bar{S}_T^{\delta_*}. \tag{S.5}$$

By application of the Itô formula to the product $\bar{V}_t = \hat{V}_t \bar{S}_t^{\delta_*}$ we obtain by the Itô formula with (S.2) and (S.3) the SDE

$$\begin{aligned}
d\bar{V}_t &= \hat{V}_t\, d\bar{S}_t^{\delta_*} + \bar{S}_t^{\delta_*}\, d\hat{V}_t + d\left[\hat{V}, \bar{S}^{\delta_*}\right]_t \\
&= \bar{V}_t \sum_{k=1}^{d} \theta_t^k\,(\theta_t^k\, dt + dW_t^k) + \bar{S}_t^{\delta_*} \sum_{k=1}^{m} x_H^k(t)\, dW_t^k + \bar{S}_t^{\delta_*} \sum_{k=1}^{d} x_H^k(t)\, \theta_t^k\, dt \\
&= \bar{S}_t^{\delta_*} \sum_{k=1}^{d} \left(x_H^k(t) + \hat{V}_t\, \theta_t^k\right)(\theta_t^k\, dt + dW_t^k) + \bar{S}_t^{\delta_*} \sum_{k=d+1}^{m} x_H^k(t)\, dW_t^k.
\end{aligned}$$

This leads under a risk neutral probability measure to the martingale representation

$$\bar{H} = \bar{V}_0 + \sum_{k=1}^{d} \int_0^T \bar{S}_t^{\delta_*} \left(x_H^k(t) + \hat{V}_t\, \theta_t^k\right) dW_\theta{}^k(t) + \sum_{k=d+1}^{m} \int_0^T \bar{S}_t^{\delta_*}\, x_H^k(t)\, dW_t^k.$$

Here $\bar{V}_t = E_\theta(\bar{H}|\mathcal{A}_t)$ with $E_\theta$ denoting expectation under $P_\theta$ and

$$W_\theta{}^k(t) = \int_0^t \theta_t^k\, dt + W_t^k$$

for $k \in \{1, 2, \ldots, d\}$ forms a Wiener process under $P_\theta$.

## Solutions for Exercises of Chapter 12

**12.1** Using the time homogenous Fokker-Planck equation the stationary density of the ARCH diffusion model is of the form

$$\bar{p}(\theta^2) = \frac{C}{\gamma^2\,\theta^4}\exp\left\{2\int_{\underline{\theta}^2}^{\theta^2}\frac{\kappa\,(\bar{\theta}^2-u)}{\gamma^2\,u^2}\,du\right\}$$
$$= \frac{C}{\gamma^2\,\theta^4}\exp\left\{\frac{2\,\kappa\,\bar{\theta}^2}{\gamma^2}\int_{\underline{\theta}^2}^{\theta^2}\frac{1}{u^2}\,du - \frac{2\,\kappa}{\gamma^2}\int_{\underline{\theta}^2}^{\theta^2}\frac{1}{u}\,du\right\}$$
$$= \frac{C}{\gamma^2\,\theta^4}\exp\left\{\frac{2\,\kappa}{\gamma^2}\left(\bar{\theta}^2\left(-\frac{1}{\theta^2}+\frac{1}{\underline{\theta}^2}\right) - \left(\ln(\theta^2)-\ln(\underline{\theta}^2)\right)\right)\right\}$$
$$= C_1\,\exp\left\{-\frac{2\,\kappa\,\bar{\theta}^2}{\gamma^2}\,\frac{1}{\theta^2}\right\}\left(\frac{1}{\theta^2}\right)^{\frac{2\kappa}{\gamma^2}+2},$$

which is an inverse gamma density with an appropriate constant $C_1 > 0$.

**12.2** The stationary density for the squared volatility needs to satisfy the expression

$$\bar{p}(\theta^2) = \frac{C}{\gamma^2\,\theta^6}\exp\left\{2\int_{\underline{\theta}^2}^{\theta^2}\frac{\kappa\,(\bar{\theta}^2-u)\,u}{\gamma^2\,u^3}\,du\right\}$$
$$= \frac{C}{\gamma^2\,\theta^6}\exp\left\{\frac{2\,\kappa}{\gamma^2}\left(\bar{\theta}^2\int_{\underline{\theta}^2}^{\theta^2}\frac{1}{u^2}\,du - \int_{\underline{\theta}^2}^{\theta^2}\frac{1}{u}\,du\right)\right\}$$
$$= \frac{C}{\gamma^2\,\theta^6}\exp\left\{\frac{2\,\kappa}{\gamma^2}\left(-\bar{\theta}^2\left(\frac{1}{\theta^2}-\frac{1}{\bar{\theta}^2}\right) - \left(\ln(\theta^2)-\ln(\underline{\theta}^2)\right)\right)\right\}$$
$$= \frac{C}{\gamma^2\,\theta^6}\exp\left\{\frac{2\,\kappa}{\gamma^2}\left(-\bar{\theta}^2\left(\frac{1}{\theta^2}-\frac{1}{\underline{\theta}^2}\right) - \ln(\theta^2)+\ln(\underline{\theta}^2)\right)\right\}$$
$$= C_1\,\exp\left\{-\frac{2\,\kappa}{\gamma^2}\,\bar{\theta}^2\,\frac{1}{\theta^2}\right\}\left(\frac{1}{\theta^2}\right)^{\frac{2\kappa}{\gamma^2}+3},$$

which is an inverse gamma density.

**12.3** For the Heston model we obtain the stationary density for the squared volatility

$$\bar{p}(\theta^2) = \frac{C}{\gamma^2\,\theta^2}\exp\left\{2\int_{\underline{\theta}^2}^{\theta^2}\frac{\kappa\,(\bar{\theta}^2-u)}{\gamma^2\,u}\,du\right\}$$
$$= \frac{C}{\gamma^2\,\theta^2}\exp\left\{\frac{2\,\kappa\,\bar{\theta}^2}{\gamma^2}\int_{\underline{\theta}^2}^{\theta^2}\frac{1}{u}\,du - \frac{2\,\kappa}{\gamma^2}\,(\theta^2-\underline{\theta}^2)\right\}$$
$$= C_1\,\exp\left\{-\frac{2\,\kappa}{\gamma^2}\,\bar{\theta}^2\right\}(\theta^2)^{\frac{2\kappa\,\bar{\theta}^2}{\gamma^2}-1},$$

which is a gamma density.

**12.4** For the Scott model the volatility $\theta_t$ has the stationary density

$$\bar{p}(\theta) = \frac{C}{\gamma^2} \exp\left\{ 2 \int_{\underline{\theta}}^{\theta} \frac{\kappa\,(\bar{\theta} - u)}{\gamma^2}\, du \right\}$$

$$= \frac{C}{\gamma^2} \exp\left\{ \frac{2\,\kappa}{\gamma^2} \left( \bar{\theta}\,(\theta - \underline{\theta}) - \frac{1}{2}(\theta^2 - \underline{\theta}^2) \right) \right\},$$

which is a Gaussian density with mean $\bar{\theta}$ and variance $\frac{\gamma^2}{2\,\kappa}$. Thus $\theta_t^2$ has a chi-square distribution with two degrees of freedom.

**12.5** It follows by the Itô formula that

$$d\theta_t^2 = \theta_t^2 \left( \kappa\,\bar{\xi} + \frac{1}{2}\,\gamma^2 - \theta_t^2\,\kappa \right) dt + \theta_t^2\,\gamma\,dW_t.$$

Therefore, the stationary density satisfies the expression

$$\bar{p}(\theta^2) = \frac{C}{\gamma^2\,\theta^4} \exp\left\{ 2 \int_{\underline{\theta}^2}^{\theta^2} \frac{u\left(\kappa\,\bar{\xi} + \frac{1}{2}\,\gamma^2 - u\,\kappa\right)}{\gamma^2\,u^2}\, du \right\}$$

$$= \frac{C}{\gamma^2\,\theta^4} \exp\left\{ \frac{2}{\gamma^2} \left( \kappa\,\bar{\xi} + \frac{1}{2}\,\gamma^2 \right) \left( \ln(\theta^2) - \ln(\underline{\theta}^2) \right) - \frac{2\,\kappa}{\gamma^2}\,(\theta^2 - \underline{\theta}^2) \right\}$$

$$= C_1 \exp\left\{ -\frac{2\,\kappa}{\gamma^2}\,\theta^2 \right\} (\theta^2)^{\frac{2}{\gamma^2}\,(\kappa\,\bar{\xi} + \frac{1}{2}\,\gamma^2) - 2}.$$

It follows that the stationary density of the squared volatility is a gamma density.

**12.6** (*) It follows by (12.2.8) and the Itô formula

$$dX_t^{-q} = -q\,X_t^{-(q+1)} \Big( (2\,(1-a)\,r\,X_t + \psi^2\,(1-a)\,(3-2a))\,dt$$
$$+ 2\,\psi\,(1-a)\,\sqrt{X_t}\,dW_t \Big)$$
$$+ \frac{1}{2}\,q\,(q+1)\,X^{-(q+2)}\,4\,\psi^2\,(1-a)^2\,X_t\,dt$$

for $t \in [0, T]$, where $X$ is a transformed squared Bessel process of dimension $\nu = \frac{3-2a}{1-a}$, see (12.2.9). Therefore, the benchmarked savings account satisfies by the Itô formula and (12.2.1) the SDE

$$\begin{aligned}
d\hat{S}_t^0 &= d\left(\frac{S_t^0}{S_t^{\delta_*}}\right) = d\left(\exp\{-r\,(\tau - t)\}\,X_t^{-q}\right)\\
&= S_t^0\,\Big(-q\,X_t^{-q}\Big[2\,(1-a)\,r\\
&\qquad + X_t^{-1}\,\psi^2\,\left((1-a)\,(3-2a) - \frac{q+1}{2}\,4\,(1-a)^2\right) - \frac{r}{9}\Big]\,dt\\
&\qquad - q\,X^{-q-\frac{1}{2}}\,2\,\psi\,(1-a)\,dW_t\Big)
\end{aligned}$$

By noting that according to (12.2.11) one has

$$q = \frac{1}{2\,(1-a)}$$

we obtain

$$d\hat{S}_t^0 = -\hat{S}_t^0\,\frac{\psi}{\sqrt{X_t}}\,dW_t.$$

Consequently, $\hat{S}^0$ is an $(\underline{\mathcal{A}}, P)$-local martingale. We have by the moments of Bessel processes (8.7.16) the finite expression

$$E\left(\left(\hat{S}_t^0\right)^2\right) = \exp\{-2\,r\,(\tau - t)\}\,E(X_t^{-2q}) < \infty$$

for $a < 1$ and $t \in [0, \tau]$. Thus, for $a < 1$ the process $\hat{S}^0$ is square integrable. Furthermore, the quadratic variation of $\hat{S}^0$ is

$$[\hat{S}^0]_t = \exp\{-2\,r\,(\tau - t)\}\,\psi^2\int_0^t X_s^{-2(q+\frac{1}{2})}\,ds$$

and its expectation yields by (8.7.14) because of $\alpha = -2q - 1 = -\frac{\nu}{2}$ an infinite value

$$E\left([\hat{S}^0]_t\right) = \exp\{-2\,r\,(\tau - t)\}\,\psi^2\int_0^t E\left(X_s^{-\frac{2-a}{1-a}}\right)\,ds = \infty.$$

By (8.7.23) $\hat{S}^0$ is a strict local martingale and so is $\hat{P}_\tau^*$.

## Solutions for Exercises of Chapter 13

**13.1** The SDE for the discounted GOP is of the form

$$d\bar{S}_t^{\delta_*} = \alpha_t^{\delta_*}\,dt + \sqrt{\bar{S}_t^{\delta_*}\,\alpha_t^{\delta_*}}\,dW_t.$$

By the Itô formula we obtain for $\ln(\bar{S}_t^{\delta_*})$ the SDE

$$d \ln\left(\bar{S}_t^{\delta_*}\right) = \frac{1}{2}\,\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}\,dt + \sqrt{\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}}\,dW_t,$$

which shows that the volatility of the discounted GOP equals $\sqrt{\frac{\alpha_t^{\delta_*}}{\bar{S}_t^{\delta_*}}}$.

**13.2** By the Itô formula we obtain for $\sqrt{\bar{S}_t^{\delta_*}}$ the SDE

$$d\sqrt{\bar{S}_t^{\delta_*}} = \left(\frac{\alpha_t^{\delta_*}}{2\sqrt{\bar{S}_t^{\delta_*}}} - \frac{1}{2}\,\frac{1}{4}\,\frac{1}{\left(\bar{S}_t^{\delta_*}\right)^{\frac{3}{2}}}\,\bar{S}_t^{\delta_*}\,\alpha_t^{\delta_*}\right) dt + \frac{\sqrt{\bar{S}_t^{\delta_*}\,\alpha_t^{\delta_*}}}{2\sqrt{\bar{S}_t^{\delta_*}}}\,dW_t$$

$$= \frac{3}{8}\,\frac{\alpha_t^{\delta_*}}{\sqrt{\bar{S}_t^{\delta_*}}}\,dt + \frac{1}{2}\sqrt{\alpha_t^{\delta_*}}\,dW_t,$$

which confirms (13.1.12).

**13.3** The differential equation for $\alpha_t^{\delta_*}$ is of the form

$$d\alpha_t^{\delta_*} = \eta_t\,\alpha_t^{\delta_*}\,dt.$$

Together with the SDE of the discounted GOP and the Itô formula it follows

$$dY_t = Y_t\left(\frac{\alpha_t}{\bar{S}_t^{\delta_*}} - \eta_t\right) dt + Y_t\sqrt{\frac{\alpha_t}{\bar{S}_t^{\delta_*}}}\,dW_t$$

$$= (1 - \eta_t\,Y_t)\,dt + \sqrt{Y_t}\,dW_t,$$

which confirms (13.2.5).

**13.4** The squared volatility of the discounted GOP equals $|\theta_t|^2 = \frac{1}{Y_t}$, and, thus the inverse of the normalized GOP. This means, we obtain by the Itô formula the SDE

$$d|\theta_t|^2 = d\left(\frac{1}{Y_t}\right) = \left(-\left(\frac{1}{Y_t}\right)^2 (1 - \eta_t\,Y_t) + \frac{Y_t}{(Y_t)^3}\right) dt - \left(\frac{1}{Y_t}\right)^2 \sqrt{Y_t}\,dW_t$$

$$= \eta_t\,\frac{1}{Y_t}\,dt - \left(\frac{1}{Y_t}\right)^{\frac{3}{2}} dW_t = \eta_t\,|\theta_t|^2\,dt - \left(|\theta_t|^2\right)^{\frac{3}{2}}\,dW_t,$$

which confirms (13.2.11).

## Solutions for Exercises of Chapter 14

**14.1** The SDE for a strictly positive portfolio $S^\delta$ is by (14.1.3) of the form

$$dS_t^\delta = S_{t-}^\delta \left( r_t\, dt + \sum_{k=1}^{d} \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k} \left( \theta_t^k\, dt + dW_t^k \right) \right).$$

By application of the Itô formula this leads for the logarithm of $S_t^\delta$ to the SDE

$$\begin{aligned}
d\ln(S_t^\delta) &= r_t\, dt + \sum_{k=1}^{d} \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\, \theta_t^k\, dt - \frac{1}{2} \sum_{k=1}^{m} \left( \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k} \right)^2 dt \\
&\quad + \sum_{k=1}^{m} \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\, dW_t^k - \sum_{k=m+1}^{d} \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k} \sqrt{h_t^k}\, dt \\
&\quad + \sum_{k=m+1}^{d} \ln \left( 1 + \sum_{j=1}^{d} \pi_{\delta,t-}^j \frac{b_t^{j,k}}{\sqrt{h_t^k}} \right) dp_t^k \\
&= \left( r_t\, dt + \sum_{k=1}^{d} \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k}\, \theta_t^k - \frac{1}{2} \sum_{k=1}^{m} \left( \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k} \right)^2 \right. \\
&\quad \left. + \sum_{k=m+1}^{d} h_t^k \left[ \ln \left( 1 + \sum_{j=1}^{d} \pi_{\delta,t}^j \frac{b_t^{j,k}}{\sqrt{h_t^k}} \right) - \sum_{j=1}^{d} \pi_{\delta,t}^j\, b_t^{j,k} \frac{1}{\sqrt{h_t^k}} \right] \right) dt \\
&\quad + \sum_{k=1}^{m} \pi_{\delta,t}^j\, b_t^{j,k}\, dW_t^k + \sum_{k=m+1}^{d} \ln \left( 1 + \sum_{j=1}^{d} \pi_{\delta,t-}^j \frac{b_t^{j,k}}{\sqrt{h_t^k}} \right) \sqrt{h_t^k}\, dW_t^k.
\end{aligned}$$

The drift of this SDE is the growth rate given in (14.1.15).

**14.2** The forward rate at time $t$ for maturity $T$ has by (14.1.32) and (14.1.31) the form

$$
\begin{aligned}
f(t,T) &= -\frac{\partial}{\partial T}\,\ln(\hat{P}(t,T)) \\
&= -\frac{\partial}{\partial T}\Bigg[\ln(\hat{P}(0,T)) - \sum_{k=1}^{m}\left(\int_0^t \sigma^k(s,T)\,dW_s^k + \frac{1}{2}\int_0^t (\sigma^k(s,T))^2\,ds\right) \\
&\quad + \sum_{k=m+1}^{d}\left(\int_0^t \sigma^k(s,T)\sqrt{h_s^{k-m}}\,ds + \int_0^t \ln\left(1-\frac{\sigma^k(s,T)}{\sqrt{h_s^{k-m}}}\right)dp_s^k\right)\Bigg] \\
&= f(0,T) + \sum_{k=1}^{m}\left(\int_0^t \frac{\partial}{\partial T}\,\sigma^k(s,T)\,dW_s^k + \frac{1}{2}\int_0^t \frac{\partial}{\partial T}\,(\sigma^k(s,T))^2\,ds\right) \\
&\quad + \sum_{k=m+1}^{d}\left(-\int_0^t \frac{\partial}{\partial T}\,\sigma^k(s,T)\sqrt{h_s^{k-m}}\,ds - \int_0^t \frac{\partial}{\partial T}\,\ln\left(1-\frac{\sigma^k(s,T)}{\sqrt{h_s^{k-m}}}\right)dp_s^k\right) \\
&= f(0,T) + \sum_{k=1}^{m}\int_0^t \frac{\partial}{\partial T}\,\sigma^k(s,T)\,(\sigma^k(s,T)\,ds + dW_s^k) \\
&\quad + \sum_{k=m+1}^{d}\int_0^t \frac{1}{1-\frac{\sigma^k(s,T)}{\sqrt{h_s^{k-m}}}}\,\frac{\partial}{\partial T}\,\sigma^k(s,T)\,(\sigma^k(s,T)\,ds + dW_s^k).
\end{aligned}
$$

**14.3** (*) (Hardy Hulley) Fix $i,j \in \{0,1,\ldots,d\}$ such that $i \neq j$, then the function

$$
\begin{aligned}
p_{s,t}^{i,j}(x_i,x_j;y_i,y_j) &= \frac{1}{2\pi y_i y_j |\sigma^i||\sigma^j|(t-s)\sqrt{1-(\varrho^{i,j})^2}} \\
&\quad\times \exp\Bigg\{-\frac{1}{2(1-(\varrho^{i,j})^2)}\Bigg[\left(\frac{\ln\left(\frac{y_i}{x_i}\right)+\frac{1}{2}|\sigma^i|^2(t-s)}{|\sigma^i|\sqrt{t-s}}\right)^2 \\
&\qquad - 2\varrho^{i,j}\,\frac{\left(\ln\left(\frac{y_i}{x_i}\right)+\frac{1}{2}|\sigma^i|^2(t-s)\right)\left(\ln\left(\frac{y_j}{x_j}\right)+\frac{1}{2}|\sigma^j|^2(t-s)\right)}{|\sigma^i||\sigma^j|(t-s)} \\
&\qquad + \left(\frac{\ln\left(\frac{y_j}{x_j}\right)+\frac{1}{2}|\sigma^j|^2(t-s)}{|\sigma^j|\sqrt{t-s}}\right)^2\Bigg]\Bigg\},
\end{aligned}
\tag{S.6}
$$

for all $x_i, x_j, y_i, y_j \in (0,\infty)$, where $s,t \in [0,\infty)$ such that $s \leq t$, is the joint transition density of $\hat{S}^{i,c}$ and $\hat{S}^{j,c}$ over the time interval $[s,t]$. The parameter $\varrho^{i,j}$ in (S.6) is determined by

$$
\varrho^{i,j} = \sum_{k=1}^{m}\frac{\sigma^{i,k}\sigma^{j,k}}{|\sigma^i||\sigma^j|}. \tag{S.7}
$$

It follows from (14.4.8) that $\varrho^{i,j}$ is the correlation between the Brownian motions $\hat{W}^i$ and $\hat{W}^j$.

To start with, we perform an auxiliary computation which allows us to price both instruments under consideration. Fix $t \in [0, \infty)$ and let $g_t$ be a non-negative $\mathcal{A}_t$-measurable random variable. We will now evaluate the following expression:

$$\frac{1}{\hat{S}_t^{j,c}} E\left(\mathbf{1}_{\{\hat{S}_T^{j,c} \geq g_t \hat{S}_T^{i,c}\}} \hat{S}_T^{j,c} \,\Big|\, \mathcal{A}_t\right) = \int_0^\infty \int_{\alpha_t x}^\infty \frac{y}{\hat{S}_t^{j,c}} p_{t,T}^{i,j}\left(\hat{S}_t^{i,c}, \hat{S}_t^{j,c}; x, y\right) dy\, dx. \tag{S.8}$$

After the change of variables we obtain

$$\bar{x} = \frac{\ln\left(\frac{x}{\hat{S}_t^{i,c}}\right) + \frac{1}{2}|\sigma^i|^2(T-t)}{|\sigma^i|\sqrt{T-t}} \tag{S.9}$$

$$\bar{y} = \frac{\ln\left(\frac{y}{\hat{S}_t^{j,c}}\right) + \frac{1}{2}|\sigma^j|^2(T-t)}{|\sigma^j|\sqrt{T-t}}, \tag{S.10}$$

and (S.8) becomes

$$\begin{aligned}\frac{1}{2\pi\sqrt{1-(\varrho^{i,j})^2}} \int_{-\infty}^\infty \int_{d(\bar{x})}^\infty \exp\Bigg\{-\frac{1}{2(1-(\varrho^{i,j})^2)}\Bigg[\left(-\bar{x} + \varrho^{i,j}|\sigma^j|\sqrt{T-t}\right)^2 \\ - 2\varrho^{i,j}\left(-\bar{x} + \varrho^{i,j}|\sigma^j|\sqrt{T-t}\right)\left(-\bar{y} + |\sigma^j|\sqrt{T-t}\right) \\ + \left(-\bar{y} + |\sigma^j|\sqrt{T-t}\right)^2\Bigg]\Bigg\}\, d\bar{y}\, d\bar{x},\end{aligned} \tag{S.11}$$

where

$$d(\bar{x}) = \frac{\ln\left(\frac{\alpha_t S_t^{i,c}}{S_t^{j,c}}\right) - \left(\frac{1}{2}|\sigma^i|^2 - \frac{1}{2}|\sigma^j|^2\right)(T-t)}{|\sigma^j|\sqrt{T-t}} + \frac{|\sigma^i|}{|\sigma^j|}\bar{x}, \tag{S.12}$$

for all $\bar{x} \in \Re$. Another transformation of variables,

$$\tilde{x} = -\bar{x} + \varrho^{i,j}|\sigma^j|\sqrt{T-t}; \tag{S.13}$$

$$\tilde{y} = -\bar{y} + |\sigma^j|\sqrt{T-t}, \tag{S.14}$$

allows us to express (S.11) as

$$\frac{1}{2\pi\sqrt{1-(\varrho^{i,j})^2}} \int_{-\infty}^\infty \int_{-\infty}^{d(\tilde{x})} \exp\left\{-\frac{1}{2(1-(\varrho^{i,j})^2)}\left[\tilde{x}^2 - 2\varrho^{i,j}\tilde{x}\tilde{y} + \tilde{y}^2\right]\right\} d\tilde{y}\, d\tilde{x}, \tag{S.15}$$

where

$$d(\tilde{x}) = \frac{\ln\left(\frac{S_t^{j,c}}{\alpha_t S_t^{i,c}}\right) + \frac{1}{2}\left(\hat{\sigma}^{i,j}\right)^2 (T-t)}{|\sigma^j|\sqrt{T-t}} + \frac{|\sigma^i|}{|\sigma^j|}\tilde{x} = a + b\tilde{x}, \tag{S.16}$$

for all $\tilde{x} \in \Re$, with

$$\hat{\sigma}^{i,j} = \sqrt{|\sigma^i|^2 - 2\varrho^{i,j}|\sigma^i||\sigma^j| + |\sigma^j|^2}. \tag{S.17}$$

After the transformation

$$\hat{y} = \tilde{y} - b\tilde{x}, \tag{S.18}$$

for all $\tilde{x} \in \Re$ and $\tilde{y} \in (-\infty, d(\tilde{x}))$, (S.15) becomes

$$\begin{aligned} \frac{1}{2\pi\sqrt{1-(\varrho^{i,j})^2}} \int_{-\infty}^{a} \int_{-\infty}^{\infty} \exp\Bigg\{ -\frac{1}{2(1-(\varrho^{i,j})^2)} \\ \times \left[\tilde{x}^2 - 2\varrho^j \tilde{x}(\hat{y}+b\tilde{x}) + (\hat{y}+b\tilde{x})^2\right] \Bigg\} \, d\tilde{x}\, d\hat{y}. \end{aligned} \tag{S.19}$$

Now, performing the change of variables

$$\hat{x} = \sqrt{\frac{1-2b\varrho^j + b^2}{1-(\varrho^{i,j})^2}} \left(\tilde{x} + \frac{b-\varrho^{i,j}}{1-2b\varrho^{i,j}+b^2}\hat{y}\right), \tag{S.20}$$

for all $\tilde{x} \in \Re$, transforms (S.19) into

$$\frac{1}{\sqrt{1-2b\varrho^{i,j}+b^2}} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left\{-\frac{1}{2}\frac{\hat{y}^2}{1-2b\varrho^{i,j}+b^2}\right\} d\hat{y}. \tag{S.21}$$

Finally, we set

$$z = \frac{\hat{y}}{\sqrt{1-2b\varrho^{i,j}+b^2}}, \tag{S.22}$$

for all $\hat{y} \in (-\infty, a)$, so that (S.21) becomes

$$\begin{aligned} \frac{1}{2\pi} \int_{-\infty}^{\frac{a}{\sqrt{1-2b\varrho^{i,j}+b^2}}} \exp\left\{-\frac{1}{2}z^2\right\} dz &= N\left(\frac{a}{\sqrt{1-2b\varrho^{i,j}+b^2}}\right) \\ &= N\left(\frac{\ln\left(\frac{S_t^{j,c}}{g_t S_t^{i,c}}\right) + \frac{1}{2}\left(\hat{\sigma}^{i,j}\right)^2 (T-t)}{\hat{\sigma}^{i,j}\sqrt{T-t}}\right), \end{aligned} \tag{S.23}$$

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

Now, to obtain (14.4.33) perform the substitutions $g_t = g(n)$ and $i = 0$ in (S.23) and substitute the resulting expression into (14.4.31), while remembering that $S^{0,c} = S^0$. Finally, perform the substitutions $g_t = g(n)^{-1}$, $i = j$ and $j = 0$ in (S.23) and substitute the resulting expression into (14.4.37), to obtain (14.4.38). It is important to remember in this case that the symmetry of the Gaussian distribution gives $N(-d_2(n)) = 1 - N(d_2(n))$, for each $n \in \mathcal{N}$, while

$$\sum_{n=0}^{\infty} \exp\{-h^{k-m}(T-t)\} \frac{(h^{k-m}(T-t))^n}{n!} = 1, \tag{S.24}$$

since this expression is the total probability of a Poisson random variable with parameter $h^{k-m}(T-t)$.

**14.4** (*) See above.

**14.5** (*) (Hardy Hulley) Firstly, by the function

$$p_4(\varrho, x; \varphi, y) = \frac{1}{2(\varphi - \varrho)} \sqrt{\frac{y}{x}} \exp\left\{-\frac{x+y}{2(\varphi - \varrho)}\right\} \times \sum_{n=0}^{\infty} \frac{1}{n!\, \Gamma(n+2)} \left(\frac{\sqrt{xy}}{2(\varphi - \varrho)}\right)^{2n+1}, \tag{S.25}$$

for all $x, y \in (0, \infty)$ and $\varrho, \varphi \in [0, \infty)$ such that $\varrho < \varphi$, is the transition density of a squared Bessel process of dimension four. Equation (S.25) is obtained from (8.7.9) with the help of the series expansion for the modified Bessel function of the second kind $I_1(\cdot)$, see Abramowitz & Stegun (1972). Note the presence in (S.25) of the gamma function $\Gamma(\cdot)$, defined in (1.2.10). It satisfies the following identity:

$$\Gamma(n) = (n-1)!\,, \tag{S.26}$$

for each $n \in \mathcal{N}$.

Now fix $i, j \in \{0, \ldots, d\}$ and $t \in [0, \infty)$ and let $g_t$ be a positive $\mathcal{A}_t$-measurable random variable. Again, we perform an auxiliary computation that enables us to price both instruments under consideration. Noting that $X^i$ and $X^j$, given by (14.4.21), are independent squared Bessel processes of dimension four, we have

$$
\begin{aligned}
&\frac{1}{\hat{S}^{j,c}_t} E\left(\mathbf{1}_{\{\hat{S}^{j,c}_T \geq g_t \hat{S}^{i,c}_T\}} \hat{S}^{j,c}_T \,\Big|\, \mathcal{A}_t\right) \\
&= X^j_{\varphi^j(t)} E\left(\mathbf{1}_{\{X^j_{\varphi^j(T)} \leq g_t^{-1} X^i_{\varphi^i(T)}\}} \frac{1}{X^j_{\varphi^j(T)}} \,\Bigg|\, \mathcal{A}_t\right) \\
&= \int_0^\infty \int_0^{g_t^{-1}x} \frac{X^j_{\varphi^j(t)}}{y} p_4\Big(\varphi^i(t), X^i_{\varphi^i(t)}; \varphi^i(T), x\Big) \\
&\qquad\qquad\qquad \times p_4\Big(\varphi^j(t), X^j_{\varphi^j(t)}; \varphi^j(T), y\Big)\, dy\, dx \\
&= \int_0^\infty \int_0^{g_t^{-1}x} \frac{1}{y} \exp\Big\{-\frac{1}{2}\lambda^j_t\Big\} \frac{1}{\varphi^i(T) - \varphi^i(t)} \exp\Big\{-\frac{1}{2}\lambda^i_t\Big\} \\
&\qquad \times \left[\sum_{l=0}^\infty \frac{1}{l!\,\Gamma(l+2) 2^{l+1}} \left(\frac{1}{2}\lambda^j_t\right)^{l+1} \left(\frac{y}{\varphi^j(T) - \varphi^j(t)}\right)^{l+1}\right. \\
&\qquad\qquad \left.\times \exp\Big\{-\frac{y}{2(\varphi^j(T) - \varphi^j(t))}\Big\}\right] \\
&\qquad \times \left[\sum_{q=0}^\infty \frac{1}{q!\,\Gamma(q+2) 2^{q+2}} \left(\frac{1}{2}\lambda^i_t\right)^{q} \left(\frac{x}{\varphi^i(T) - \varphi^i(t)}\right)^{q+1}\right. \\
&\qquad\qquad \left.\times \exp\Big\{-\frac{x}{2(\varphi^i(T) - \varphi^i(t))}\Big\}\right] dy\, dx \\
&= \exp\Big\{-\frac{1}{2}\big(\lambda^j_t + \lambda^i_t\big)\Big\} \sum_{l=0}^\infty \frac{\big(\frac{1}{2}\lambda^j_t\big)^{l+1}}{l!\,\Gamma(l+2) 2^{l+1}} \sum_{q=0}^\infty \frac{\big(\frac{1}{2}\lambda^i_t\big)^{q}}{q!\,\Gamma(q+2) 2^{q+2}} \\
&\qquad \times \int_0^\infty \int_0^{\bar{g}_t \bar{x}} \exp\Big\{-\frac{1}{2}(\bar{x} + \bar{y})\Big\} \bar{y}^l \bar{x}^{q+1}\, d\bar{y}\, d\bar{x}.
\end{aligned}
\tag{S.27}
$$

Here we have made the substitutions

$$\bar{x} := \frac{x}{\varphi^i(T) - \varphi^i(t)}; \tag{S.28}$$

$$\bar{y} := \frac{y}{\varphi^j(T) - \varphi^j(t)}. \tag{S.29}$$

The constant in the upper limit of the inner integral in (S.27) is thus given by

$$\bar{g}_t := \frac{\varphi^i(T) - \varphi^i(t)}{g_t}. \tag{S.30}$$

The random variables $\lambda^i_t$ and $\lambda^j_t$ are given by (14.4.24).

With the aid of another change of variables, namely

$$\tilde{y} := \frac{\bar{y}}{\bar{x}}, \tag{S.31}$$

(S.27) now becomes

$$\begin{aligned}
&\exp\Big\{-\frac{1}{2}\big(\lambda_t^j+\lambda_t^i\big)\Big\}\sum_{l=0}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^j\big)^{l+1}}{l!\,\Gamma(l+2)2^{l+1}}\sum_{q=0}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!\,\Gamma(q+2)2^{q+2}}\\
&\qquad\times\int_0^{\bar{g}_t}\int_0^{\infty}\exp\Big\{-\frac{1}{2}\bar{x}(1+\tilde{y})\Big\}\tilde{y}^l\bar{x}^{q+l+2}\,d\bar{x}\,d\tilde{y}\\
&\quad=\exp\Big\{-\frac{1}{2}\big(\lambda_t^j+\lambda_t^i\big)\Big\}\sum_{l=0}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^j\big)^{l+1}}{l!\,\Gamma(l+2)2^{l+1}}\sum_{q=0}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!\,\Gamma(q+2)2^{q+2}}\\
&\qquad\times\Gamma(q+l+3)2^{q+l+3}\int_0^{\bar{g}_t}\frac{\tilde{y}^l}{(1+\tilde{y})^{q+l+3}}\,d\tilde{y}\\
&\quad=\exp\Big\{-\frac{1}{2}\big(\lambda_t^j+\lambda_t^i\big)\Big\}\sum_{m=1}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^j\big)^{m}}{m!}\sum_{q=0}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!}\frac{\Gamma(q+m+2)}{\Gamma(m)\Gamma(q+2)}\\
&\qquad\times\int_0^{\bar{g}_t}\frac{\tilde{y}^{m-1}}{(1+\tilde{y})^{q+m+2}}\,d\tilde{y}\\
&\quad=\exp\Big\{-\frac{1}{2}\big(\lambda_t^j+\lambda_t^i\big)\Big\}\sum_{m=1}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^j\big)^{m}}{m!}\sum_{q=0}^{\infty}\frac{\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!}\frac{\Gamma(q+m+2)}{\Gamma(m)\Gamma(q+2)}\\
&\qquad\times\frac{\bar{g}_t^m}{m}\,{}_2F_1\big(m,q+m+2;m+1;-\bar{g}_t\big)\\
&\quad=\sum_{m=1}^{\infty}\frac{\exp\{-\frac{1}{2}\lambda_t^j\}\big(\frac{1}{2}\lambda_t^j\big)^{m}}{m!}\sum_{q=0}^{\infty}\frac{\exp\{-\frac{1}{2}\lambda_t^i\}\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!}\frac{\Gamma(q+m+2)}{\Gamma(m+1)\Gamma(q+2)}\bar{g}_t^m\\
&\qquad\times{}_2F_1\big(m,q+m+2;m+1;-\bar{g}_t\big).
\end{aligned}\tag{S.32}$$

The first equality in (S.32) follows from the definition of the gamma function in (1.2.10). The second equality follows from some algebra and (S.26). The third equality was obtained with the help of Mathematica's symbolic integration facility. The final equality is another application of (S.26). The hypergeometric function ${}_2F_1(a,b;c;z)$ is described in Abramowitz & Stegun (1972). Now, note that

$$\begin{aligned}
&\frac{\exp\{-\frac{1}{2}\lambda_t^j\}\big(\frac{1}{2}\lambda_t^j\big)^{0}}{0!}\sum_{q=0}^{\infty}\frac{\exp\{-\frac{1}{2}\lambda_t^i\}\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!}\frac{\Gamma(q+0+2)}{\Gamma(0+1)\Gamma(q+2)}\bar{g}_t^0\\
&\qquad\qquad\times{}_2F_1\big(0,q+0+2;0+1;-\bar{g}_t\big)\\
&\quad=\exp\Big\{-\frac{1}{2}\lambda_t^j\Big\}\sum_{q=0}^{\infty}\frac{\exp\{-\frac{1}{2}\lambda_t^i\}\big(\frac{1}{2}\lambda_t^i\big)^{q}}{q!}\\
&\quad=\exp\Big\{-\frac{1}{2}\lambda_t^j\Big\}.
\end{aligned}\tag{S.33}$$

The first equality follows from (S.26) and the properties of the hypergeometric function. For the second equality, note that $\sum_{q=0}^{\infty} \frac{\exp\left\{-\frac{1}{2}\lambda_t^i\right\}\left(\frac{1}{2}\lambda_t^i\right)^q}{q!}$ is the total probability of a Poisson random variable with parameter $\frac{1}{2}\lambda_t^i$.

Thus, putting (S.32) and (S.33) together, we see that (S.27) can be expressed as

$$\begin{aligned}
&\left[\sum_{m=0}^{\infty} \frac{\exp\left\{-\frac{1}{2}\lambda_t^j\right\}\left(\frac{1}{2}\lambda_t^j\right)^m}{m!} \sum_{q=0}^{\infty} \frac{\exp\left\{-\frac{1}{2}\lambda_t^i\right\}\left(\frac{1}{2}\lambda_t^i\right)^q}{q!} \frac{\Gamma(q+m+2)}{\Gamma(m+1)\Gamma(q+2)} \bar{g}_t^m \right.\\
&\qquad \left. \times\, {}_2F_1\left(m, q+m+2; m+1; -\bar{g}_t\right)\right] - \exp\left\{-\frac{1}{2}\lambda_t^j\right\}\\
&= \left[\sum_{m=0}^{\infty} \frac{\exp\left\{-\frac{1}{2}\lambda_t^j\right\}\left(\frac{1}{2}\lambda_t^j\right)^m}{m!} \sum_{q=0}^{\infty} \frac{\exp\left\{-\frac{1}{2}\lambda_t^i\right\}\left(\frac{1}{2}\lambda_t^i\right)^q}{q!} \frac{\Gamma(q+m+2)}{\Gamma(m)\Gamma(q+2)} \right.\\
&\qquad \left. \times \int_0^{\bar{g}_t} \frac{\tilde{y}^{m-1}}{(1+\tilde{y})^{q+m+2}}\, d\tilde{y}\right] - \exp\left\{-\frac{1}{2}\lambda_t^j\right\}\\
&= P\left(\frac{\chi_0'^2(\lambda_t^j)}{\chi_4'^2(\lambda_t^i)} \leq \bar{g}_t\right) - \exp\left\{-\frac{1}{2}\lambda_t^j\right\},
\end{aligned} \tag{S.34}$$

according to Johnson et al. (1995), (30.49), p. 499, where $\chi_\nu'^2(\lambda)$ denotes a non-central chi-square distributed random variable with $\nu$ degrees of freedom and non-centrality $\lambda$. Following the lead of Johnson et al. (1995), we express the distribution function of the ratio of non-central chi-square random variables $\chi_{\nu_1}'^2(\lambda_1)/\chi_{\nu_2}'^2(\lambda_2)$ as $G''_{\nu_1,\nu_2}(\cdot\,; \lambda_1, \lambda_2)$, whence (S.34) becomes

$$G''_{0,4}\left(\frac{\varphi^i(T) - \varphi^i(t)}{g_t}; \lambda_t^j, \lambda_t^i\right) - \exp\left\{-\frac{1}{2}\lambda_t^j\right\}, \tag{S.35}$$

by (S.30).

Now to obtain (14.4.36) perform the substitutions $g_t = g(n)$ and $i = 0$ in (S.35) and substitute the resulting expression into (14.4.31). Finally, perform the substitutions $g_t = g(n)^{-1}$, $i = j$ and $j = 0$ in (S.35) and substitute the resulting expression into (14.4.37), to obtain (14.4.10).

**14.6** (*) See above.

### Solutions for Exercises of Chapter 15

**15.1** By the Lyapunov inequality the variance $\mathrm{Var}((X_i)^{\frac{1}{2}})$ is bounded. That is,

$$\begin{aligned}
\mathrm{Var}\left((X_i)^{\frac{1}{2}}\right) &= E\left(\left((X_i)^{\frac{1}{2}} - E\left((X_i)^{\frac{1}{2}}\right)\right)^2\right) \leq E\left(\left((X_i)^{\frac{1}{2}}\right)^2\right)\\
&= E\left(|X_i|\right) \leq \sqrt{E\left((X_i)^2\right)} = \sqrt{\sigma^2 + \mu^2} < \infty.
\end{aligned}$$

Therefore, the Monte Carlo estimator $\hat{\varrho}_n = \frac{1}{n}\sum_{i=1}^n (X_i)^{\frac{1}{2}}$ is strongly consistent for estimating $\varrho = E((X_i)^{\frac{1}{2}})$ since

$$\begin{aligned}\mathrm{Var}\,(\hat{\varrho}_n) = E\left((\hat{\varrho}_n - \varrho)^2\right) &= E\left(\left(\frac{1}{n}\sum_{i=1}^n \left((X_i)^{\frac{1}{2}} - \varrho\right)\right)^2\right) \\ &= \frac{1}{n^2}\sum_{i=1}^n E\left(\left((X_i)^{\frac{1}{2}} - \varrho\right)^2\right) \\ &= \frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^n \mathrm{Var}\left((X_i)^{\frac{1}{2}}\right)\right) \\ &= \frac{1}{n}\,\mathrm{Var}\left((X_i)^{\frac{1}{2}}\right) = \frac{1}{n}\sqrt{\sigma^2 + \mu^2}.\end{aligned}$$

Consequently, the variance of the estimator $\hat{\varrho}_n$ decreases proportionally to $\frac{1}{n}$ and $\mathrm{Var}(\hat{\varrho}_n)$ converges almost surely to zero.

**15.2** By application of the Wagner-Platen expansion one obtains directly

$$X_{t_0+h} - X_{t_0} = a\,X_{t_0}\,dt + b\,X_{t_0}\,(W_{t_0+h} - W_{t_0}) + R.$$

**15.3** The Euler scheme is given by

$$Y_{n+1} = Y_n + (\mu\,Y_n + \eta)\,\Delta + \gamma\,Y_n\,\Delta W,$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$. The Milstein scheme has the form

$$Y_{n+1} = Y_n + (\mu\,Y_n + \eta)\,\Delta + \gamma\,Y_n\,\Delta W + \frac{\gamma^2}{2}\,Y_n((\Delta W)^2 - \Delta).$$

**15.4** Due to the additive noise of the Vasicek model the Euler and Milstein schemes are identical and of the form

$$Y_{n+1} = Y_n + \gamma(\bar{r} - Y_n)\,\Delta + \beta\,\Delta W,$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$.

**15.5** The explicit strong order 1.0 scheme has the form

$$\begin{aligned}Y_{n+1} &= Y_n + Y_n\,\mu\,\Delta + Y_n\,\sigma\,\Delta W \\ &\quad + \frac{\sigma\,Y_n}{2\sqrt{\Delta}}\left(\mu\,\Delta + \sigma\sqrt{\Delta}\right)((\Delta W)^2 - \Delta),\end{aligned}$$

where $\Delta W = W_{\tau_{n+1}} - W_{\tau_n}$.

**15.6** (*) It follows that the diffusion coefficients for the first Wiener process $W^1$ are

$$b^{1,1} = 1 \qquad \text{and} \qquad b^{2,1} = 0$$

and that of the second Wiener process are

$$b^{1,2} = 0 \qquad \text{and} \qquad b^{2,2} = X_t^1.$$

Therefore it follows that

$$L^1 \, b^{2,2} = b^{1,1} \, \frac{\partial}{\partial x^1} \, b^{2,2} + b^{2,1} \, \frac{\partial}{\partial x^2} \, b^{2,2} = 1$$

and

$$L^2 \, b^{2,1} = b^{1,2} \, \frac{\partial}{\partial x^1} \, b^{2,1} + b^{2,2} \, \frac{\partial}{\partial x^2} \, b^{2,1} = 0.$$

Since the above values are not equal, the SDE is not commutative.

**15.7** (*) The Milstein scheme applied to the given SDE is of the form

$$\begin{aligned} Y_{n+1}^1 &= Y_n^1 + \Delta W^1 \\ Y_{n+1}^2 &= Y_n^2 + Y_n^1 \, \Delta W^2 + I_{(1,2)} \end{aligned}$$

with

$$I_{(1,2)} = \int_{\tau_n}^{\tau_{n+1}} \int_{\tau_n}^{s_2} dW_{s_1}^1 \, dW_{s_2}^2.$$

$$\begin{aligned} \Delta W^1 &= W_{\tau_{n+1}}^1 - W_{\tau_n}^1 \\ \Delta W^2 &= W_{\tau_{n+1}}^2 - W_{\tau_n}^2 \end{aligned}$$

**15.8** Due to symmetry one obtains

$$E\left(\Delta \hat{W}\right) = E\left(\left(\Delta \hat{W}\right)^3\right) = 0.$$

Furthermore, it follows

$$E\left(\left(\Delta \hat{W}\right)^2\right) = \frac{1}{2}\,\Delta + \frac{1}{2}\,\Delta = \Delta$$

and thus

$$E\left(\left(\Delta \hat{W}\right)^2\right) - \Delta = 0.$$

This proves that

$$\left|E\left(\Delta \hat{W}\right)\right| + \left|E\left(\left(\Delta \hat{W}\right)^3\right)\right| + \left|E\left(\left(\Delta \hat{W}\right)^2 - \Delta\right)\right| = 0 \le K\,\Delta^2.$$

**15.9** We have the expectation

$$E\left(1 + Z + \frac{1}{2}\,(Z)^2\right) = \frac{3}{2}$$

and it follows that

$$E(V_N^+) = E(V_N^-) = E(\hat{V}_N) = \frac{3}{2}.$$

Therefore, $\hat{V}_N$ is unbiased.

We calculate the variance of $V_N^+$, which is

$$\begin{aligned}
\mathrm{Var}(V_N^+) &= E\left(\left(\frac{1}{N}\sum_{k=1}^{N}\left(1 + Z(\omega_k) + \frac{1}{2}(Z(\omega_k))^2\right) - \frac{3}{2}\right)^2\right) \\
&= E\left(\left(\frac{1}{N}\sum_{k=1}^{N}\left(Z(\omega_k) + \frac{1}{2}\Big((Z(\omega_k))^2 - 1\Big)\right)\right)^2\right) \\
&= \frac{1}{N}\,E\left(\left(Z + \frac{1}{2}\Big((Z)^2 - 1\Big)\right)^2\right) \\
&= \frac{1}{N}\left(E\left((Z)^2\right) + \frac{1}{4}\Big(E\left((Z)^4\right) - 2\,E\left((Z)^2\right) + 1\Big)\right) \\
&= \frac{1}{N}\left(1 + \frac{1}{4}\,(3 - 2 + 1)\right) = \frac{3}{2\,N}.
\end{aligned}$$

Alternatively, we obtain

$$\begin{aligned}
\mathrm{Var}\left(\hat{V}_N\right) = E\Bigg(\Bigg(\frac{1}{2}\Bigg(&\frac{1}{N}\sum_{k=1}^{N}\left(1 + Z(\omega_k) + \frac{1}{2}\,(Z(\omega_k))^2\right) \\
&+ \frac{1}{N}\sum_{k=1}^{N}\left(1 - Z(\omega_k) + \frac{1}{2}\,(Z(\omega_k))^2\right)\Bigg) - \frac{3}{2}\Bigg)^2\Bigg) \\
= E&\left(\left(\frac{1}{N}\sum_{k=1}^{N}\frac{1}{2}\Big((Z(\omega_k))^2 - 1\Big)\right)^2\right) \\
= \frac{1}{N}&\,\frac{1}{4}\,E\left(\Big((Z)^2 - 1\Big)^2\right) = \frac{1}{4\,N}\left(E\left((Z)^4\right) - 2\,E\left((Z)^2\right) + 1\right) \\
= \frac{1}{4\,N}&\,(3 - 2 + 1) = \frac{1}{2\,N}.
\end{aligned}$$

This shows that the antithetic method provides a Monte Carlo estimate with a third of the variance of a raw Monte Carlo estimate.

**15.10** For $E(V_N^*) = \gamma = 1$ we have $E(\tilde{V}_N) = \frac{2}{3}$, and $\tilde{V}_N$ is an unbiased estimator. The variance of $\tilde{V}_N$ is then obtained as

$$\mathrm{Var}\left(\tilde{V}_N\right) = E\left(\left(\frac{1}{N}\sum_{k=1}^{N}\left(1 + Z(\omega_k) + \frac{1}{2}(Z(\omega_k))^2\right)\right.\right.$$

$$\left.\left. + \alpha\left(1 - \frac{1}{N}\sum_{k=1}^{N}(1 + Z(\omega_k))\right) - \frac{3}{2}\right)^2\right)$$

$$= E\left(\left(\frac{1}{N}\sum_{k=1}^{N}\left(Z(\omega_k)\,(1-\alpha) + \frac{1}{2}\left((Z(\omega_k))^2 - 1\right)\right)\right)^2\right)$$

$$= \frac{1}{N}\,E\left(\left(Z\,(1-\alpha) + \frac{1}{2}\left((Z)^2 - 1\right)\right)^2\right)$$

$$= \frac{1}{N}\left(E\Big((Z)^2\,(1-\alpha)^2\Big) + \frac{1}{4}\Big(E\left((Z)^4\right) - 2\,E\left((Z)^2\right) + 1\Big)\right)$$

$$= \frac{1}{N}\left((1-\alpha)^2 + \frac{1}{4}\,(3 - 2 + 1)\right) = \frac{1}{N}\left(\frac{1}{2} + (1-\alpha)^2\right).$$

It turns out that the minimum variance can be achieved for $\alpha = \alpha_{\min} = 1$, which yields $\mathrm{Var}(\tilde{V}_N) = \frac{1}{2\,N}$.

**15.11** For the multi-period binomial tree we have the benchmarked return

$$u = \exp\{\sigma\sqrt{\Delta}\} - 1$$

with probability $p = \frac{-d}{u-d}$ and the benchmarked return

$$d = \exp\{-\sigma\sqrt{\Delta}\} - 1$$

with the remaining probability $1 - p$. The binomial European put price at time $t = 0$ is then given by the expression

$$
\begin{aligned}
S_0^{\delta_H} = \hat{S}_0^{\delta_H} &= E\left(\left(\hat{K} - \hat{S}_T\right)^+ \mid \mathcal{A}_0\right) \\
&= \sum_{k=0}^{n_T} \frac{n_T\,!}{k\,!\,(n_T - k)\,!}\, p^k\,(1-p)^{n_T-k} \left(\hat{K} - (1+u)^k (1+d)^{n_T-k}\,\hat{S}_0\right)^+ \\
&= \hat{K} \sum_{k=0}^{\bar{k}} \frac{n_T\,!}{k\,!\,(n_T - k)\,!}\, p^k\,(1-p)^{n_T-k} \\
&\quad - S_0 \sum_{k=0}^{\bar{k}} \frac{n_T\,!}{k\,!\,(n_T - k)\,!}\, p^k\,(1-p)^{n_T-k}\,(1+u)^k (1+d)^{n_T-k},
\end{aligned}
$$

where $\bar{k}$ denotes the first integer $k$ for which $S_0\,(1+u)^k(1+d)^{n_T-k} < \hat{K}$.

**15.12** Let us introduce for the Box-Muller random variables

$$
\begin{aligned}
Y_1 &= \cos(2\,\pi\,X_2)\,\sqrt{-2\,\ln X_1} \\
Y_2 &= \sin(2\,\pi\,X_2)\,\sqrt{-2\,\ln X_1},
\end{aligned}
$$

with $X_1, X_2 \sim U(0,1)$ uniformly distributed and independent, the functions $x_1(y_1, y_2) = \exp\{-\frac{1}{2}(y_1^2 + y_2^2)\}$ and $x_2(y_1, y_2) = \frac{1}{2\,\pi}\,\arctan(\frac{y_2}{y_1})$. If we denote by $p(x_1, x_2)$ the joint density of $(X_1, X_2)$, the joint density $q(y_1, y_2)$ of $(Y_1, Y_2)$ is given by

$$
\begin{aligned}
q(y_1, y_2) &= p(x_1(y_1, y_2),\, x_2(y_1, y_2)) \left|\det\left[\frac{\partial x_i}{\partial y_i}\right]\right| \\
&= 1\,\frac{1}{2\pi}\,e^{-\frac{1}{2}}(y_1^2 + y_2^2) = \frac{1}{2\pi}\,e^{-\frac{1}{2}y_1^2}\,\frac{1}{2\pi}\,e^{-\frac{1}{2}y_2^2},
\end{aligned}
$$

which is the density of two independent Gaussian random variables.

# References


Abramowitz, M. & Stegun, I. A. (eds) (1972). *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, Dover, New York.

Ait-Sahalia, Y. (1996). Testing continuous-time models for the spot interest rate, *Rev. Financial Studies* **9**: 385–426.

Ait-Sahalia, Y. & Lo, A. (2000). Non-parametric estimation of state-price densities implicit in financial asset prices, *J. Finance* **53**: 499–548.

Albanese, C. & Chen, O. X. (2005). Discrete credit barrier models. *Quant. Finance* **5**(3): 247–256.

Andersen, L. & Andreasen, J. (2000). Volatility skews and extensions of the LIBOR market model, *Appl. Math. Finance* **7**: 1–32.

Ansel, J. P. & Stricker, C. (1994). Couverture des actifs contingents, *Ann. Inst. H. Poincaré Probab. Statist.* **30**: 303–315.

Arrow, K. J. (1965). *Aspects of the Theory of Risk-Bearing*, Helsinki: Yrjö Hahnsson Foundation.

Artzner, P., Delbaen, F., Eber, J. M. & Heath, D. (1997). Thinking coherently, *Risk* **10**: 68–71.

Bachelier, L. (1900). Théorie de la spéculation, *Annales de l'Ecole Normale Supérieure, Series 3* **17**: 21–86.

Bajeux-Besnainou, I. & Portait, I. (1997). The numeraire portfolio: A new perspective on financial theory, *The European Journal of Finance* **3**: 291–309.

Bakshi, G., Cao, C. & Chen, Z. (1997). Empirical performance of alternative option pricing models, *J. Finance* **LII**: 2003–2049.

Barndorff-Nielsen, O. (1977). Exponentially decreasing distributions for the logarithm of particle size, *Proceedings of the Royal Society of London, Series A* **353**: 401–419.

Barndorff-Nielsen, O. (1978). Hyperbolic distributions and distributions on hyperbolae, *Scand. J. Statist.* **5**: 151–157.

Barndorff-Nielsen, O. (1995). Normal-Inverse Gaussian processes and the modelling of stock returns, *Technical report*, University of Aarhus. 300.

Barndorff-Nielsen, O. (1998). Processes of normal inverse Gaussian type, *Finance Stoch.* **2**(1): 41–68.

Barndorff-Nielsen, O. & Blaesild, P. (1981). Hyperbolic distributions and ramifications: Contributions to theory and application, *in* C. Taillie, G. P. Patil & B. A. Baldessari (eds), *Statistical Distributions in Scientific Work*, Vol. 4, Reidel, Dordrecht, pp. 19–44.



E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

Barndorff-Nielsen, O. & Shephard, N. (2001). Modelling by Lévy processes for financial econometrics, *in* O. E. Barndorff-Nielsen, T. Mikosch & S. Resnick (eds), *Lévy Processes - Theory and Applications*, Birkhäuser, Boston, pp. 283–318.

Barndorff-Nielsen, O. & Shephard, N. (2003). Realised power variation and stochastic volatility models, *Bernoulli* **9**: 243–265.

Barraquand, J. (1993). Numerical valuation of high dimensional multivariate European securities. (working paper).

Barrett, R., Berry, M., Chan, T. F., Demmel, J., Dongarra, J., Eijkhout, V., Pozo, R., Romine, C. & van der Vorst, H. (1994). *Templates for the Solution of Linear Systems: Building Blocks for Iterative Methods*, SIAM, Philadelphia.

Basle (1995). *Planned Supplement to the Capital Accord to Incorporate Market Risk*, Basle Committee on Banking and Supervision, Basle, Switzerland.

Becherer (2001). The numeraire portfolio for unbounded semimartingales, *Finance Stoch.* **5**: 327–341.

Beckers, S. (1980). The constant elasticity of variance model and its implications for option pricing, *J. Finance* **35**(3): 661–673.

Björk, T., Kabanov, Y. & Runggaldier, W. (1997). Bond market structure in the presence of marked point processes, *Math. Finance* **7**: 211–239.

Black, F. (1976). Studies in stock price volatility changes, *Proceedings of the 1976 Business Meeting of the Business and Economic Statistics Section, American Statistical Association*, pp. 177–181.

Black, F. (1995). Interest rates as options, *J. Finance* **1**(7): 1371–1376.

Black, F., Derman, E. & Toy, W. (1990). A one-factor model of interest rates and its application to treasury bond options, *Financial Analysts J.* **46**(1): 33–39.

Black, F. & Karasinski, P. (1991). Bond and option pricing when short rates are lognormal, *Financial Analysts J.* pp. 52–59.

Black, F. & Scholes, M. (1973). The pricing of options and corporate liabilities, *J. Political Economy* **81**: 637–654.

Blattberg, R. C. & Gonedes, N. (1974). A comparison of the stable and Student distributions as statistical models for stock prices, *J. Business* **47**: 244–280.

Bochner, S. (1955). *Harmonic Analysis and the Theory of Probability*, University of California Press, Berkeley, CA.

Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity, *J. Econometrics* **31**: 307–327.

Bollerslev, T., Chou, R. & Kroner, K. (1992). ARCH modelling in finance: A review of the theory and empirical evidence, *J. Econometrics* **52**: 5–59.

Borodin, A. N. & Salminen, P. (2002). *Handbook of Brownian Motion - Facts and Formulae*, 2nd edn, Birkhäuser.

Bouchaud, J. P., Sagna, N., Cont, R., El Karoui, N. & Potters, M. (1999). Phenomenology of the interest rate curve, *Appl. Math. Finance* **6**: 209–241.

Boyle, P. P. (1977). A Monte Carlo approach, *J. Financial Economics* **4**: 323–338.

Breeden, D. T. & Litzenberger, R. (1978). Prices of state-contingent claims implicit in option prices, *J. Business* **51**: 621–651.

Breiman, L. (1960). Investment policies for expanding business optimal in a long run sense, *Naval Research Logistics Quarterly* **7**(4): 647–651.

Breiman, L. (1961). Optimal gambling systems for favorable games, *Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability*, Vol. I, pp. 65–78.

Breymann, W., Dias, A. & Embrechts, P. (2003). Dependence structures for multivariate high-frequency data in finance, *Quant. Finance.* **3**(1): 1–14.

Breymann, W., Kelly, L. & Platen, E. (2006). Intraday empirical analysis and modeling of diversified stock indices, *Asia-Pacific Financial Markets* **12**(1): 1–28.

Brigo, D. & Mercurio, F. (2005). *Interest Rate Models - Theory and Practice*, 2nd edn, Springer Finance.

Brigo, D., Mercurio, F. & Rapisarda, F. (2004). Smile at the uncertainty, *Risk* **17**(5): 97–101.

Browne, S. (1999). The risks and rewards of minimizing shortfall probability, *J. Portfolio Manag.* **25**(4): 76–85.

Bru, M.-F. (1991). Wishart processes, *J. Theoret. Probab.* **4**(4): 725–751.

Bruti-Liberati, N., Nikitopoulos-Sklibosios, C. & Platen, E. (2006). First order strong approximations of jump diffusions, *Monte Carlo Methods Appl.* **12**(3-4): 191–209.

Bruti-Liberati, N., Nikitopoulos-Sklibosios, C. & Platen, E. (2009). Real-world jump-diffusion term structure models, *Quant. Finance* 1–15, iFirst.

Bruti-Liberati, N. & Platen, E. (2004). On the efficiency of simplified weak Taylor schemes for Monte Carlo simulation in finance, *Computational Science - ICCS 2004*, Vol. 3039 of *Lecture Notes in Comput. Sci.*, Springer, pp. 771–778.

Bruti-Liberati, N. & Platen, E. (2007a). Approximation of jump diffusions in finance and economics, *Computational Economics* **29**(3-4): 283–312.

Bruti-Liberati, N. & Platen, E. (2007b). Strong approximations of stochastic differential equations with jumps, *J. Comput. Appl. Math.* **205**(2): 982–1001.

Buchen, P. W. (2004). The pricing of dual-expiry exotics, *Quant. Finance.* **4**(1): 101–108.

Buchen, P. W. & Konstandatos, O. (2005). A new method of pricing lookback options, *Math. Finance* **15**(2): 245–259.

Bühlmann, H. (1970). *Mathematical Methods in Risk Theory*, Springer.

Bühlmann, H. & Platen, E. (2005). A discrete time benchmark approach for insurance and finance, *ASTIN Bulletin* **33**(2): 153–172.

Carr, P., Geman, H., Madan, D. & Yor, M. (2003). Stochastic volatility for Lévy processes, *Math. Finance* **13**(3): 345–382.

Carr, P. & Wu, L. (2003). Finite moment log stable processes and option pricing, *J. Finance* **58**: 753–777.

Cherny, A. S. (2000). On the strong and weak solutions of stochastic differential equations governing Bessel processes, *Stochastics Stochastics Rep.* **70**: 213–219.

Chesney, M. & Scott, L. O. (1989). Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model, *J. Financial and Quantitative Analysis* **24**: 267–284.

Christensen, M. M. & Larsen, K. (2007). No arbitrage and the growth optimal portfolio, *Stochastic Anal. Appl.* **25**(1): 255–280.

Christensen, M. M. & Platen, E. (2005). A general benchmark model for stochastic jump sizes, *Stochastic Anal. Appl.* **23**(5): 1017–1044.

Christensen, M. M. & Platen, E. (2007). Sharpe ratio maximization and expected utility when asset prices have jumps, *Int. J. Theor. Appl. Finance* **10**(8): 1339–1364.

Clark, P. K. (1973). A subordinated stochastic process model with finite variance for speculative prices, *Econometrica* **41**: 135–159.

Clewlow, L. & Xu, X. (1994). The dynamics of stochastic volatility, *FORC Preprint 94/53*, Financial Options Research Centre, University of Warwick.

Cochrane, J. H. (2001). *Asset Pricing*, Princeton University Press.

Constatinides, G. M. (1992). A theory of the nominal structure of interest rates, *Rev. Financial Studies* **5**: 531–552.

Cont, R. & da Fonseca, J. (2002). Dynamics of implied volatility surfaces, *Quant. Finance.* **2**: 45–60.

Cont, R. & Tankov, P. (2004). *Financial Modelling with Jump Processes*, Financial Mathematics Series, Chapman & Hall/CRC.

Corsi, F., Zumbach, G., Müller, U. & Dacorogna, M. (2001). Consistent high-precision volatility from high-frequency data, *Economic Notes* **30**(2): 183–204.

Courtadon, G. (1982). The pricing of options on default-free bonds, *J. Financial and Quantitative Analysis* **17**: 75–100.

Cover, T. (1991). Universal portfolios, *Math. Finance* **1**: 1–29.

Cox, J. C. (1975). Notes on option pricing I: constant elasticity of variance diffusions. Stanford University, (working paper, unpublished).

Cox, J. C. (1996). The constant elasticity of variance option pricing model, *J. Portfolio Manag.* (Special Issue): 15–17.

Cox, J. C., Ingersoll, J. E. & Ross, S. A. (1985). A theory of the term structure of interest rates, *Econometrica* **53**: 385–407.

Cox, J. C. & Ross, S. A. (1976). The valuation of options for alternative stochastic processes, *J. Financial Economics* **3**: 145–166.

Cox, J. C., Ross, S. A. & Rubinstein, M. (1979). Option pricing: A simplified approach, *J. Financial Economics* **7**: 229–263.

Craddock, M. & Platen, E. (2004). Symmetry group methods for fundamental solutions, *J. of Differential Equations* **207**(2): 285–302.

Curran, M. (1994). Rata gems, *Risk* **7**(3).

Dacorogna, M., Müller, U., Pictet, O. V. & De Vries, C. G. (2001). Extremal forex returns in extremely large data sets, *Extremes* **4**(2): 105–127.

Davis, M. H. A. (1997). Option pricing in incomplete markets, *in* M. A. H. Dempster & S. R. Pliska (eds), *Mathematics of Derivative Securities*, Cambridge University Press, pp. 227–254.

Delbaen, F. & Schachermayer, W. (1994). A general version of the fundamental theorem of asset pricing, *Math. Ann.* **300**: 463–520.

Delbaen, F. & Schachermayer, W. (1995). Arbitrage possibilities in Bessel processes and their relation to local martingales, *Probab. Theory Related Fields* **102**: 357–366.

Delbaen, F. & Schachermayer, W. (1998). The fundamental theorem of asset pricing for unbounded stochastic processes, *Math. Ann.* **312**: 215–250.

Delbaen, F. & Schachermayer, W. (2006). *The Mathematics of Arbitrage*, Springer Finance, Springer.

Delbaen, F. & Shirakawa, H. (1997). Squared Bessel processes and their applications to the square root interest rate model. Preprint. Department of Industrial Engineering and Management, Tokio Institute of Technology.

Delbaen, F. & Shirakawa, H. (2002). A note on option pricing for the constant elasticity of variance model, *Asia-Pacific Financial Markets* **9**(2): 85–99.

Derman, E. & Kani, I. (1994a). Riding on a smile, *Risk* **7**: 32–39.

Derman, E. & Kani, I. (1994b). The volatility smile and its implied tree, *Goldman Sachs Quantitative Strategies Research Notes.*

Dickson, D. C. M. (1992). On the distribution of the surplus prior to ruin, *Insurance: Mathematics and Economics* **11**: 191–207.

Dickson, D. C. M. (1993). On the distribution of the claim causing ruin, *Insurance: Mathematics and Economics* **12**: 143–154.

Dimson, E., Marsh, P. & Staunton, M. (2002). *Triumph of the Optimists: 101 Years of Global Investment Returns*, Princeton University Press.

Doob, J. L. (1953). *Stochastic Processes*, Wiley, New York.

Doob, J. L. (1994). *Measure Theory*, Vol. 143 of *Graduate Texts in Mathematics*, Springer.

Dothan, L. U. (1978). On the term structure of interest rates, *J. Financial Economics* **6**: 59–69.

Duffie, D. (2001). *Dynamic Asset Pricing Theory*, 3rd edn, Princeton, University Press.

Duffie, D. & Kan, R. (1994). Multi-factor term structure models, *Philos. Trans. Roy. Soc. London Ser. A* **347**: 577–580.

Duffie, D., Pan, J. & Singleton, K. (2000). Transform analysis and option pricing for affine jump diffusions, *Econometrica* **68**: 1343–1376.

Dumas, B., Fleming, J. & Whaley, R. (1998). Implied volatility functions: Empirical tests, *J. Finance* **53**: 2059–2106.

Dupire, B. (1992). Arbitrage pricing with stochastic volatility, *Proceedings of AFFI Conference, Paris.*

Dupire, B. (1993). Model art, *Risk* **6**: 118–124.

Dupire, B. (1994). Pricing with a smile, *Risk* **7**: 18–20.

Eberlein, E. (2002). Applications of generalized Lévy motions in finance, *in* O. E. Barndorff-Nielsen, T. Mikosch & S. Resnick (eds), *Lévy Processes: Theory and Application*, Birkhäuser Boston, pp. 319–336.

Eberlein, E. & Keller, U. (1995). Hyperbolic distributions in finance, *Bernoulli* **1**: 281–299.

Einstein, A. (1905). On the movement of small particles suspended in stationary liquid demanded by molecular-kinetic theory of heat, *Ann. Phys.* **17**.

Elliott, R. J. (1982). *Stochastic Calculus and Applications*, Springer.

Elliott, R. J. & Kopp, P. E. (2005). *Mathematics of Financial Markets*, Springer Finance, 2nd edn, Springer, New York.

Elliott, R. J. & van der Hoek, J. (2003). A general fractional white noise theory and applications to finance, *Math. Finance* **13**: 301–330.

Embrechts, P., Klüppelberg, C. & Mikosch, T. (1997). *Modelling Extremal Events. For Insurance and Finance*, Vol. 33 of *Appl. Math.*, Springer.

Embrechts, P., McNeil, A. & Straumann, D. (2002). Correlation and dependencies in risk management: Properties and pitfalls, *Risk Management: Value at Risk and Beyond*, Cambridge Univ. Press, pp. 176–223.

Engle, R. F. (1982). Autoregressive conditional heteroskedasticity with estimates of the variance of U.K. inflation, *Econometrica* **50**(4): 987–1007.

Engle, R. F. & Bollerslev, T. (1986). Modelling the persistence of conditional variances. With comments and a reply by the authors, *Econometric Rev.* **5**(1): 1–87.

Fama, E. F. (1963). Mandelbrot and the stable Paretian hypothesis, *J. Business* **36**: 420–429. Reprinted in Cootner (1964), Chapter 14, 297-306.

Fama, E. F. (1965). The behavior of stock-market prices, *J. Business* **38**: 34–105.

Fang, K. T., Kotz, S. & Ng, K. W. (1990). *Symmetric Multivariate and Related Distributions*, Chapman Hall, London.

Feller, W. (1968). *An Introduction to Probability Theory and Its Applications*, Vol. 1, 3rd edn, Wiley, New York.

Fergusson, K. & Platen, E. (2006). On the distributional characterization of log-returns of a world stock index, *Appl. Math. Finance* **13**(1): 19–38.

Fernholz, E. R. (2002). *Stochastic Portfolio Theory*, Vol. 48 of *Appl. Math.*, Springer.

Fernholz, E. R., Karatzas, I. & Kardaras, C. (2005). Diversity and relative arbitrage in equity markets, *Finance Stoch.* **9**(1): 1–27.

Fischer, P. & Platen, E. (1999). Applications of the balanced method to stochastic differential equations in filtering, *Monte Carlo Methods Appl.* **5**(1): 19–38.

Fishman, G. S. (1996). *Monte Carlo: Concepts, Algorithms and Applications*, Springer Ser. Oper. Res., Springer.

Fleming, W. H. & Sheu, S.-J. (1999). Optimal long term growth rate of expected utility of wealth, *Ann. Appl. Probab.* **9**: 871–903.

Föllmer, H. (1981). Calcul d'Itô sans probabilités, *Seminar on Probability*, XV, Vol. 850 of *Lecture Notes in Math.*, Springer, Berlin, pp. 143–150. (Univ. Strasbourg, Strasbourg, 1979/1980), (in French).

Föllmer, H. & Schiedt, A. (2002). *Stochastic Finance. An Introduction In Discrete Time*, Vol. 27 of *de Gruyter Studies in Mathematics*, Walter de Gruyter & Co., Berlin.

Föllmer, H. & Schweizer, M. (1991). Hedging of contingent claims under incomplete information, *in* M. H. A. Davis & R. J. Elliott (eds), *Applied Stochastic Analysis*, Vol. 5 of *Stochastics Monogr.*, Gordon and Breach, London/New York, pp. 389–414.

Föllmer, H. & Schweizer, M. (1993). A microeconomic approach to diffusion models for stock prices, *Math. Finance* **3**: 1–23.

Föllmer, H. & Sondermann, D. (1986). Hedging of non-redundant contingent claims, *in* W. Hildebrandt & A. Mas-Colell (eds), *Contributions to Mathematical Economics*, North Holland, pp. 205–223.

Frey, R. (1997). Derivative asset analysis in models with level-dependent and stochastic volatility. Mathematics of Finance, Part II, *CWI Quarterly* **10**(1): 1–34.

Friedman, A. (1975). *Stochastic Differential Equations and Applications, Vol. I*, Vol. 28 of *Probability and Mathematical Statistics*, Academic Press, New York.

Gaines, J. G. & Lyons, T. J. (1994). Random generation of stochastic area integrals, *SIAM J. Appl. Math.* **54**(4): 1132–1146.

Geman, H., Madan, D. & Yor, M. (2001). Asset prices are Brownian motion: only in business time, *Quantitative Analysis in Financial Markets*, World Sci. Publishing, pp. 103–146.

Geman, S., El Karoui, N. & Rochet, J. C. (1995). Changes of numeraire, changes of probability measures and pricing of options, *J. Appl. Probab.* **32**: 443–458.

Gerber, H. U. (1979). *An Introduction to Mathematical Risk Theory*, Vol. 8 of *S. S. Heubner Foundation Monograph Series*, Richard D. Irwin, Homewood.

Gerber, H. U. (1997). *Life Insurance Mathematics*, Springer, Berlin. 3rd edition.

Gerber, H. U. & Shiu, E. S. W. (1997). The joint distribution of the time of ruin, the surplus immediately before ruin, and the deficit at ruin, *Insurance: Mathematics and Economics* **21**: 129–137.

Ghysels, E., Harvey, A. & Renault, E. (1996). Stochastic volatility, *Statistical Methods in Finance*, Vol. 14 of *Handbook of Statist.*, North-Holland, pp. 119–191.

Gibson, M. S. (2001). Incorporating event risk into Value at Risk. Discussion Paper Federal Reserve Board, Washington (http://www.gloriamundi.org).

Glasserman, P. (2004). *Monte Carlo Methods in Financial Engineering*, Vol. 53 of *Appl. Math.*, Springer.

Glasserman, P. & Merener, N. (2003). Numerical solution of jump-diffusion LIBOR market models, *Finance Stoch.* **7**(1): 1–27.

Göing-Jaeschke, A. & Yor, M. (2003). A survey and some generalizations of Bessel processes, *Bernoulli* **9**(2): 313–349.

Gorovoi, V. & Linetsky, V. (2004). Black's model of interest rates as options, eigenfunction expansions and Japanese interest rates, *Math. Finance* **14**(1): 49–78.

Grandell, J. (1991). *Aspects of Risk Theory*, Springer, New York.

Hakansson, N. H. (1971). Multi-period mean-variance analysis: towards a general theory of portfolio choice, *J. Finance* **26**: 857–884.

Hakansson, N. H. & Ziemba, W. T. (1995). Capital growth theory, *in* R. Jarrow, V. Maksimovic & W. T. Ziemba (eds), *Handbooks in Operations Research and Management Science: Finance*, Vol. 9, Elsevier Science, pp. 65–86.

Hansen, L. P. (1982). Large sample properties of the generalized method of moments, *Econometrica* **50**: 1029–1054.

Hansen, L. P. & Jagannathan, R. (1991). Implications of security market data for models of dynamic economies, *J. Political Economy* **99**: 225–262.

Harrison, J. M. & Kreps, D. M. (1979). Martingale and arbitrage in multiperiod securities markets, *J. Economic Theory* **20**: 381–408.

Harrison, J. M. & Pliska, S. R. (1981). Martingales and stochastic integrals in the theory of continuous trading, *Stochastic Process. Appl.* **11**(3): 215–260.

Harrison, J. M. & Pliska, S. R. (1983). A stochastic calculus model of continuous trading: Complete markets, *Stochastic Process. Appl.* **15**(3): 313–316.

Heath, D., Jarrow, R. & Morton, A. (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claim valuation, *Econometrica* **60**(1): 77–105.

Heath, D., Hurst, S. R. & Platen, E. (2001). Modelling the stochastic dynamics of volatility for equity indices, *Asia-Pacific Financial Markets* **8**: 179–195.

Heath, D. & Platen, E. (2002a). Consistent pricing and hedging for a modified constant elasticity of variance model, *Quant. Finance.* **2**(6): 459–467.

Heath, D. & Platen, E. (2002b). Perfect hedging of index derivatives under a minimal market model, *Int. J. Theor. Appl. Finance* **5**(7): 757–774.

Heath, D. & Platen, E. (2002c). A variance reduction technique based on integral representations, *Quant. Finance.* **2**(5): 362–369.

Heath, D. & Platen, E. (2003). Pricing of index options under a minimal market model with lognormal scaling, *Quant. Finance.* **3**(6): 442–450.

Heath, D. & Platen, E. (2005a). Currency derivatives under a minimal market model with random scaling, *Int. J. Theor. Appl. Finance* **8**(8): 1157–1177.

Heath, D. & Platen, E. (2005b). Understanding the implied volatility surface for options on a diversified index, *Asia-Pacific Financial Markets* **11**(1): 55–77.

Heath, D. & Platen, E. (2006). Local volatility function models under a benchmark approach, *Quant. Finance.* **6**(3): 197–2006.

Heath, D., Platen, E. & Schweizer, M. (2001). A comparison of two quadratic approaches to hedging in incomplete markets, *Math. Finance* **11**(4): 385–413.

Heath, D. & Schweizer, M. (2000). Martingales versus PDEs in finance: An equivalence result with examples, *J. Appl. Probab.* **37**(4): 947–957.

Heston, S. L. (1993). A closed-form solution for options with stochastic volatility with applications to bond and currency options, *Rev. Financial Studies* **6**(2): 327–343.

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application. A General Approach to Optimal Parameter Estimation*, Springer Ser. Statist., Springer.

Heyde, C. C. (1999). A risky asset model with strong dependence through fractal activity time, *J. Appl. Probab.* **36**: 1234–1239.

Heyde, C. C. & Liu, S. (2001). Empirical realities for a minimal description risky asset model. The need for fractal features, *J. Korean Math. Soc.* **38**: 1047–1059.

Ho, T. S. Y. & Lee, S. B. (1986). Term structure movements and pricing of interest rate claims, *J. Finance* **41**: 1011–1029.

Hofmann, N., Platen, E. & Schweizer, M. (1992). Option pricing under incompleteness and stochastic volatility, *Math. Finance* **2**(3): 153–187.

Hull, J. (2000). *Options, Futures and Other Derivatives*, Prentice-Hall, New Jersey.

Hull, J. & White, A. (1987). The pricing of options on assets with stochastic volatilities, *J. Finance* **42**: 281–300.

Hull, J. & White, A. (1988). The use of control variate techniques in option pricing, *J. Financial and Quantitative Analysis* **23**(3): 237–251.

Hull, J. & White, A. (1990). Pricing interest rate derivative securities, *Rev. Financial Studies* **3**: 573–592.

Hulley, H., Miller, S. & Platen, E. (2005). Benchmarking and fair pricing applied to two market models, *Kyoto Economic Review* **74**(1): 85–118.

Hurst, S. R. (1997). *On the Stochastic Dynamics of Stock Market Volatility*, PhD thesis, ANU, Canberra.

Hurst, S. R. & Platen, E. (1997). The marginal distributions of returns and volatility, *in* Y. Dodge (ed.), *$L_1$-Statistical Procedures and Related Topics*, Vol. 31 of *IMS Lecture Notes - Monograph Series*, Institute of Mathematical Statistics Hayward, California, pp. 301–314.

Hurst, S. R., Platen, E. & Rachev, S. T. (1999). Option pricing for a logstable asset price model, *Math. Comput. Modelling* **29**(10-12): 105–119.

Ikeda, N. & Watanabe, S. (1989). *Stochastic Differential Equations and Diffusion Processes*, 2nd edn, North-Holland. (first edition (1981)).

Ingersoll, J. E. (2000). Digital contracts: Simple tools for pricing complex derivatives, *The Journal of Business* **73**(1): 67–88.

Jäckel, P. (2002). *Monte Carlo Methods in Finance*, Wiley.

Jacod, J. & Shiryaev, A. N. (2003). *Limit Theorems for Stochastic Processes*, 2nd edn, Springer.

Jamshidian, F. (1989). An exact bond option pricing formula, *J. Finance* **44**: 205–209.

Jeanblanc, M., Yor, M. & Chesney, M. (2009). *Mathematical Methods for Financial Markets*, Springer Finance, Springer.

Johnson, H. & Shanno, D. (1987). Option pricing when the variance is changing, *J. Financial and Quantitative Analysis* **22**: 143–151.

Johnson, N. L., Kotz, S. & Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2 of *Wiley Series in Probability and Mathematical Statistics*, 2nd edn, John Wiley & Sons.

Joy, C., Boyle, P. P. & Tan, K. S. (1996). Quasi Monte Carlo methods in numerical finance, *Management Science* **42**(6): 926–938.

Kahaner, D., Moler, C. & Nash, S. (1989). *Numerical Methods and Software.*, Prentice Hall Series in Computational Mathematics, Englewood Cliffs.

Karatzas, I. & Shreve, S. E. (1991). *Brownian Motion and Stochastic Calculus*, 2nd edn, Springer.

Karatzas, I. & Shreve, S. E. (1998). *Methods of Mathematical Finance*, Vol. 39 of *Appl. Math.*, Springer.

Kelly, J. R. (1956). A new interpretation of information rate, *Bell Syst. Techn. J.* **35**: 917–926.

Kelly, L. (1999). An assessment of Black-Scholes hedging of index options. University of Technology, Sydney; Honours Thesis Bachelor of Mathematics & Finance.

Kelly, L., Platen, E. & Sørensen, M. (2004). Estimation for discretely observed diffusions using transform functions, *Stochastic Methods and Their Applications*, Vol. 41A of *J. Appl. Probab.*, Applied Prob. Trust, pp. 99–118.

Kessler, M. (1997). Estimation of an ergodic diffusion from discrete observations, *Scand. J. Statist.* **24**: 211–229.

Khanna, A. & Kulldorff, M. (1999). A generalization of the mutual fund theorem, *Finance Stoch.* **3**(2): 167–185.

Kloeden, P. E. & Platen, E. (1999). *Numerical Solution of Stochastic Differential Equations*, Vol. 23 of *Appl. Math.*, Springer. Third corrected printing.

Kloeden, P. E., Platen, E. & Schurz, H. (2003). *Numerical Solution of SDE's Through Computer Experiments*, Universitext, Springer. Third corrected printing.

Korn, R. (1997). *Optimal Portfolios*, World Scientific.

Kou, S. G. (2002). A jump diffusion model for option pricing, *Management Science* **48**: 1086–1101.

Krylov, N. V. (1980). *Controlled Diffusion Processes*, Vol. 14 of *Appl. Math.*, Springer.

Kubilius, K. & Platen, E. (2002). Rate of weak convergence of the Euler approximation for diffusion processes with jumps, *Monte Carlo Methods Appl.* **8**(1): 83–96.

Küchler, U., Neumann, K., Sørensen, M. & Streller, A. (1999). Stock returns and hyperbolic distributions, *Math. Comput. Modelling* **29**: 1–15.

Kunita, H. (1984). Stochastic differential equations and stochastic flows of diffeomorphisms, *École d'été de probabilités de Saint-Flour, XII*, Vol. 1097 of *Lecture Notes in Math.*, Springer, pp. 143–303.

Lamberton, D. & Lapeyre, B. (1996). *Introduction to Stochastic Calculus Applied to Finance*, Chapman & Hall, London. Translation from French.

Latané, H. (1959). Criteria for choice among risky ventures, *J. Political Economy* **38**: 145–155.

Le, T. (2005). Stochastic market volatility models, *Applied. Financial Economics Letters* **1**: 177–188.

Le, T. & Platen, E. (2006). Approximating the growth optimal portfolio with a diversified world stock index, *J. Risk Finance* **7**(5): 559–574.

Ledoit, O., Santa-Clara, P. & Wolf, M. (2003). Flexible multivariate GARCH modeling with an application to international stock markets, *Rev. of Economic Studies* **85**(3): 735–747.

Lewis, A. L. (2000). *Option Valuation Under Stochastic Volatility*, Finance Press, Newport Beach.

Lintner, J. (1965). The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets, *Rev. Econom. Statist.* **47**: 13–37.

Lo, C. F., Yuen, P. H. & Hui, C. H. (2000). Constant elasticity of variance option pricing model with time dependent parameters, *Int. J. Theor. Appl. Finance* **3**(4): 661–674.

Loewenstein, M. & Willard, G. A. (2000). Local martingales, arbitrage, and viability: Free snacks and cheap thrills, *Econometric Theory* **16**(1): 135–161.

Long, J. B. (1990). The numeraire portfolio, *J. Financial Economics* **26**: 29–69.

Longstaff, F. A. (1989). A nonlinear general equilibrium model of the term structure of interest rates, *J. Financial Economics* **23**: 195–224.

Luenberger, D. G. (1969). *Optimization by Vector Space Methods*, Wiley, New York.

Lukacs, E. (1960). *Characteristic Functions*, number 5 in *Griffin's Statistical Monographs & Courses*, Hafner Publishing Co., New York.

Madan, D. & Milne, F. (1991). Option pricing with (V.G.) martingale components, *Math. Finance* **1**: 39–56.

Madan, D. & Seneta, E. (1990). The variance gamma (V.G.) model for share market returns, *J. Business* **63**: 511–524.

Maghsoodi, Y. (1996). Mean-square efficient numerical solution of jump-diffusion stochastic differential equations, *SANKHYA A* **58**(1): 25–47.

Maghsoodi, Y. (1998). Exact solutions and doubly efficient approximations of jump-diffusion Itô equations, *Stochastic Anal. Appl.* **16**(6): 1049–1072.

Maghsoodi, Y. & Harris, C. J. (1987). In-probability approximation and simulation of nonlinear jump-diffusion stochastic differential equations, *IMA J. Math. Control Inform.* **4**(1): 65–92.

Mandelbrot, B. (1963). The variation of certain speculative prices, *J. Business* **36**: 394–419. Reprinted in Cootner (1964), Chapter 15, 307-337.

Mandelbrot, B. (1967). The variation of some other speculative prices, *J. Business* **40**: 393–413.

Mandelbrot, B. & Taylor, H. M. (1967). On the distribution of stock price differences, *Math. Oper. Res.* **15**: 1057–1062.

Margrabe, W. (1978). The value of an option to exchange one asset for another, *The Journal of Finance* **33**(1): 177–186.

Markowitz, H. (1952). Portfolio selection, *J. Finance* **VII**(1): 77–91.

Markowitz, H. (1959). *Portfolio Selection: Efficient Diversification of Investment*, Wiley, New York.

Markowitz, H. (1976). Investment for the long run: New evidence for an old rule, *J. Finance* **XXXI**(5): 1273–1286.

Markowitz, H. & Usmen, N. (1996a). The likelihood of various stock market return distributions, Part 1: Principles of inference, *J. Risk & Uncertainty* **13**(3): 207–219.

Markowitz, H. & Usmen, N. (1996b). The likelihood of various stock market return distributions, Part 2: Empirical results, *J. Risk & Uncertainty* **13**(3): 221–247.

Marsh, T. A. & Rosenfeld, E. R. (1983). Stochastic processes for interest rates and equilibrium bond prices, *J. Financial Economics* **38**: 635–650.

Maruyama, G. (1955). Continuous Markov processes and stochastic equations, *Rend. Circ. Mat. Palermo* **4**: 48–90.

Matacz, A. & Bouchaud, J. P. (2000). An empirical investigation of the forward interest rate term structure, *Int. J. Theor. Appl. Finance* **3**(4): 703–729.

McNeil, A., Frey, R. & Embrechts, P. (2005). *Quantitative Risk Management*, Princeton University Press.

Melino, A. & Turnbull, S. (1990). Pricing foreign currency options with stochastic volatility, *J. Econometrics* **45**: 239–265.

Mendel, J. (1995). *Lessons in Estimation Theory for Signal Processing, Communications and Control*, Prentice Hall.

Merton, R. C. (1973a). An intertemporal capital asset pricing model, *Econometrica* **41**: 867–888.

Merton, R. C. (1973b). Theory of rational option pricing, *Bell J. Econ. Management Sci.* **4**: 141–183.

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous, *J. Financial Economics* **2**: 125–144.

Merton, R. C. (1992). *Continuous-Time Finance*, Blackwell, Oxford.

Mikulevicius, R. & Platen, E. (1988). Time discrete Taylor approximations for Ito processes with jump component, *Math. Nachr.* **138**: 93–104.

Miller, S. & Platen, E. (2005). A two-factor model for low interest rate regimes, *Asia-Pacific Financial Markets* **11**(1): 107–133.

Miller, S. & Platen, E. (2008). Analytic pricing of contingent claims under the real world measure, *Int. J. Theor. Appl. Finance* **11**(8): 841–867.

Milstein, G. N. (1974). Approximate integration of stochastic differential equations, *Theory Probab. Appl.* **19**: 557–562.

Milstein, G. N. (1995). *Numerical Integration of Stochastic Differential Equations*, Mathematics and Its Applications, Kluwer.

Milstein, G. N., Platen, E. & Schurz, H. (1998). Balanced implicit methods for stiff stochastic systems, *SIAM J. Numer. Anal.* **35**(3): 1010–1019.

Miltersen, K. R. & Schwartz, E. (1998). Pricing of options on commodity futures with stochastic term structure of convenience yields and interest rates, *J. Financial and Quantitative Analysis* **33**: 33–59.

Miyahara, Y. & Novikov, A. A. (2002). Geometric Lévy process pricing model, *Proc. Steklov Inst. Math.* **2**(237): 176–191.

Mossin, J. (1966). Equilibrium in a capital asset market, *Econometrica* **3**: 768–783.

Musiela, M. & Rutkowski, M. (2005). *Martingale Methods in Financial Modelling*, Vol. 36 of *Appl. Math.*, 2nd edn, Springer.

Nelsen, R. (1999). *An Introduction to Copulas*, Springer, New York.

Nelson, D. B. (1990). ARCH models as diffusion approximations, *J. Econometrics* **45**: 7–38.

Neyman, J. & Pearson, E. S. (1928). On the use and interpolation of certain test criteria for purposes of statistical inference, *Biometrika* **20A**: 175–240 and 263–294.

Niederreiter, H. (1992). *Random Number Generation and Quasi-Monte-Carlo Methods*, SIAM, Philadelphia, PA.

Nielsen, L. T. & Vassalou, M. (2004). Sharpe ratios and alphas in continuous time, *J. Financial and Quantitative Analysis* **39**(1): 141–152.

Novikov, A. A. (1972). On an identity for stochastic integrals, *Theory Probab. Appl.* **17**: 717–720.

Osborne, M. F. M. (1959). Brownian motion in the stock market, *Math. Oper. Res.* **7**: 145–173.

Paskov, S. & Traub, J. (1995). Faster valuation of financial derivatives, *J. Portfolio Manag.* pp. 113–120.

Pearson, N. & Sun, T. S. (1989). A test of the Cox, Ingersoll, Ross model of the term structure of interest rates using the method of moments. Sloan School of Management, Massachusetts Institute of Technology (working paper).

Picard, P. (1994). On some measures of the severity of ruin in the classical Poisson model, *Insurance: Mathematics and Economics* **14**: 107–115.

Pivitt, B. (1999). Accessing the Intel® Random Number Generator with CDSA, *Technical report*, Intel® Platform Security Division.

Platen, E. (1982a). An approximation method for a class of Itô processes with jump component, *Liet. Mat. Rink.* **22**(2): 124–136.

Platen, E. (1982b). A generalized Taylor formula for solutions of stochastic differential equations, *SANKHYA A* **44**(2): 163–172.

Platen, E. (1984). Beiträge zur zeitdiskreten Approximation von Itôprozessen. Habilitation, Academy of Sciences, Berlin.

Platen, E. (1992). Higher-order weak approximation of Itô diffusions by Markov chains, *Probab. Engrg. Inform. Sci.* **6**: 391–408.

Platen, E. (1995). On weak implicit and predictor-corrector methods, *Math. Comput. Simulation* **38**: 69–76.

Platen, E. (1997). A non-linear stochastic volatility model, *Technical report*, Australian National University, Canberra, Financial Mathematics Research Reports. FMRR 005-97.

Platen, E. (1999). A short term interest rate model, *Finance Stoch.* **3**(2): 215–225.

Platen, E. (2001). A minimal financial market model, *Trends in Mathematics*, Birkhäuser, pp. 293–301.

Platen, E. (2002). Arbitrage in continuous complete markets, *Adv. in Appl. Probab.* **34**(3): 540–558.

Platen, E. (2004a). A benchmark framework for risk management, *Stochastic Processes and Applications to Mathematical Finance*, Proceedings of the Ritsumeikan Intern. Symposium, World Scientific, pp. 305–335.

Platen, E. (2004b). A class of complete benchmark models with intensity based jumps, *J. Appl. Probab.* **41**: 19–34.

Platen, E. (2004c). Modeling the volatility and expected value of a diversified world index, *Int. J. Theor. Appl. Finance* **7**(4): 511–529.

Platen, E. (2005a). An alternative interest rate term structure model, *Int. J. Theor. Appl. Finance* **8**(6): 717–735.

Platen, E. (2005b). Diversified portfolios with jumps in a benchmark framework, *Asia-Pacific Financial Markets* **11**(1): 1–22.

Platen, E. (2005c). On the role of the growth optimal portfolio in finance, *Australian Economic Papers* **44**(4): 365–388.

Platen, E. (2006a). A benchmark approach to asset management, *J. Asset Management* **6**(6): 390–405.

Platen, E. (2006b). Capital asset pricing for markets with intensity based jumps, *International Conference Stochastic Finance 2004*, Springer, pp. 157–182.

Platen, E. (2006c). A parsimonious financial market model. (working paper, University of Technology, Sydney).

Platen, E. & Rebolledo, R. (1985). Weak convergence of semimartingales and discretization methods, *Stochastic Process. Appl.* **20**: 41–58.

Platen, E. & Rebolledo, R. (1996). Principles for modelling financial markets, *J. Appl. Probab.* **33**: 163–172.

Platen, E. & Schweizer, M. (1998). On feedback effects from hedging derivatives, *Math. Finance* **8**(1): 67–84.

Platen, E. & Stahl, G. (2003). A structure for general and specific market risk, *Computational Statistics* **18**(3): 355–373.

Platen, E. & Wagner, W. (1982). On a Taylor formula for a class of Itô processes, *Probab. Math. Statist.* **3**(1): 37–51.

Pliska, S. R. (1997). *Introduction to Mathematical Finance - Discrete Time Models*, Blackwell Publishers.

Praetz, P. D. (1972). The distribution of share price changes, *J. Business* **45**: 49–55.

Prakasa Rao, B. L. S. (1999). *Statistical Inference for Diffusion Type Processes*, Arnold Publishers, Oxford University Press.

Pratt, J. W. (1964). Risk aversion in the small and in the large, *Econometrica* **32**: 122–136.

Protter, P. (2004). *Stochastic Integration and Differential Equations*, 2nd edn, Springer.

Rao, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd edn, Wiley, New York.

Raviart, P. A. & Thomas, J. M. (1983). *Introduction to the Numerical Analysis of Partial Differential Equations*, Collection of Applied Mathematics for the Master's Degree, Masson, Paris. (in French).

Revuz, D. & Yor, M. (1999). *Continuous Martingales and Brownian Motion*, 3rd edn, Springer.

Richtmeyer, R. & Morton, K. (1967). *Difference Methods for Initial-Value Problems*, Interscience, New York.

Ripley, B. D. (1983). *Stochastic Simulation*, Wiley, New York.

Rogers, L. C. G. (1997). The potential approach to the term structure of interest rates and their exchange rates, *Math. Finance* **7**: 157–176.

Rogers, L. C. G. & Williams, D. (2000). *Diffusions, Markov Processes and Martingales: Itô Calculus*, Vol. 2 of *Cambridge Mathematical Library*, 2nd edn, Cambridge University Press.

Ross, S. A. (1976). The arbitrage theory of capital asset pricing, *J. Economic Theory* **13**: 341–360.

Ross, S. M. (1990). *A Course in Simulation*, Macmillan.

Rubinstein, M. (1976). The strong case for the generalized logarithmic utility model as the premier model of financial markets, *J. Finance* **31**: 551–571.

Rubinstein, M. (1985). Nonparametric tests of alternative option pricing models using all reported trades and quotes on the 30 most active CBOE option classes from August 23, 1976 through August 31, 1978, *J. Finance* **11**: 455–480.

Samuelson, P. A. (1955). Brownian motion in the stock market. (Unpublished manuscript).

Samuelson, P. A. (1957). Intertemporal price equilibrium: A prologue to the theory of speculation, *Weltwirtschaftliches Archiv* **79**: 181–221.

Samuelson, P. A. (1963). Risk and uncertainty: A fallacy of large numbers, *Scienta* **6**: 1–6.

Samuelson, P. A. (1965a). Proof that properly anticipated prices fluctuate randomly, *Industr. Manag. Rev.* **6**: 41–50.

Samuelson, P. A. (1965b). Rational theory of warrant pricing, *Industrial Management Review* **6**(2): 13–32.

Samuelson, P. A. (1969). Life time portfolio selection, *Rev. Econom. Statist.* **51**: 239–246.

Samuelson, P. A. (1971). The fallacy of maximising the geometric mean in long sequences of investing or gambling, *Proceedings of the National Academy of Science* pp. 2493–2496.

Samuelson, P. A. (1979). Why we should not make mean log of wealth big though years to act are long, *J. Banking and Finance* **3**: 305–307.

Sandmann, K. & Sondermann, D. (1994). On the stability of lognormal interest rate models. Discussion paper No. B-263, University Bonn.

Schönbucher, P. J. (1999). A market model for stochastic implied volatility, *Philos. Trans. Roy. Soc. London Ser. A* **357**: 2071–2092.

Schönbucher, P. J. (2003). *Credit Derivatives Pricing Models*, Wiley, Chichester.

Schroder, M. (1989). Computing the constant elasticity of variance option pricing formula, *J. Finance* **44**(1): 211–219.

Schweizer, M. (1995). On the minimal martingale measure and the Föllmer-Schweizer decomposition, *Stochastic Anal. Appl.* **13**: 573–599.

Scott, L. O. (1987). Option pricing when the variance changes randomly: Theory, estimation and an application, *J. Financial and Quantitative Analysis* **22**: 419–438.

Sharpe, W. F. (1964). Capital asset prices: A theory of market equilibrium under conditions of risk, *J. Finance* **19**: 425–442.

Sharpe, W. F. (1966). Mutual fund performance, *J. Business* **39**: 119–138.

Shaw, W. (1998). *Pricing Derivatives with Mathematica*, Cambridge University Press.

Shaw, W. (2005). New methods for simulating the Student $t$ distribution - Direct use of the inverse cumulative distribution. Nomura Centre for Quantitative Finance, Oxford (working paper).

Shiga, T. & Watanabe, S. (1973). Bessel diffusion as a one parameter family of diffsuion processes, *Z. Wahrsch. Verw. Gebiete* **27**: 37–46.

Shiryaev, A. N. (1984). *Probability*, Springer.

Sklar, A. (1959). Fonction de repartition à $n$-dimensions et leur marges, *Publications de l'Institute Statistique de l'Université de Paris* **8**: 229–231.

Smith, G. D. (1985). *Numerical Solution of Partial Differential Equations: Finite Difference Methods*, 3rd edn, Clarendon Press, Oxford.

Stein, E. M. & Stein, J. C. (1991). Stock price distributions with stochastic volatility: An analytic approach, *Rev. Financial Studies* **4**: 727–752.

Stroock, D. W. & Varadhan, S. R. S. (1982). *Multidimensional Diffusion Processes*, Vol. 233 of *Grundlehren Math. Wiss.*, Springer.

Stuart, A. & Ord, J. K. (1994). *Kendall's Advanced Theory of Statistics. Vol. 1. Distribution Theory*, 6th edn, Edward Arnold, London.

Stutzer, M. J. (2000). A portfolio performance index, *Financial Analysts J.* **May/June**: 52–61.

Talay, D. & Tubaro, L. (1990). Expansion of the global error for numerical schemes solving stochastic differential equations, *Stochastic Anal. Appl.* **8**(4): 483–509.

Tanaka, H. (1963). Note on continuous additive functionals of the 1-dimensional Brownian path, *Z. Wahrscheinlichkeitstheorie* **1**: 251–257.

Tavella, D. & Randall, C. (2000). *Pricing Financial Instruments. The Finite Difference Method*, Wiley, New York.

Taylor, S. J. & Xu, X. (1994). The magnitude of implied volatility smiles: Theory and empirical evidence for exchange rates, *Rev. Futures Markets* **13**: 355–380.

Thorp, E. O. (1961). A favourable strategy for twenty-one, *Proceedings of the National Academy of Sciences*, Vol. 47, pp. 110–112.

Thorp, E. O. (1972). Portfolio choice and the Kelly criterion, *Proceedings of the 1971 Business and Economics Section of the American Statistical Association*, Vol. 21, pp. 5–224.

Tobin, J. (1958a). Estimation of relationships for limited dependent variables, *Econometrica* **26**: 24–36.

Tobin, J. (1958b). Liquidity preference as behaviour towards risk, *Rev. of Economic Studies* **25**(2): 65–86.

van der Hoek, J. & Elliott, R. J. (2006). *Binomial Methods in Finance*, Springer Finance, Springer.

Vasicek, O. A. (1977). An equilibrium characterization of the term structure, *J. Financial Economics* **5**: 177–188.

von Neumann, J. & Morgenstern, O. (1953). *The Theory of Games and Economic Behavior*, 3rd edn, Princeton University Press, NJ.

Wagner, W. & Platen, E. (1978). Approximation of Itô integral equations. Preprint ZIMM, Akad. Wissenschaften, DDR, Berlin.

Wei, L. & Wu, R. (2002). The joint distributions of several important actuarial diagnostics in the classical risk model, *Insurance: Mathematics and Economics* **30**: 451–462.

Wiggins, J. B. (1987). Option values under stochastic volatility. Theory and empirical estimates, *J. Financial Economics* **19**: 351–372.

Williams, J. B. (1936). Speculation and the carryover, *The Quarterly Journal of Economics* **50**(3): 436–455.

Wilmott, P., Dewynne, J. & Howison, S. (1993). *Option Pricing: Mathematical Models and Computation*, Oxford Financial Press.

Yamada, T. & Watanabe, S. (1971). On the uniqueness of solutions of stochastic differential equations, *J. Math. Kyoto Univ.* **11**: 155–167.

Zhao, Y. & Ziemba, W. T. (2003). Intertemporal mean-variance efficiency with a Markovian state price density. (working paper, University of British Columbia).

Ziemba, W. T. & Mulvey, J. M. (1998). *Worldwide Asset and Liability Modeling*, Cambridge University Press.

# Author Index




E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,

# Index



E. Platen, D. Heath, *A Benchmark Approach to Quantitative Finance*,
Springer Finance,