OXFORD

fixed income modelling

CLAUS MUNK

FIXED INCOME MODELLING

This page intentionally left blank

Fixed Income Modelling

CLAUS MUNK

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford OX2 6DP

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide in

Oxford New York

Auckland Cape Town Dar es Salaam Hong Kong Karachi
Kuala Lumpur Madrid Melbourne Mexico City Nairobi
New Delhi Shanghai Taipei Toronto

With offices in

Argentina Austria Brazil Chile Czech Republic France Greece
Guatemala Hungary Italy Japan Poland Portugal Singapore
South Korea Switzerland Thailand Turkey Ukraine Vietnam

Oxford is a registered trade mark of Oxford University Press
in the UK and in certain other countries

Published in the United States
by Oxford University Press Inc., New York

First published 2011

British Library Cataloguing in Publication Data

Data available

Library of Congress Cataloging in Publication Data

Library of Congress Control Number: 2011929415

Typeset by SPI Publisher Services, Pondicherry, India
Printed in Great Britain
on acid-free paper by
MPG Books Group, Bodmin and King's Lynn

ISBN 978–0–19–957508–4

1 3 5 7 9 10 8 6 4 2

Preface

This book provides a unified presentation of dynamic term structure models and their applications to the pricing and risk management of fixed income securities. The book can serve as a textbook in a specialized course on fixed income or in a broader course on derivatives in continuation of an introductory derivatives textbook. The targeted students are Master of Science students specializing in finance and/or economics, quantitatively oriented MBA students, and first- or second-year finance PhD students. Students are expected to have taken a broad intermediate finance course and to have some formal training in mathematics and statistics so that they can handle basic probability theory, general algebra, and vectors and matrices. Students will benefit from earlier exposure to basic option pricing models, such as the binomial model and the Black–Scholes model for stock option pricing. In addition to serving as a textbook, the book will also be a useful reference for researchers and finance professionals.

The overall purpose of the book is that, after reading through it, the reader will

- have a solid understanding of the economics of the term structure of interest rates and how various interest rate models are founded (or not founded) in economic theory,
- be familiar with the basic fixed income securities and their properties and uses as well as the relations between those securities,
- have an in-depth understanding of classical affine models, HJM models, and LIBOR market models, how to apply those models for interest rate risk management and for the pricing of various widely traded fixed income securities, and how the various models compare, and
- be able to apply various numerical techniques for fixed income pricing and risk management.

There are already plenty of books covering various aspects of fixed income markets and models. The key distinctive features of this book are the following:

- A balanced presentation offering both formal mathematical modelling and economic intuition and understanding. A separate chapter gives a thorough and accessible introduction to stochastic processes and the stochastic calculus needed for the modern financial modelling approach used in the book.
- A separate chapter explaining how the term structure of interest rates relates to macro-economic variables such as inflation and aggregate consumption and production, and to what extent the concrete interest rate models are founded in general economic theory.
- The book provides a detailed analysis of the pricing of the most important fixed income securities in various models, but does not try to give a

complete overview of the hundreds of more or less exotic real-life products. The focus is on the main pricing principles and techniques and their application to the most important types of securities, then the reader should be able to adapt those principles and techniques to more or less any specific product.

- The book has a detailed coverage of the main classes of models and the main concrete models in each class, but does not try to give an exhaustive presentation of the countless models suggested in the literature.
- A separate chapter on interest rate risk measurement and management provides a critical review of the traditional approach to that topic and also introduces and discusses interest rate risk measures that are better founded in the modern term structure models and how those measures are applied in risk management.
- A chapter on mortgage-backed securities describes this increasingly important class of contracts and presents and discusses some simple models for the pricing of these fairly complex assets.
- Each chapter ends with a number of problems.

Readers looking for more detailed descriptions of fixed income markets, products, and basic concepts are referred to Fabozzi (2010). More information about the practical application and implementation of some dynamic term structure models can be found in Brigo and Mercurio (2006). References to other relevant textbooks and original research articles are supplied throughout the book.

The book is based on a set of lecture notes that I began writing more than ten years ago and then gradually developed into a coherent and comprehensive treatment of fixed income modelling. A substantial part of the book was written while I was still at the Department of Business and Economics at the University of Southern Denmark. The book was completed after I joined the School of Economics and Management and the Department of Mathematical Sciences at Aarhus University in January 2009. I am indebted to both institutions for providing such stimulating teaching and research environments and for their generous support over the years. I am grateful to Lene Holbæk at the University of Southern Denmark and Thomas Stephansen at Aarhus University for their excellent secretarial assistance. I highly appreciate the comments on and corrections to earlier drafts that I have received from a number of students exposed to the material and from former and current colleagues. In particular, I thank Simon Lysbjerg Hansen for his extensive comments and suggestions of improvements to Chapter 13. I am very grateful to Oxford University Press and all the people I have been in touch with there for their willingness to publish the book and their patience and professional assistance.

I also want to take the opportunity to express my gratitude to Kristian Risgaard Miltersen and Peter Ove Christensen. They did a great job introducing me to modern finance theory and supervising my master's thesis on dynamic term structure models and my PhD thesis on optimal consumption and portfolio choice. Throughout my career, Peter has been an invaluable mentor and academic role model and for that I am extremely grateful. I also thank all the people who I have

worked with on research projects over the years and from whom I have learned so much, in particular Holger Kraft and Carsten Sørensen. Finally, I am deeply grateful to my wife Lene for her continuing love and support.

Claus Munk

Aarhus
April 2011

This page intentionally left blank

Contents

List of Figures

List of Tables

1

Introduction and Overview

1.1 WHAT IS FIXED INCOME ANALYSIS?

This book develops and studies techniques and models that are helpful in the analysis of fixed income securities. It is difficult to give a clear-cut and universally accepted definition of the term 'fixed income security'. Certainly, the class of fixed income securities includes securities where the issuer promises one or several fixed, predetermined payments at given points in time. This is the case for standard deposit arrangements and bonds. However, we will also consider several related securities as being fixed income securities, although the payoffs of such a security are typically not fixed and known at the time when the investor purchases the security, but depend on the future development in some particular interest rate or the price of some basic fixed income security. In this broader sense of the term, the many different interest rate and bond derivatives are also considered fixed income securities, for example options and futures on bonds or interest rates, caps and floors, swaps and swaptions.

The prices of many fixed income securities are often expressed in terms of various interest rates and yields, so understanding fixed income pricing is equivalent to understanding interest rate behaviour. The key concept in the analysis of fixed income securities and interest rate behaviour is really **the term structure of interest rates**. The interest rate on a loan will normally depend on the maturity of the loan, and in the bond markets there will often be differences between the yields on short-term bonds and long-term bonds. Loosely, the term structure of interest rates is defined as the dependence between interest rates and Maturities. We will be more concrete later on.

We split the overall analysis into two parts which are clearly related to each other. The first part focuses on the economics of the term structure of interest rates, in the sense that the aim is to explore the relations between interest rates and other macroeconomic variables such as aggregate consumption, production, inflation, and money supply. This will help us understand the level of bond prices and interest rates and the shape of the term structure of interest rates at a given point in time and it will give us some tools for understanding and studying the reactions of interest rates and prices to macroeconomic news and shocks. The second part of the analysis focuses on developing tools and models for the pricing and risk management of the many different fixed income securities. Such models are used in all modern financial institutions that trade fixed income securities or are otherwise concerned with the dynamics of interest rates.

In this chapter we will first introduce some basic concepts and terminology and discuss how the term structure of interest rates can be represented in various equivalent ways. In Section 1.3 we take a closer look at the bond and money markets across the world. Among other things we will discuss the size of different markets, the distinction between domestic and international markets, and who the issuers of bonds are. Section 1.4 briefly introduces some fixed income derivatives. Finally, a detailed outline of the rest of the book is given in Section 1.5.

1.2 BASIC BOND MARKET TERMINOLOGY

The simplest fixed income securities are bonds. A bond is nothing but a tradable loan agreement. The issuer sells a contract promising the holder a predetermined payment schedule. Bonds are issued by governments, private and public corporations, and financial institutions. Most bonds are subsequently traded at organized exchanges or over-the-counter (OTC). Bond investors include pension funds and other financial institutions, the central banks, corporations, and households. Bonds are traded with various maturities and with various types of payment schedule. In the so-called money markets major financial institutions offer various bond-like loan agreements of a maturity of less than one year. Below, we will introduce some basic concepts and terminology.

1.2.1 Bond types

We distinguish between zero-coupon bonds and coupon bonds. A **zero-coupon bond** is the simplest possible bond. It promises a single payment at a single future date, the maturity date of the bond. Bonds which promise more than one payment when issued are referred to as **coupon bonds**. We will assume throughout that the face value of any bond is equal to 1 unit of account unless stated otherwise. For concreteness, we will often refer to the unit of account as a 'dollar'. Suppose that at some date t a zero-coupon bond with maturity $T \geq t$ is traded in the financial markets at a price of B_t^T. This price reflects the market **discount factor** for sure time T payments. If many zero-coupon bonds with different maturities are traded, we can form the function $T \mapsto B_t^T$, which we call the market **discount function** prevailing at time t. Note that $B_t^t = 1$, since the value of getting 1 dollar right away is 1 dollar, of course. Presumably, all investors will prefer getting 1 dollar at some time T rather than at a later time S. Therefore, the discount function should be decreasing, that is

$$1 \geq B_t^T \geq B_t^S \geq 0, \quad t \leq T \leq S.$$

A coupon bond has multiple payment dates, which we will generally denote by $T_1, T_2, \ldots, T_n$. Without loss of generality we assume that $T_1 < T_2 < \cdots < T_n$. The payment at date T_i is denoted by Y_i. For almost all traded coupon bonds the payments occur at regular intervals so that, for all i, $T_{i+1} - T_i = \delta$ for some fixed δ. If we measure time in years, typical bonds have $\delta \in \{0.25, 0.5, 1\}$

corresponding to quarterly, semi-annual, or annual payments. The size of each of the payments is determined by the face value, the coupon rate, and the amortization principle of the bond. The face value is also known as the par value or principal of the bond, and the coupon rate is also called the nominal rate or stated interest rate. In many cases, the coupon rate is quoted as an annual rate even when payments occur more frequently. If a bond with a payment frequency of δ has a quoted coupon rate of R, this means that the periodic coupon rate is δR.

Most coupon bonds are so-called **bullet bonds** or **straight-coupon bonds** where all the payments before the final payment are equal to the product of the coupon rate and the face value. The final payment at the maturity date is the sum of the same interest rate payment and the face value. If R denotes the periodic coupon rate, the payments per unit of face value are therefore

$$Y_i = \begin{cases} R, & i = 1, \dots, n-1 \\ 1+R, & i = n \end{cases}$$

Of course for $R = 0$ we are back to the zero-coupon bond.

Other bonds are so-called **annuity bonds**, which are constructed so that the total payment is equal for all payment dates. Each payment is the sum of an interest payment and a partial repayment of the face value. The outstanding debt and the interest payment are gradually decreasing over the life of an annuity, so that the repayment increases over time. Let R again denote the periodic coupon rate. Assuming a face value of 1, the constant periodic payment is

$$Y_i = Y \equiv \frac{R}{1-(1+R)^{-n}}, \quad i = 1, \dots, n.$$

The outstanding debt of the annuity immediately after the i'th payment is

$$D_i = Y \frac{1-(1+R)^{-(n-i)}}{R},$$

the interest part of the i'th payment is

$$I_i = R D_{i-1} = R \frac{1-(1+R)^{-(n-i+1)}}{1-(1+R)^{-n}},$$

and the repayment part of the i'th payment is

$$X_i = Y\,(1+R)^{-(n-i+1)}$$

so that $X_i + I_i = Y$.

Some bonds are so-called **serial bonds** where the face value is paid back in equal instalments. The payment at a given payment date is then the sum of the instalment and the interest rate on the outstanding debt. The interest rate payments, and hence the total payments, will therefore decrease over the life of the bond. With a face value of 1, each instalment or repayment is $X_i = 1/n$, $i = 1, \dots, n$. Immediately after the i'th payment date, the outstanding debt must be $D_i = (n-i)/n = 1 - (i/n)$. The interest payment at T_i is therefore $I_i = R D_{i-1} = R\,(1 - (i-1)/n)$. Consequently, the total payment at T_i must be

$$Y_i = X_i + I_i = \frac{1}{n} + R\left(1 - \frac{i-1}{n}\right).$$

Finally, few bonds are **perpetuities** or **consols** that last forever and only pay interest, that is $Y_i = R$, $i = 1, 2, \dots$. The face value of a perpetuity is never repaid.

Most coupon bonds have a fixed coupon rate, but a small minority of bonds have coupon rates that are reset periodically over the life of the bond. Such bonds are called *floating rate bonds*. Typically, the coupon rate effective for the payment at the end of one period is set at the beginning of the period at the current market interest rate for that period, for example to the 6-month interest rate for a floating rate bond with semi-annual payments. We will look more closely at the valuation of floating rate bonds in Section 1.2.5.

A coupon bond can be seen as a portfolio of zero-coupon bonds, namely a portfolio of Y_1 zero-coupon bonds maturing at T_1, Y_2 zero-coupon bonds maturing at T_2, and so on. If all these zero-coupon bonds are traded in the market, the price of the coupon bond at any time t must be

$$B_t = \sum_{T_i > t} Y_i B_t^{T_i}, \tag{1.1}$$

where the sum is over all future payment dates of the coupon bond. If this relation does not hold, there will be a clear arbitrage opportunity in the market. The absence of arbitrage is a cornerstone of financial asset pricing, since a market in which prices allow for the construction of arbitrage opportunities cannot be in equilibrium. More on arbitrage and asset pricing theory in Chapter 4.

Example 1.1 Consider a bullet bond with a face value of 100, a coupon rate of 7%, annual payments, and exactly 3 years to maturity. Suppose zero-coupon bonds are traded with face values of 1 dollar and time-to-maturity of 1, 2, and 3 years, respectively. Assume that the prices of these zero-coupon bonds are $B_t^{t+1} = 0.94$, $B_t^{t+2} = 0.90$, and $B_t^{t+3} = 0.87$. According to (1.1), the price of the bullet bond must then be

$$B_t = 7 \cdot 0.94 + 7 \cdot 0.90 + 107 \cdot 0.87 = 105.97.$$

If the price is lower than 105.97, risk-free profits can be locked in by buying the bullet bond and selling 7 1-year, 7 2-year, and 107 3-year zero-coupon bonds. If the price of the bullet bond is higher than 105.97, sell the bullet bond and buy 7 1-year, 7 2-year, and 107 3-year zero-coupon bonds.

If not all the relevant zero-coupon bonds are traded, we cannot justify the relation (1.1) as a result of the no-arbitrage principle. Still, it is a valuable relation. Suppose that an investor has determined (from private or macroeconomic information) a discount function showing the value *she* attributes to payments at different future points in time. Then she can value all sure cash flows in a consistent way by substituting that discount function into (1.1).

The market prices of all bonds reflect a market discount function, which is the result of the supply of and demand for the bonds of all market participants. We can think of the market discount function as a very complex average of the individual discount functions of the market participants. In most markets only few zero-coupon bonds are traded, so that information about the discount function must be inferred from market prices of coupon bonds. We discuss ways of doing that in Chapter 2.

1.2.2 Bond yields and zero-coupon rates

Although discount factors provide full information about how to discount amounts back and forth, it is pretty hard to relate to a 5-year discount factor of 0.7835. It is far easier to relate to the information that the 5-year interest rate is 5%. Interest rates are always quoted on an annual basis, that is as some percentage per year. However, to apply and assess the magnitude of an interest rate, we also need to know the compounding frequency of that rate. More frequent compounding of a given interest rate per year results in higher 'effective' interest rates. Furthermore, we need to know at which time the interest rate is set or observed and for which period of time the interest rate applies. First we consider spot rates which apply to a period beginning at the time the rate is set. In the next subsection, we consider forward rates which apply to a future period of time.

The **yield** of a bond is the discount rate which has the property that the present value of the future payments discounted at that rate is equal to the current price of the bond. The convention in many bond markets is to quote rates using annual compounding. For a coupon bond with current price B_t and payments $Y_1, \dots, Y_n$ at time $T_1, \dots, T_n$, respectively, the annually compounded yield is then the number $\hat{y}_t^B$ satisfying the equation

$$B_t = \sum_{T_i > t} Y_i \left(1 + \hat{y}_t^B\right)^{-(T_i - t)}.$$

Note that the same discount rate is applied to all payments. In particular, for a zero-coupon bond with a payment of 1 at time T, the annually compounded yield $\hat{y}_t^T$ at time $t \leq T$ is such that

$$B_t^T = (1 + \hat{y}_t^T)^{-(T-t)}$$

and, consequently,

$$\hat{y}_t^T = \left(B_t^T\right)^{-1/(T-t)} - 1.$$

We call $\hat{y}_t^T$ the time t **zero-coupon yield**, **zero-coupon rate**, or **spot rate** for date T. The zero-coupon rates as a function of maturity is called the **zero-coupon yield curve** or simply the **yield curve**. It is one way to express the term structure of interest rates. Due to the one-to-one relationship between zero-coupon bond prices and zero-coupon rates, the discount function $T \mapsto B_t^T$ and the zero-coupon yield curve $T \mapsto \hat{y}_t^T$ carry exactly the same information.

Figure 1.1 shows yields on U.S. government bonds of maturities of 1, 5, and 10 years over the period from January 1954 to February 2010. Note the high vari-

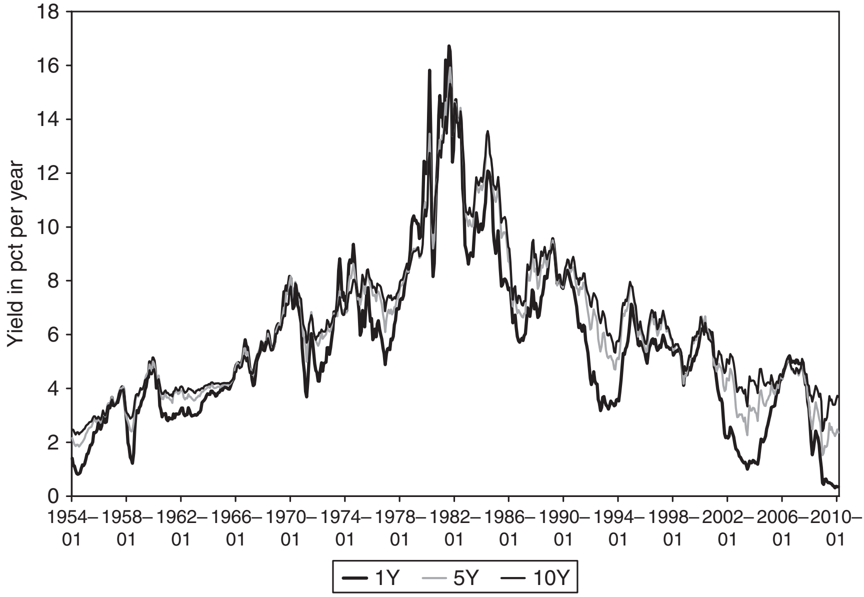

Fig. 1.1: Yields of nominal U.S. Treasury bonds of maturities of 1 year, 5 years, and 10 years from January 1954 to February 2010.

Source: the homepage of the Federal Reserve (`www.federalreserve.gov`) at 2 March, 2010.

ability of the level of interest rates over the full period, whereas over shorter periods yields are quite persistent. Short-maturity yields are more volatile than long-maturity yields. Most of the time, the 1-year yield is less than the 5-year yield which again is less than the 10-year yield indicating that the yield curve is typically upward-sloping. This is also reflected by Fig. 1.2, which shows U.S. yield curves from January in the years 1985, 1990, 1995, 2000, 2005, and 2010. The yield curve from January 1990 is almost flat, but the other yield curves are upward-sloping at least up to maturities of 5–10 years after which the curves are relatively flat. However, in some rather short time periods, short-maturity yields have been higher than long-maturity yields indicating a downward-sloping or *inverted* yield curve. Sometimes the yield curve is non-monotonic and may exhibit a 'hump' (first increasing to a maximum, then decreasing) or a 'trough' (first decreasing to a minimum, then increasing) or have some even more complex shape.

For some bonds and loans interest rates are quoted using semi-annual, quarterly, or monthly compounding. An interest rate of R per year compounded m times a year, corresponds to a discount factor of $(1 + R/m)^{-m}$ over a year. The annually compounded interest rate that corresponds to an interest rate of R compounded m times a year is $(1 + R/m)^m - 1$. This is sometimes called the "effective" interest rate corresponding to the nominal interest rate R. This convention is typically applied for interest rates set for loans at the international money markets, the most commonly used being the LIBOR (London InterBank Offered Rate) rates that are fixed in London. The compounding period equals the maturity of the loan with

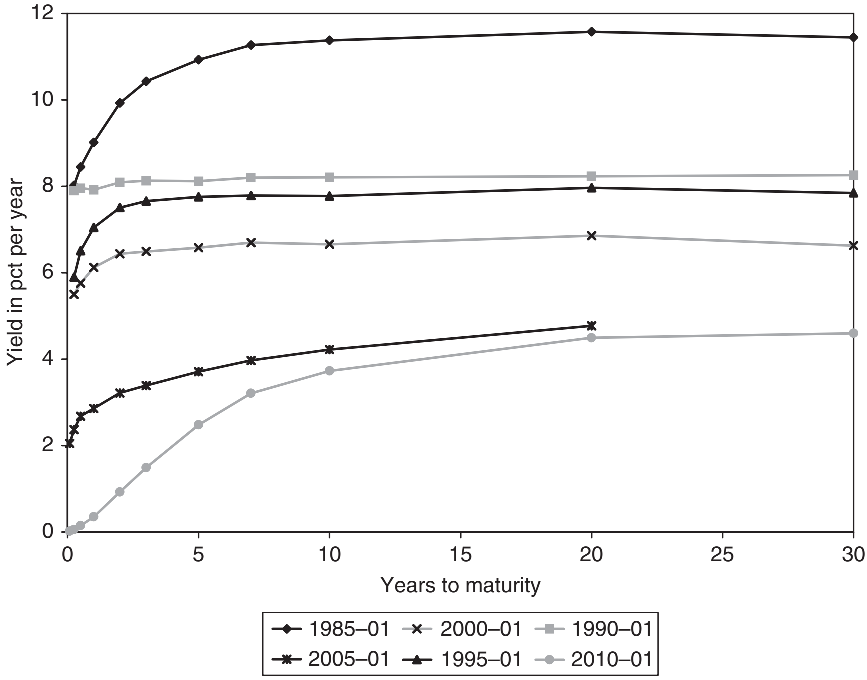

Fig. 1.2: Yield curves determined from nominal U.S. Treasury bonds in January of 1985, 1990, 1995, 2000, 2005, and 2010. The curves are drawn by connecting reported yields for maturities of 1, 3, and 6 months as well as 1, 2, 3, 5, 7, 10, 20, and 30 years. For the curves from 1985, 1990, 1995, and 2000 there is no data for the 1-month yield. For the 1990 curve the 20-year yield is obtained by interpolating the reported 10-year and 30-year yields. For the 2005 curve there is no data for the 30-year yield.

Source: the homepage of the Federal Reserve (`www.federalreserve.gov`) at 2 March, 2010.

3, 6, or 12 months as the most frequently used maturities. If the quoted annualized rate for say a 3-month loan is $l_t^{t+0.25}$, it means that the 3-month interest rate is $l_t^{t+0.25}/4 = 0.25 l_t^{t+0.25}$ so that the present value of 1 dollar paid 3 months from now is

$$B_t^{t+0.25} = \frac{1}{1 + 0.25\, l_t^{t+0.25}}.$$

Hence, the 3-month rate is

$$l_t^{t+0.25} = \frac{1}{0.25}\left(\frac{1}{B_t^{t+0.25}} - 1\right).$$

More generally, the relations are

$$B_t^T = \frac{1}{1 + l_t^T (T - t)} \tag{1.2}$$

and

$$l_t^T = \frac{1}{T-t}\left(\frac{1}{B_t^T} - 1\right).$$

We shall use the term **LIBOR rates** for interest rates that are quoted in this way. Note that if we had a full LIBOR rate curve $T \mapsto l_t^T$, this would carry exactly the same definition as the discount function $T \mapsto B_t^T$. Some fixed income securities provide payoffs that depend on future values of LIBOR rates. In order to price such securities it is natural to model the dynamics of LIBOR rates and this is exactly what is done in one popular class of term structure models (see Chapter 11).

Increasing the compounding frequency m, the effective annual return of 1 dollar invested at the interest rate R per year increases to e^R, due to the mathematical result saying that

$$\lim_{m\to\infty}\left(1 + \frac{R}{m}\right)^m = e^R.$$

A nominal, continuously compounded interest rate R is equivalent to an annually compounded interest rate of $e^R - 1$ (which is bigger than R). Similarly, the zero-coupon bond price B_t^T is related to the continuously compounded zero-coupon rate y_t^T by

$$B_t^T = e^{-y_t^T(T-t)} \tag{1.3}$$

so that

$$y_t^T = -\frac{1}{T-t}\ln B_t^T.$$

The function $T \mapsto y_t^T$ is also a zero-coupon yield curve that contains exactly the same information as the discount function $T \mapsto B_t^T$ and also the same information as the annually compounded yield curve $T \mapsto \hat{y}_t^T$ (or the yield curve with any other compounding frequency). We have the following relation between the continuously compounded and the annually compounded zero-coupon rates:

$$y_t^T = \ln(1 + \hat{y}_t^T).$$

For mathematical convenience we will focus on the continuously compounded yields in most models.

1.2.3 Forward rates

While a zero-coupon or spot rate reflects the price on a loan between today and a given future date, a **forward rate** reflects the price on a loan between two future dates. The annually compounded relevant forward rate at time t for the period between time T and time S is denoted by $\hat{f}_t^{T,S}$. Here, we have $t \le T < S$. This is the rate which is appropriate at time t for discounting between time T and S. We can think of discounting from time S back to time t by first discounting from time S to time T and then discounting from time T to time t. We must therefore have that

$$\left(1+\hat{y}_t^S\right)^{-(S-t)} = \left(1+\hat{y}_t^T\right)^{-(T-t)}\left(1+\hat{f}_t^{T,S}\right)^{-(S-T)}, \tag{1.4}$$

from which we find that

$$\hat{f}_t^{T,S} = \frac{(1+\hat{y}_t^T)^{-(T-t)/(S-T)}}{(1+\hat{y}_t^S)^{-(S-t)/(S-T)}} - 1.$$

We can also write (1.4) in terms of zero-coupon bond prices as

$$B_t^S = B_t^T\left(1+\hat{f}_t^{T,S}\right)^{-(S-T)}, \tag{1.5}$$

so that the forward rate is given by

$$\hat{f}_t^{T,S} = \left(\frac{B_t^T}{B_t^S}\right)^{1/(S-T)} - 1.$$

Note that since $B_t^t = 1$, we have

$$\hat{f}_t^{t,S} = \left(\frac{B_t^t}{B_t^S}\right)^{1/(S-t)} - 1 = \left(B_t^S\right)^{-1/(S-t)} - 1 = \hat{y}_t^S,$$

that is the forward rate for a period starting today equals the zero-coupon rate or spot rate for the same period.

Again, we may use periodic compounding. For example, a 6-month forward LIBOR rate of $L_t^{T,T+0.5}$ valid for the period $[T, T+0.5]$ means that the discount factor is

$$B_t^{T+0.5} = B_t^T\left(1+0.5\,L_t^{T,T+0.5}\right)^{-1}$$

so that

$$L_t^{T,T+0.5} = \frac{1}{0.5}\left(\frac{B_t^T}{B_t^{T+0.5}} - 1\right).$$

More generally, the time t forward LIBOR rate for the period $[T, S]$ is given by

$$L_t^{T,S} = \frac{1}{S-T}\left(\frac{B_t^T}{B_t^S} - 1\right). \tag{1.6}$$

If $f_t^{T,S}$ denotes the continuously compounded forward rate prevailing at time t for the period between T and S, we must have that

$$B_t^S = B_t^T e^{-f_t^{T,S}(S-T)},$$

by analogy with (1.5). Consequently,

$$f_t^{T,S} = -\frac{\ln B_t^S - \ln B_t^T}{S-T}. \tag{1.7}$$

Using (1.3), we get the following relation between zero-coupon rates and forward rates under continuous compounding:

$$f_t^{T,S} = \frac{y_t^S(S-t) - y_t^T(T-t)}{S-T}. \tag{1.8}$$

In the following chapters, we shall often focus on forward rates for future periods of infinitesimal length. The forward rate for an infinitesimal period starting at time T is simply referred to as the forward rate for time T and is defined as $f_t^T = \lim_{S\to T} f_t^{T,S}$. The function $T \mapsto f_t^T$ is called the **term structure of forward rates** or the **forward rate curve**. Letting $S \to T$ in the expression (1.7), we get

$$f_t^T = -\frac{\partial \ln B_t^T}{\partial T} = -\frac{\partial B_t^T/\partial T}{B_t^T}, \tag{1.9}$$

assuming that the discount function $T \mapsto B_t^T$ is differentiable. Conversely,

$$B_t^T = e^{-\int_t^T f_t^u\, du}. \tag{1.10}$$

Note that a full term structure of forward rates $T \mapsto f_t^T$ contains the same information as the discount function $T \mapsto B_t^T$.

Applying (1.8), the relation between the infinitesimal forward rate and the spot rates can be written as

$$f_t^T = \frac{\partial [y_t^T(T-t)]}{\partial T} = y_t^T + \frac{\partial y_t^T}{\partial T}(T-t)$$

under the assumption of a differentiable term structure of spot rates $T \mapsto y_t^T$. The forward rate reflects the slope of the zero-coupon yield curve. In particular, the forward rate f_t^T and the zero-coupon rate y_t^T will coincide if and only if the zero-coupon yield curve has a horizontal tangent at T. Conversely, we see from (1.10) and (1.3) that

$$y_t^T = \frac{1}{T-t}\int_t^T f_t^u\, du, \tag{1.11}$$

that is the zero-coupon rate is an average of the forward rates.

1.2.4 The term structure of interest rates in different disguises

We emphasize that discount factors, spot rates, and forward rates (with any compounding frequency) are perfectly equivalent ways of expressing the same information. If a complete yield curve of, say, quarterly compounded spot rates is given, we can compute the discount function and spot rates and forward rates for any given period and with any given compounding frequency. If a complete term structure of forward rates is known, we can compute discount functions and spot rates, and so on. Academics frequently apply continuous compounding since the mathematics involved in many relevant computations is more elegant when exponentials are used, but continuously compounded rates can easily be transformed to any other compounding frequency.

There are even more ways of representing the term structure of interest rates. Since most bonds are bullet bonds, many traders and analysts are used to thinking in terms of yields of bullet bonds rather than in terms of discount factors or zero-coupon rates. The **par yield** for a given maturity is the coupon rate that causes a bullet bond of the given maturity to have a price equal to its face value. Again we have to fix the coupon period of the bond. U.S. treasury bonds typically have semi-annual coupons which are therefore often used when computing par yields. Given a discount function $T \mapsto B_t^T$, the n-year par yield is the value of c that solves the equation

$$\sum_{i=1}^{2n} \left(\frac{c}{2}\right) B_t^{t+0.5i} + B_t^{t+n} = 1.$$

It reflects the current market interest rate for an n-year bullet bond. The par yield is closely related to the so-called swap rate, which is a key concept in the swap markets, compare Section 6.5.

1.2.5 Floating rate bonds

Floating rate bonds have coupon rates that are reset periodically over the life of the bond. We will consider the most common floating rate bond, which is a bullet bond, where the coupon rate effective for the payment at the end of one period is set at the beginning of the period at the current market interest rate for that period.

Assume again that the payment dates of the bond are $T_1 < \cdots < T_n$, where $T_i - T_{i-1} = \delta$ for all i. The annualized coupon rate valid for the period $[T_{i-1}, T_i]$ is the δ-period market rate at date T_{i-1} computed with a compounding frequency of δ. We will denote this interest rate by $l_{T_{i-1}}^{T_i}$, although the rate is not necessarily a LIBOR rate, but can also be a Treasury rate. If the face value of the bond is H, the payment at time T_i $(i = 1, 2, \ldots, n-1)$ equals $H\delta l_{T_{i-1}}^{T_i}$, while the final payment at time T_n equals $H(1 + \delta l_{T_{n-1}}^{T_n})$. If we define $T_0 = T_1 - \delta$, the dates $T_0, T_1, \ldots, T_{n-1}$ are often referred to as the reset dates of the bond.

We will argue that immediately after each reset date, the value of the bond will equal its face value. To see this, first note that immediately after the last reset date T_{n-1}, the bond is equivalent to a zero-coupon bond with a coupon rate equal to the market interest rate for the last coupon period. By definition of that market interest rate, the time T_{n-1} value of the bond will be exactly equal to the face value H. In mathematical terms, the market discount factor to apply for the discounting of time T_n payments back to time T_{n-1} is $(1 + \delta l_{T_{n-1}}^{T_n})^{-1}$, so the time T_{n-1} value of a payment of $H(1+\delta l_{T_{n-1}}^{T_n})$ at time T_n is precisely H. Immediately after the next-to-last reset date T_{n-2}, we know that we will receive a payment of $H\delta l_{T_{n-2}}^{T_{n-1}}$ at time T_{n-1} and that the time T_{n-1} value of the following payment (received at T_n) equals H. We therefore have to discount the sum $H\delta l_{T_{n-2}}^{T_{n-1}} + H = H(1+\delta l_{T_{n-2}}^{T_{n-1}})$ from T_{n-1} back to T_{n-2}. The discounted value is exactly H. Continuing this procedure, we get that immediately after a reset of the coupon rate, the floating rate bond is valued at par. Note that it is crucial for this result that the coupon rate is adjusted to the interest rate considered by the market to be 'fair'.

We can also derive the value of the floating rate bond between two payment dates. Suppose we are interested in the value at some time t between T_0 and T_n. Introduce the notation

$$i(t) = \min\{i \in \{1, 2, \dots, n\} : T_i > t\}, \tag{1.12}$$

so that $T_{i(t)}$ is the nearest following payment date after time t. We know that the following payment at time $T_{i(t)}$ equals $H\delta l_{T_{i(t)-1}}^{T_{i(t)}}$ and that the value at time $T_{i(t)}$ of all the remaining payments will equal H. The value of the bond at time t will then be

$$B_t^{\text{fl}} = H(1+\delta l_{T_{i(t)-1}}^{T_{i(t)}}) B_t^{T_{i(t)}}, \quad T_0 \le t < T_n. \tag{1.13}$$

This expression also holds at payment dates $t = T_i$, where it results in H, which is the value excluding the payment at that date.

Relatively few floating rate bonds are traded, but the results above are also very useful for the analysis of interest rate swaps studied in Section 6.5.

1.3 BOND MARKETS AND MONEY MARKETS

This section will give an overview of the bond and money markets across the world. We can distinguish between national markets and international markets. In the national market of a country, bonds primarily issued by domestic issuers and aimed at domestic investors are traded, although some bonds issued by certain foreign governments or corporations or international associations are often also traded. The bonds issued in a given national market must comply with the regulation of that particular country. The international bond market is often referred to as the Eurobond market. A Eurobond can have an issuer located in one country, be listed on the exchange in a second country, and be denominated in the currency of a third country. Eurobonds are usually underwritten by an international syndicate and offered to investors in several countries simultaneously. The Eurobond market is less regulated than most national markets. Eurobonds are typically listed on one national exchange, but most of the trading in these bonds takes place in rather well-organized OTC markets. Other Eurobonds are issued as a private placement with financial institutions. Eurobonds are typically issued by international institutions, governments, or large multi-national corporations.

The Bank for International Settlements (BIS) regularly publishes statistics on financial markets across the world. BIS distinguishes between domestic debt and international debt securities. The term 'debt securities' covers both bonds and money market contracts. The term 'domestic' means that the security is issued in the local currency by residents in that country and targeted at resident investors. All other debt securities are classified by BIS as 'international'. Based on BIS statistics published in Bank for International Settlements (2010), henceforth referred to as BIS (2010), Table 1.1 ranks domestic markets for debt securities according to the amounts outstanding in September 2009. The United States has by far the largest domestic bond market, while Japan is a clear number two. The size of the domestic bond market relative to GDP varies significantly across countries. For example, both Belgium and Denmark have larger domestic bond markets relative to GDP, whereas China and the United Kingdom have smaller bond markets when compared to their GDP. In several countries, including the U.S., the domestic bond market has a size similar to the stock market.

Table 1.2 lists the countries most active when it comes to issuing international debt securities. The domestic markets are significantly larger than the international markets, and international bond markets are much larger than international money markets. Relative to their GDP, European countries such as Germany, the United Kingdom, and the Netherlands have a large share of the international bond and money markets, whereas U.S.-based issuers are relatively inactive. This is also reflected by Table 1.3 which shows that the Euro is the most frequently used currency in the international markets for debt securities, but the U.S. dollar is also used very often.

Table 1.1: The largest domestic markets for debt securities divided by issuer category as of September 2009.

Country	Amounts outstanding (billion USD)	Fraction of world market (%)	Fraction of domestic market (%)		
			governments	financial institut.	corporate issuers
United States	25,105	39.0	36.5	52.3	11.2
Japan	11,602	18.0	83.6	9.6	6.8
Italy	3,770	5.8	54.5	32.3	13.2
France	3,189	4.9	53.1	38.1	8.9
Germany	2,927	4.5	53.3	34.9	11.8
China	2,413	3.7	58.7	28.9	12.4
Spain	2,071	3.2	34.0	30.7	35.3
United Kingdom	1,566	2.4	72.9	25.7	1.4
Canada	1,260	2.0	68.7	20.6	10.7
Brazil	1,227	1.9	65.3	34.0	0.7
South Korea	1,071	1.7	39.7	29.9	30.4
The Netherlands	1,003	1.6	38.4	51.5	10.1
Australia	843	1.3	24.7	70.9	4.4
Belgium	724	1.1	59.9	31.2	8.9
Denmark	589	0.9	16.7	83.0	0.3
All countries	64,448	100.0	52.7	36.4	10.9

Source: Tables 16A-B in BIS (2010).

Table 1.2: International debt securities by nationality of issuer as of December 2009. The numbers are amounts outstanding in billions of USD.

Country	Total	Security (%)		Issuer (%)		
		bonds, notes	money market	governments	financial institut.	corporate issuers
United States	6,712	99.0	1.0	0.2	81.5	18.4
United Kingdom	3,174	96.3	3.7	1.6	89.2	9.2
Germany	2,932	96.1	3.9	10.3	85.2	4.5
France	2,017	95.0	5.0	2.8	77.4	19.9
Spain	1,824	95.1	4.9	7.8	89.1	3.1
Italy	1,402	96.9	3.1	17.5	75.8	6.7
The Netherlands	1,285	93.2	6.8	1.8	92.8	5.4
Ireland	589	91.4	8.6	10.6	87.7	1.7
Belgium	586	95.3	4.7	25.0	69.4	5.6
Canada	569	99.0	1.0	17.6	55.7	26.7
Intl. organizations	802	99.0	1.0	NA	NA	NA
All countries	27,010	96.5	3.5	8.4	77.3	11.3

Source: Tables 12A-D and 15A-B in BIS (2010).

Table 1.3: International debt securities by currency. The numbers are amounts outstanding in billions of USD as of December 2009.

Currency	Bonds and notes	Money market
Euro	12,386	443
US dollar	9,429	320
Pound sterling	2,149	99
Yen	691	17
Swiss franc	366	21
Canadian dollar	306	1
Australian dollar	268	10
Swedish krona	69	2
Hong Kong dollar	61	9
Norwegian krone	54	1
Other currencies	300	10
Total	26,078	932

Source: Tables 13A-B in BIS (2010).

Tables 1.1 and 1.2 split up the different markets according to three categories of issuers: governments, financial institutions, and corporate issuers. On average, close to 53% of the debt securities traded in domestic markets are issued by governments, 36% by financial institutions, and 11% by corporate issuers. Note the large difference across countries. Some domestic markets (for example Japan and the United Kingdom) are dominated by government bonds, others (for example Denmark and Australia) by bonds issued by financial institutions, and corporate bonds are also very common in some countries (for example Spain and South Korea) while virtually non-existent in other countries. The international markets are dominated by financial institutions who stand behind approximately 77% of the issues, 11% are issued by corporations, 8% by governments, and 3% by international organizations. Some governments (for example Belgium, Canada, and Italy) often issue bonds on the international market, while others (for example the United States and the United Kingdom) rarely do so. Let us look more closely at the different issuers and the type of debt securities they typically issue.

Government bonds are bonds issued by the government to finance and refinance the public debt. In most countries, such bonds can be considered to be free of default risk, and interest rates in the government bond market are then a benchmark against which the interest rates on other bonds are measured. However, in some economically or politically unstable countries, the default risk on government bonds cannot be ignored.[1] In the U.S., government bonds are issued by the Department of the Treasury and are referred to as Treasury securities or just Treasuries. These securities are divided into three categories: bills, notes, and bonds. **Treasury bills** (or simply T-bills) are short-term securities that mature in one year or less from their issue date. T-bills are zero-coupon bonds since they have a single payment equal to the face value. **Treasury notes and bonds** are coupon-bearing bullet bonds with semi-annual payments. The only difference between notes and bonds is the time-to-maturity when first issued. Treasury notes are issued with a time-to-maturity of 1–10 years, while Treasury bonds mature in more than 10 years and up to 30 years from their issue date. The Treasury sells two types of notes and bonds: fixed-principal and inflation-indexed. The fixed-principal type promises given dollar payments in the future, whereas the dollar payments of the inflation-indexed type are adjusted to reflect inflation in consumer prices.[2] Finally, the U.S. Treasury also issue so-called **savings bonds** to individuals and certain organizations, but these bonds are not subsequently tradable.

While Treasury notes and bonds are issued as coupon bonds, the Treasury Department introduced the so-called STRIPS program in 1985 that lets investors hold and trade the individual interest and principal components of most Treasury notes and bonds as separate securities.[3] These separate securities, which are usually referred to as STRIPs, are zero-coupon bonds. Market participants create STRIPs by separating the interest and principal parts of a Treasury note or bond. For example, a 10-year Treasury note consists of 20 semi-annual interest payments and a principal payment payable at maturity. When this security is 'stripped', each of the 20 interest payments and the principal payment become separate securities and can be held and transferred separately.

In some countries, including the U.S., bonds issued by various public institutions, for example utility companies, railway companies, export support funds, and so on are backed by the government, so that the default risk on such bonds is the risk that the government defaults. In addition, some bonds are issued by government-sponsored entities created to facilitate borrowing and reduce bor-

[1] This risk is not hypothetical. Tomz and Wright (2007) report that 106 countries have defaulted on their debt a total of 250 times in the period 1820–2004. For more on default risk, see Chapter 13.

[2] The principal value of an inflation-indexed note or bond is adjusted before each payment date according to the change in the consumer price index. Since the semi-annual interest payments are computed as the product of the fixed coupon rate and the current principal, all the payments of an inflation-indexed note or bond are inflation-adjusted.

[3] STRIPS is short for Separate Trading of Registered Interest and Principal of Securities.

rowing costs for, for example farmers, homeowners, and students. However, these bonds are typically not backed by the government and are therefore exposed to the risk of default of the issuing organization. Bonds may also be issued by local governments. In the U.S. such bonds are known as municipal bonds.

In the United States and some other countries, corporations will traditionally raise large amounts of capital by issuing bonds, so-called **corporate bonds**. In other countries, for example Germany and Japan, corporations borrow funds primarily through bank loans, so that the market for corporate bonds is limited. For corporate bonds, investors cannot ignore the possibility that the issuer defaults and cannot meet the obligations represented by the bonds. Bond investors can either perform their own analysis of the creditworthiness of the issuer or rely on the analysis of professional rating agencies such as Moody's Investors Service or Standard & Poor's Corporation. These agencies designate letter codes to bond issuers both in the U.S. and in other countries. Investors will typically treat bonds with the same rating as having (nearly) the same default risk. Due to the default risk, corporate bonds are traded at lower prices than similar (default-free) government bonds. The management of the issuing corporation can effectively transfer wealth from bond-holders to equity-holders, for example by increasing dividends, taking on more risky investment projects, or issuing new bonds with the same or even higher priority in case of default. Corporate bonds are often issued with bond covenants or bond indentures that restrict management from implementing such actions. Default risk, credit ratings, and the valuation of corporate bonds will be thoroughly discussed in Chapter 13.

U.S. corporate bonds are typically issued with maturities of 10–30 years and are often callable bonds, so that the issuer has the right to buy back the bonds on certain terms (at given points in time and for a given price). Some corporate bonds are convertible bonds meaning that the bond-holders may convert the bonds into stocks of the issuing corporation on predetermined terms. Although most corporate bonds are listed on a national exchange, much of the trading in these bonds is in the OTC market.

When commercial banks and other financial institutions issue bonds, the promised payments are sometimes linked to the payments on a pool of loans that the issuing institution has provided to households or firms. An important example is the class of **mortgage-backed bonds** which constitutes a large part of some bond markets, for example in the U.S., Germany, Denmark, Sweden, and Switzerland. A mortgage is a loan that can (partly) finance the borrower's purchase of a given real estate property, which is then used as collateral for the loan. Mortgages can be residential (family houses, apartments, and so on) or non-residential (corporations, farms, and so on). The issuer of the loan (the lender) is a financial institution. Typical mortgages have a maturity between 15 and 30 years and are annuities in the sense that the total scheduled payment (interest plus repayment) at all payment dates are identical. Fixed-rate mortgages have a fixed interest rate, while adjustable-rate mortgages have an interest rate which is reset periodically according to some reference rate. A characteristic feature of most mortgages is the prepayment option. At any payment date in the life of the loan, the borrower has the right to pay off all or part of the outstanding debt. This can occur due to a sale of the underlying real estate property, but can also occur after a drop in market interest rates, since the borrower then has the chance to get a cheaper loan.

Mortgages are pooled either by the issuers or other institutions, who then issue mortgage-backed securities that have an ownership interest in a given pool of mortgage loans. The most common type of mortgage-backed securities is the so-called **pass-through**, where the pooling institution simply collects the payments from borrowers with loans in a given pool and 'passes through' the cash flow to investors less some servicing and guaranteeing fees. Many pass-throughs have payment schemes equal to the payment schemes of bonds, for example pass-throughs issued on the basis of a pool of fixed-rate annuity mortgage loans have a payment schedule equal to that of annuity bonds. However, when borrowers in the pool prepay their mortgage, these prepayments are also passed through to the security-holders, so that their payments will be different from annuities. In general, owners of pass-through securities must take into account the risk that the mortgage borrowers in the pool default on their loans. In the U.S. most pass-throughs have been issued by three organizations that guarantee the payments to the securities even if borrowers default. These organizations are the Government National Mortgage Association (called 'Ginnie Mae'), the Federal Home Loan Mortgage Corporation ('Freddie Mac'), and the Federal National Mortgage Association ('Fannie Mae'). The securities issued by these institutions have generally been considered free of default risk. There is more on mortgages and mortgage-backed securities in Chapter 14.

The **money market** is a market for borrowing and lending large amounts over a period of up to one year. The major players in the money market are financial institutions and large private corporations. The debt contracts issued in the money market are mainly zero-coupon loans, which have a single repayment date. The loans are implemented by the issuance of various instruments. Large corporations, both financial corporations and others, often finance short-term liquidity needs by issuing so-called **commercial papers**. Another standard money market contract is a **repurchase agreement** or simply **repo**. One party of this contract sells a certain asset, such as a short-term Treasury bill, to the other party and promises to buy back that asset at a given future date at the market price at that date. A repo is effectively a collateralized loan, where the underlying asset serves as collateral. As central banks in other countries, the Federal Reserve in the U.S. participates actively in the repo market to implement its monetary policy. The interest rate on repos is called the repo rate. Other popular instruments in the money market are certificates of deposit and foreign exchange swaps, but the money markets also include standard deposits, forward rate agreements, and trading in Treasury bills and short-lived asset-backed securities.

The central bank is a key player in the money market. Banks keep deposits in the central bank to comply with reserve requirements, to facilitate financial transactions, and to manage short-term liquidity. If one bank is in need of additional reserves, it can borrow money (usually overnight) from another bank with surplus reserves at the central bank. The interest rate on the loan is negotiated between the two banks. In the U.S., the weighted average of these interest rates across all such transactions is referred to as the federal funds (or just Fed funds) effective rate. This rate is a crucial determinant of the interest rates offered and charged by banks to their customers. The Fed funds rate is also very important for floating rate bonds and the prices and yields of short-term bonds. The governors of the Federal Reserve set and regularly reconsider a target Fed funds rate based on their view of the

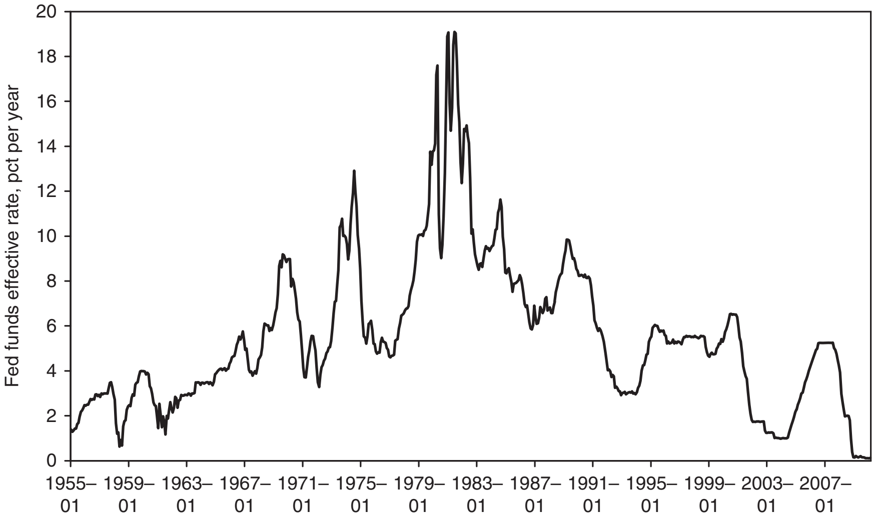

Fig. 1.3: The Fed funds effective rate from July 1954 to February 2010.
Source: the homepage of the Federal Reserve (`www.federalreserve.gov`) at 2 March, 2010.

current and future economic conditions. The Federal Reserve then buys and sells securities in open market operations to manage the liquidity in the market, thereby also affecting the Fed funds effective rate. Figure 1.3 depicts the Fed funds effective rate in the period from July 1954 to February 2010. Banks may obtain temporary credit directly from the Federal Reserve at the so-called 'discount window'. The interest rate charged by the Fed on such credit is called the federal discount rate, but since such borrowing is quite uncommon nowadays, the federal discount rate serves more as a signaling device for the targets of the Federal Reserve.

Many of the contracts used in the money markets are benchmarked to (that is priced by reference to) the London Interbank Offered Rate (LIBOR) for the appropriate term and currency. In the Euromarket, deposits are negotiated for various terms and currencies, but most deposits are in U.S. dollars or the Euro for a period of 1, 3, or 6 months. Interest rates set on unsecured deposits at the London interbank market are called LIBOR rates.

More details on U.S. bond markets can be found in Fabozzi (2010), while Batten et al. (2004) provide detailed information on European bond and money markets.

1.4 FIXED INCOME DERIVATIVES

A wide variety of fixed income derivatives are traded around the world. In this section we provide a brief introduction to the markets for such securities. In the pricing models we develop in later chapters we will look for prices of some of the most popular fixed income derivatives. Chapter 6 contains more details on a number of fixed income derivatives, what cash flow they offer, how the different derivatives are related, and so on. Credit-related derivatives are discussed in Chapter 13, where key statistics for those markets are also provided.

A **forward** is the simplest derivative. A forward contract is an agreement between two parties on a given transaction at a given future point in time and at a price that is already fixed when the agreement is made. For example, a forward on a bond is a contract where the parties agree to trade a given bond at a future point in time for a price which is already fixed today. This fixed price is usually set so that the value of the contract at the time of inception is equal to zero. In that case no money changes hands before the delivery date. A closely related contract is the so-called **forward rate agreement (FRA)**. Here the two parties agree that one party will borrow money from the other party over some period beginning at a given future date and the interest rate for that loan is fixed already when this FRA is entered. In other words, the interest rate for the future period is locked in. FRAs are quite popular instruments in the money markets.

As a forward contract, a **futures** contract is an agreement upon a specified future transaction, for example a trade of a given security. The special feature of a future is that changes in its value are settled continuously throughout the life of the contract (usually once every trading day). This so-called **marking-to-market** ensures that the value of the contract (that is the value of the future payments) is zero immediately following a settlement. This procedure makes it practically possible to trade futures at organized exchanges, since there is no need to keep track of when the futures position was originally taken. Futures on government bonds are traded at many leading exchanges. A very popular exchange-traded derivative is the so-called **Eurodollar futures**, which is basically the futures equivalent of a forward rate agreement.

An **option** gives the holder the right to make some specified future transaction at terms that are already fixed. A call option gives the holder the right to buy a given security at a given price at or before a given date. Conversely, a put option gives the holder the right to sell a given security. If the option gives the right to make the transaction at only one given date, the option is said to be European-style. If the right can be exercised at any point in time up to some given date, the option is said to be American-style. Both European- and American-style options are traded. Options on government bonds are traded at several exchanges and also on the OTC-markets. In addition, many bonds are issued with 'embedded' options. For example, many mortgage-backed bonds and corporate bonds are callable, in the sense that the issuer has the right to buy back the bond at a pre-specified price. To value such bonds, we must be able to value the option element.

Various interest rate options are also traded in the fixed income markets. The most popular are **caps and floors**. A cap is designed to protect an investor who has borrowed funds on a floating interest rate basis against the risk of paying very high interest rates. Therefore the cap basically gives you the right to borrow at some given rate. A cap can be seen as a portfolio of interest rate call options. Conversely, a floor is designed to protect an investor who has lent funds on a floating rate basis against receiving very low interest rates. A floor is a portfolio of interest rate put options. Various exotic versions of caps and floors are also quite popular.

A **swap** is an exchange of two cash flow streams that are determined by certain interest rates. In the simplest and most common interest rate swap, a **plain vanilla** swap, two parties exchange a stream of fixed interest rate payments and a stream

Table 1.4: Derivatives traded on organized exchanges. All amounts are in billions of U.S. dollars. The amount outstanding is of December 2009, whereas the turnover figures are for the fourth quarter of 2009.

Instruments/ Location	Futures		Options	
	Amount outstanding	Turnover	Amount outstanding	Turnover
All markets	21,749	307,315	51,388	137,182
Interest rate	20,623	276,215	46,435	106,523
Currency	164	7,677	147	582
Equity index	962	23,423	4,807	30,077
North America	10,716	156,160	23,875	55,216
Europe	8,054	129,016	26,331	62,937
Asia-Pacific	2,446	18,567	310	17,235
Other markets	532	3,573	873	1,793

Source: Table 23A in BIS (2010).

of floating interest rate payments. There are also currency swaps where streams of payments in different currencies are exchanged. In addition, many exotic swaps with special features are widely used. The international OTC swap markets are huge, both in terms of transactions and outstanding contracts. The **credit default swap** or just CDS is a widely used swap contract, in which the buyer makes a series of payments to the seller and, in exchange, receives a payoff if there is a default (or another 'credit event') on a certain bond or loan issued by a third party. More on credit default swaps and other credit-related securities can be found in Chapter 13.

A **swaption** is an option on a swap, that is it gives the holder the right, but not the obligation, to enter into a specific swap with pre-specified terms at or before a given future date. Both European- and American-style swaptions are traded.

The Bank for International Settlements (BIS) also publishes statistics on derivative trading around the world. Table 1.4 provides some interesting statistics on the size of derivatives markets at organized exchanges. The markets for interest rate derivatives are much larger than the markets for currency- or equity-linked derivatives. The option markets generally dominate futures markets measured by the amounts outstanding, but ranked according to turnover futures markets are larger than options markets.

The BIS statistics also contain information about the size of OTC markets for derivatives. BIS estimates that in June 2009 the total amount outstanding on OTC derivative markets was 604,622 billion of U.S. dollars, of which single-currency interest rate derivatives account for 72.3%, currency derivatives account for 8.1%, credit default swaps for 6.0%, equity-linked derivatives for 1.1%, commodity contracts for 0.6%, while the remainder cannot be split into any of these categories, compare Table 19 in BIS (2010). Table 1.5 shows how the interest rate derivatives market can be disaggregated according to instrument and maturity. Approximately 36.7% of these OTC-traded interest rate derivatives are denominated in Euros, 35.3% in U.S. dollars, 13.1% in yen, 7.5% in pounds sterling, and the remaining 7.4% in other currencies, compare Table 21B in BIS (2010).

Table 1.5: Amounts outstanding (billions of U.S. dollars) on OTC single-currency interest rate derivatives as of June 2009.

Contracts	Total	Maturity in years		
		≤ 1	1–5	≥ 5
All interest rates	437,198	159,143	128,301	149,754
Forward rate agreements	46,798	150,630	111,431	126,623
Swaps	341,886			
Options	48,513	8,513	16,870	23,130

Source: Tables 21A and 21C in BIS (2010).

1.5 AN OVERVIEW OF THE BOOK

The key element in our analysis will be the term structure of interest rates. The cleanest picture of the link between interest rates and maturities is given by a zero-coupon yield curve. In many markets only a few zero-coupon bonds are traded, so that we have to extract an estimate of the zero-coupon yield curve from prices of the traded coupon bonds. We will discuss methods for doing that in Chapter 2.

Risk is a central concept in fixed income modelling. The prices and yields of bonds are affected by the expectation and uncertainty about future values of certain macroeconomic variables. The prices of fixed income derivatives reflect the present value of the future payoff, which is generally dependent on the value of some bond price or interest rate at a later date. Consequently, we need to model the behaviour of uncertain variables or objects over time. This is done in terms of stochastic processes. A stochastic process is basically a collection of random variables, namely one random variable for each of the points in time at which we are interested in the value of this object. To understand and work with modern fixed income models therefore requires some knowledge about stochastic processes, their properties, and how to do relevant calculations involving stochastic processes. Chapter 3 provides the information about stochastic processes that is needed for our purposes.

This book focuses on the pricing of fixed income securities. However, the pricing of fixed income securities follows the same general principles as the pricing of all other financial assets. Chapter 4 reviews some of the important results on asset pricing theory. In particular, we define and relate the key concepts of arbitrage, state-price deflators, and risk-neutral probability measures. The connections to market completeness and individual investors' behaviour are also addressed. All these results will be applied in the following chapters to the term structure of interest rates and the pricing of fixed income securities.

In Chapter 5 we study the links between the term structure of interest rates and macroeconomic variables such as aggregate consumption, production, and inflation. The term structure of interest rates reflects the prices of bonds of various maturities and, as always, prices are set to align supply and demand. An investor with a clear preference for current capital to finance investments or current consumption can borrow by issuing a bond to an investor with a clear preference for future consumption opportunities. The price of a bond of a given maturity will therefore depend on the attractiveness of the real investment opportunities and on the individuals' preferences for consumption over the maturity of the bond. Following this intuition we develop relations between interest rates, aggregate consumption, and aggregate production. We also explore the relations between nominal interest rates, real interest rates, and inflation. Finally, the chapter reviews some of the traditional hypotheses on the shape of the yield curve, for example the expectation hypotheses, and discuss their relevance (or, rather, irrelevance) for modern fixed income analysis.

Chapter 6 provides an overview of the most popular fixed income derivatives, such as futures and options on bonds, Eurodollar futures, caps and floors, and swaps and swaptions. We will look at the characteristics of these securities and what we can say about their prices without setting up any concrete term structure model.

Starting with Chapter 7 we focus on dynamic term structure models developed for the pricing of fixed income securities and the management of interest rate risk. Chapter 7 goes through so-called one-factor diffusion models. This type of model was the first to be applied in the literature and dates back at least to 1970. The one-factor models of Vasicek and Cox, Ingersoll, and Ross are still frequently applied both in practice and in academic research. They have a lot of realistic features and deliver simple pricing formulas for many fixed income securities. Chapter 8 explores multi-factor diffusion models which have several advantages over one-factor models, but are also more complicated to analyse and apply.

The diffusion models deliver prices both for bonds and derivatives. However, the model price for a given bond may not be identical to the actually observed price of the bond. If you want to price a derivative on that bond, this seems problematic. If the model does not get the price of the underlying security right, why trust the model's price of the derivative? In Chapter 9 we illustrate how one-factor diffusion models can be extended to be consistent with current market information, such as bond prices and volatilities. A more direct route to ensuring consistency is explored in Chapter 10 that introduces and analyses so-called Heath–Jarrow–Morton models. They are characterized by taking the current market term structure of interest rates as given and then modelling the evolution of the entire term structure in an arbitrage-free way. We will explore the relation between these models and the factor models studied in earlier chapters.

Yet another class of models is the subject of Chapter 11. These 'market models' are designed for the pricing and hedging of specific products that are traded on a large scale in the international markets, namely caps, floors, and swaptions. These models have become increasingly popular in recent years.

In Chapters 6–11 we focus on the pricing of various fixed income securities. However, it is also extremely important to be able to measure and manage interest rate risk. Interest rate risk measures of individual securities are needed in order to obtain an overview of the total interest rate risk of the investors' portfolio and to identify the contribution of each security to this total risk. Many institutional investors are required to produce such risk measures for regulatory authorities and for publication in their accounting reports. In addition, such risk measures constitute an important input to the portfolio management. Interest rate risk management is the topic of Chapter 12. First, some traditional interest rate risk measures are reviewed and criticized. Then we turn to risk measures defined in relation to the dynamic term structure models studied in the previous chapters.

The following chapters deal with some securities that require special attention. In Chapter 13 we discuss the pricing of corporate bonds and other fixed income securities where the default risk of the issuer cannot be ignored. The subject of Chapter 14 is how to construct models for the pricing and risk management of mortgage-backed securities. The main concern is how to adjust the models studied in earlier chapters to take the prepayment options involved in mortgages into account. Chapter 15 focuses on the consequences that stochastic variations in interest rates have for the valuation of securities with payments that are not directly related to interest rates, such as stock options and currency options.

Finally, Chapter 16 describes and illustrates several numerical techniques that are often applied in cases where explicit pricing and hedging formulas are not available.

1.6 EXERCISES

Exercise 1.1 Show that if the discount function does *not* satisfy the condition

$$B_t^T \geq B_t^S, \quad t \leq T < S,$$

then negative forward rates will exist. Are non-negative forward rates likely to exist? Explain!

Exercise 1.2 Consider two bullet bonds, both with annual payments and exactly 4 years to maturity. The first bond has a coupon rate of 6% and is traded at a price of 101.00. The other bond has a coupon rate of 4% and is traded at a price of 93.20. What is the 4-year discount factor? What is the 4-year zero-coupon interest rate?

Exercise 1.3 Consider a bond market in which the annually compounded zero-coupon yields of maturities from 1 to 5 years are

$$\hat{y}^1 = 5\%, \quad \hat{y}^2 = 6\%, \quad \hat{y}^3 = 6.8\%, \quad \hat{y}^4 = 7.4\%, \quad \hat{y}^5 = 7.5\%.$$

What are the corresponding discount factors B^T, $T = 1, 2, \ldots, 5$? What are the 1-year forward rates $\hat{f}^{T,T+1}$, $T = 0, 1, \ldots, 4$?

Exercise 1.4 Consider a coupon bond with payments Y_i at time $T_i = i$, $i = 1, \ldots, n$, such that is there is a payment of Y_1 in 1 year, a payment of Y_2 in 2 years, and so on. Suppose you discount all future payments with a constant, annually compounded interest rate of r. Let B denote the present value, that is

$$B = \sum_{i=1}^{n} Y_i (1 + r)^{-i}.$$

(a) Show that if the bond is a *bullet bond* with a coupon rate of R and a face value of 1, then

$$B = \frac{R}{r} + \left(1 - \frac{R}{r}\right)(1 + r)^{-n}.$$

(b) Show that if the bond is an *annuity bond* with a coupon rate of R and a face value of 1, then

$$B = \frac{\alpha(n,r)}{\alpha(n,R)},$$

where $\alpha(N,\rho) = \rho^{-1}\left(1-(1+\rho)^{-N}\right)$.

(c) Show that if the bond is a *serial bond* with a coupon rate of R and a face value of 1, then

$$B = \frac{R}{r} + \frac{1}{n}\left(1-\frac{R}{r}\right)\alpha(n,r).$$

Exercise 1.5 What bonds are currently traded in your domestic market? Try to find information about historical interest rates in your country, either yields on government bonds or official interest rates fixed by the central bank or both.

2

Extracting Yield Curves from Bond Prices

2.1 INTRODUCTION

As discussed in Chapter 1, the clearest picture of the term structure of interest rates is obtained by looking at the yields of zero-coupon bonds of different maturities. However, in most countries almost all traded bonds are coupon bonds, not zero-coupon bonds. This chapter discusses methods to extract or estimate a zero-coupon yield curve from the prices of coupon bonds at a given point in time. In the United States Treasury coupon STRIPs are traded and these are indeed zero-coupon bonds so they could be used directly to construct a zero-coupon yield curve. But the liquidity of the STRIPs is smaller than that of Treasury coupon bonds and the pricing of STRIPs might therefore include an illiquidity discount compared to the coupon bonds. In addition, there is a tax disadvantage to STRIPs for some investors. As a consequence, most analysts and dealers prefer to work with a zero-coupon yield curve extracted from the prices of the coupon bonds.

Section 2.2 considers the so-called bootstrapping technique. It is sometimes possible to construct zero-coupon bonds by forming certain portfolios of coupon bonds. If so, we can deduce an arbitrage-free price of the zero-coupon bond and transform it into a zero-coupon yield. This is the basic idea of the bootstrapping approach. The bootstrapping approach can deliver a decent estimate of the whole zero-coupon yield curve in bond markets with sufficiently many coupon bonds having a variety of maturities and overlapping payment dates. In other markets, alternative methods are called for. In any case, the bootstrapping technique can at most deliver zero-coupon yields for the payment dates of the bonds considered. In order to obtain a full, continuous yield curve interpolation between adjacent payment dates is required.

We study two alternatives to bootstrapping in Sections 2.3 and 2.4. Both are based on the assumption that the discount function is of a given functional form with some unknown parameters. The values of these parameters are then estimated by least-squares methods to obtain the best possible agreement between observed bond prices and theoretical bond prices computed using the functional form. Typically, the assumed functional forms are either polynomials or exponential functions of maturity or some combination. This is consistent with the usual perception that discount functions and yield curves are continuous and smooth. If the yield for a given maturity was much higher than the yield for another maturity very close to the first, most bond owners would probably shift from bonds with the low-yield maturity to bonds with the high-yield maturity. Conversely, bond issuers (borrowers) would shift to the low-yield maturity. These changes in supply and demand will cause the gap between the yields for the two maturities to shrink. Hence, the equilibrium yield curve should be continuous and smooth.

We focus here on two of the most frequently applied parametrization techniques, namely cubic splines and the Nelson–Siegel parametrization. An overview of some of the many other approaches suggested in the literature can be seen in Anderson et al. (1996, Ch. 2), James and Webber (2000, Ch. 15), and Hagan and West (2006). For some more recent procedures, see Tanggaard (1997), Jaschke (1998), Linton et al. (2001), and Andersen (2007).

2.2 BOOTSTRAPPING

In many bond markets only very few zero-coupon bonds are issued and traded (all bonds issued as coupon bonds will eventually become a zero-coupon bond after their next-to-last payment date). Usually, such zero-coupon bonds have a very short maturity. To obtain knowledge of the market zero-coupon yields for longer maturities, we have to extract information from the prices of traded coupon bonds. In some markets it is possible to construct some longer-term zero-coupon bonds by forming portfolios of traded coupon bonds. Market prices of these 'synthetical' zero-coupon bonds and the associated zero-coupon yields can then be derived. Let us first take a simple example.

Example 2.1 Consider a market where two bullet bonds are traded, a 10% bond expiring in 1 year and a 5% bond expiring in 2 years. Both have annual payments and a face value of 100. The 1-year bond has the payment structure of a zero-coupon bond: 110 dollars in 1 year and nothing at all other points in time. A share of 1/110 of this bond corresponds exactly to a zero-coupon bond paying 1 dollar in a year. If the price of the 1-year bullet bond is 100, the 1-year discount factor is given by

$$B_t^{t+1} = \frac{1}{110} \cdot 100 \approx 0.9091.$$

The 2-year bond provides payments of 5 dollars in 1 year and 105 dollars in 2 years. Hence, it can be seen as a portfolio of 5 1-year zero-coupon bonds and 105 2-year zero-coupon bonds, all with a face value of 1 dollar. The price of the 2-year bullet bond is therefore

$$B_{2,t} = 5B_t^{t+1} + 105B_t^{t+2},$$

see (1.1). Isolating B_t^{t+2}, we get

$$B_t^{t+2} = \frac{1}{105}B_{2,t} - \frac{5}{105}B_t^{t+1}. \tag{2.1}$$

If, for example, the price of the two-year bullet bond is 90, the 2-year discount factor will be

$$B_t^{t+2} = \frac{1}{105} \cdot 90 - \frac{5}{105} \cdot 0.9091 \approx 0.8139.$$

From (2.1) we see that we can construct a 2-year zero-coupon bond as a portfolio of 1/105 units of the 2-year bullet bond and −5/105 units of the 1-year zero-coupon bond. This is equivalent to a portfolio of 1/105 units of the 2-year bullet bond and $-5/(105 \cdot 110)$ units of the 1-year bullet bond. Given the discount factors, zero-coupon rates and forward rates can be calculated as shown in Section 1.2.

The example above can easily be generalized to more periods. Suppose we have M bonds with maturities of $1, 2, \ldots, M$ periods, respectively, one payment date each period, and identical payment dates. Then we can construct successively zero-coupon bonds for each of these maturities and hence compute the market discount factors $B_t^{t+1}, B_t^{t+2}, \ldots, B_t^{t+M}$. First, B_t^{t+1} is computed using the shortest bond. Then, B_t^{t+2} is computed using the next-to-shortest bond and the already computed value of B_t^{t+1}, etc. Given the discount factors $B_t^{t+1}, B_t^{t+2}, \ldots, B_t^{t+M}$, we can compute the zero-coupon interest rates and hence the zero-coupon yield curve up to time $t + M$ (for the M selected maturities). This approach is called **bootstrapping** or **yield curve stripping**.

Bootstrapping also applies to the case where the maturities of the M bonds are not all different and regularly increasing as above. As long as the M bonds together have at most M different payment dates and each bond has at most one payment date where none of the bonds provide payments, then we can construct zero-coupon bonds for each of these payment dates and compute the associated discount factors and rates. Let us denote the payment of bond i ($i = 1, \ldots, M$) at time $t + j$ ($j = 1, \ldots, M$) by Y_{ij}. Some of these payments may well be zero, for example if the bond matures before time $t + M$. Let $B_{i,t}$ denote the price of bond i. From (1.1) we have that the discount factors $B_t^{t+1}, B_t^{t+2}, \ldots, B_t^{t+M}$ must satisfy the system of equations

$$\begin{pmatrix} B_{1,t} \\ B_{2,t} \\ \vdots \\ B_{M,t} \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \ldots & Y_{1M} \\ Y_{21} & Y_{22} & \ldots & Y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{M1} & Y_{M2} & \ldots & Y_{MM} \end{pmatrix} \begin{pmatrix} B_t^{t+1} \\ B_t^{t+2} \\ \vdots \\ B_t^{t+M} \end{pmatrix}. \tag{2.2}$$

The conditions on the bonds ensure that the payment matrix of this equation system is non-singular so that a unique solution will exist.

For each of the payment dates $t + j$, we can construct a portfolio of the M bonds, which is equivalent to a zero-coupon bond with a payment of 1 at time $t + j$. Denote by $x_i(j)$ the number of units of bond i which enter the portfolio replicating the zero-coupon bond maturing at $t + j$. Then we must have that

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{21} & \ldots & \ldots & \ldots & Y_{M1} \\ Y_{12} & Y_{22} & \ldots & \ldots & \ldots & Y_{M2} \\ \vdots & \vdots & \ddots & & & \vdots \\ Y_{1j} & Y_{2j} & \ldots & \ldots & \ldots & Y_{Mj} \\ \vdots & \vdots & & & \ddots & \vdots \\ Y_{1M} & Y_{2M} & \ldots & \ldots & \ldots & Y_{MM} \end{pmatrix} \begin{pmatrix} x_1(j) \\ x_2(j) \\ \vdots \\ x_j(j) \\ \vdots \\ x_M(j) \end{pmatrix}, \tag{2.3}$$

where the 1 on the left-hand side of the equation is at the j'th entry of the vector. Of course, there will be the following relation between the solution $(B_t^{t+1}, \dots, B_t^{t+M})$ to (2.2) and the solution $(x_1(j), \dots, x_M(j))$ to (2.3):[1]

$$\sum_{i=1}^{M} x_i(j) B_{i,t} = B_t^{t+j}. \tag{2.4}$$

Thus, first the zero-coupon bonds can be constructed, that is (2.3) is solved for each $j = 1, \dots, M$, and next (2.4) can be applied to compute the discount factors.

Example 2.2 In Example 2.1 we considered a 2-year 5% bullet bond. Assume now that a 2-year 8% serial bond with the same payment dates is traded. The payments from this bond are 58 dollars in 1 year and 54 dollars in 2 years. Assume that the price of the serial bond is 98 dollars. From these two bonds we can set up the following equation system to solve for the discount factors B_t^{t+1} and B_t^{t+2}:

$$\begin{pmatrix} 90 \\ 98 \end{pmatrix} = \begin{pmatrix} 5 & 105 \\ 58 & 54 \end{pmatrix} \begin{pmatrix} B_t^{t+1} \\ B_t^{t+2} \end{pmatrix}.$$

The solution is $B_t^{t+1} \approx 0.9330$ and $B_t^{t+2} \approx 0.8127$.

More generally, if there are M traded bonds having in total N different payment dates, the system (2.2) becomes one of M equations in N unknowns. If $M > N$, the system may not have any solution, since it may be impossible to find discount factors consistent with the prices of all M bonds. If no such solution can be found, there will be an arbitrage opportunity.

Example 2.3 In the Examples 2.1 and 2.2 we have considered three bonds: a 1-year bullet bond, a 2-year bullet bond, and a 2-year serial bond. In total, these three bonds have two different payment dates. According to the prices and payments of these three bonds, the discount factors B_t^{t+1} and B_t^{t+2} must satisfy the following three equations:

$$\begin{aligned} 100 &= 110 B_t^{t+1}, \\ 90 &= 5 B_t^{t+1} + 105 B_t^{t+2}, \\ 98 &= 58 B_t^{t+1} + 54 B_t^{t+2}. \end{aligned}$$

No solution exists. In Example 2.1 we found that the solution to the first two equations is

$$B_t^{t+1} \approx 0.9091 \qquad \text{and} \qquad B_t^{t+2} \approx 0.8139.$$

[1] In matrix notation, Equation (2.2) can be written as $\boldsymbol{B}_{\text{cpn}} = \Upsilon \boldsymbol{B}_{\text{zero}}$ and Equation (2.3) can be written as $\boldsymbol{e}_j = \Upsilon^\top \boldsymbol{x}(j)$, where $\boldsymbol{e}_j$ is the vector on the left hand side of (2.3), the symbol $^\top$ indicates transposition, and the other symbols are self-explanatory. Hence,

$$\boldsymbol{x}(j)^\top \boldsymbol{B}_{\text{cpn}} = \boldsymbol{x}(j)^\top \Upsilon \boldsymbol{B}_{\text{zero}} = \boldsymbol{e}_j^\top \boldsymbol{B}_{\text{zero}} = B_t^{t+j},$$

which is equivalent to (2.4).

In contrast, we found in Example 2.2 that the solution to the last two equations is

$$B_t^{t+1} \approx 0.9330 \qquad \text{and} \qquad B_t^{t+2} \approx 0.8127.$$

If the first solution is correct, the price on the serial bond should be

$$58 \cdot 0.9091 + 54 \cdot 0.8139 \approx 96.68, \tag{2.5}$$

but it is not. The serial bond is mispriced relative to the two bullet bonds. More precisely, the serial bond is too expensive. We can exploit this by selling the serial bond and buying a portfolio of the two bullet bonds that *replicates* the serial bond, that is provides the same cash flow. We know that the serial bond is equivalent to a portfolio of 58 1-year zero-coupon bonds and 54 2-year zero-coupon bonds, all with a face value of 1 dollar. In Example 2.1 we found that the 1-year zero-coupon bond is equivalent to 1/110 units of the 1-year bullet bond, and that the 2-year zero-coupon bond is equivalent to a portfolio of $-5/(105 \cdot 110)$ units of the 1-year bullet bond and 1/105 units of the 2-year bullet bond. It follows that the serial bond is equivalent to a portfolio consisting of

$$58 \cdot \frac{1}{110} - 54 \cdot \frac{5}{105 \cdot 110} \approx 0.5039$$

units of the 1-year bullet bond and

$$54 \cdot \frac{1}{105} \approx 0.5143$$

units of the 2-year bullet bond. This portfolio will give exactly the same cash flow as the serial bond, that is 58 dollars in one year and 54 dollars in 2 years. The price of the portfolio is

$$0.5039 \cdot 100 + 0.5143 \cdot 90 \approx 96.68,$$

which is exactly the price found in (2.5).

In some markets, the government bonds are issued with many different payment dates. The system (2.2) will then typically have fewer equations than unknowns. In that case there are many solutions to the equation system, that is many sets of discount factors can be consistent both with observed prices and the no-arbitrage pricing principle.

2.3 CUBIC SPLINES

Bootstrapping can only provide knowledge of the discount factors for (some of) the payment dates of the traded bonds. In many situations information about market discount factors for other future dates will be valuable. In this section and the next, we will consider methods to estimate the entire discount function $T \mapsto B_t^T$ (at least up to some large T). To simplify the notation in what follows, we fix the current time t and let $\bar{B}(\tau)$ denote the discount factor for the next τ periods, that is $\bar{B}(\tau) = B_t^{t+\tau}$. Hence, the function $\bar{B}(\tau)$ for $\tau \in [0, \infty)$ represents the time t market discount function. In particular, $\bar{B}(0) = 1$. We will use a similar notation for zero-coupon rates and forward rates: $\bar{y}(\tau) = y_t^{t+\tau}$ and $\bar{f}(\tau) = f_t^{t+\tau}$. The methods studied in this and the following sections are both based on the assumption that the discount function $\tau \mapsto \bar{B}(\tau)$ can be described by some functional form involving some unknown parameters. The parameter values are chosen to get a close match between the observed bond prices and the theoretical bond prices computed using the assumed discount function.

The approach studied in this section is a version of the cubic splines approach introduced by McCulloch (1971) and later modified by McCulloch (1975) and Litzenberger and Rolfo (1984). The word spline indicates that the maturity axis is divided into subintervals and that separate functions (of the same type) are used to describe the discount function in the different subintervals. The reasoning for doing this is that it can be quite hard to fit a relatively simple functional form to prices of a large number of bonds with very different maturities. To ensure a continuous and smooth term structure of interest rates, one must impose certain conditions for the maturities separating the subintervals.

Given prices for M bonds with time-to-maturities of $T_1 \leq T_2 \leq \cdots \leq T_M$, divide the maturity axis into subintervals defined by the 'knot points' $0 = \tau_0 < \tau_1 < \cdots < \tau_k = T_M$. A spline approximation of the discount function $\bar{B}(\tau)$ is based on an expression like

$$\bar{B}(\tau) = \sum_{j=0}^{k-1} G_j(\tau) I_j(\tau),$$

where the G_j's are basis functions, and the I_j's are the step functions

$$I_j(\tau) = \begin{cases} 1, & \text{if } \tau \geq \tau_j, \\ 0, & \text{otherwise.} \end{cases}$$

Hence,

$$\bar{B}(\tau) = \begin{cases} G_0(\tau), & \text{for } \tau \in [\tau_0, \tau_1), \\ G_0(\tau) + G_1(\tau), & \text{for } \tau \in [\tau_1, \tau_2), \\ \dots & \dots \\ G_0(\tau) + G_1(\tau) + \cdots + G_{k-1}(\tau), & \text{for } \tau \geq \tau_{k-1}. \end{cases}$$

We demand that the G_j's are continuous and differentiable and ensure a smooth transition in the knot points τ_j. A polynomial spline is a spline where the basis functions are polynomials. Let us consider a cubic spline, where

$$G_j(\tau) = \alpha_j + \beta_j(\tau - \tau_j) + \gamma_j(\tau - \tau_j)^2 + \delta_j(\tau - \tau_j)^3,$$

and α_j, β_j, γ_j, and δ_j are constants.

For $\tau \in [0, \tau_1)$, we have

$$\bar{B}(\tau) = \alpha_0 + \beta_0\tau + \gamma_0\tau^2 + \delta_0\tau^3. \tag{2.6}$$

Since $\bar{B}(0) = 1$, we must have $\alpha_0 = 1$. For $\tau \in [\tau_1, \tau_2)$, we have

$$\bar{B}(\tau) = \left(1 + \beta_0 \tau + \gamma_0 \tau^2 + \delta_0 \tau^3\right) + \left(\alpha_1 + \beta_1(\tau - \tau_1) + \gamma_1(\tau - \tau_1)^2 + \delta_1(\tau - \tau_1)^3\right). \tag{2.7}$$

To get a smooth transition between (2.6) and (2.7) in the point $\tau = \tau_1$, we demand that

$$\bar{B}(\tau_1-) = \bar{B}(\tau_1+), \tag{2.8}$$

$$\bar{B}'(\tau_1-) = \bar{B}'(\tau_1+), \tag{2.9}$$

$$\bar{B}''(\tau_1-) = \bar{B}''(\tau_1+), \tag{2.10}$$

where $\bar{B}(\tau_1-) = \lim_{\tau \to \tau_1, \tau < \tau_1} \bar{B}(\tau)$, $\bar{B}(\tau_1+) = \lim_{\tau \to \tau_1, \tau > \tau_1} \bar{B}(\tau)$, etc. The condition (2.8) ensures that the discount function is continuous in the knot point τ_1. The condition (2.9) ensures that the graph of the discount function has no kink at τ_1 by restricting the first-order derivative to approach the same value whether t approaches τ_1 from below or from above. The condition (2.10) requires the same to be true for the second-order derivative, which ensures an even smoother behaviour of the graph around the knot point τ_1.

The condition (2.8) implies $\alpha_1 = 0$. Differentiating (2.6) and (2.7), we find

$$\bar{B}'(\tau) = \beta_0 + 2\gamma_0 \tau + 3\delta_0 \tau^2, \quad 0 \le \tau < \tau_1,$$

and

$$\bar{B}'(\tau) = \beta_0 + 2\gamma_0 \tau + 3\delta_0 \tau^2 + \beta_1 + 2\gamma_1(\tau - \tau_1) + 3\delta_1(\tau - \tau_1)^2, \quad \tau_1 \le \tau < \tau_2.$$

The condition (2.9) now implies $\beta_1 = 0$. Differentiating again, we get

$$\bar{B}''(\tau) = 2\gamma_0 + 6\delta_0 \tau, \quad 0 \le \tau < \tau_1,$$

and

$$\bar{B}''(\tau) = 2\gamma_0 + 6\delta_0 \tau + 2\gamma_1 + 6\delta_1(\tau - \tau_1), \quad \tau_1 \le \tau < \tau_2.$$

Consequently, the condition (2.10) implies $\gamma_1 = 0$. Similarly, it can be shown that $\alpha_j = \beta_j = \gamma_j = 0$ for all $j = 1, \dots, k-1$. The cubic spline is therefore reduced to

$$\bar{B}(\tau) = 1 + \beta_0 \tau + \gamma_0 \tau^2 + \delta_0 \tau^3 + \sum_{j=1}^{k-1} \delta_j (\tau - \tau_j)^3 I_j(\tau). \tag{2.11}$$

Let $t_1, t_2, \dots, t_N$ denote the time distance from today (date t) to each of the payment dates in the set of all payment dates of the bonds in the data set. Let Y_{in} denote the payment of bond i in t_n periods. From the no-arbitrage pricing relation (1.1), we should have that

$$B_i = \sum_{n=1}^{N} Y_{in} \bar{B}(t_n),$$

where B_i is the current market price of bond i. Since not all the zero-coupon bonds involved in this equation are traded, we will allow for a deviation ε_i so that

$$B_i = \sum_{n=1}^{N} Y_{in} \bar{B}(t_n) + \varepsilon_i. \tag{2.12}$$

We assume that ε_i is normally distributed with mean zero and variance σ^2 (assumed to be the same for all bonds) and that the deviations for different bonds are mutually independent. We want to pick parameter values that minimize the sum of squared deviations $\sum_{i=1}^{M} \varepsilon_i^2$.

Substituting (2.11) into (2.12) yields

$$B_i = \sum_{n=1}^{N} Y_{in} \left\{ 1 + \beta_0 t_n + \gamma_0 t_n^2 + \delta_0 t_n^3 + \sum_{j=1}^{k-1} \delta_j (t_n - \tau_j)^3 I_j(\tau) \right\} + \varepsilon_i,$$

which implies that

$$\begin{aligned} B_i - \sum_{n=1}^{N} Y_{in} &= \beta_0 \sum_{n=1}^{N} Y_{in} t_n + \gamma_0 \sum_{n=1}^{N} Y_{in} t_n^2 + \delta_0 \sum_{n=1}^{N} Y_{in} t_n^3 \\ &\quad + \sum_{j=1}^{k-1} \delta_j \sum_{n=1}^{N} Y_{in} (t_n - \tau_j)^3 I_j(t_n) + \varepsilon_i. \end{aligned}$$

Given the prices and payment schemes of the M bonds, the $k+2$ parameters $\beta_0, \gamma_0, \delta_0, \delta_1, \dots, \delta_{k-1}$ can now be estimated using ordinary least squares. Substituting the estimated parameters into (2.11), we get an estimated discount function, from which estimated zero-coupon yield curves and forward rate curves can be derived as explained in Chapter 1.

It remains to describe how the number of subintervals k and the knot points τ_j are to be chosen. McCulloch (1971, 1975) suggests letting k be the nearest integer to $\sqrt{M}$ and defining the knot points by

$$\tau_j = T_{h_j} + \theta_j (T_{h_j+1} - T_{h_j}),$$

where $h_j = \lfloor j \cdot M/k \rfloor$ (here the brackets mean the integer part) and $\theta_j = j \cdot M/k - h_j$. In particular, $\tau_k = T_M$. Alternatively, the knot points can be placed at, for example, 1 year, 5 years, and 10 years, so that the intervals broadly correspond to the short-term, intermediate-term, and long-term segments of the market, in line with the idea of 'preferred habitats' discussed in Section 5.7.

There are several potentially undesirable properties of the cubic splines approach. First, the value of the discount function for some maturities can often be determined by pure no-arbitrage arguments as utilized in the bootstrapping approach. The discount function estimated with cubic splines will not necessarily match those values so applications of the estimated function will not respect the fundamental no-arbitrage pricing principle.

Second, there is no guarantee whatsoever that the discount function estimated using cubic splines has an economically credible form. In particular, the discount function should be positive and decreasing (which will ensure positive forward rates), but there is nothing in the approach ensuring that. Nevertheless, in most cases the estimated discount function turns out to exhibit those features, at least for maturities up to the longest bond maturity in the data set, T_M. As the maturity approaches infinity, the cubic spline discount function will approach either plus or minus infinity depending on the sign of the coefficient of the third order term. Of course, both properties are unacceptable, but the method cannot be expected to provide reasonable values beyond the maturities covered by the bonds in the data set.

Third, the zero-coupon yield curve and forward rate curve derived from an estimated discount function can have unrealistic shapes. The zero-coupon rates will often increase or decrease significantly for maturities approaching T_M, see Shea (1984, 1985). The derived forward rate curve will typically be quite rugged especially near the knot points, and the curve tends to be very sensitive to the precise location of the knot points. Therefore, yield curves and especially forward rate curves estimated using cubic splines should be applied with caution.

Fourth, small variations in the input bond prices may have a substantial effect on the estimated discount function and yield curve. In particular, a change in the input price of a short-maturity bond may even affect the long-maturity end of the estimated curves.

As discussed by Hagan and West (2006) and Andersen (2007), among others, some of these problems can be mitigated by applying different types of polynomials and by applying the cubic splines parametrization not directly to the discount function but to some transformation of the discount function or the yield curve. For example, the yield curves published by the U.S. Treasury are computed using a so-called quasi-cubic Hermite spline function using the maturities of the on-the-run (that is the most recently issued and thus typically also the most liquid) Treasury securities as knot points.[2] The competition between the many alternative methods is still open and it seems impossible to find one method that clearly dominates all other methods.

2.4 THE NELSON–SIEGEL PARAMETRIZATION

Nelson and Siegel (1987) proposed a simple parametrization of the term structure of interest rates, which has become quite popular. The approach is based on the following parametrization of the forward rates:

$$\bar{f}(\tau) = \beta_0 + \beta_1 e^{-\tau/\theta} + \beta_2 \frac{\tau}{\theta} e^{-\tau/\theta}, \tag{2.13}$$

[2] The information was taken from `www.treas.gov/offices/domestic-finance/debt-management/interest-rate` on 2 March 2010.

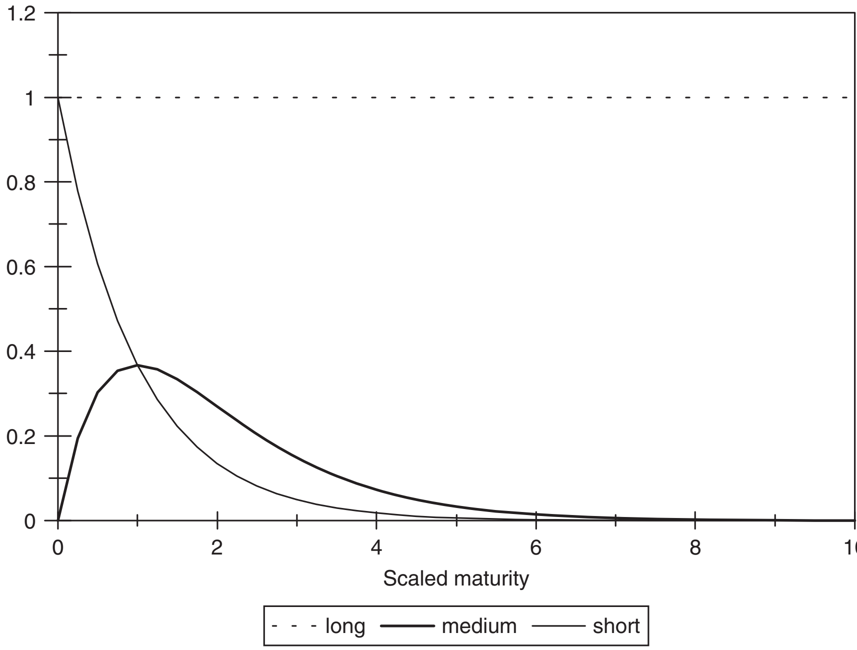

Fig. 2.1: The three curves which the Nelson–Siegel parametrization combines.

where β_0, β_1, β_2, and θ are constants to be estimated. The same constants are assumed to apply for all maturities, so no splines are involved. The simple functional form ensures a smooth and yet quite flexible curve. Figure 2.1 shows the graphs of the three functions that constitute (2.13). The flat curve (corresponding to the constant term β_0) will by itself determine the long-term forward rates, the term $\beta_1 e^{-\tau/\theta}$ is mostly affecting the short-term forward rates, while the term $\beta_2 \tau/\theta e^{-\tau/\theta}$ is important for medium-term forward rates. The value of the parameter θ determines how large a maturity interval the non-constant terms will affect. The value of the parameters β_0, β_1, and β_2 determine the relative weighting of the three curves.

According to (1.11), the term structure of zero-coupon rates is given by

$$\bar{y}(\tau) = \frac{1}{\tau}\int_0^\tau \bar{f}(u)\,du = \beta_0 + \beta_1 \frac{1 - e^{-\tau/\theta}}{\tau/\theta} + \beta_2 \left(\frac{1 - e^{-\tau/\theta}}{\tau/\theta} - e^{-\tau/\theta} \right),$$

which we will rewrite as

$$\bar{y}(\tau) = a + b\frac{1 - e^{-\tau/\theta}}{\tau/\theta} + ce^{-\tau/\theta}. \tag{2.14}$$

Figure 2.2 depicts the possible forms of the zero-coupon yield curve for different values of a, b, and c. By varying the parameter θ, the curves can be stretched or compressed in the horizontal dimension.

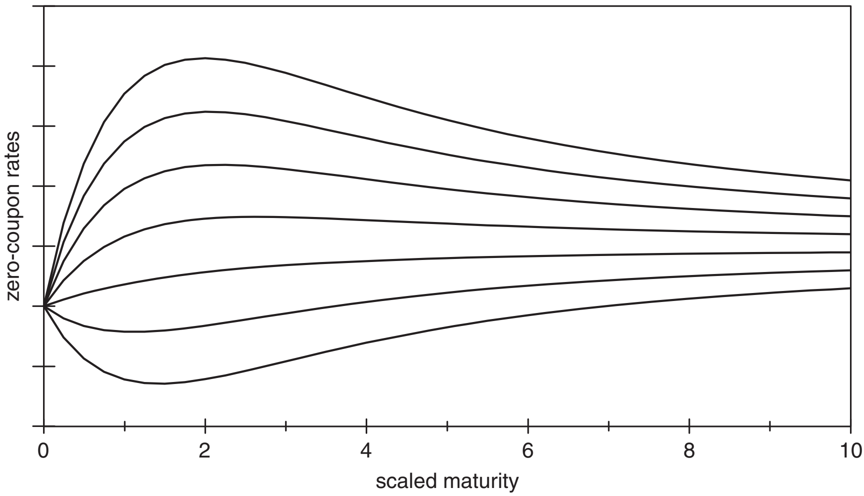

Fig. 2.2: Possible forms of the zero-coupon yield curve using the Nelson–Siegel parametrization.

Suppose we could directly observe zero-coupon rates $\bar{y}(T_i)$ for different maturities T_i, $i = 1, \ldots, M$. Then, for any given θ, we could estimate the parameters a, b, and c using simple linear regression on the model

$$\bar{y}(\tau) = a + b\frac{1 - e^{-\tau/\theta}}{\tau/\theta} + ce^{-\tau/\theta} + \varepsilon_i,$$

where $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \ldots, M$, are independent error terms. Doing this for various choices of θ, we could pick the θ and the corresponding regression estimates of a, b and c that result in the highest R^2, that is that best explain the data. This is exactly the procedure used by Nelson and Siegel on data on short-term zero-coupon bonds in the U.S. market.

When the data set involves coupon bonds, the estimation procedure is slightly more complicated. The discount function associated with the forward rate structure in (2.14) is given by

$$\bar{B}(\tau) = \exp\left\{-a\tau - b\theta\left(1 - e^{-\tau/\theta}\right) - c\tau e^{-\tau/\theta}\right\}.$$

Substituting this into (2.12), we get

$$B_i = \sum_{n=1}^{N} Y_{in} \exp\left\{-at_n - b\theta\left(1 - e^{-t_n/\theta}\right) - ct_n e^{-t_n/\theta}\right\} + \varepsilon_i.$$

Since this is a non-linear expression in the unknown parameters, the estimation must be based on generalized least squares, that is non-linear regression techniques.

According to Christensen, Diebold, and Rudebusch (2009), the empirically reasonable range for θ is from 1 to 2 for U.S. Treasury yield data. However, the slope and the curvature factors then decay rapidly to zero as the maturity increases. Hence, it is difficult to fit medium- and long-term yields. To allow for additional flexibility, Svensson (1995) suggested adding another curvature term so that the zero-coupon yield curve is parameterized by

$$\begin{aligned}\bar{y}(\tau) &= \beta_0 + \beta_1 \frac{1-e^{-\tau/\theta}}{\tau/\theta} + \beta_2 \left(\frac{1-e^{-\tau/\theta}}{\tau/\theta} - e^{-\tau/\theta} \right) \\ &\quad + \beta_3 \left(\frac{1-e^{-\tau/\theta'}}{\tau/\theta'} - e^{-\tau/\theta} \right),\end{aligned}$$

where θ' is sometimes restricted to be equal to θ. Other ad-hoc extensions are sometimes considered to further improve the fit, at the cost of increased complexity.

2.5 ADDITIONAL REMARKS ON YIELD CURVE ESTIMATION

Above we looked at two of the many estimation procedures based on a given parameterized form of either the discount function, the zero-coupon yield curve, or the forward rate curve. A clear disadvantage of both methods is that the estimated discount function is not necessarily consistent with those (probably few) discount factors that can be derived from market prices assuming only no-arbitrage. The procedures do not punish deviations from no-arbitrage values.

A more essential disadvantage of all such estimation procedures is that they only consider the term structure of interest rates at one particular point in time. Estimations at two different dates are completely independent and do not take into account the possible dynamics of the term structure over time. As we shall see in Chapters 7 and 8, there are many dynamic term structure models which also provide a parameterized form for the term structure at any given date. Applying such models, the estimation can (and should) be based on bond price observations at different dates. Typically, the possible forms of the term structure in such models resemble those of the Nelson–Siegel approach. We will return to this discussion in Chapter 7.

Finally, we will emphasize that the estimated term structure of interest rates should be used with caution. An obvious use of the estimated yield curve is to value fixed income securities. In particular, the coupon bonds in the data set used in the estimation can be priced using the estimated discount function. For some of the bonds the price according to the estimated curve will be lower (higher) than the market price. Therefore, one might think such bonds are overvalued (undervalued) by the market (in an estimation like (2.12) this can be seen directly from the residual ε_i). It would seem a good strategy to sell the overvalued and buy the undervalued bonds. However, such a strategy is not a risk-free arbitrage, but a risky strategy, since the applied discount function is not derived from the no-arbitrage principle only, but depends on the assumed parametric form and the other bonds in the data set. With another parameterized form or a different set of bonds the estimated discount function and, hence, the assessment of over- and undervaluation can be different.

2.6 EXERCISES

Exercise 2.1 Find a list of current price quotes on government bonds at an exchange in your country. Derive as many discount factors and zero-coupon yields as possible using only the no-arbitrage pricing principle, that is the bootstrapping approach.

Exercise 2.2 Consider a bond market in which 10 bullet bonds are traded. They all have a face value of 1,000, a coupon rate of 5%, annual payments, and exactly one year to the next payment date. The bonds mature in $1, 2, \ldots, 10$ years. The current prices are given in the following table.

Maturity	Price	Maturity	Price
1	1019.42	6	1042.56
2	1032.90	7	1045.52
3	1041.18	8	1052.10
4	1044.60	9	1054.82
5	1043.55	10	1053.99

Use the bootstrapping approach to determine the discount factors B^T, the annually compounded zero-coupon yields $\hat{y}^T$, and the annually compounded one-year forward rates $\hat{f}^{T-1,T}$ for $T = 1, 2, \ldots, 10$.

3

Stochastic Processes and Stochastic Calculus

3.1 INTRODUCTION

Most interest rates and asset prices vary over time in a non-deterministic way. We can observe the price of a given asset today, but the price of the same asset at any future point in time will typically be unknown, in other words a random variable. In order to describe the uncertain evolution in the price of the asset over time, we need a collection of random variables, namely one random variable for each point in time. Such a collection of random variables is called a stochastic process. Modern finance models therefore apply stochastic processes to represent the evolution in prices and rates over time. This is also the case for the dynamic interest rate models presented in this book.

This chapter gives an introduction to stochastic processes and the mathematical tools needed to do calculations with stochastic processes, the so-called stochastic calculus. We will omit many technical details that are not important for a reasonable level of understanding and focus on processes and results that will become important in later chapters. For more details and proofs, the reader is referred to textbooks on stochastic processes such as, for example, Øksendal (2003) and Karatzas and Shreve (1988), and to more extensive and formal introductions to stochastic processes in the mathematical finance textbooks of Dothan (1990), Duffie (2001), and Björk (2009).

The outline of the remainder of the chapter is as follows. In Section 3.2 we define the concept of a stochastic process more formally and introduce much of the terminology used. We define a particular process, the so-called Brownian motion, in Section 3.3. This will be the basic building block in the definition of other processes. In Section 3.4 we introduce the class of diffusion processes, which contains most of the processes used in popular fixed income models. Section 3.5 gives a short introduction to the more general class of Itô processes. Both diffusions and Itô processes involve stochastic integrals, which are discussed in Section 3.6. In Section 3.7 we state the very important Itô's Lemma, which is frequently applied when handling stochastic processes. Three diffusions that are widely used in finance models are introduced and studied in Section 3.8. Section 3.9 discusses multi-dimensional processes. Finally, Section 3.10 explains the change of probability measure which turns out to be highly relevant in the models studied in later chapters.

3.2 WHAT IS A STOCHASTIC PROCESS?

3.2.1 Probability spaces and information filtrations

The basic object for studies of uncertain events is a **probability space**, which is a triple $(\Omega, \mathcal{F}, \mathbb{P})$. Let us look at each of the three elements.

Ω is the **state space**, which is the set of possible states or outcomes of the uncertain object. For example, if one studies the outcome of a throw of a die (meaning the number of 'dots' on top of the die), the state space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. In our finance models an outcome is a realization of all relevant uncertain objects over the entire time interval studied in the model. Only one outcome, the 'true' outcome, will be realized.

$\mathcal{F}$ is the set of events to which a probability can be assigned, that is the set of 'probabilizable' events. Here, an **event** is a set of possible outcomes, that is a subset of the state space. In the example with the die, some events are $\{1, 2, 3\}$, $\{4, 5\}$, $\{1, 3, 5\}$, $\{6\}$, and $\{1, 2, 3, 4, 5, 6\}$. In a finance model an event is some set of realizations of the uncertain object. For example, in a model of the uncertain dynamics of a given asset price over a period of 10 years, one event is that the asset price one year into the future is above 100. Since $\mathcal{F}$ is a set of events, it is really a set of subsets of the state space. It is required that

(i) the entire state space can be assigned a probability, that is $\Omega \in \mathcal{F}$;
(ii) if some event $F \subseteq \Omega$ can be assigned a probability, so can its complement $F^c \equiv \Omega \setminus F$, that is $F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$; and
(iii) given a sequence of probabilizable events, the union is also probabilizable, that is $F_1, F_2, \dots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} F_i \in \mathcal{F}$.

Often $\mathcal{F}$ is referred to as a **sigma-algebra**.

$\mathbb{P}$ is a **probability measure**, which formally is a function from the sigma-algebra $\mathcal{F}$ into the interval $[0, 1]$. To each event $F \in \mathcal{F}$, the probability measure assigns a number $\mathbb{P}(F)$ in the interval $[0, 1]$. This number is called the $\mathbb{P}$-probability (or simply the probability) of F. A probability measure must satisfy the following conditions:

(i) $\mathbb{P}(\Omega) = 1$ and $\mathbb{P}(\emptyset) = 0$, where $\emptyset$ denotes the empty set;
(ii) the probability of the state being in the union of disjoint sets is equal to the sum of the probabilities for each of the sets, that is given $F_1, F_2, \dots \in \mathcal{F}$ with $F_i \cap F_j = \emptyset$ for all $i \neq j$, we have $\mathbb{P}(\cup_{i=1}^{\infty} F_i) = \sum_{i=1}^{\infty} \mathbb{P}(F_i)$.

Many different probability measures can be defined on the same sigma-algebra, $\mathcal{F}$, of events. In the example of the die, a probability measure $\mathbb{P}$ corresponding to the idea that the die is 'fair' is defined by

$$\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \dots = \mathbb{P}(\{6\}) = 1/6.$$

Another probability measure, $\mathbb{Q}$, can be defined by

$$\mathbb{Q}(\{1\}) = 1/12, \quad \mathbb{Q}(\{2\}) = \dots = \mathbb{Q}(\{5\}) = 1/6, \quad \mathbb{Q}(\{6\}) = 3/12,$$

which may be appropriate if the die is believed to be 'unfair' in a particular way.

Two probability measures $\mathbb{P}$ and $\mathbb{Q}$ defined on the same state space Ω and sigma-algebra $\mathcal{F}$ are called **equivalent** if the two measures assign probability zero to exactly the same events, that is if $\mathbb{P}(A) = 0 \Leftrightarrow \mathbb{Q}(A) = 0$. The two probability measures in the die example are equivalent. In the stochastic models of financial markets, switching between equivalent probability measures turns out to be important.

In our models of the uncertain evolution of financial markets, the uncertainty is resolved gradually over time. At each date we can observe values of prices and rates that were previously uncertain, so we learn more and more about the true outcome. We need to keep track of the information flow. Let us again consider the throw of a die so that the state space is $\Omega = \{1, 2, 3, 4, 5, 6\}$ and the set $\mathcal{F}$ of probabilizable events consists of all subsets of Ω. Suppose now that the outcome of the throw of the die is not resolved at once, but sequentially. In the beginning, at 'time 0', we know nothing about the true outcome so it can be any element in Ω. Then, at 'time 1', you will be told that the outcome is either in the set $\{1, 2\}$, in the set $\{3, 4, 5\}$, or in the set $\{6\}$. Of course, in the latter case you will know exactly the true outcome, but in the first two cases there is still uncertainty about the true outcome. Later on, at 'time 2', the true outcome will be announced.

We can represent the information available at a given point in time by a **partition** of Ω. By a partition of a given set, we simply mean a collection of disjoint subsets of Ω so that the union of these subsets equals the entire set Ω. At time 0, we only know that one of the six elements in Ω will be realized. This corresponds to the (trivial) partition $F_0 = \{\Omega\}$. The information at time 1 can be represented by the partition

$$F_1 = \big\{\{1, 2\}, \{3, 4, 5\}, \{6\}\big\}.$$

At time 2 we know exactly the true outcome, corresponding to the partition

$$F_2 = \big\{\{1\}, \{2\}, \{3\}, \{4\}, \{5\}, \{6\}\big\}.$$

As time passes we receive more and more information about the true path. This is reflected by the fact that the partitions become finer and finer in the sense that every set in F_1 is a subset of some set in F_0 and every set in F_2 is a subset of some set in F_1. The information flow in this simple example can then be represented by the sequence (F_0, F_1, F_2) of partitions of Ω. In more general models, the information flow can be represented by a sequence $(F_t)_{t \in \mathcal{T}}$ of partitions, where $\mathcal{T}$ is the set of relevant points in time in the model. Each F_t consists of disjoint events and the interpretation is that at time t we will know which of these events the true outcome belongs to. The fact that we learn more and more about the true outcome implies that the partitions will be increasingly fine, meaning that, for $u > t$, every element in F_t is a union of elements in F_u.

An alternative way of representing the information flow is in terms of an **information filtration**. Given a partition F_t of Ω, we can define $\mathcal{F}_t$ as the set of all unions of sets in F_t, including the 'empty union', that is the empty set $\emptyset$. Where F_t contains the **disjoint** 'decidable' events at time t, $\mathcal{F}_t$ contains *all* 'decidable' events at time t. Each $\mathcal{F}_t$ is a sigma-algebra. For our example above we get

$$\mathcal{F}_0 = \{\emptyset, \Omega\},$$
$$\mathcal{F}_1 = \{\emptyset, \{1,2\}, \{3,4,5\}, \{6\}, \{1,2,3,4,5\}, \{1,2,6\}, \{3,4,5,6\}, \Omega\},$$

whereas $\mathcal{F}_2$ becomes the collection of *all* possible subsets of Ω. The sequence $\mathbf{F} = (\mathcal{F}_0, \mathcal{F}_1, \mathcal{F}_2)$ is called an information filtration. In models involving the set $\mathcal{T}$ of points in time, the information filtration is written as $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$. The fact that we accumulate information dictates that $\mathcal{F}_t \subset \mathcal{F}_{t'}$ whenever $t < t'$, that is every set in $\mathcal{F}_t$ is also in $\mathcal{F}_{t'}$.

Above we constructed an information filtration from a sequence of partitions. We can also go from a filtration to a sequence of partitions. In each $\mathcal{F}_t$, simply remove all sets that are unions of other sets in $\mathcal{F}_t$. Therefore there is a one-to-one relationship between information filtration and a sequence of partitions. When we go to models with an infinite state space, the information filtration representation is preferable. Hence, our formal model of uncertainty and information is a **filtered probability space** $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F})$, where $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space and $\mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}}$ is an information filtration. We will always assume that all the uncertainty is resolved over time. Hence, $\mathcal{F}_T = \mathcal{F}$ in an economy where the terminal time point is T. We will also assume that to begin with we know nothing about the future realizations of uncertainty, that is $\mathcal{F}_0$ is the trivial sigma-algebra consisting of only the full state space Ω and the empty set $\emptyset$.

It might seem frightening to have to specify a certain filtered probability space in which the behaviour of interest rates, bond prices, and so on can be studied. However, in the models we are going to consider, the relevant filtered probability space will be implicitly defined via assumptions about the way the key variables can evolve over time.

In our models we will often deal with expectations of random variables, for example the expectation of the (discounted) payoff of an asset at a future point in time. In the computation of such an expectation we should take the information currently available into account. Hence we need to consider conditional expectations. One can generally write the expectation of a random variable X given the σ-algebra $\mathcal{F}_t$ as $\mathrm{E}[X|\mathcal{F}_t]$. For our purposes the σ-algebra $\mathcal{F}_t$ will always represent the information at time t and we will write $\mathrm{E}_t[X]$ instead of $\mathrm{E}[X|\mathcal{F}_t]$. Since we assume that the information at time 0 is trivial, conditioning on time 0 information is the same as not conditioning on any information, hence $\mathrm{E}_0[X] = \mathrm{E}[X]$. If we assume that all uncertainty is resolved at time T, we have $\mathrm{E}_T[X] = X$. We will sometimes use the following result:

Theorem 3.1 (The Law of Iterated Expectations) *If $\mathcal{F}$ and $\mathcal{G}$ are two σ-algebras with $\mathcal{F} \subseteq \mathcal{G}$ and X is a random variable, then $\mathrm{E}\left[\mathrm{E}[X|\mathcal{G}] \,|\, \mathcal{F}\right] = \mathrm{E}[X|\mathcal{F}]$. In particular, if $(\mathcal{F}_t)_{t\in\mathcal{T}}$ is an information filtration and $t' > t$, we have*

$$\mathrm{E}_t\left[\mathrm{E}_{t'}[X]\right] = \mathrm{E}_t[X].$$

Loosely speaking, the theorem says that what you expect today of some variable that will be realized in two days is equal to what you expect today that you will expect tomorrow about the same variable. This is a very intuitive result. For a more formal statement and proof, see Øksendal (2003).

We can define conditional variances, covariances, and correlations from the conditional expectation exactly as one defines (unconditional) variances, covariances, and correlations from (unconditional) expectations:

$$\mathrm{Var}_t[X] = \mathrm{E}_t\left[(X - \mathrm{E}_t[X])^2\right] = \mathrm{E}_t[X^2] - (\mathrm{E}_t[X])^2,$$

$$\mathrm{Cov}_t[X, Y] = \mathrm{E}_t\left[(X - \mathrm{E}_t[X])(Y - \mathrm{E}_t[Y])\right] = \mathrm{E}_t[XY] - \mathrm{E}_t[X]\,\mathrm{E}_t[Y],$$

$$\mathrm{Corr}_t[X, Y] = \frac{\mathrm{Cov}_t[X, Y]}{\sqrt{\mathrm{Var}_t[X]\,\mathrm{Var}_t[Y]}}.$$

Again the conditioning on time t information is indicated by a t subscript.

3.2.2 Random variables and stochastic processes

A random variable is a function from Ω into $\mathbb{R}^K$ for some integer K. The random variable $x : \Omega \to \mathbb{R}^K$ associates to each outcome $\omega \in \Omega$ a value $x(\omega) \in \mathbb{R}^K$. Sometimes we will emphasize the dimension and say that the random variable is K-dimensional. With sequential resolution of the uncertainty the values of some random variables will be known before all uncertainty is resolved.

In the die example with sequential information from before, suppose that your friend George will pay you 10 dollars if the die shows either three, four, or five dots and nothing in other cases. The payment from George is a random variable x. Of course, at time 2 you will know the true outcome, so the payment x will be known at time 2. We say that x is time 2 measurable or $\mathcal{F}_2$-measurable. At time 1 you will also know the payment x because you will be told either that the true outcome is in $\{1, 2\}$, in which case the payment will be 0, or that the true outcome is in $\{3, 4, 5\}$, in which case the payment will be 10, or that the true outcome is 6, in which case the payment will be 0. So the random variable x is also $\mathcal{F}_1$-measurable. Of course, at time 0 you will not know what payment you will get so x is not $\mathcal{F}_0$-measurable. Suppose your friend John promises to pay you 10 dollars if the die shows 4 or 5 and nothing otherwise. Represent the payment from John by the random variable y. Then y is surely $\mathcal{F}_2$-measurable. However, y is not $\mathcal{F}_1$-measurable, because if at time 1 you learn that the true outcome is in $\{3, 4, 5\}$, you still will not know whether you get the 10 dollars or not.

A stochastic process x is a collection of random variables, namely one random variable for each relevant point in time. We write this as $x = (x_t)_{t\in\mathcal{T}}$, where each x_t is a random variable. We still have an underlying filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}})$ representing uncertainty and information flow. We will only consider processes x that are **adapted** in the sense that for every $t \in \mathcal{T}$ the random variable x_t is $\mathcal{F}_t$-measurable. This is just to say that the time t value of the process will be known at time t. Some models consider the dynamic investment decisions of utility-maximizing investors (or other dynamic decisions under uncertainty). The investment decision is represented by a portfolio process characterizing the portfolio to be held at given points in time depending on the information of the investor at that date. Hence, it is natural to require that the portfolio process is adapted to the information filtration. You cannot base investment decisions on information you have not yet received.

By observing a given stochastic process x adapted to a given filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}})$, we obtain some information about the true state. In fact, we can define an information filtration $\mathbf{F}^x = \left(\mathcal{F}_t^x\right)_{t\in\mathcal{T}}$ generated by x. Here, $\mathcal{F}_t^x$ represents the information that can be deduced by knowing the values x_s for $s \leq t$ (for technical reasons, this sigma-algebra is 'completed' by including all sets of $\mathcal{F}$ that have zero $\mathbb{P}$-probability). $\mathbf{F}^x$ is the smallest sigma-algebra with respect to which x is adapted. By construction, $\mathcal{F}_t^x \subseteq \mathcal{F}_t$.

3.2.3 Other important concepts and terminology

Let $x = (x_t)_{t\in\mathcal{T}}$ denote a stochastic process defined on a filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F} = (\mathcal{F}_t)_{t\in\mathcal{T}})$. Each possible outcome $\omega \in \Omega$ will fully determine the value of the process at all points in time. We refer to this collection $(x_t(\omega))_{t\in\mathcal{T}}$ of realized values as a **(sample) path** of the process.

As time goes by, we can observe the evolution in the object which the stochastic process describes. At any given time t', the previous values $(x_t)_{t\leq t'}$ will be known. These values constitute the **history** of the process up to time t'. The future values are (typically) still stochastic.

As time passes and we obtain new information about the true outcome, we will typically revise our expectations of the future values of the process or, more precisely, revise the probability distribution we attribute to the value of the process at any future point in time. Suppose we stand at time t and consider the value of a process x at a future time $t' > t$. The distribution of the value of $x_{t'}$ is characterized by probabilities $\mathbb{P}(x_{t'} \in A)$ for different sets A. If for all $t, t' \in \mathcal{T}$ with $t < t'$ and all A, we have that

$$\mathbb{P}\left(x_{t'} \in A \mid (x_s)_{s\in[0,t]}\right) = \mathbb{P}\left(x_{t'} \in A \mid x_t\right),$$

then x is called a **Markov process**. Broadly speaking, this condition says that, given the present, the future is independent of the past. The history contains no information about the future value that cannot be extracted from the current value. Markov processes are often used in financial models to describe the evolution in prices of financial assets, since the Markov property is consistent with the so-called weak form of market efficiency, which says that extraordinary returns cannot be achieved by use of the precise historical evolution in the price of an asset.[1] If extraordinary returns could be obtained in this manner, all investors would try to profit from them, so that prices would change immediately to a level where the extraordinary return is non-existent. Therefore, it is reasonable to model prices by Markov processes. In addition, models based on Markov processes are often more tractable than models with non-Markov processes.

A stochastic process is said to be a **martingale** if, at all points in time, the expected change in the value of the process over any given future period is equal to zero. In other words, the expected future value of the process is equal to the current

[1] This does not conflict with the fact that the historical evolution is often used to identify some characteristic properties of the process, for example for estimation of means and variances.

value of the process. Because expectations depend on the probability measure, the concept of a martingale should be seen in connection with the applied probability measure. More rigorously, a stochastic process $x = (x_t)_{t \geq 0}$ is a $\mathbb{P}$-martingale if for all $t \in \mathcal{T}$ we have that

$$\mathrm{E}_t^{\mathbb{P}}\left[x_{t'}\right] = x_t, \quad \text{for all } t' \in \mathcal{T} \text{ with } t' > t.$$

Here, $\mathrm{E}_t^{\mathbb{P}}$ denotes the expected value computed under the $\mathbb{P}$-probabilities given the information available at time t, that is, given the history of the process up to and including time t. Sometimes the probability measure will be clear from the context and can be notationally suppressed.

We assume, furthermore, that all the random variables x_t take on values in the same set $\mathcal{S}$, which we call the **value space** of the process. More precisely this means that $\mathcal{S}$ is the smallest set with the property that $\mathbb{P}(\{x_t \in \mathcal{S}\}) = 1$. If $\mathcal{S} \subseteq \mathbb{R}$, we call the process a one-dimensional, real-valued process. If $\mathcal{S}$ is a subset of $\mathbb{R}^K$ (but not a subset of $\mathbb{R}^{K-1}$), the process is called a K-dimensional, real-valued process, which can also be thought of as a collection of K one-dimensional, real-valued processes. Note that as long as we restrict ourselves to equivalent probability measures, the value space will not be affected by changes in the probability measure.

3.2.4 Different types of stochastic processes

A stochastic process for the state of an object at every point in time in a given interval is called a **continuous-time stochastic process**. This corresponds to the case where the set $\mathcal{T}$ takes the form of an interval $[0, T]$ or $[0, \infty)$. In contrast, a stochastic process for the state of an object at countably many separated points in time is called a **discrete-time stochastic process**. This is, for example, the case when $\mathcal{T} = \{0, \Delta t, 2\Delta t, \dots, T \equiv N\Delta t\}$ or $\mathcal{T} = \{0, \Delta t, 2\Delta t, \dots\}$ for some $\Delta t > 0$. If the process can take on all values in a given interval (for example all real numbers), the process is called a **continuous-variable stochastic process**. On the other hand, if the state can take on only countably many different values, the process is called a **discrete-variable stochastic process**.

What type of processes should we use in our models for asset pricing in general and fixed income analysis in particular? Our choice will be guided both by realism and tractability. First, let us consider the time dimension. The investors in the financial markets can trade at more or less any point in time. Due to practical considerations and transaction costs, no investor will trade continuously. However, it is not possible in advance to pick a fairly moderate number of points in time where all trades take place. Also, with many investors there will be some trades at almost any point in time, so that prices and interest rates and so on will also change almost continuously. Therefore, it seems to be a better approximation of real life to describe such economic variables by continuous-time stochastic processes than by discrete-time stochastic processes. Continuous-time stochastic processes are in many aspects also easier to handle than discrete-time stochastic processes.

Next, consider the value dimension. Strictly speaking, most economic variables can only take on countably many values in practice. Stock prices are multiples of the smallest possible unit (0.01 currency units in many countries), and interest rates are only stated with a given number of decimals. But since the possible values are very close together, it seems reasonable to use continuous-variable processes in the modelling of these objects. In addition, the mathematics involved in the analysis of continuous-variable processes is simpler and more elegant than the mathematics for discrete-variable processes. Integrals are easier to deal with than sums, derivatives are easier to handle than differences, and so on. Some interest rate models were originally formulated using discrete-time, discrete-variable processes as, for example, the binomial tree models introduced by Ho and Lee (1986) and Black et al. (1990). For many years, all significant model developments have applied continuous-time, continuous-variable processes, and such continuous-time term structure models are now standard in the financial industry and in academic work. In sum, we will use continuous-time, continuous-variable stochastic processes throughout to describe the evolution in prices and rates. Therefore the remaining sections of this chapter will be devoted to that type of stochastic processes.

It should be noted that discrete-time and/or discrete-variable processes also have their virtues. First, many concepts and results are more easily understood or illustrated in a simple framework. Second, even if we have low-frequency data for many financial variables, we do not have continuous data. When it comes to estimation of parameters in financial models, continuous-time processes often have to be approximated by discrete-time processes. Third, although explicit results on asset prices, optimal investment strategies, and so on are easier to obtain with continuous-time models, not all relevant questions can be explicitly answered. Some problems are solved numerically by computer algorithms and also for that purpose it is often necessary to approximate continuous-time, continuous-variable processes with discrete-time, discrete-variable processes (see Chapter 16).

3.2.5 How to write up stochastic processes

Many financial models describe the movements and co-movements of various variables simultaneously. In fixed income models we are interested in the dynamic behaviour of yields of bonds of different maturities, prices of different bonds and options, and so on. The standard modelling procedure is to assume that there is some common exogenous shock that affects all the relevant variables and then model the response of all these variables to that shock. First, consider a discrete-time framework with time set $\mathcal{T} = \{0, t_1, t_2, \dots, t_N \equiv T\}$ where $t_n = n\Delta t$. The shock over any period $[t_n, t_{n+1}]$ is represented by a random variable $\varepsilon_{t_{n+1}}$, which in general may be multi-dimensional, but let us for now just focus on the one-dimensional case. The sequence of shocks $\varepsilon_{t_1}, \varepsilon_{t_2}, \dots, \varepsilon_{t_N}$ constitutes the basic or the underlying uncertainty in the model. Since the shock should represent some unexpected information, assume that every ε_{t_n} has mean zero.

A stochastic process $x = (x_t)_{t\in\mathcal{T}}$ representing the dynamics of a price, an interest rate, or another interesting variable can then be defined by the initial value x_0 and the increments $\Delta x_{t_{n+1}} \equiv x_{t_{n+1}} - x_{t_n}$, $n = 0, \dots, N-1$, which are typically assumed to be of the form

$$\Delta x_{t_{n+1}} = \mu_{t_n}\Delta t + \sigma_{t_n}\varepsilon_{t_{n+1}}. \tag{3.1}$$

In general μ_{t_n} and σ_{t_n} can themselves be stochastic, but must be known at time t_n, that is they must be $\mathcal{F}_{t_n}$-measurable random variables. In fact, we can form adapted processes $\mu = (\mu_t)_{t\in\mathcal{T}}$ and $\sigma = (\sigma_t)_{t\in\mathcal{T}}$. Given the information available at time t_n, the only random variable on the right-hand side of (3.1) is $\varepsilon_{t_{n+1}}$, which is assumed to have mean zero and some variance $\text{Var}[\varepsilon_{t_{n+1}}]$. Hence, the mean and variance of $\Delta x_{t_{n+1}}$, conditional on time t_n information, are

$$\mathrm{E}_{t_n}[\Delta x_{t_{n+1}}] = \mu_{t_n}\Delta t, \qquad \text{Var}_{t_n}[\Delta x_{t_{n+1}}] = \sigma_{t_n}^2 \,\text{Var}[\varepsilon_{t_{n+1}}].$$

We can see that μ_{t_n} has the interpretation of the expected change in x per time period.

If the shocks $\varepsilon_{t_1}, \dots, \eta_{t_N}$ are the only source of randomness in all the quantities we care about, then the relevant information filtration is exactly $\mathbf{F}^{\varepsilon} = \left(\mathcal{F}_t^{\varepsilon}\right)_{t\in\mathcal{T}}$, that is $\mathcal{F}_t = \mathcal{F}_t^{\varepsilon}$. In that case μ_{t_n} and σ_{t_n} are required to be measurable with respect to $\mathcal{F}_{t_n}^{\varepsilon}$, that is they can depend on the realizations of $\varepsilon_{t_1}, \dots, \varepsilon_{t_n}$. If σ_{t_n} is non-zero at all times and for all states, we can invert (3.1) to get

$$\varepsilon_{t_{n+1}} = \frac{\Delta x_{t_{n+1}} - \mu_{t_n}\Delta t}{\sigma_{t_n}}.$$

It is then clear that we learn exactly the same from observing the x-process as observing the exogenous shocks directly, that is $\mathbf{F}^x = \mathbf{F}^{\varepsilon} = \mathbf{F}$. We can fix the set of probabilizable events $\mathcal{F}$ to $\mathcal{F}_T^{\varepsilon} = \mathcal{F}_T^x$. The probability measure $\mathbb{P}$ will be defined by specifying the probability distribution of each of the shocks ε_{t_n}.

From the sequence $\varepsilon_{t_1}, \varepsilon_{t_2}, \dots, \varepsilon_{t_N}$ of exogenous shocks we can define a stochastic process $z = (z_t)_{t\in\mathcal{T}}$ by letting $z_0 = 0$ and $z_{t_n} = \varepsilon_{t_1} + \dots + \varepsilon_{t_n}$. Consequently, $\varepsilon_{t_{n+1}} = z_{t_{n+1}} - z_{t_n} \equiv \Delta z_{t_{n+1}}$. Now the process z captures the basic uncertainty in the model. The information filtration of the model is then defined by the information that can be extracted from observing the path of z. Without loss of generality we can assume that $\text{Var}[\Delta z_{t_{n+1}}] = \text{Var}[\varepsilon_{t_{n+1}}] = \Delta t$ for any period $[t_n, t_{n+1}]$. With the z-notation we can rewrite (3.1) as

$$\Delta x_{t_{n+1}} = \mu_{t_n}\Delta t + \sigma_{t_n}\Delta z_{t_{n+1}} \tag{3.2}$$

and now $\text{Var}_{t_n}[\Delta x_{t_{n+1}}] = \sigma_{t_n}^2\Delta t$ so that $\sigma_{t_n}^2$ can be interpreted as the variance of the change in x per time period.

The distribution of $\Delta x_{t_{n+1}}$ will be determined by the distribution assumed for the shocks $\varepsilon_{t_{n+1}} = \Delta z_{t_{n+1}}$. If the shocks are assumed to be normally distributed, the increment $\Delta x_{t_{n+1}}$ will be normally distributed conditional on time t information, but not necessarily if we condition on earlier or no information.

We can loosely think of a continuous-time model as the result of taking a discrete-time model and letting Δt go to zero. In that spirit we will often define a continuous-time stochastic process $x = (x_t)_{t\in\mathcal{T}}$ by writing

$$dx_t = \mu_t\, dt + \sigma_t\, dz_t \tag{3.3}$$

which is to be thought of as the limit of (3.2) as $\Delta t \to 0$. Hence, dx_t represents the change in x over the infinitesimal (infinitely short) period after time t. Similarly for dz_t. The interpretations of μ_t and σ_t are also similar to the discrete-time case. While (3.3) might seem very intuitive, it does not really make much sense to talk about the change of something over a period of infinitesimal length. The expression (3.3) really means that the change in the value of x over any time interval $[t, t'] \subseteq \mathcal{T}$ is given by

$$x_{t'} - x_t = \int_t^{t'} \mu_u \, du + \int_t^{t'} \sigma_u \, dz_u.$$

The problem is that the right-hand side of this equation will not make sense before we define the two integrals. The integral $\int_t^{t'} \mu_u \, du$ is simply defined as the random variable whose value in any state $\omega \in \Omega$ is given by $\int_t^{t'} \mu_u(\omega) \, du$, which is an ordinary integral of a real-valued function of time. If μ is adapted, the value of the integral $\int_t^{t'} \mu_u \, du$ will become known at time t'. The definition of the integral $\int_t^{t'} \sigma_u \, dz_u$ is much more delicate. We will return to this issue in Section 3.6.

In almost all the continuous-time models studied in this book we will assume that the basic exogenous shocks are normally distributed, that is that the change in the shock process z over any time interval is normally distributed. A process z with this property is the so-called standard Brownian motion. In the next section we will formally define this process and study some of its properties. Then in later sections we will build various processes x from the basic process e.

3.3 BROWNIAN MOTIONS

All the stochastic processes we shall apply in the financial models in the following chapters build upon a particular class of processes, the so-called Brownian motions. A (one-dimensional) stochastic process $z = (z_t)_{t \geq 0}$ is called a **standard Brownian motion**, if it satisfies the following conditions:

(i) $z_0 = 0$,
(ii) for all $t, t' \geq 0$ with $t < t'$: $z_{t'} - z_t \sim N(0, t' - t)$ [normally distributed increments],
(iii) for all $0 \leq t_0 < t_1 < \cdots < t_n$, the random variables $z_{t_1} - z_{t_0}, \ldots, z_{t_n} - z_{t_{n-1}}$ are mutually independent [independent increments],
(iv) z has continuous paths.

Here $N(a, b)$ denotes the normal distribution with mean a and variance b.

If we suppose that a standard Brownian motion z represents the basic exogenous shock to an economy over a time interval $[0, T]$, then the relevant filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, \mathbf{F})$ is implicitly given as follows. The state space Ω is the set of all possible paths $(z_t)_{t \in [0,T]}$. The information filtration is the one generated by z, that is $\mathbf{F} = \mathbf{F}^z$. The set of probabilizable events $\mathcal{F}$ is equal to $\mathcal{F}_T^z$. The probability measure $\mathbb{P}$ is defined by the requirement that

$$\mathbb{P}\left(\frac{z_{t'} - z_t}{\sqrt{t' - t}} < h \right) = N(h) \equiv \int_{-\infty}^{h} \frac{1}{\sqrt{2\pi}} e^{-a^2/2} \, da$$

for all $t < t'$ and all $h \in \mathbb{R}$, where $N(\cdot)$ denotes the cumulative distribution function for an $N(0,1)$-distributed random stochastic variable.

Note that a standard Brownian motion is a Markov process, since the increment from today to any future point in time is independent of the history of the process. A standard Brownian motion is also a martingale, since the expected change in the value of the process is zero.

The name Brownian motion is in honor of the Scottish botanist Robert Brown, who in 1828 observed the apparently random movements of pollen submerged in water. The often-used name 'Wiener process' is due to Norbert Wiener, who in the 1920s was the first to show the existence of a stochastic process with these properties and who initiated a mathematically rigorous analysis of the process. As early as the year 1900, the standard Brownian motion was used in a model for stock price movements by the French researcher Louis Bachelier, who derived the first option pricing formula, see Bachelier (1900).

The choice of using standard Brownian motions to represent the underlying uncertainty has an important consequence. All the processes defined by equations of the form (3.3) will then have continuous paths, that is there will be no jumps. Stochastic processes which have paths with discontinuities also exist. The jumps of such processes are often modelled by Poisson processes or related processes. It is well-known that large, sudden movements in financial variables occur from time to time, for example, in connection with stock market crashes. There may be many explanations of such large movements, for example, a large unexpected change in the productivity in a particular industry or the economy in general, perhaps due to a technological breakthrough. Another source of sudden, large movements is changes in the political or economic environment, such as unforseen interventions by the government or central bank. Stock market crashes are sometimes explained by the bursting of a bubble. Whether such sudden, large movements can be explained by a sequence of small continuous movements in the same direction, or jumps have to be included in the models, is an empirical question which is still open. Large movements over a short period of time seem to be less frequent in interest rates and bond prices than in stock prices.

There are numerous financial models of stock markets that allow for jumps in stock prices, for example Merton (1976) discusses the pricing of stock options in such a framework. There are still relatively very few models allowing for jumps in interest rates, but in recent years jump processes have gained popularity in interest rate modelling as well (references will be given in later chapters). Of course, models for corporate bonds must be able to handle the possible default of the issuing company, which in some cases comes as a surprise to the financial market. Therefore, such models will typically involve jump processes. We will study such models in Chapter 13, but in the other chapters we will focus on default-free contracts and use processes with continuous sample paths.

The defining characteristics of a standard Brownian motion look very nice, but they have some drastic consequences. It can be shown that the paths of a standard Brownian motion are nowhere differentiable, which broadly speaking means that the paths bend at all points in time and are therefore strictly speaking impossible to illustrate. However, one can get an idea of the paths by simulating the values of the process at different times. If $\varepsilon_1, \dots, \varepsilon_n$ are independent draws from a standard

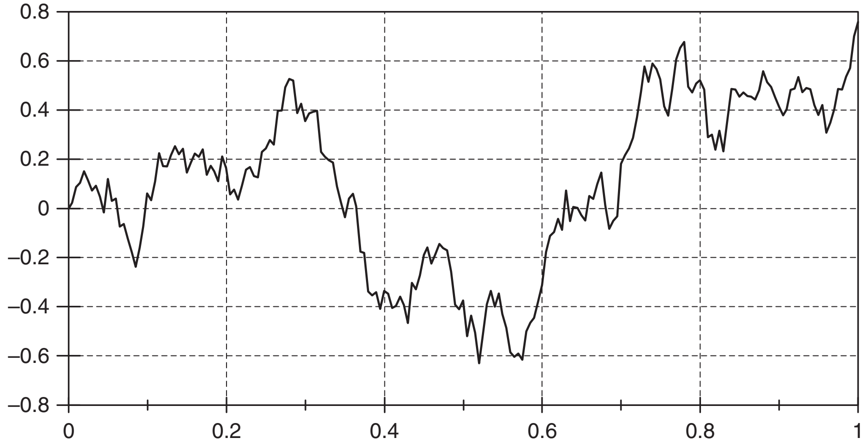

Fig. 3.1: A simulated path of a standard Brownian motion based on 200 subintervals.

$N(0,1)$ distribution, we can simulate the value of the standard Brownian motion at time $0 \equiv t_0 < t_1 < t_2 < \cdots < t_n$ as follows:

$$z_{t_i} = z_{t_{i-1}} + \varepsilon_i \sqrt{t_i - t_{i-1}}, \quad i = 1, \ldots, n.$$

With more time points and hence shorter intervals we get a more realistic impression of the paths of the process. Figure 3.1 shows a simulated path for a standard Brownian motion over the interval $[0,1]$ based on a partition of the interval into 200 subintervals of equal length. More details on the simulation of stochastic processes are given in Chapter 16. Note that since a normally distributed random variable can take on infinitely many values, a standard Brownian motion has infinitely many paths that each have a zero probability of occurring. The figure shows just one possible path.

Another property of a standard Brownian motion is that the expected length of the path over any future time interval (no matter how short) is infinite. In addition, the expected number of times a standard Brownian motion takes on any given value in any given time interval is also infinite. Intuitively, these properties are due to the fact that the size of the increment of a standard Brownian motion over an interval of length Δt is proportional to $\sqrt{\Delta t}$, in the sense that the standard deviation of the increment equals $\sqrt{\Delta t}$. When Δt is close to zero, $\sqrt{\Delta t}$ is significantly larger than Δt, so the changes are large relative to the length of the time interval over which the changes are measured.

The expected change in an object described by a standard Brownian motion equals zero and the variance of the change over a given time interval equals the length of the interval. This can easily be generalized. As before let $z = (z_t)_{t \geq 0}$ be a one-dimensional standard Brownian motion and define a new stochastic process $x = (x_t)_{t \geq 0}$ by

$$x_t = x_0 + \mu t + \sigma z_t, \quad t \geq 0,$$

where x_0, μ, and σ are constants. The constant x_0 is the initial value for the process x. It follows from the properties of the standard Brownian motion that, seen from time 0, the value x_t is normally distributed with mean $x_0 + \mu t$ and variance $\sigma^2 t$, that is $x_t \sim N(x_0 + \mu t, \sigma^2 t)$.

The change in the value of the process between two arbitrary points in time t and t', where $t < t'$, is given by

$$x_{t'} - x_t = \mu(t' - t) + \sigma(z_{t'} - z_t).$$

The change over an infinitesimally short interval $[t, t + \Delta t]$ with $\Delta t \to 0$ is often written as

$$dx_t = \mu \, dt + \sigma \, dz_t, \tag{3.4}$$

where dz_t can loosely be interpreted as a $N(0, dt)$-distributed random variable. As discussed earlier, this must really be interpreted as a limit of the expression

$$x_{t+\Delta t} - x_t = \mu \Delta t + \sigma(z_{t+\Delta t} - z_t)$$

for $\Delta t \to 0$. The process x is called a **generalized Brownian motion**, or an arithmetic Brownian motion, or a generalized Wiener process. The parameter μ reflects the expected change in the process per unit of time and is called the drift rate or simply the **drift** of the process. The parameter σ reflects the uncertainty about the future values of the process. More precisely, σ^2 reflects the variance of the change in the process per unit of time and is often called the **variance rate** of the process. σ is a measure for the standard deviation of the change per unit of time and is referred to as the **volatility** of the process.

A generalized Brownian motion inherits many of the characteristic properties of a standard Brownian motion. For example, a generalized Brownian motion is also a Markov process, and the paths of a generalized Brownian motion are also continuous and nowhere differentiable. However, a generalized Brownian motion is not a martingale unless $\mu = 0$. The paths can be simulated by choosing time points $0 \equiv t_0 < t_1 < \cdots < t_n$ and iteratively computing

$$x_{t_i} = x_{t_{i-1}} + \mu(t_i - t_{i-1}) + \varepsilon_i \sigma \sqrt{t_i - t_{i-1}}, \quad i = 1, \ldots, n,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are independent draws from a standard normal distribution. Figures 3.2 and 3.3 show simulated paths for different values of the parameters μ and σ. The straight lines represent the deterministic trend of the process, which corresponds to imposing the condition $\sigma = 0$ and hence ignoring the uncertainty. Both figures are drawn using the same sequence of random numbers ε_i, so that they are directly comparable. The parameter μ determines the trend, and the parameter σ determines the size of the fluctuations around the trend.

If the parameters μ and σ are allowed to be time-varying in a deterministic way, the process x is said to be a **time-inhomogeneous** generalized Brownian motion. In differential terms such a process can be written as defined by

$$dx_t = \mu(t) \, dt + \sigma(t) \, dz_t. \tag{3.5}$$

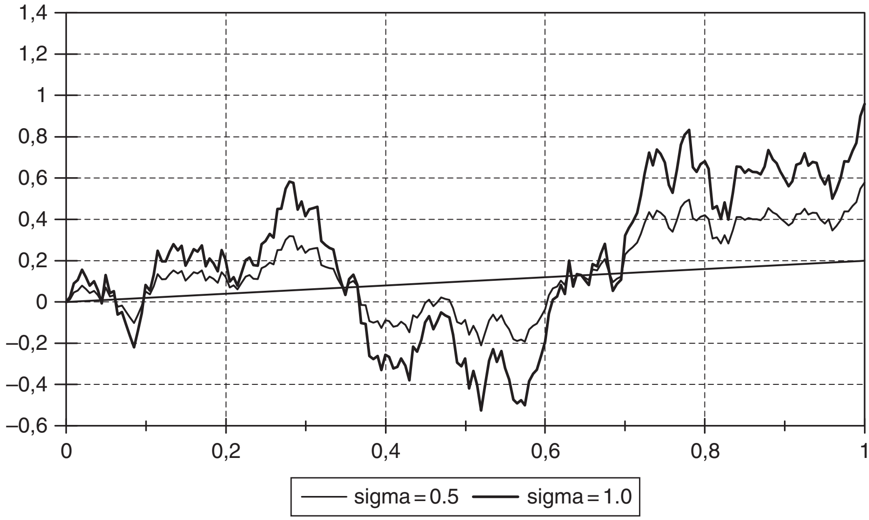

Fig. 3.2: Simulation of a generalized Brownian motion with $\mu = 0.2$ and $\sigma = 0.5$ or $\sigma = 1.0$. The straight line shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals.

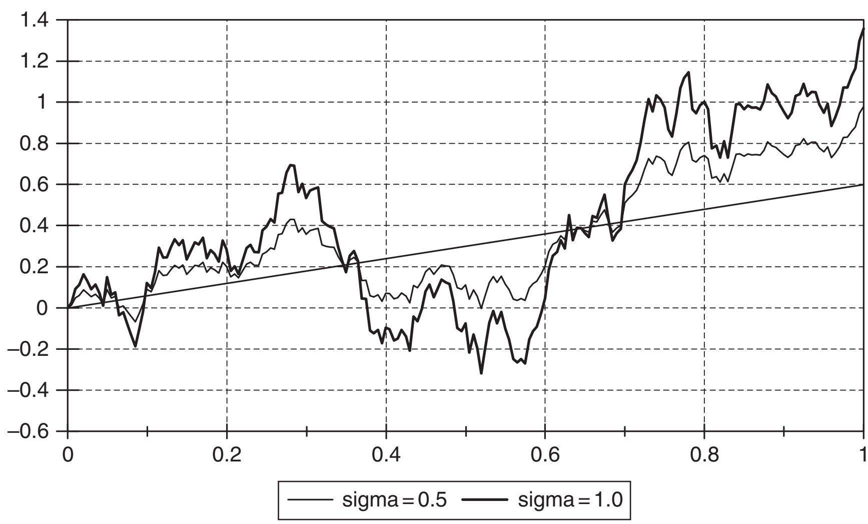

Fig. 3.3: Simulation of a generalized Brownian motion with $\mu = 0.6$ and $\sigma = 0.5$ or $\sigma = 1.0$. The straight line shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals.

Over a very short interval $[t, t+\Delta t]$ the expected change is approximately $\mu(t)\Delta t$, and the variance of the change is approximately $\sigma(t)^2\Delta t$. More precisely, the increment over any interval $[t, t']$ is given by

$$x_{t'} - x_t = \int_t^{t'} \mu(u)\, du + \int_t^{t'} \sigma(u)\, dz_u.$$

The last integral is a so-called stochastic integral, which we will define and describe in a later section. There we will also state a theorem, which implies that, seen from time t, the integral $\int_t^{t'} \sigma(u)\, dz_u$ is a normally distributed random variable with mean zero and variance $\int_t^{t'} \sigma(u)^2\, du$.

3.4 DIFFUSION PROCESSES

For both standard Brownian motions and generalized Brownian motions, the future value is normally distributed and can therefore take on any real value, that is the value space is equal to $\mathbb{R}$. Many economic variables can only have values in a certain subset of $\mathbb{R}$. For example, prices of financial assets with limited liability are non-negative. The evolution in such variables cannot be well represented by the stochastic processes studied so far. In many situations we will instead use so-called diffusion processes.

A (one-dimensional) **diffusion process** is a stochastic process $x = (x_t)_{t\geq 0}$ for which the change over an infinitesimally short time interval $[t, t+dt]$ can be written as

$$dx_t = \mu(x_t, t)\, dt + \sigma(x_t, t)\, dz_t, \tag{3.6}$$

where z is a standard Brownian motion, but where the drift μ and the volatility σ are now functions of time and the current value of the process.[2] This expression generalizes (3.4), where μ and σ were assumed to be constants, and (3.5), where μ and σ were functions of time only. An equation like (3.6), where the stochastic process enters both sides of the equality, is called a **stochastic differential equation**. Hence, a diffusion process is a solution to a stochastic differential equation.

If both functions μ and σ are independent of time, the diffusion is said to be **time-homogeneous**, otherwise it is said to be time-inhomogeneous. For a time-homogeneous diffusion process, the distribution of the future value will only depend on the current value of the process and how far into the future we are looking—not on the particular point in time we are standing at. For example, the distribution of $x_{t+\delta}$ given $x_t = x$ will only depend on x and δ, but not on t. This is not the case for a time-inhomogeneous diffusion, where the distribution will also depend on t.

[2] For the process x to be mathematically meaningful, the functions $\mu(x,t)$ and $\sigma(x,t)$ must satisfy certain conditions. See, for example Øksendal (2003, Ch. 7) and Duffie (2001, App. E).

In the expression (3.6) one may think of dz_t as being $N(0, dt)$-distributed, so that the mean and variance of the change over an infinitesimally short interval $[t, t + dt]$ are given by

$$\mathrm{E}_t[dx_t] = \mu(x_t, t)\, dt, \qquad \mathrm{Var}_t[dx_t] = \sigma(x_t, t)^2\, dt,$$

where E_t and Var_t denote the mean and variance, respectively, conditionally on the available information at time t. To be more precise, the change in a diffusion process over any interval $[t, t']$ is

$$x_{t'} - x_t = \int_t^{t'} \mu(x_u, u)\, du + \int_t^{t'} \sigma(x_u, u)\, dz_u, \tag{3.7}$$

where $\int_t^{t'} \sigma(x_u, u)\, dz_u$ is a stochastic integral, which we will discuss in Section 3.6. However, we will continue to use the simple and intuitive differential notation (3.6). The drift rate $\mu(x_t, t)$ and the variance rate $\sigma(x_t, t)^2$ are really the limits

$$\mu(x_t, t) = \lim_{\Delta t \to 0} \frac{\mathrm{E}_t\left[x_{t+\Delta t} - x_t\right]}{\Delta t},$$

$$\sigma(x_t, t)^2 = \lim_{\Delta t \to 0} \frac{\mathrm{Var}_t\left[x_{t+\Delta t} - x_t\right]}{\Delta t}.$$

A diffusion process is a Markov process as can be seen from (3.6), since both the drift and the volatility only depend on the current value of the process and not on previous values. A diffusion process is not a martingale, unless the drift $\mu(x_t, t)$ is zero for all x_t and t. A diffusion process will have continuous, but nowhere differentiable, paths. The value space for a diffusion process and the distribution of future values will depend on the functions μ and σ. If $\sigma(x, t)$ is continuous and non-zero, the information generated by x will be identical to the information generated by z, that is $\mathbf{F}^x = \mathbf{F}^z$.

In Section 3.8 we will give some important examples of diffusion processes which we shall use in later chapters to model the evolution of some economic variables.

3.5 ITÔ PROCESSES

It is possible to define even more general continuous-variable stochastic processes than those in the class of diffusion processes. A (one-dimensional) stochastic process x_t is said to be an **Itô process**, if the local increments are of the form

$$dx_t = \mu_t\, dt + \sigma_t\, dz_t, \tag{3.8}$$

where the drift μ and the volatility σ themselves are stochastic processes. A diffusion process is the special case where the values of the drift μ_t and the volatility σ_t are given by t and x_t. For a general Itô process, the drift and volatility may also depend on past values of the x process. Or the drift and volatility can depend on another exogenous shock, for example, another standard Brownian motion than z. It follows that Itô processes are generally not Markov processes. They are generally not martingales either, unless μ_t is identically equal to zero (and σ_t satisfies some technical conditions). The processes μ and σ must satisfy certain regularity conditions for the x process to be well-defined. We will refer the reader to Øksendal (2003, Ch. 4).

The expression (3.8) gives an intuitive understanding of the evolution of an Itô process, but it is more precise to state the evolution in the integral form

$$x_{t'} - x_t = \int_t^{t'} \mu_u \, du + \int_t^{t'} \sigma_u \, dz_u, \tag{3.9}$$

where the last term again is a stochastic integral.

3.6 STOCHASTIC INTEGRALS

3.6.1 Definition and properties of stochastic integrals

In (3.7) and (3.9) and similar expressions a term of the form $\int_t^{t'} \sigma_u \, dz_u$ appears. An integral of this type is called a stochastic integral or an Itô integral. We will only consider stochastic integrals where the 'integrator' z is a standard Brownian motion, although stochastic integrals involving more general processes can also be defined. For given $t < t'$, the stochastic integral $\int_t^{t'} \sigma_u \, dz_u$ is a random variable. Assuming that σ_u is known at time u, the value of the integral becomes known at time t'. The process σ is called the integrand.

The stochastic integral can be defined for very general integrands. The simplest integrands are those that are piecewise constant. Assume that there are points in time $t \equiv t_0 < t_1 < \cdots < t_n \equiv t'$, so that σ_u is constant on each subinterval $[t_i, t_{i+1})$. The stochastic integral is then defined by

$$\int_t^{t'} \sigma_u \, dz_u = \sum_{i=0}^{n-1} \sigma_{t_i} \left(z_{t_{i+1}} - z_{t_i} \right).$$

If the integrand process σ is not piecewise constant, a sequence of piecewise constant processes $\sigma^{(1)}, \sigma^{(2)}, \dots$ exists, which converges to σ. For each of the processes $\sigma^{(m)}$, the integral $\int_t^{t'} \sigma_u^{(m)} \, dz_u$ is defined as above. The integral $\int_t^{t'} \sigma_u \, dz_u$ is then defined as a limit of the integrals of the approximating processes:

$$\int_t^{t'} \sigma_u \, dz_u = \lim_{m \to \infty} \int_t^{t'} \sigma_u^{(m)} \, dz_u.$$

We will not discuss exactly how this limit is to be understood and which integrand processes we can allow. Again the interested reader is referred to Øksendal (2003). The distribution of the integral $\int_t^{t'} \sigma_u \, dz_u$ will, of course, depend on the integrand process and can generally not be completely characterized, but the following theorem gives the mean and the variance of the integral:

Theorem 3.2 *If* $\sigma = (\sigma_t)$ *satisfies some regularity conditions, the stochastic integral* $\int_t^{t'} \sigma_u \, dz_u$ *has the following properties:*

$$\mathrm{E}_t\left[\int_t^{t'} \sigma_u \, dz_u\right] = 0,$$

$$\mathrm{Var}_t\left[\int_t^{t'} \sigma_u \, dz_u\right] = \int_t^{t'} \mathrm{E}_t[\sigma_u^2] \, du.$$

Proof: Suppose that σ is piecewise constant and divide the interval $[t, t']$ into subintervals defined by the time points $t \equiv t_0 < t_1 < \cdots < t_n \equiv t'$ so that σ is constant on each subinterval $[t_i, t_{i+1})$ with a value σ_{t_i} which is known at time t_i. Then

$$\mathrm{E}_t\left[\int_t^{t'} \sigma_u \, dz_u\right] = \sum_{i=0}^{n-1} \mathrm{E}_t\left[\sigma_{t_i}\left(z_{t_{i+1}} - z_{t_i}\right)\right] = \sum_{i=0}^{n-1} \mathrm{E}_t\left[\sigma_{t_i} \mathrm{E}_{t_i}\left[\left(z_{t_{i+1}} - z_{t_i}\right)\right]\right] = 0,$$

using the Law of Iterated Expectations. For the variance we have

$$\begin{aligned}
\mathrm{Var}_t\left[\int_t^{t'} \sigma_u \, dz_u\right] &= \mathrm{E}_t\left[\left(\int_t^{t'} \sigma_u \, dz_u\right)^2\right] - \left(\mathrm{E}_t\left[\int_t^{t'} \sigma_u \, dz_u\right]\right)^2 \\
&= \mathrm{E}_t\left[\left(\int_t^{t'} \sigma_u \, dz_u\right)^2\right]
\end{aligned}$$

and

$$\begin{aligned}
\mathrm{E}_t\left[\left(\int_t^{t'} \sigma_u \, dz_u\right)^2\right] &= \mathrm{E}_t\left[\sum_{i=0}^{n-1}\sum_{j=0}^{n-1} \sigma_{t_i}\sigma_{t_j}(z_{t_{i+1}} - z_{t_i})(z_{t_{j+1}} - z_{t_j})\right] \\
&= \sum_{i=0}^{n-1} \mathrm{E}_t\left[\sigma_{t_i}^2 (z_{t_{i+1}} - z_{t_i})^2\right] = \sum_{i=0}^{n-1} \mathrm{E}_t\left[\sigma_{t_i}^2\right](t_{i+1} - t_i) \\
&= \int_t^{t'} \mathrm{E}_t[\sigma_u^2] \, du.
\end{aligned}$$

If σ is not piecewise constant, we can approximate it by a piecewise constant process and take appropriate limits. We skip the details. □

If the integrand is a deterministic function of time, $\sigma(u)$, the integral will be normally distributed, so that the following result holds:

Theorem 3.3 *If $\sigma(u)$ is a deterministic function of time, the random variable $\int_t^{t'} \sigma(u)\, dz_u$ is normally distributed with mean zero and variance $\int_t^{t'} \sigma(u)^2\, du$.*

Proof: We present a sketch of the proof. Dividing the interval $[t, t']$ into subintervals defined by the time points $t \equiv t_0 < t_1 < \cdots < t_n \equiv t'$, we can approximate the integral with a sum,

$$\int_t^{t'} \sigma(u)\, dz_u \approx \sum_{i=0}^{n-1} \sigma(t_i) \left(z_{t_{i+1}} - z_{t_i}\right).$$

The increment of the Brownian motion over any subinterval is normally distributed with mean zero and a variance equal to the length of the subinterval. Furthermore, the different terms in the sum are mutually independent. It is well-known that a sum of normally distributed random variables is itself normally distributed, and that the mean of the sum is equal to the sum of the means, which in the present case yields zero. Due to the independence of the terms in the sum, the variance of the sum is also equal to the sum of the variances, that is

$$\begin{aligned}
\operatorname{Var}_t\left(\sum_{i=0}^{n-1} \sigma(t_i)\left(z_{t_{i+1}} - z_{t_i}\right)\right) &= \sum_{i=0}^{n-1} \sigma(t_i)^2 \operatorname{Var}_t\left(z_{t_{i+1}} - z_{t_i}\right) \\
&= \sum_{i=0}^{n-1} \sigma(t_i)^2 (t_{i+1} - t_i),
\end{aligned}$$

which is an approximation of the integral $\int_t^{t'} \sigma(u)^2\, du$. The result now follows from an appropriate limit where the subintervals shrink to zero length. □

Note that the process $y = (y_t)_{t \geq 0}$ defined by $y_t = \int_0^t \sigma_u\, dz_u$ is a martingale (under regularity conditions on σ), since

$$\begin{aligned}
\mathrm{E}_t[y_{t'}] &= \mathrm{E}_t\left[\int_0^{t'} \sigma_u\, dz_u\right] = \mathrm{E}_t\left[\int_0^{t} \sigma_u\, dz_u + \int_t^{t'} \sigma_u\, dz_u\right] \\
&= \mathrm{E}_t\left[\int_0^{t} \sigma_u\, dz_u\right] + \mathrm{E}_t\left[\int_t^{t'} \sigma_u\, dz_u\right] = \int_0^t \sigma_u\, dz_u = y_t,
\end{aligned}$$

so that the expected future value is equal to the current value. More generally $y_t = y_0 + \int_0^t \sigma_u\, dz_u$ for some constant y_0, is a martingale. The converse is also true in the sense that any martingale can be expressed as a stochastic integral. This is the so-called martingale representation theorem:

Theorem 3.4 *Suppose the process $M = (M_t)$ is a martingale with respect to a filtered probability space implicitly defined by the standard Brownian motion $z = (z_t)_{t \in [0,T]}$ so that, in particular, the information filtration is $\mathbf{F} = \mathbf{F}^z$. Then a unique adapted process $\theta = (\theta_t)$ exists such that*

$$M_t = M_0 + \int_0^t \theta_u \, dz_u$$

for all t.

For a mathematically more precise statement of the result and a proof, see Øksendal (2003, Thm. 4.3.4).

3.6.2 Leibnitz' rule for stochastic integrals

Leibnitz' differentiation rule for ordinary integrals is as follows: If $f(t, s)$ is a deterministic function, and we define $Y(t) = \int_t^T f(t, s) \, ds$, then

$$Y'(t) = -f(t, t) + \int_t^T \frac{\partial f}{\partial t}(t, s) \, ds.$$

If we use the notation $Y'(t) = \frac{dY}{dt}$ and $\frac{\partial f}{\partial t} = \frac{df}{dt}$, we can rewrite this result as

$$dY = -f(t, t) \, dt + \left(\int_t^T \frac{df}{dt}(t, s) \, ds \right) dt,$$

and formally cancelling the dt-terms, we get

$$dY = -f(t, t) \, dt + \int_t^T df(t, s) \, ds.$$

We will now consider a similar result in the case where $f(t, s)$ and, hence, $Y(t)$ are stochastic processes. We will make use of this result in Chapter 10 (and only in that chapter).

Theorem 3.5 *For any $s \in [t_0, T]$, let $f^s = (f_t^s)_{t \in [t_0, s]}$ be the Itô process defined by the dynamics*

$$df_t^s = \alpha_t^s \, dt + \beta_t^s \, dz_t,$$

where α^s and β^s are sufficiently well-behaved stochastic processes. Then the dynamics of the stochastic process $Y_t = \int_t^T f_t^s \, ds$ is given by

$$dY_t = \left[\left(\int_t^T \alpha_t^s \, ds \right) - f_t^t \right] dt + \left(\int_t^T \beta_t^s \, ds \right) dz_t.$$

Since the result is usually not included in standard textbooks on stochastic calculus, a sketch of the proof is included. The proof applies the generalized Fubini rule for stochastic processes, which was stated and demonstrated in the appendix of Heath et al. (1992). The Fubini rule says that the order of integration in double integrals can be reversed, if the integrand is a sufficiently well-behaved function—we will assume that this is indeed the case.

Proof: Given any arbitrary $t_1 \in [t_0, T]$, since

$$f_{t_1}^s = f_{t_0}^s + \int_{t_0}^{t_1} \alpha_t^s \, dt + \int_{t_0}^{t_1} \beta_t^s \, dz_t,$$

we get

$$\begin{aligned}
Y_{t_1} &= \int_{t_1}^{T} f_{t_0}^s \, ds + \int_{t_1}^{T} \left[\int_{t_0}^{t_1} \alpha_t^s \, dt \right] ds + \int_{t_1}^{T} \left[\int_{t_0}^{t_1} \beta_t^s \, dz_t \right] ds \\
&= \int_{t_1}^{T} f_{t_0}^s \, ds + \int_{t_0}^{t_1} \left[\int_{t_1}^{T} \alpha_t^s \, ds \right] dt + \int_{t_0}^{t_1} \left[\int_{t_1}^{T} \beta_t^s \, ds \right] dz_t \\
&= Y_{t_0} + \int_{t_0}^{t_1} \left[\int_{t}^{T} \alpha_t^s \, ds \right] dt + \int_{t_0}^{t_1} \left[\int_{t}^{T} \beta_t^s \, ds \right] dz_t \\
&\quad - \int_{t_0}^{t_1} f_{t_0}^s \, ds - \int_{t_0}^{t_1} \left[\int_{t}^{t_1} \alpha_t^s \, ds \right] dt - \int_{t_0}^{t_1} \left[\int_{t}^{t_1} \beta_t^s \, ds \right] dz_t \\
&= Y_{t_0} + \int_{t_0}^{t_1} \left[\int_{t}^{T} \alpha_t^s \, ds \right] dt + \int_{t_0}^{t_1} \left[\int_{t}^{T} \beta_t^s \, ds \right] dz_t \\
&\quad - \int_{t_0}^{t_1} f_{t_0}^s \, ds - \int_{t_0}^{t_1} \left[\int_{t_0}^{s} \alpha_t^s \, dt \right] ds - \int_{t_0}^{t_1} \left[\int_{t_0}^{s} \beta_t^s \, dz_t \right] ds \\
&= Y_{t_0} + \int_{t_0}^{t_1} \left[\int_{t}^{T} \alpha_t^s \, ds \right] dt + \int_{t_0}^{t_1} \left[\int_{t}^{T} \beta_t^s \, ds \right] dz_t - \int_{t_0}^{t_1} f_s^s \, ds \\
&= Y_{t_0} + \int_{t_0}^{t_1} \left[\left(\int_{t}^{T} \alpha_t^s \, ds \right) - f_t^t \right] dt + \int_{t_0}^{t_1} \left[\int_{t}^{T} \beta_t^s \, ds \right] dz_t,
\end{aligned}$$

where the Fubini rule was employed in the second and fourth equality. The result now follows from the final expression. □

3.7 ITÔ'S LEMMA

In our dynamic models of the term structure of interest rates, we will take as given a stochastic process for the dynamics of some basic quantity such as the short-term interest rate. Many other quantities of interest will be functions of that basic variable. To determine the dynamics of these other variables, we shall apply Itô's Lemma, which is basically the chain rule for stochastic processes. We will state the result for a function of a general Itô process, although we will most frequently apply the result for the special case of a function of a diffusion process.

Theorem 3.6 *Let $x = (x_t)_{t \geq 0}$ be a real-valued Itô process with dynamics*

$$dx_t = \mu_t \, dt + \sigma_t \, dz_t,$$

where μ and σ are real-valued processes, and z is a one-dimensional standard Brownian motion. Let $g(x,t)$ be a real-valued function which is two times continuously differentiable in x and continuously differentiable in t. Then the process $y = (y_t)_{t\geq 0}$ defined by

$$y_t = g(x_t, t)$$

is an Itô process with dynamics

$$dy_t = \left(\frac{\partial g}{\partial t}(x_t, t) + \frac{\partial g}{\partial x}(x_t, t)\mu_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(x_t, t)\sigma_t^2 \right) dt + \frac{\partial g}{\partial x}(x_t, t)\sigma_t \, dz_t.$$

The proof is based on a Taylor expansion of $g(x_t, t)$ combined with appropriate limits, but a formal proof is beyond the scope of this book. Once again, we refer to Øksendal (2003, Ch. 4) and similar textbooks. The result can also be written in the following way, which may be easier to remember:

$$dy_t = \frac{\partial g}{\partial t}(x_t, t)\, dt + \frac{\partial g}{\partial x}(x_t, t)\, dx_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(x_t, t)(dx_t)^2. \tag{3.10}$$

Here, in the computation of $(dx_t)^2$, one must apply the rules $(dt)^2 = dt \cdot dz_t = 0$ and $(dz_t)^2 = dt$, so that

$$(dx_t)^2 = (\mu_t \, dt + \sigma_t \, dz_t)^2 = \mu_t^2 (dt)^2 + 2\mu_t \sigma_t \, dt \cdot dz_t + \sigma_t^2 (dz_t)^2 = \sigma_t^2 \, dt.$$

The intuition behind these rules is as follows: When dt is close to zero, $(dt)^2$ is far less than dt and can therefore be ignored. Since $dz_t \sim N(0, dt)$, we get $\mathrm{E}[dt \cdot dz_t] = dt \cdot \mathrm{E}[dz_t] = 0$ and $\mathrm{Var}[dt \cdot dz_t] = (dt)^2 \,\mathrm{Var}[dz_t] = (dt)^3$, which is also very small compared to dt and is therefore ignorable. Finally, we have $\mathrm{E}[(dz_t)^2] = \mathrm{Var}[dz_t] - (\mathrm{E}[dz_t])^2 = dt$, and it can be shown that[3] $\mathrm{Var}[(dz_t)^2] = 2(dt)^2$. For dt close to zero, the variance is therefore much less than the mean, so $(dz_t)^2$ can be approximated by its mean dt.

In standard mathematics, the differential of a function $y = g(x,t)$ where x and t are real variables is defined as $dy = \frac{\partial g}{\partial t}\, dt + \frac{\partial g}{\partial x}\, dx$. When x is an Itô process, (3.10) shows that we have to add a second-order term.

In Section 3.8, we give examples of the application of Itô's Lemma. We will use Itô's Lemma extensively throughout the rest of the book.

3.8 IMPORTANT DIFFUSION PROCESSES

In this section we will discuss particular examples of diffusion processes that are frequently applied in modern financial models, such as those we consider in the following chapters.

[3] This is based on the computation $\mathrm{Var}[(z_{t+\Delta t} - z_t)^2] = \mathrm{E}[(z_{t+\Delta t} - z_t)^4] - \left(\mathrm{E}[(z_{t+\Delta t} - z_t)^2]\right)^2 = 3(\Delta t)^2 - (\Delta t)^2 = 2(\Delta t)^2$ and a passage to the limit.

3.8.1 Geometric Brownian motions

A stochastic process $x = (x_t)_{t \geq 0}$ is said to be a **geometric Brownian motion** if it is a solution to the stochastic differential equation

$$dx_t = \mu x_t \, dt + \sigma x_t \, dz_t, \tag{3.11}$$

where μ and σ are constants. The initial value for the process is assumed to be positive, $x_0 > 0$. A geometric Brownian motion is the particular diffusion process that is obtained from (3.6) by inserting $\mu(x_t, t) = \mu x_t$ and $\sigma(x_t, t) = \sigma x_t$. Paths can be simulated by computing

$$x_{t_i} = x_{t_{i-1}} + \mu x_{t_{i-1}}(t_i - t_{i-1}) + \sigma x_{t_{i-1}} \varepsilon_i \sqrt{t_i - t_{i-1}}.$$

Figure 3.4 shows a single simulated path for $\sigma = 0.2$ and a path for $\sigma = 0.5$. For both paths we have used $\mu = 0.1$ and $x_0 = 100$, and the same sequence of random numbers. For more on the simulation of a geometric Brownian motion, see Section 16.3.

The expression (3.11) can be rewritten as

$$\frac{dx_t}{x_t} = \mu \, dt + \sigma \, dz_t,$$

which is the relative (percentage) change in the value of the process over the next infinitesimally short time interval $[t, t + dt]$. If x_t is the price of a traded asset, then dx_t/x_t is the rate of return on the asset over the next instant. The constant μ is the

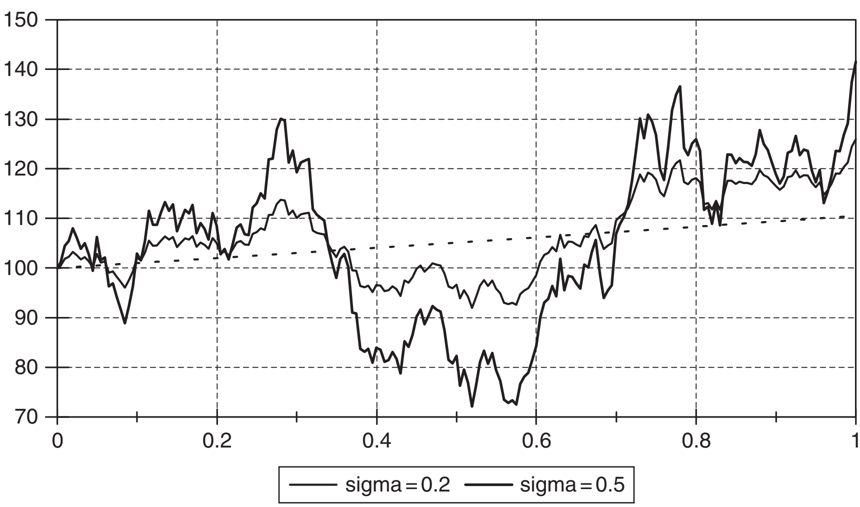

Fig. 3.4: Simulation of a geometric Brownian motion with initial value $x_0 = 100$, relative drift rate $\mu = 0.1$, and a relative volatility of $\sigma = 0.2$ and $\sigma = 0.5$, respectively. The smooth curve shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals of equal length, and the same sequence of random numbers has been used for the two σ-values.

expected rate of return per period, while σ is the standard deviation of the rate of return per period. In this context it is often μ which is called the drift (rather than μx_t) and σ which is called the volatility (rather than σx_t). Strictly speaking, one must distinguish between the relative drift and volatility (μ and σ, respectively) and the absolute drift and volatility (μx_t and σx_t, respectively). An asset with a constant expected rate of return and a constant relative volatility has a price that follows a geometric Brownian motion. For example, such an assumption is used for the stock price in the famous Black–Scholes–Merton model for stock option pricing, see Section 4.8, and a geometric Brownian motion is also used to describe the evolution in the short-term interest rate in some models of the term structure of interest rate, see Section 7.7.

Next, we will find an explicit expression for x_t, that is we will find a solution to the stochastic differential equation (3.11). We can then also determine the distribution of the future value of the process. We apply Itô's Lemma with the function $g(x,t) = \ln x$ and define the process $y_t = g(x_t, t) = \ln x_t$. Since

$$\frac{\partial g}{\partial t}(x_t,t) = 0, \qquad \frac{\partial g}{\partial x}(x_t,t) = \frac{1}{x_t}, \qquad \frac{\partial^2 g}{\partial x^2}(x_t,t) = -\frac{1}{x_t^2},$$

we get from Theorem 3.6 that

$$dy_t = \left(0 + \frac{1}{x_t}\mu x_t - \frac{1}{2}\frac{1}{x_t^2}\sigma^2 x_t^2\right) dt + \frac{1}{x_t}\sigma x_t\, dz_t = \left(\mu - \frac{1}{2}\sigma^2\right) dt + \sigma\, dz_t.$$

Hence, the process $y_t = \ln x_t$ is a generalized Brownian motion. In particular, we have

$$y_{t'} - y_t = \left(\mu - \frac{1}{2}\sigma^2\right)(t'-t) + \sigma(z_{t'} - z_t),$$

which implies that

$$\ln x_{t'} = \ln x_t + \left(\mu - \frac{1}{2}\sigma^2\right)(t'-t) + \sigma(z_{t'} - z_t).$$

Taking exponentials on both sides, we get

$$x_{t'} = x_t \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)(t'-t) + \sigma(z_{t'} - z_t)\right\}. \tag{3.12}$$

This is true for all $t' > t \geq 0$. In particular,

$$x_t = x_0 \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)t + \sigma z_t\right\}.$$

Since exponentials are always positive, we see that x_t can only have positive values, so that the value space of a geometric Brownian motion is $\mathcal{S} = (0,\infty)$.

Suppose now that we stand at time t and have observed the current value x_t of a geometric Brownian motion. Which probability distribution is then appropriate for the uncertain future value, say at time t'? Since $z_{t'} - z_t \sim N(0, t'-t)$, we see from (3.12) that the future value $x_{t'}$ (given x_t) will be lognormally distributed. The probability density function for $x_{t'}$ (given x_t) is

$$f(x) = \frac{1}{x\sqrt{2\pi\sigma^2(t'-t)}} \exp\left\{-\frac{1}{2\sigma^2(t'-t)}\left(\ln\left(\frac{x}{x_t}\right)\right.\right.$$
$$\left.\left. -\left(\mu - \frac{1}{2}\sigma^2\right)(t'-t)\right)^2\right\}, \quad x > 0,$$

and the mean and variance are

$$\mathrm{E}_t[x_{t'}] = x_t e^{\mu(t'-t)},$$
$$\mathrm{Var}_t[x_{t'}] = x_t^2 e^{2\mu(t'-t)}\left[e^{\sigma^2(t'-t)} - 1\right],$$

see Appendix A.

The geometric Brownian motion in (3.11) is time-homogeneous, since neither the drift nor the volatility are time-dependent. We will also make use of the time-inhomogeneous variant, which is characterized by the dynamics

$$dx_t = \mu(t)x_t\, dt + \sigma(t)x_t\, dz_t,$$

where μ and σ are deterministic functions of time. Following the same procedure as for the time-homogeneous geometric Brownian motion, one can show that the inhomogeneous variant satisfies

$$x_{t'} = x_t \exp\left\{\int_t^{t'}\left(\mu(u) - \frac{1}{2}\sigma(u)^2\right) du + \int_t^{t'} \sigma(u)\, dz_u\right\}.$$

According to Theorem 3.3, $\int_t^{t'} \sigma(u)\, dz_u$ is normally distributed with mean zero and variance $\int_t^{t'} \sigma(u)^2\, du$. Therefore, the future value of the time-inhomogeneous geometric Brownian motion is also lognormally distributed. In addition, we have

$$\mathrm{E}_t[x_{t'}] = x_t e^{\int_t^{t'} \mu(u)\, du},$$
$$\mathrm{Var}_t[x_{t'}] = x_t^2 e^{2\int_t^{t'} \mu(u)\, du}\left(e^{\int_t^{t'} \sigma(u)^2\, du} - 1\right).$$

3.8.2 Ornstein–Uhlenbeck processes

Another stochastic process we shall apply in models of the term structure of interest rates is the so-called Ornstein–Uhlenbeck process. A stochastic process $x = (x_t)_{t\geq 0}$ is said to be an **Ornstein–Uhlenbeck process**, if its dynamics is of the form

$$dx_t = [\varphi - \kappa x_t]\, dt + \beta\, dz_t, \tag{3.13}$$

where φ, β, and κ are constants with $\kappa > 0$. Alternatively, this can be written as

$$dx_t = \kappa\,[\theta - x_t]\;dt + \beta\,dz_t,$$

where $\theta = \varphi/\kappa$. An Ornstein–Uhlenbeck process exhibits *mean reversion* in the sense that the drift is positive when $x_t < \theta$ and negative when $x_t > \theta$. The process is therefore always pulled towards a long-term level of θ. However, the random shock to the process through the term $\beta\,dz_t$ may cause the process to move further away from θ. The parameter κ controls the size of the expected adjustment towards the long-term level and is often referred to as the mean reversion parameter or the speed of adjustment.

To determine the distribution of the future value of an Ornstein–Uhlenbeck process we proceed as for the geometric Brownian motion. We will define a new process y_t as some function of x_t such that $y = (y_t)_{t\geq 0}$ is a generalized Brownian motion. It turns out that this is satisfied for $y_t = g(x_t, t)$, where $g(x,t) = e^{\kappa t}x$. From Itô's Lemma we get

$$\begin{aligned}
dy_t &= \left[\frac{\partial g}{\partial t}(x_t,t) + \frac{\partial g}{\partial x}(x_t,t)\,(\varphi - \kappa x_t) + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(x_t,t)\beta^2\right]dt + \frac{\partial g}{\partial x}(x_t,t)\beta\,dz_t \\
&= \left[\kappa e^{\kappa t}x_t + e^{\kappa t}\,(\varphi - \kappa x_t)\right]dt + e^{\kappa t}\beta\,dz_t \\
&= \varphi e^{\kappa t}\,dt + \beta e^{\kappa t}\,dz_t.
\end{aligned}$$

This implies that

$$y_{t'} = y_t + \int_t^{t'} \varphi e^{\kappa u}\,du + \int_t^{t'} \beta e^{\kappa u}\,dz_u.$$

After substitution of the definition of y_t and $y_{t'}$ and a multiplication by $e^{-\kappa t'}$, we arrive at the expression

$$\begin{aligned}
x_{t'} &= e^{-\kappa(t'-t)}x_t + \int_t^{t'} \varphi e^{-\kappa(t'-u)}\,du + \int_t^{t'} \beta e^{-\kappa(t'-u)}\,dz_u \\
&= e^{-\kappa(t'-t)}x_t + \theta\left(1 - e^{-\kappa(t'-t)}\right) + \int_t^{t'} \beta e^{-\kappa(t'-u)}\,dz_u.
\end{aligned}$$

This holds for all $t' > t \geq 0$. In particular, we get that the solution to the stochastic differential equation (3.13) can be written as

$$x_t = e^{-\kappa t}x_0 + \theta\left(1 - e^{-\kappa t}\right) + \int_0^t \beta e^{-\kappa(t-u)}\,dz_u.$$

According to Theorem 3.3, the integral $\int_t^{t'} \beta e^{-\kappa(t'-u)}\,dz_u$ is normally distributed with mean zero and variance $\int_t^{t'} \beta^2 e^{-2\kappa(t'-u)}\,du = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa(t'-t)}\right)$. We can thus conclude that $x_{t'}$ (given x_t) is normally distributed, with mean and variance given by

$$\mathrm{E}_t[x_{t'}] = e^{-\kappa(t'-t)}x_t + \theta\left(1 - e^{-\kappa(t'-t)}\right), \tag{3.14}$$

$$\mathrm{Var}_t[x_{t'}] = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa(t'-t)}\right). \tag{3.15}$$

The value space of an Ornstein–Uhlenbeck process is $\mathbb{R}$. For $t' \to \infty$, the mean approaches θ, and the variance approaches $\beta^2/(2\kappa)$. For $\kappa \to \infty$, the mean approaches θ, and the variance approaches 0. For $\kappa \to 0$, the mean approaches the current value x_t, and the variance approaches $\beta^2(t'-t)$. The distance between the level of the process and the long-term level is expected to be halved over a period of $t'-t = (\ln 2)/\kappa$, since $\mathrm{E}_t[x_{t'}] - \theta = \frac{1}{2}(x_t - \theta)$ implies that $e^{-\kappa(t'-t)} = \frac{1}{2}$ and, hence, $t'-t = (\ln 2)/\kappa$.

The effect of the different parameters can also be evaluated by looking at the paths of the process, which can be simulated by

$$x_{t_i} = x_{t_{i-1}} + \kappa[\theta - x_{t_{i-1}}](t_i - t_{i-1}) + \beta\varepsilon_i\sqrt{t_i - t_{i-1}}.$$

Figure 3.5 shows a single path for different combinations of x_0, κ, θ, and β. In each subfigure one of the parameters is varied and the others fixed. The base values of the parameters are $x_0 = 0.08$, $\theta = 0.08$, $\kappa = \ln 2 \approx 0.69$, and $\beta = 0.03$. All paths are computed using the same sequence of random numbers $\varepsilon_1, \dots, \varepsilon_n$ and are therefore directly comparable. None of the paths shown involve negative values of the process, but other paths will (see Fig. 3.6). As a matter of fact, it can be shown that an Ornstein–Uhlenbeck process with probability 1 will sooner or later become negative. For more on the simulation of Ornstein–Uhlenbeck processes, see Section 16.3.

We will also apply the time-inhomogeneous Ornstein–Uhlenbeck process, where the constants φ and β are replaced by deterministic functions:

$$dx_t = [\varphi(t) - \kappa x_t]\ dt + \beta(t)\,dz_t = \kappa\ [\theta(t) - x_t]\ dt + \beta(t)\,dz_t.$$

Following the same line of analysis as above, it can be shown that the future value $x_{t'}$ given x_t is normally distributed with mean and variance given by

$$\mathrm{E}_t[x_{t'}] = e^{-\kappa(t'-t)}x_t + \int_t^{t'} \varphi(u)e^{-\kappa(t'-u)}\,du,$$

$$\mathrm{Var}_t[x_{t'}] = \int_t^{t'} \beta(u)^2 e^{-2\kappa(t'-u)}\,du.$$

One can also allow κ to depend on time, but we will not make use of that extension.

One of the earliest (but still frequently applied) dynamic models of the term structure of interest rates is based on the assumption that the short-term interest rate follows an Ornstein–Uhlenbeck process, see Section 7.4. In an extension of that model, the short-term interest rate is assumed to follow a time-inhomogeneous Ornstein–Uhlenbeck process, see Section 9.4.

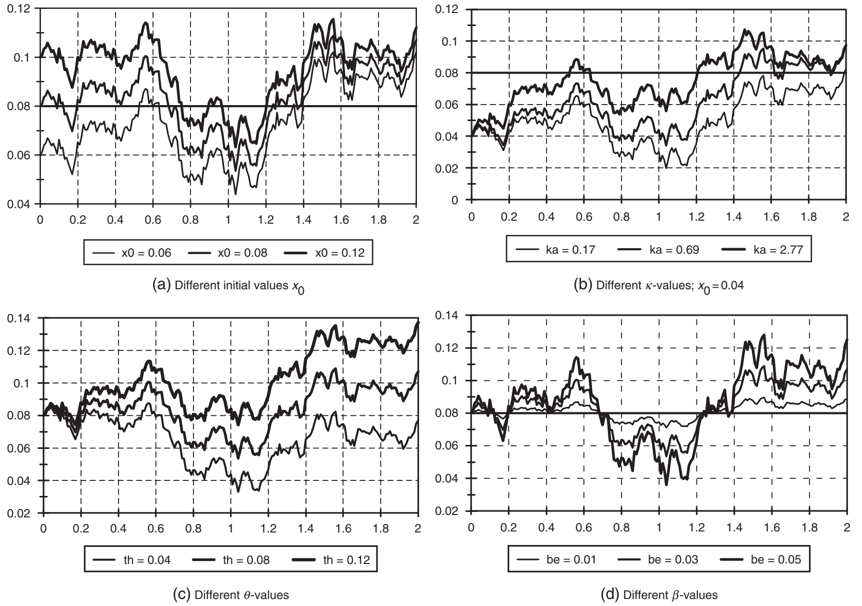

Fig. 3.5: Simulated paths for an Ornstein–Uhlenbeck process. The basic parameter values are $x_0 = \theta = 0.08$, $\kappa = \ln 2 \approx 0.69$, and $\beta = 0.03$.

3.8.3 Square-root processes

Another stochastic process frequently applied in term structure models is the so-called square-root process. A one-dimensional stochastic process $x = (x_t)_{t \geq 0}$ is said to be a **square-root process**, if its dynamics is of the form

$$dx_t = [\varphi - \kappa x_t]\ dt + \beta\sqrt{x_t}\,dz_t = \kappa\ [\theta - x_t]\ dt + \beta\sqrt{x_t}\,dz_t, \tag{3.16}$$

where $\varphi = \kappa\theta$. Here, φ, θ, β, and κ are positive constants. We assume that the initial value of the process x_0 is positive, so that the square root function can be applied. The only difference to the dynamics of an Ornstein–Uhlenbeck process is the term $\sqrt{x_t}$ in the volatility. The variance rate is now $\beta^2 x_t$ which is proportional to the level of the process. A square-root process also exhibits mean reversion.

A square-root process can only take on non-negative values. To see this, note that if the value should become zero, then the drift is positive and the volatility zero, and therefore the value of the process will with certainty become positive immediately after (zero is a so-called reflecting barrier). It can be shown that if $2\varphi \geq \beta^2$, the positive drift at low values of the process is so big relative to the volatility that the process cannot even reach zero, but stays strictly positive.[4] Hence, the value space for a square-root process is either $\mathcal{S} = [0, \infty)$ or $\mathcal{S} = (0, \infty)$.

[4] To show this, the results of Karlin and Taylor (1981, p. 226ff) can be applied.

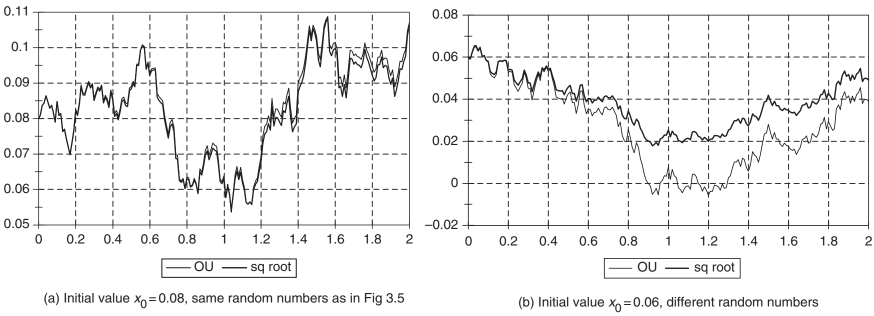

Fig. 3.6: A comparison of simulated paths for an Ornstein–Uhlenbeck process and a square-root process. For both processes, the parameters $\theta = 0.08$ and $\kappa = \ln 2 \approx 0.69$ are used, while β is set to 0.03 for the Ornstein–Uhlenbeck process and to $0.03/\sqrt{0.08} \approx 0.1061$ for the square-root process.

Paths for the square-root process can be simulated by successively calculating

$$x_{t_i} = x_{t_{i-1}} + \kappa[\theta - x_{t_{i-1}}](t_i - t_{i-1}) + \beta\sqrt{x_{t_{i-1}}}\varepsilon_i\sqrt{t_i - t_{i-1}}.$$

Variations in the different parameters will have similar effects as for the Ornstein–Uhlenbeck process, which is illustrated in Fig. 3.5. Instead, let us compare the paths for a square-root process and an Ornstein–Uhlenbeck process using the same drift parameters κ and θ, but where the β-parameter for the Ornstein–Uhlenbeck process is set equal to the β-parameter for the square-root process multiplied by the square root of θ, which ensures that the processes will have the same variance rate at the long-term level. Figure 3.6 compares two pairs of paths of the processes. In part (a), the initial value is set equal to the long-term level, and the two paths continue to be very close to each other. In part (b), the initial value is lower than the long-term level, so that the variance rates of the two processes differ from the beginning. For the given sequence of random numbers, the Ornstein-Uhlenbeck process becomes negative, while the square-root process of course stays positive. In this case there is a clear difference between the paths of the two processes. For more on the simulation of a square-root process, see Section 16.3.

Since a square-root process cannot become negative, the future values of the process cannot be normally distributed. In order to find the actual distribution, let us try the same trick as for the Ornstein–Uhlenbeck process, that is we look at $y_t = e^{\kappa t}x_t$. By Itô's Lemma,

$$\begin{aligned} dy_t &= \kappa e^{\kappa t}x_t\,dt + e^{\kappa t}(\varphi - \kappa x_t)\,dt + e^{\kappa t}\beta\sqrt{x_t}\,dz_t \\ &= \varphi e^{\kappa t}\,dt + \beta e^{\kappa t}\sqrt{x_t}\,dz_t, \end{aligned}$$

so that

$$y_{t'} = y_t + \int_t^{t'} \varphi e^{\kappa u}\,du + \int_t^{t'} \beta e^{\kappa u}\sqrt{x_u}\,dz_u.$$

Computing the ordinary integral and substituting the definition of y, we get

$$x_{t'} = x_t e^{-\kappa(t'-t)} + \theta\left(1 - e^{-\kappa(t'-t)}\right) + \beta \int_t^{t'} e^{-\kappa(t'-u)} \sqrt{x_u}\, dz_u.$$

Since x enters the stochastic integral we cannot immediately determine the distribution of $x_{t'}$ given x_t from this equation. We can, however, use it to obtain the mean and variance of $x_{t'}$. Due to the fact that the stochastic integral has mean zero, see Theorem 3.2, we easily get

$$\mathrm{E}_t[x_{t'}] = e^{-\kappa(t'-t)} x_t + \theta\left(1 - e^{-\kappa(t'-t)}\right) = \theta + (x_t - \theta)\, e^{-\kappa(t'-t)}.$$

To compute the variance we apply the second equation of Theorem 3.2:

$$\begin{aligned}
\mathrm{Var}_t[x_{t'}] &= \mathrm{Var}_t\left[\beta \int_t^{t'} e^{-\kappa(t'-u)} \sqrt{x_u}\, dz_u\right] \\
&= \beta^2 \int_t^{t'} e^{-2\kappa(t'-u)}\, \mathrm{E}_t[x_u]\, du \\
&= \beta^2 \int_t^{t'} e^{-2\kappa(t'-u)} \left(\theta + (x_t - \theta)\, e^{-\kappa(u-t)}\right) du \\
&= \beta^2 \theta \int_t^{t'} e^{-2\kappa(t'-u)}\, du + \beta^2 (x_t - \theta)\, e^{-2\kappa t' + \kappa t} \int_t^{t'} e^{\kappa u}\, du \\
&= \frac{\beta^2 \theta}{2\kappa}\left(1 - e^{-2\kappa(t'-t)}\right) + \frac{\beta^2}{\kappa}(x_t - \theta)\left(e^{-\kappa(t'-t)} - e^{-2\kappa(t'-t)}\right) \\
&= \frac{\beta^2 x_t}{\kappa}\left(e^{-\kappa(t'-t)} - e^{-2\kappa(t'-t)}\right) + \frac{\beta^2 \theta}{2\kappa}\left(1 - e^{-\kappa(t'-t)}\right)^2.
\end{aligned}$$

Note that the mean is identical to the mean for an Ornstein–Uhlenbeck process, whereas the variance is more complicated for the square-root process. For $t' \to \infty$, the mean approaches θ, and the variance approaches $\theta\beta^2/(2\kappa)$. For $\kappa \to \infty$, the mean approaches θ, and the variance approaches 0. For $\kappa \to 0$, the mean approaches the current value x_t, and the variance approaches $\beta^2 x_t (t' - t)$.

It can be shown that, conditional on the value x_t, the value $x_{t'}$ with $t' > t$ is given by the non-central χ^2-distribution. A non-central χ^2-distribution is characterized by a number a of degrees of freedom and a non-centrality parameter b and is denoted by $\chi^2(a, b)$. More precisely, the distribution of $x_{t'}$ given x_t is identical to the distribution of the random variable $Y/c(t' - t)$ where c is the deterministic function

$$c(\tau) = \frac{4\kappa}{\beta^2\left(1 - e^{-\kappa\tau}\right)}$$

and Y is a $\chi^2(a, b(t'-t))$-distributed random variable with

$$a = \frac{4\varphi}{\beta^2}, \qquad b(\tau) = x_t c(\tau) e^{-\kappa\tau}.$$

The density function for a $\chi^2(a,b)$-distributed random variable is

$$\begin{aligned} f_{\chi^2(a,b)}(y) &= \sum_{i=0}^{\infty} \frac{e^{-b/2}(b/2)^i}{i!} f_{\chi^2(a+2i)}(y) \\ &= \sum_{i=0}^{\infty} \frac{e^{-b/2}(b/2)^i}{i!} \frac{(1/2)^{i+a/2}}{\Gamma(i+a/2)} y^{i-1+a/2} e^{-y/2}, \end{aligned}$$

where $f_{\chi^2(a+2i)}$ is the density function for a central χ^2-distribution with $a+2i$ degrees of freedom. Inserting this density in the first sum will give the second sum. Here Γ denotes the so-called gamma function defined as $\Gamma(m) = \int_0^\infty x^{m-1} e^{-x}\, dx$. The probability density function for the value of $x_{t'}$ conditional on x_t is then

$$f(x) = c(t'-t) f_{\chi^2(a,b(t'-t))}\big(c(t'-t)x\big).$$

The mean and variance of a $\chi^2(a,b)$-distributed random variable are $a+b$ and $2(a+2b)$, respectively. This opens another way of deriving the mean and variance of $x_{t'}$ given x_t. We leave it for the reader to verify that this procedure will yield the same results as given above.

A frequently applied dynamic model of the term structure of interest rates is based on the assumption that the short-term interest rate follows a square-root process, see Section 7.5. Since interest rates are positive and empirically seem to have a variance rate which is positively correlated to the interest rate level, the square-root process gives a more realistic description of interest rates than the Ornstein–Uhlenbeck process. On the other hand, models based on square-root processes are more complicated to analyse than models based on Ornstein–Uhlenbeck processes.

3.9 MULTI-DIMENSIONAL PROCESSES

So far we have only considered one-dimensional processes, that is processes with a value space which is $\mathbb{R}$ or a subset of $\mathbb{R}$. In many cases we want to keep track of several processes, for example price processes for different assets, and we will often be interested in covariances and correlations between different processes.

In a continuous-time model where the exogenous shock process $z = (z_t)_{t\in[0,T]}$ is one-dimensional, the instantaneous increments of any two processes will be perfectly correlated. For example, if we consider the two Itô processes x and y defined by

$$dx_t = \mu_{xt}\, dt + \sigma_{xt}\, dz_t, \qquad dy_t = \mu_{yt}\, dt + \sigma_{yt}\, dz_t,$$

then $\text{Cov}_t[dx_t, dy_t] = \sigma_{xt}\sigma_{yt}\,dt$ so that the instantaneous correlation becomes

$$\text{Corr}_t[dx_t, dy_t] = \frac{\text{Cov}_t[dx_t, dy_t]}{\sqrt{\text{Var}_t[dx_t]\,\text{Var}_t[dy_t]}} = \frac{\sigma_{xt}\sigma_{yt}\,dt}{\sqrt{\sigma_{xt}^2\,dt\;\sigma_{yt}^2\,dt}} = 1.$$

Increments over any non-infinitesimal time interval are generally not perfectly correlated, that is for any $h > 0$ a correlation like $\text{Corr}_t[x_{t+h} - x_t, y_{t+h} - y_t]$ is typically different from one but close to one for small h.

To obtain non-perfectly correlated changes over the shortest time period considered by the model we need an exogenous shock of a dimension higher than one, that is a shock **vector**. One can without loss of generality assume that the different components of this shock vector are mutually independent and generate non-perfect correlations between the relevant processes by varying the sensitivities of those processes towards the different exogenous shocks. We will first consider the case of two processes and later generalize.

3.9.1 Two-dimensional processes

In the example above, we can avoid the perfect correlation by introducing a second standard Brownian motion so that

$$dx_t = \mu_{xt}\,dt + \sigma_{x1t}\,dz_{1t} + \sigma_{x2t}\,dz_{2t}, \qquad dy_t = \mu_{yt}\,dt + \sigma_{y1t}\,dz_{1t} + \sigma_{y2t}\,dz_{2t},$$

where $z_1 = (z_{1t})$ and $z_2 = (z_{2t})$ are independent standard Brownian motions. This generates an instantaneous covariance of $\text{Cov}_t[dx_t, dy_t] = \left(\sigma_{x1t}\sigma_{y1t} + \sigma_{x2t}\sigma_{y2t}\right) dt$, instantaneous variances of $\text{Var}_t[dx_t] = (\sigma_{x1t}^2 + \sigma_{x2t}^2)\,dt$ and $\text{Var}_t[dy_t] = \left(\sigma_{y1t}^2 + \sigma_{y2t}^2\right) dt$, and thus an instantaneous correlation of

$$\text{Corr}_t[dx_t, dy_t] = \frac{\sigma_{x1t}\sigma_{y1t} + \sigma_{x2t}\sigma_{y2t}}{\sqrt{\left(\sigma_{x1t}^2 + \sigma_{x2t}^2\right)\left(\sigma_{y1t}^2 + \sigma_{y2t}^2\right)}},$$

which can be anywhere in the interval $[-1, +1]$.

The shock coefficients σ_{x1t}, σ_{x2t}, σ_{y1t}, and σ_{y2t} are determining the two instantaneous variances and the instantaneous correlation. But many combinations of the four shock coefficients will give rise to the same variances and correlation. We have one degree of freedom in fixing the shock coefficients. For example, we can put $\sigma_{x2t} \equiv 0$, which has the nice implication that it will simplify various expressions and interpretations. If we thus write the dynamics of x and y as

$$dx_t = \mu_{xt}\,dt + \sigma_{xt}\,dz_{1t}, \qquad dy_t = \mu_{yt}\,dt + \sigma_{yt}\left[\rho_t\,dz_{1t} + \sqrt{1-\rho_t^2}\,dz_{2t}\right],$$

σ_{xt}^2 and σ_{yt}^2 are the variance rates of x_t and y_t, respectively, while the covariance is $\text{Cov}_t[dx_t, dy_t] = \rho_t\sigma_{xt}\sigma_{yt}$. If σ_{xt} and σ_{yt} are both positive, then ρ_t will be the instantaneous correlation between the two processes x and y.

In many continuous-time models, one stochastic process is defined in terms of a function of two other, not necessarily perfectly correlated, stochastic processes. For that purpose we need the following two-dimensional version of Itô's Lemma.

Theorem 3.7 *Suppose $x = (x_t)$ and $y = (y_t)$ are two stochastic processes with dynamics*

$$\begin{aligned} dx_t &= \mu_{xt}\, dt + \sigma_{x1t}\, dz_{1t} + \sigma_{x2t}\, dz_{2t}, \\ dy_t &= \mu_{yt}\, dt + \sigma_{y1t}\, dz_{1t} + \sigma_{y2t}\, dz_{2t}, \end{aligned} \tag{3.17}$$

where $z_1 = (z_{1t})$ and $z_2 = (z_{2t})$ are independent standard Brownian motions. Let $g(x, y, t)$ be a real-valued function for which all the derivatives $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x}$, $\frac{\partial g}{\partial y}$, $\frac{\partial^2 g}{\partial x^2}$, $\frac{\partial^2 g}{\partial y^2}$, and $\frac{\partial^2 g}{\partial x \partial y}$ exist and are continuous. Then the process $W = (W_t)$ defined by $W_t = g(x_t, y_t, t)$ is an Itô process with

$$\begin{aligned} dW_t = &\left(\frac{\partial g}{\partial t} + \frac{\partial g}{\partial x}\mu_{xt} + \frac{\partial g}{\partial y}\mu_{yt} + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\left(\sigma_{x1t}^2 + \sigma_{x2t}^2\right) + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}\left(\sigma_{y1t}^2 + \sigma_{y2t}^2\right) \right. \\ &\left. + \frac{\partial^2 g}{\partial x \partial y}\left(\sigma_{x1t}\sigma_{y1t} + \sigma_{x2t}\sigma_{y2t}\right) \right) dt \\ &+ \left(\frac{\partial g}{\partial x}\sigma_{x1t} + \frac{\partial g}{\partial y}\sigma_{y1t} \right) dz_{1t} + \left(\frac{\partial g}{\partial x}\sigma_{x2t} + \frac{\partial g}{\partial y}\sigma_{y2t} \right) dz_{2t}, \end{aligned}$$

where the dependence of all the partial derivatives on (x_t, y_t, t) has been notationally suppressed.

Alternatively, the result can be written more compactly as

$$\begin{aligned} dW_t = &\frac{\partial g}{\partial t}\, dt + \frac{\partial g}{\partial x}\, dx_t + \frac{\partial g}{\partial y}\, dy_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}\,(dx_t)^2 + \frac{1}{2}\frac{\partial^2 g}{\partial y^2}\,(dy_t)^2 \\ &+ \frac{\partial^2 g}{\partial x \partial y}\,(dx_t)(dy_t), \end{aligned}$$

where it is understood that $(dt)^2 = dt \cdot dz_{1t} = dt \cdot dz_{2t} = dz_{1t} \cdot dz_{2t} = 0$.

Example 3.1 Suppose that the dynamics of x and y are given by (3.17) and $W_t = x_t y_t$. In order to find the dynamics of W, we apply the above version of Itô's Lemma with the function $g(x, y) = xy$. The relevant partial derivatives are

$$\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial x} = y, \quad \frac{\partial g}{\partial y} = x, \quad \frac{\partial^2 g}{\partial x^2} = 0, \quad \frac{\partial^2 g}{\partial y^2} = 0, \quad \frac{\partial^2 g}{\partial x \partial y} = 1.$$

Hence,

$$dW_t = y_t\, dx_t + x_t\, dy_t + (dx_t)(dy_t).$$

In particular, if the dynamics of x and y are written in the form

$$
\begin{aligned}
dx_t &= x_t \left[m_{xt}\, dt + v_{x1t}\, dz_{1t} + v_{x2t}\, dz_{2t}\right], \\
dy_t &= y_t \left[m_{yt}\, dt + v_{y1t}\, dz_{1t} + v_{y2t}\, dz_{2t}\right],
\end{aligned} \tag{3.18}
$$

we get

$$
\begin{aligned}
dW_t = W_t \Big[&\left(m_{xt} + m_{yt} + v_{x1t}v_{y1t} + v_{x2t}v_{y2t}\right)\, dt + \left(v_{x1t} + v_{y1t}\right)\, dz_{1t} \\
&+ \left(v_{x2t} + v_{y2t}\right)\, dz_{2t}\Big].
\end{aligned}
$$

For the special case, where both x and y are geometric Brownian motion so that m_x, m_y, v_{x1}, v_{x2}, v_{y1}, and v_{y2} are all constants, it follows that $W_t = x_t y_t$ is also a geometric Brownian motion.

Example 3.2 Define $W_t = x_t/y_t$. In this case we need to apply Itô's Lemma with the function $g(x, y) = x/y$ which has derivatives

$$
\frac{\partial g}{\partial t} = 0, \quad \frac{\partial g}{\partial x} = \frac{1}{y}, \quad \frac{\partial g}{\partial y} = -\frac{x}{y^2}, \quad \frac{\partial^2 g}{\partial x^2} = 0, \quad \frac{\partial^2 g}{\partial y^2} = 2\frac{x}{y^3}, \quad \frac{\partial^2 g}{\partial x \partial y} = -\frac{1}{y^2}.
$$

Then

$$
\begin{aligned}
dW_t &= \frac{1}{y_t}\, dx_t - \frac{x_t}{y_t^2}\, dy_t + \frac{x_t}{y_t^3}(dy_t)^2 - \frac{1}{y_t^2}(dx_t)(dy_t) \\
&= W_t \left[\frac{dx_t}{x_t} - \frac{dy_t}{y_t} + \left(\frac{dy_t}{y_t}\right)^2 - \frac{dx_t}{x_t}\frac{dy_t}{y_t} \right].
\end{aligned}
$$

In particular, if the dynamics of x and y are given by (3.18), the dynamics of $W_t = x_t/y_t$ becomes

$$
\begin{aligned}
dW_t = W_t \Big[&\Big(m_{xt} - m_{yt} + (v_{y1t}^2 + v_{y2t}^2) - (v_{x1t}v_{y1t} + v_{x2t}v_{y2t})\Big)\, dt \\
&+ \left(v_{x1t} - v_{y1t}\right)\, dz_{1t} + \left(v_{x2t} - v_{y2t}\right)\, dz_{2t}\Big].
\end{aligned}
$$

Note that for the special case, where both x and y are geometric Brownian motions, $W = x/y$ is also a geometric Brownian motion.

We can apply the two-dimensional version of Itô's Lemma to prove the following useful result relating expected discounted values and the drift rate.

Theorem 3.8 *Under suitable regularity conditions, the relative drift rate of an Itô process $x = (x_t)$ is given by the process $m = (m_t)$ if and only if $x_t = \mathrm{E}_t[x_T \exp\{-\int_t^T m_s\, ds\}]$.*

Proof: Suppose first that the relative drift rate is given by m so that $dx_t = x_t[m_t\, dt + v_t\, dz_t]$. Let us use Itô's Lemma to identify the dynamics of the process $W_t = x_t \exp\{-\int_0^t m_s\, ds\}$ or $W_t = x_t y_t$, where $y_t = \exp\{-\int_0^t m_s\, ds\}$. Note that $dy_t = -y_t m_t\, dt$ so that y is a locally deterministic stochastic process. From Example 3.1, the dynamics of W becomes

$$dW_t = W_t\left[(m_t - m_t + 0)\ dt + v_t\ dz_t\right] = W_t v_t\ dz_t.$$

Since W has zero drift, it is a martingale. It follows that $W_t = \mathrm{E}_t[W_T]$, i.e., $x_t \exp\{-\int_0^t m_s\,ds\} = \mathrm{E}_t[x_T \exp\{-\int_0^T m_s\,ds\}]$ and hence $x_t = \mathrm{E}_t[x_T \exp\{-\int_t^T m_s\,ds\}]$.

If, on the other hand, $x_t = \mathrm{E}_t[x_T \exp\{-\int_t^T m_s\,ds\}]$ for all t, then the absolute drift of x follows from this computation:

$$\begin{aligned}
\frac{1}{\Delta t}\,\mathrm{E}_t[x_{t+\Delta t} - x_t] &= \frac{1}{\Delta t}\,\mathrm{E}_t\left[\left(\mathrm{E}_{t+\Delta t}\left[x_T e^{-\int_{t+\Delta t}^T m_s\,ds}\right]\right) - \left(\mathrm{E}_t\left[x_T e^{-\int_t^T m_s\,ds}\right]\right)\right]\\
&= \frac{1}{\Delta t}\,\mathrm{E}_t\left[x_T e^{-\int_{t+\Delta t}^T m_s\,ds} - x_T e^{-\int_t^T m_s\,ds}\right]\\
&= \mathrm{E}_t\left[x_T e^{-\int_t^T m_s\,ds}\,\frac{e^{\int_t^{t+\Delta t} m_s\,ds} - 1}{\Delta t}\right]\\
&\to m_t\,\mathrm{E}_t\left[x_T e^{-\int_t^T m_s\,ds}\right] = m_t x_t,
\end{aligned}$$

so that the relative drift rate equals m_t. □

3.9.2 K-dimensional processes

Simultaneously modelling the dynamics of a lot of economic quantities requires the use of a lot of shocks to those quantities. For that purpose we will represent shocks to the economy by a vector standard Brownian motion. We define this below and state Itô's Lemma for processes of a general dimension.

A K-**dimensional standard Brownian motion** $\boldsymbol{z} = (z_1, \dots, z_K)^\top$ is a stochastic process for which the individual components z_i are mutually independent one-dimensional standard Brownian motions. If we let $\mathbf{0} = (0, \dots, 0)^\top$ denote the zero vector in $\mathbb{R}^K$ and let $\underline{\underline{I}}$ denote the identity matrix of dimension $K \times K$ (the matrix with ones in the diagonal and zeros in all other entries), then we can write the defining properties of a K-dimensional Brownian motion $\boldsymbol{z}$ as follows:

(i) $\boldsymbol{z}_0 = \mathbf{0}$,
(ii) for all $t, t' \geq 0$ with $t < t'$: $\boldsymbol{z}_{t'} - \boldsymbol{z}_t \sim \boldsymbol{N}(\mathbf{0}, (t' - t)\underline{\underline{I}})$ [normally distributed increments],
(iii) for all $0 \leq t_0 < t_1 < \dots < t_n$, the random variables $\boldsymbol{z}_{t_1} - \boldsymbol{z}_{t_0}, \dots, \boldsymbol{z}_{t_n} - \boldsymbol{z}_{t_{n-1}}$ are mutually independent [independent increments],
(iv) $\boldsymbol{z}$ has continuous sample paths in $\mathbb{R}^K$.

Here, $\boldsymbol{N}(\boldsymbol{a}, \underline{\underline{b}})$ denotes a K-dimensional normal distribution with mean vector $\boldsymbol{a}$ and variance-covariance matrix $\underline{\underline{b}}$.

A K-**dimensional diffusion process** $\boldsymbol{x} = (x_1, \dots, x_K)^\top$ is a process with increments of the form

$$d\boldsymbol{x}_t = \boldsymbol{\mu}(\boldsymbol{x}_t, t)\,dt + \underline{\underline{\sigma}}(\boldsymbol{x}_t, t)\,d\boldsymbol{z}_t,$$

where $\boldsymbol{\mu}$ is a function from $\mathbb{R}^K \times \mathbb{R}_+$ into $\mathbb{R}^K$, and $\underline{\underline{\sigma}}$ is a function from $\mathbb{R}^K \times \mathbb{R}_+$ into the space of $K \times K$-matrices. As before, $\boldsymbol{z}$ is a K-dimensional standard Brownian motion. The evolution of the multi-dimensional diffusion can also be written componentwise as

$$\begin{aligned} dx_{it} &= \mu_i(\boldsymbol{x}_t, t)\, dt + \boldsymbol{\sigma}_i(x_t, t)^\top\, d\boldsymbol{z}_t \\ &= \mu_i(\boldsymbol{x}_t, t)\, dt + \sum_{k=1}^K \sigma_{ik}(\boldsymbol{x}_t, t)\, dz_{kt}, \quad i = 1, \dots, K, \end{aligned}$$

where $\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)^\top$ is the i'th row of the matrix $\underline{\underline{\sigma}}(\boldsymbol{x}_t, t)$, and $\sigma_{ik}(\boldsymbol{x}_t, t)$ is the (i, k)'th entry (that is the entry in row i, column k). Since $dz_{1t}, \dots, dz_{Kt}$ are mutually independent and all $N(0, dt)$ distributed, the expected change in the i'th component process over an infinitesimal period is

$$\mathrm{E}_t[dx_{it}] = \mu_i(\boldsymbol{x}_t, t)\, dt, \quad i = 1, \dots, K,$$

so that μ_i can be interpreted as the drift of the i'th component. Furthermore, the covariance between changes in the i'th and the j'th component processes over an infinitesimal period becomes

$$\begin{aligned} \mathrm{Cov}_t[dx_{it}, dx_{jt}] &= \mathrm{Cov}_t\left[\sum_{k=1}^K \sigma_{ik}(\boldsymbol{x}_t, t)\, dz_{kt}, \sum_{l=1}^K \sigma_{jl}(\boldsymbol{x}_t, t)\, dz_{lt}\right] \\ &= \sum_{k=1}^K \sum_{l=1}^K \sigma_{ik}(\boldsymbol{x}_t, t)\sigma_{jl}(\boldsymbol{x}_t, t)\, \mathrm{Cov}_t[dz_{kt}, dz_{lt}] \\ &= \sum_{k=1}^K \sigma_{ik}(\boldsymbol{x}_t, t)\sigma_{jk}(\boldsymbol{x}_t, t)\, dt \\ &= \boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)^\top \boldsymbol{\sigma}_j(\boldsymbol{x}_t, t)\, dt, \quad i, j = 1, \dots, K, \end{aligned}$$

where we have applied the usual rules for covariances and the independence of the components of $\boldsymbol{z}$. In particular, the variance of the change in the i'th component process of an infinitesimal period is given by

$$\mathrm{Var}_t[dx_{it}] = \mathrm{Cov}_t[dx_{it}, dx_{it}] = \sum_{k=1}^K \sigma_{ik}(\boldsymbol{x}_t, t)^2\, dt = \|\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)\|^2\, dt, \; i = 1, \dots, K.$$

The volatility of the i'th component is given by $\|\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)\|$. The variance-covariance matrix of changes of $\boldsymbol{x}_t$ over the next instant is $\underline{\underline{\Sigma}}(\boldsymbol{x}_t, t)\, dt = \underline{\underline{\sigma}}(\boldsymbol{x}_t, t)\underline{\underline{\sigma}}(\boldsymbol{x}_t, t)^\top\, dt$. The correlation between instantaneous increments in two component processes is

$$\mathrm{Corr}_t[dx_{it}, dx_{jt}] = \frac{\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)^\top \boldsymbol{\sigma}_j(\boldsymbol{x}_t, t)\, dt}{\sqrt{\|\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)\|^2\, dt\, \|\boldsymbol{\sigma}_j(\boldsymbol{x}_t, t)\|^2\, dt}} = \frac{\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)^\top \boldsymbol{\sigma}_j(\boldsymbol{x}_t, t)}{\|\boldsymbol{\sigma}_i(\boldsymbol{x}_t, t)\|\, \|\boldsymbol{\sigma}_j(\boldsymbol{x}_t, t)\|},$$

which can be any number in $[-1, 1]$ depending on the elements of $\boldsymbol{\sigma}_i$ and $\boldsymbol{\sigma}_j$.

Similarly, we can define a K-**dimensional Itô process** $\boldsymbol{x} = (x_1, \dots, x_K)^\top$ to be a process with increments of the form

$$d\boldsymbol{x}_t = \boldsymbol{\mu}_t \, dt + \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t,$$

where $\boldsymbol{\mu} = (\boldsymbol{\mu}_t)$ is a K-dimensional stochastic process and $\underline{\underline{\sigma}} = (\underline{\underline{\sigma}}_t)$ is a stochastic process with values in the space of $K \times K$-matrices.

Next, we state a multi-dimensional version of Itô's Lemma, where a one-dimensional process is defined as a function of time and a multi-dimensional process.

Theorem 3.9 *Let $\boldsymbol{x} = (\boldsymbol{x}_t)_{t \geq 0}$ be an Itô process in $\mathbb{R}^K$ with dynamics $d\boldsymbol{x}_t = \boldsymbol{\mu}_t \, dt + \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t$ or, equivalently,*

$$dx_{it} = \mu_{it} \, dt + \boldsymbol{\sigma}_{it}^\top \, d\boldsymbol{z}_t = \mu_{it} \, dt + \sum_{k=1}^{K} \sigma_{ikt} \, dz_{kt}, \quad i = 1, \dots, K,$$

where $z_1, \dots, z_K$ are independent standard Brownian motions, and μ_i and σ_{ik} are well-behaved stochastic processes.

Let $g(\boldsymbol{x}, t)$ be a real-valued function for which all the derivatives $\frac{\partial g}{\partial t}$, $\frac{\partial g}{\partial x_i}$, and $\frac{\partial^2 g}{\partial x_i \partial x_j}$ exist and are continuous. Then the process $y = (y_t)_{t \geq 0}$ defined by $y_t = g(\boldsymbol{x}_t, t)$ is also an Itô process with dynamics

$$dy_t = \left(\frac{\partial g}{\partial t}(\boldsymbol{x}_t, t) + \sum_{i=1}^{K} \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t, t) \mu_{it} + \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\partial^2 g}{\partial x_i \partial x_j}(\boldsymbol{x}_t, t) \gamma_{ijt} \right) dt$$
$$+ \sum_{i=1}^{K} \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t, t) \sigma_{i1t} \, dz_{1t} + \dots + \sum_{i=1}^{K} \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t, t) \sigma_{iKt} \, dz_{Kt},$$

where $\gamma_{ij} = \sigma_{i1}\sigma_{j1} + \dots + \sigma_{iK}\sigma_{jK}$ is the covariance between the processes x_i and x_j.

The result can also be written as

$$dy_t = \frac{\partial g}{\partial t}(\boldsymbol{x}_t, t) \, dt + \sum_{i=1}^{K} \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t, t) \, dx_{it} + \frac{1}{2} \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\partial^2 g}{\partial x_i \partial x_j}(\boldsymbol{x}_t, t)(dx_{it})(dx_{jt}),$$

where in the computation of $(dx_{it})(dx_{jt})$ one must use the rules $(dt)^2 = dt \cdot dz_{it} = 0$ for all i, $dz_{it} \cdot dz_{jt} = 0$ for $i \neq j$, and $(dz_{it})^2 = dt$ for all i. Alternatively, the result can be expressed using vector and matrix notation:

$$dy_t = \left(\frac{\partial g}{\partial t}(\boldsymbol{x}_t, t) + \left(\frac{\partial g}{\partial \boldsymbol{x}}(\boldsymbol{x}_t, t) \right)^\top \boldsymbol{\mu}_t + \frac{1}{2} \operatorname{tr} \left(\underline{\underline{\sigma}}_t^\top \left[\frac{\partial^2 g}{\partial \boldsymbol{x}^2}(\boldsymbol{x}_t, t) \right] \underline{\underline{\sigma}}_t \right) \right) dt$$
$$+ \left(\frac{\partial g}{\partial \boldsymbol{x}}(\boldsymbol{x}_t, t) \right)^\top \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t,$$

where

$$\frac{\partial g}{\partial \boldsymbol{x}}(\boldsymbol{x}_t, t) = \begin{pmatrix} \frac{\partial g}{\partial x_1}(\boldsymbol{x}_t, t) \\ \cdots \\ \frac{\partial g}{\partial x_k}(\boldsymbol{x}_t, t) \end{pmatrix},$$

$$\frac{\partial^2 g}{\partial \boldsymbol{x}^2}(\boldsymbol{x}_t, t) = \begin{pmatrix} \frac{\partial^2 g}{\partial x_1^2}(\boldsymbol{x}_t, t) & \frac{\partial^2 g}{\partial x_1 \partial x_2}(\boldsymbol{x}_t, t) & \dots & \frac{\partial^2 g}{\partial x_1 \partial x_K}(\boldsymbol{x}_t, t) \\ \frac{\partial^2 g}{\partial x_2 \partial x_1}(\boldsymbol{x}_t, t) & \frac{\partial^2 g}{\partial x_2^2}(\boldsymbol{x}_t, t) & \dots & \frac{\partial^2 g}{\partial x_2 \partial x_K}(\boldsymbol{x}_t, t) \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 g}{\partial x_K \partial x_1}(\boldsymbol{x}_t, t) & \frac{\partial^2 g}{\partial x_K \partial x_2}(\boldsymbol{x}_t, t) & \dots & \frac{\partial^2 g}{\partial x_K^2}(\boldsymbol{x}_t, t) \end{pmatrix},$$

and tr denotes the trace of a quadratic matrix, that is the sum of the diagonal elements. For example, $\operatorname{tr}(\underline{\underline{A}}) = \sum_{i=1}^{K} A_{ii}$.

The probabilistic properties of a K-dimensional diffusion process is completely specified by the drift function $\boldsymbol{\mu}$ and the variance-covariance function $\underline{\underline{\Sigma}}$. The values of the variance-covariance function are symmetric and positive-definite matrices. Above we had $\underline{\underline{\Sigma}} = \underline{\underline{\sigma}}\,\underline{\underline{\sigma}}^{\top}$ for a general $(K \times K)$-matrix $\underline{\underline{\sigma}}$. But from linear algebra it is well-known that a symmetric and positive-definite matrix can be written as $\underline{\underline{\hat{\sigma}}}\,\underline{\underline{\hat{\sigma}}}^{\top}$ for a lower-triangular matrix $\underline{\underline{\hat{\sigma}}}$, that is a matrix with $\hat{\sigma}_{ik} = 0$ for $k > i$. This is the so-called Cholesky decomposition. Hence, we may write the dynamics as

$$\begin{aligned} dx_{1t} &= \mu_1(\boldsymbol{x}_t, t)\, dt + \hat{\sigma}_{11}(\boldsymbol{x}_t, t)\, dz_{1t} \\ dx_{2t} &= \mu_2(\boldsymbol{x}_t, t)\, dt + \hat{\sigma}_{21}(\boldsymbol{x}_t, t)\, dz_{1t} + \hat{\sigma}_{22}(\boldsymbol{x}_t, t)\, dz_{2t} \\ &\vdots \\ dx_{Kt} &= \mu_K(\boldsymbol{x}_t, t)\, dt + \hat{\sigma}_{K1}(\boldsymbol{x}_t, t)\, dz_{1t} + \hat{\sigma}_{K2}(\boldsymbol{x}_t, t)\, dz_{2t} + \dots + \hat{\sigma}_{KK}(\boldsymbol{x}_t, t)\, dz_{Kt} \end{aligned} \tag{3.19}$$

We can think of building up the model by starting with x_1. The shocks to x_1 are represented by the standard Brownian motion z_1 and its coefficient $\hat{\sigma}_{11}$ is the volatility of x_1. Then we extend the model to include x_2. Unless the infinitesimal changes to x_1 and x_2 are always perfectly correlated we need to introduce another standard Brownian motion, z_2. The coefficient $\hat{\sigma}_{21}$ is fixed to match the covariance between changes to x_1 and x_2 and then $\hat{\sigma}_{22}$ can be chosen so that $\sqrt{\hat{\sigma}_{21}^2 + \hat{\sigma}_{22}^2}$ equals the volatility of x_2. The model may be extended to include additional processes in the same manner.

Some authors prefer to write the dynamics in an alternative way with a single standard Brownian motion $\hat{z}_i$ for each component x_i such as

$$\begin{aligned}
dx_{1t} &= \mu_1(\boldsymbol{x}_t, t)\, dt + V_1(\boldsymbol{x}_t, t)\, d\hat{z}_{1t} \\
dx_{2t} &= \mu_2(\boldsymbol{x}_t, t)\, dt + V_2(\boldsymbol{x}_t, t)\, d\hat{z}_{2t} \\
&\vdots \\
dx_{Kt} &= \mu_K(\boldsymbol{x}_t, t)\, dt + V_K(\boldsymbol{x}_t, t)\, d\hat{z}_{Kt}
\end{aligned} \tag{3.20}$$

Clearly, the coefficient $V_i(\boldsymbol{x}_t, t)$ is then the volatility of x_i. To capture an instantaneous non-zero correlation between the different components the standard Brownian motions $\hat{z}_1, \dots, \hat{z}_K$ have to be mutually correlated. Let ρ_{ij} be the correlation between $\hat{z}_i$ and $\hat{z}_j$. If (3.20) and (3.19) are meant to represent the same dynamics, we must have

$$V_i = \sqrt{\hat{\sigma}_{i1}^2 + \dots + \hat{\sigma}_{ii}^2}, \quad i = 1, \dots, K,$$

$$\rho_{ii} = 1; \quad \rho_{ij} = \frac{\sum_{k=1}^{i} \hat{\sigma}_{ik}\hat{\sigma}_{jk}}{V_i V_j}, \; \rho_{ji} = \rho_{ij}, \quad i < j.$$

3.10 CHANGE OF PROBABILITY MEASURE

When we represent the evolution of a given economic variable by a stochastic process and discuss the distributional properties of this process, we have implicitly fixed a probability measure $\mathbb{P}$. For example, when we use the square-root process $x = (x_t)$ in (3.16) for the dynamics of a particular interest rate, we have taken as given a probability measure $\mathbb{P}$ under which the stochastic process $z = (z_t)$ is a standard Brownian motion. Since the process x is presumably meant to represent the uncertain dynamics of the interest rate in the world we live in, we refer to the measure $\mathbb{P}$ as the real-world probability measure. Of course, it is the real-world dynamics and distributional properties of economic variables that we are ultimately interested in. Nevertheless, it turns out that in order to compute and understand prices and rates it is often convenient to look at the dynamics and distributional properties of these variables assuming a world that is different from the world we live in. The prime example is a hypothetical world in which investors are assumed to be risk-neutral instead of risk-averse. Loosely speaking, a different world is represented mathematically by a different probability measure. Hence, we need to be able to analyse stochastic variables and processes under different probability measures. In this section we will briefly discuss how we can change the probability measure.

Consider first a state space with finitely many elements, $\Omega = \{\omega_1, \dots, \omega_n\}$. As before, the set of events, that is subsets of Ω, that can be assigned a probability, is denoted by $\mathcal{F}$. Let us assume that the single-element sets $\{\omega_i\}$, $i = 1, \dots, n$, belong to $\mathcal{F}$. In this case we can represent a probability measure $\mathbb{P}$ by a vector $(p_1, \dots, p_n)$ of probabilities assigned to each of the individual elements:

$$p_i = \mathbb{P}(\{\omega_i\}), \quad i = 1, \dots, n.$$

Of course, we must have that $p_i \in [0,1]$ and that $\sum_{i=1}^n p_i = 1$. The probability assigned to any other event can be computed from these basic probabilities. For example, the probability of the event $\{\omega_2, \omega_4\}$ is given by

$$\mathbb{P}(\{\omega_2, \omega_4\}) = \mathbb{P}(\{\omega_2\} \cup \{\omega_4\}) = \mathbb{P}(\{\omega_2\}) + \mathbb{P}(\{\omega_4\}) = p_2 + p_4.$$

Another probability measure $\mathbb{Q}$ on $\mathcal{F}$ is similarly given by a vector $(q_1, \dots, q_n)$ with $q_i \in [0,1]$ and $\sum_{i=1}^n q_i = 1$. We are only interested in equivalent probability measures. In this setting, the two measures $\mathbb{P}$ and $\mathbb{Q}$ will be equivalent whenever $p_i > 0 \Leftrightarrow q_i > 0$ for all $i = 1, \dots, n$. With a finite state space there is no point in including states that occur with zero probability so we can assume that all p_i, and therefore all q_i, are strictly positive.

We can represent the change of probability measure from $\mathbb{P}$ to $\mathbb{Q}$ by the vector $\xi = (\xi_1, \dots, \xi_n)$, where

$$\xi_i = \frac{q_i}{p_i}, \quad i = 1, \dots, n.$$

We can think of ξ as a random variable that will take on the value ξ_i if the state ω_i is realized. Sometimes ξ is called the Radon–Nikodym derivative of $\mathbb{Q}$ with respect to $\mathbb{P}$ and is denoted by $d\mathbb{Q}/d\mathbb{P}$. Note that $\xi_i > 0$ for all i and that the $\mathbb{P}$-expectation of $\xi = d\mathbb{Q}/d\mathbb{P}$ is

$$\mathrm{E}^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = \mathrm{E}^{\mathbb{P}}[\xi] = \sum_{i=1}^n p_i \xi_i = \sum_{i=1}^n p_i \frac{q_i}{p_i} = \sum_{i=1}^n q_i = 1.$$

Consider a random variable x that takes on the value x_i if state i is realized. The expected value of x under the measure $\mathbb{Q}$ is given by

$$\mathrm{E}^{\mathbb{Q}}[x] = \sum_{i=1}^n q_i x_i = \sum_{i=1}^n p_i \frac{q_i}{p_i} x_i = \sum_{i=1}^n p_i \xi_i x_i = \mathrm{E}^{\mathbb{P}}[\xi x].$$

Now let us consider the case where the state space Ω is infinite. Also in this case the change from a probability measure $\mathbb{P}$ to an equivalent probability measure $\mathbb{Q}$ is represented by a strictly positive random variable $\xi = d\mathbb{Q}/d\mathbb{P}$ with $\mathrm{E}^{\mathbb{P}}[\xi] = 1$. Again the expected value under the measure $\mathbb{Q}$ of a random variable x is given by $\mathrm{E}^{\mathbb{Q}}[x] = \mathrm{E}^{\mathbb{P}}[\xi x]$, since

$$\mathrm{E}^{\mathbb{Q}}[x] = \int_\Omega x \, d\mathbb{Q} = \int_\Omega x \frac{d\mathbb{Q}}{d\mathbb{P}} \, d\mathbb{P} = \int_\Omega x\xi \, d\mathbb{P} = \mathrm{E}^{\mathbb{P}}[\xi x].$$

In our economic models we will model the dynamics of uncertain objects over some time span $[0, T]$. For example, we might be interested in determining bond prices with maturities up to T years. Then we are interested in the stochastic process on this time interval, that is $x = (x_t)_{t \in [0,T]}$. The state space Ω is the set of possible paths of the relevant processes over the period $[0, T]$ so that all the relevant uncertainty has been resolved at time T and the values of all relevant random variables will be known at time T. The Radon–Nikodym derivative $\xi = d\mathbb{Q}/d\mathbb{P}$ is also a random variable and is therefore known at time T and usually not before time T. To indicate this the Radon–Nikodym derivative is often denoted by $\xi_T = \frac{d\mathbb{Q}}{d\mathbb{P}}$.

We can define a stochastic process $\xi = (\xi_t)_{t\in[0,T]}$ by setting

$$\xi_t = \mathrm{E}_t^{\mathbb{P}}\left[\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = \mathrm{E}_t^{\mathbb{P}}\left[\xi_T\right].$$

This definition is consistent with ξ_T being identical to $d\mathbb{Q}/d\mathbb{P}$, since all uncertainty is resolved at time T so that the time T expectation of any variable is just equal to the variable. Note that the process ξ is a $\mathbb{P}$-martingale, since for any $t < t' \leq T$ we have

$$\mathrm{E}_t^{\mathbb{P}}\left[\xi_{t'}\right] = \mathrm{E}_t^{\mathbb{P}}\left[\mathrm{E}_{t'}^{\mathbb{P}}\left[\xi_T\right]\right] = \mathrm{E}_t^{\mathbb{P}}\left[\xi_T\right] = \xi_t.$$

Here the first and the third equalities follow from the definition of ξ. The second equality follows from the Law of Iterated Expectations, Theorem 3.1. The following result turns out to be very useful in our dynamic models of the economy. Let $x = (x_t)_{t\in[0,T]}$ be any stochastic process. Then we have

$$\mathrm{E}_t^{\mathbb{Q}}\left[x_{t'}\right] = \mathrm{E}_t^{\mathbb{P}}\left[\frac{\xi_{t'}}{\xi_t} x_{t'}\right]. \tag{3.21}$$

This is called Bayes' Formula. For a proof, see Björk (2009, Prop. B.41).

Suppose that the underlying uncertainty is represented by a standard Brownian motion $z = (z_t)$ (under the real-world probability measure $\mathbb{P}$), as will be the case in all the models we will consider. Let $\lambda = (\lambda_t)_{t\in[0,T]}$ be any sufficiently well-behaved stochastic process.[5] Here, z and λ must have the same dimension. For notational simplicity, we assume in the following that they are one-dimensional, but the results generalize naturally to the multi-dimensional case. We can generate an equivalent probability measure $\mathbb{Q}^\lambda$ in the following way. Define the process $\xi^\lambda = (\xi_t^\lambda)_{t\in[0,T]}$ by

$$\xi_t^\lambda = \exp\left\{-\int_0^t \lambda_s\, dz_s - \frac{1}{2}\int_0^t \lambda_s^2\, ds\right\}. \tag{3.22}$$

Then $\xi_0^\lambda = 1$, ξ^λ is strictly positive, and it can be shown that ξ^λ is a $\mathbb{P}$-martingale (see Exercise 3.11) so that $\mathrm{E}^{\mathbb{P}}[\xi_T^\lambda] = \xi_0^\lambda = 1$. Consequently, an equivalent probability measure $\mathbb{Q}^\lambda$ can be defined by the Radon–Nikodym derivative

$$\frac{d\mathbb{Q}^\lambda}{d\mathbb{P}} = \xi_T^\lambda = \exp\left\{-\int_0^T \lambda_s\, dz_s - \frac{1}{2}\int_0^T \lambda_s^2\, ds\right\}.$$

[5] Sufficient conditions are that λ is square-integrable in the sense that $\int_0^T \lambda_t^2\, dt$ is finite with probability 1 and that λ satisfies Novikov's condition, that is the expectation $\mathrm{E}^{\mathbb{P}}\left[\exp\left\{\frac{1}{2}\int_0^T \lambda_t^2\, dt\right\}\right]$ is finite.

From (3.21), we get that

$$\mathrm{E}_t^{\mathbb{Q}^\lambda}[x_{t'}] = \mathrm{E}_t^{\mathbb{P}}\left[\frac{\xi_{t'}^\lambda}{\xi_t^\lambda} x_{t'}\right] = \mathrm{E}_t^{\mathbb{P}}\left[x_{t'} \exp\left\{-\int_t^{t'} \lambda_s \, dz_s - \frac{1}{2}\int_t^{t'} \lambda_s^2 \, ds\right\}\right]$$

for any stochastic process $x = (x_t)_{t\in[0,T]}$. A central result is Girsanov's Theorem:

Theorem 3.10 (Girsanov) *The process $z^\lambda = (z_t^\lambda)_{t\in[0,T]}$ defined by*

$$z_t^\lambda = z_t + \int_0^t \lambda_s \, ds, \quad 0 \le t \le T,$$

is a standard Brownian motion under the probability measure $\mathbb{Q}^\lambda$. In differential notation,

$$dz_t^\lambda = dz_t + \lambda_t \, dt.$$

This theorem has the attractive consequence that the effects on a stochastic process of changing the probability measure from $\mathbb{P}$ to some $\mathbb{Q}^\lambda$ are captured by a simple adjustment of the drift. If $x = (x_t)$ is an Itô process with dynamics

$$dx_t = \mu_t \, dt + \sigma_t \, dz_t,$$

then

$$dx_t = \mu_t \, dt + \sigma_t \left(dz_t^\lambda - \lambda_t \, dt\right) = (\mu_t - \sigma_t\lambda_t) \, dt + \sigma_t \, dz_t^\lambda.$$

Hence, $\mu - \sigma\lambda$ is the drift under the probability measure $\mathbb{Q}^\lambda$, which is different from the drift under the original measure $\mathbb{P}$ unless σ or λ are identically equal to zero. In contrast, the volatility remains the same as under the original measure.

In many financial models, the relevant change of measure is such that the distribution under $\mathbb{Q}^\lambda$ of the future value of the central processes is of the same class as under the original $\mathbb{P}$ measure, but with different moments. For example, consider the Ornstein–Uhlenbeck process

$$dx_t = (\varphi - \kappa x_t) \, dt + \sigma \, dz_t$$

and perform the change of measure given by a constant $\lambda_t = \lambda$. Then the dynamics of x under the measure $\mathbb{Q}^\lambda$ is given by

$$dx_t = \left(\hat{\varphi} - \kappa x_t\right) \, dt + \sigma \, dz_t^\lambda,$$

where $\hat{\varphi} = \varphi - \sigma\lambda$. Consequently, the future values of x are normally distributed both under $\mathbb{P}$ and $\mathbb{Q}^\lambda$. From (3.14) and (3.15), we see that the variance of $x_{t'}$ (given x_t) is the same under $\mathbb{Q}^\lambda$ and $\mathbb{P}$, but the expected values will differ (recall that $\theta = \varphi/\kappa$):

$$\mathrm{E}_t^{\mathbb{P}}[x_{t'}] = e^{-\kappa(t'-t)} x_t + \frac{\varphi}{\kappa}\left(1 - e^{-\kappa(t'-t)}\right),$$

$$\mathrm{E}_t^{\mathbb{Q}^\lambda}[x_{t'}] = e^{-\kappa(t'-t)} x_t + \frac{\hat{\varphi}}{\kappa}\left(1 - e^{-\kappa(t'-t)}\right).$$

However, in general, a shift of probability measure may change not only some or all moments of future values, but also the distributional class.

3.11 EXERCISES

Exercise 3.1 Suppose $x = (x_t)$ is a geometric Brownian motion, $dx_t = \mu x_t\, dt + \sigma x_t\, dz_t$. What is the dynamics of the process $y = (y_t)$ defined by $y_t = (x_t)^n$? What can you say about the distribution of future values of the y process?

Exercise 3.2 Let y be a random variable and define a stochastic process $x = (x_t)$ by $x_t = \mathrm{E}_t[y]$. Show that x is a martingale.

Exercise 3.3 (Adapted from Björk (2009).) Define the process $y = (y_t)$ by $y_t = z_t^4$, where $z = (z_t)$ is a standard Brownian motion. Find the dynamics of y. Show that

$$y_t = 6 \int_0^t z_s^2\, ds + 4 \int_0^t z_s^3\, dz_s.$$

Show that $\mathrm{E}[y_t] \equiv \mathrm{E}[z_t^4] = 3t^2$, where $\mathrm{E}[\ ]$ denotes the expectation given the information at time 0.

Exercise 3.4 (Adapted from Björk (2009).) Define the process $y = (y_t)$ by $y_t = e^{az_t}$, where a is a constant and $z = (z_t)$ is a standard Brownian motion. Find the dynamics of y. Show that

$$y_t = 1 + \frac{1}{2} a^2 \int_0^t y_s\, ds + a \int_0^t y_s\, dz_s.$$

Define $m(t) = \mathrm{E}[y_t]$. Show that m satisfies the ordinary differential equation

$$m'(t) = \frac{1}{2} a^2 m(t), \quad m(0) = 1.$$

Show that $m(t) = e^{a^2 t/2}$ and conclude that

$$\mathrm{E}\left[e^{az_t}\right] = e^{a^2 t/2}.$$

Exercise 3.5 Consider the two general stochastic processes $x_1 = (x_{1t})$ and $x_2 = (x_{2t})$ defined by the dynamics

$$dx_{1t} = \mu_{1t}\, dt + \sigma_{1t}\, dz_{1t},$$

$$dx_{2t} = \mu_{2t}\, dt + \rho_t \sigma_{2t}\, dz_{1t} + \sqrt{1 - \rho_t^2}\,\sigma_{2t}\, dz_{2t},$$

where z_1 and z_2 are independent one-dimensional standard Brownian motions. Interpret μ_{it}, σ_{it}, and ρ_t. Define the processes $y = (y_t)$ and $w = (w_t)$ by $y_t = x_{1t}x_{2t}$ and $w_t = x_{1t}/x_{2t}$. What are the dynamics of y and w? Concretize your answer for the special case where x_1 and x_2 are geometric Brownian motions with constant correlation, that is $\mu_{it} = \mu_i x_{it}$, $\sigma_{it} = \sigma_i x_{it}$, and $\rho_t = \rho$ with μ_i, σ_i, and ρ being constants.

Exercise 3.6 Find the dynamics of the process ξ^λ defined in (3.22).

4

A Review of General Asset Pricing Theory

4.1 INTRODUCTION

Bonds and other fixed income securities have some special characteristics that make them distinctively different from other financial assets such as stocks and stock market derivatives. However, in the end, all financial assets serve the same purpose: shifting consumption opportunities across time and states. Hence, the pricing of fixed income securities follows the same general principles as the pricing of all other financial assets. In this chapter we will discuss some important general concepts and results in asset pricing theory that will then be applied in the following chapters to the term structure of interest rates and the pricing of fixed income securities.

The fundamental concepts of asset pricing theory are arbitrage, state prices, risk-neutral probability measures, market prices of risk, and market completeness. Asset pricing models aim at characterizing equilibrium prices of financial assets. A market is in equilibrium if the prices are such that the market clears (that is supply equals demand) and every investor has picked a trading strategy in the financial assets that is optimal given his preferences and budget constraints and given the prices prevailing in the market. An arbitrage is a trading strategy that generates a risk-free profit, that is gives something for nothing. If an investor has the opportunity to invest in an arbitrage, he will surely do so, and hence change his original trading strategy. A market in which prices allow arbitrage is therefore not in equilibrium. When searching for equilibrium prices we can thus limit ourselves to no-arbitrage prices. In Section 4.2 we introduce our general model of assets and define the concept of an arbitrage more formally.

In typical financial markets thousands of different assets are traded. The price of each asset will, of course, depend on the future payoffs of the asset. In order to price the assets in a financial market, one strategy would be to specify the future payoffs of all assets in all possible states of the world and then try to figure out which set of prices would rule out arbitrage. However, this would surely be a quite complicated procedure. Instead we try first to determine how a general future payoff stream should be valued in order to rule out arbitrage and then this general arbitrage-free pricing mechanism can be applied to the payoffs of any particular asset. We will show how to capture the general arbitrage-free pricing mechanisms in a market in three different, but equivalent objects: a state-price deflator, a risk-neutral probability measure, and a market price of risk. Once one of these objects has been specified, any payoff stream can be priced. We discuss these objects and the relations between them and no-arbitrage pricing in Section 4.3. We will also see that the general pricing mechanism is closely related to the marginal utilities of consumption of the agents investing in the market.

While the risk-neutral probability measure is a standard object for summarizing an arbitrage-free price system, we show in Section 4.4 that we might as well use other probability measures for the same purpose. When it comes to derivative pricing, it is often computationally convenient to use a carefully selected probability measure.

In Section 4.5, we make a distinction between markets which are complete and markets which are incomplete. Basically, a market is complete if all risks are traded in the sense that agents can obtain any desired exposure to the shocks to the economy. In general markets many state-price deflators (or risk-neutral probability measures or market prices of risk) will be consistent with absence of arbitrage. We will see that in a complete, arbitrage-free market there will be a unique state-price deflator (or risk-neutral probability measure or market price of risk). We show in Section 4.6 that in a complete market, we may assume that the economy is inhabited by a single, representative agent. We will apply this in the next chapter in order to link the term structure of interest rate to aggregate consumption.

For notational simplicity we will first develop the main results under the assumption that the available assets only pay dividends at some time T, where all relevant uncertainty is resolved. In Section 4.7 we show how to generalize the results to the more realistic case with dividends at other points in time.

Finally, Section 4.8 considers the special class of diffusion models which covers many popular term structure models and also the famous Black–Scholes–Merton model for stock option pricing. Assuming that the relevant information for the pricing of a given asset is captured by a (preferably low-dimensional) diffusion process, the price of the asset can be found by solving a partial differential equation.

Our analysis is set in the framework of continuous-time stochastic models. Most of the general asset pricing concepts and results were originally developed in discrete-time models, where interpretations and proofs are sometimes easier to understand. Some classic references are Arrow (1951, 1953, 1964, 1971), Debreu (1953, 1954, 1959), Negishi (1960), and Ross (1978). As already discussed in Section 3.2.4 continuous-time models are often more elegant and tractable, and a continuous-time setting can be argued to be more realistic than a discrete-time setting. Moreover, most term structure models are formulated in continuous time, so we really need the continuous-time versions of the general asset pricing concepts and results. Many of the definitions and results in the continuous-time framework are originally due to Harrison and Kreps (1979) and Harrison and Pliska (1981, 1983). Asset pricing theory is developed in more detail in a number of textbooks. Presentations focusing on discrete-time asset pricing include Huang and Litzenberger (1988), LeRoy and Werner (2001), Cochrane (2005), and Skiadas (2009), whereas Karatzas and Shreve (1998) provides a mathematically rigorous view on continuous-time asset pricing. There are also asset pricing texts covering both discrete-time and continuous-time models such as Ingersoll (1987), Duffie (2001), Pennacchi (2008), and Munk (2010).

4.2 ASSETS, TRADING STRATEGIES, AND ARBITRAGE

We will set up a model for an economy over a certain time period $[0, T]$, where T represents some terminal point in time in the sense that we do not model what happens after time T. We assume that the basic uncertainty in the economy is represented by the evolution of a d-dimensional standard Brownian motion, $z = (z_t)_{t\in[0,T]}$. Think of dz_t as a vector of d exogenous shocks to the economy at time t. All the uncertainty that affects the dividends of the assets or the investment decisions of investors stems from these exogenous shocks. This includes

- *financial* uncertainty about the dividends of the assets, the evolution of prices and interest rates, future expected returns, volatilities, and correlations;
- *non-financial* uncertainty, for example, about the prices of consumption goods and the future labour income of the agents.

The state space Ω is in this case the set of all paths of the Brownian motion z. Note that since a Brownian motion has infinitely many possible paths, we have an infinite state space. The information filtration $\mathbf{F} = (\mathcal{F}_t)_{t\in[0,T]}$ represents the information that can be extracted from observing z so that $\mathbf{F} = \mathbf{F}^z$, the smallest filtration with respect to which the process z is adapted.

For notational simplicity we shall first develop the main results for the case where the available assets pay no dividends before time T. Later we will discuss the necessary modifications in the presence of intermediate dividends.

4.2.1 Assets

We model a financial market with 1 instantaneously risk-free and N risky assets. Let us first describe the instantaneously risk-free asset. Let r_t denote the continuously compounded, instantaneously risk-free interest rate at time t so that the rate of return over an infinitesimal interval $[t, t + dt]$ is $r_t\, dt$. The instantaneously risk-free asset is a continuous roll-over of such instantaneously risk-free investments. We shall refer to this asset as **the bank account**. Let $A = (A_t)$ denote the price process of the bank account. The increment to the balance of the account over an infinitesimal interval $[t, t + dt]$ is known at time t to be

$$dA_t = A_t r_t\, dt.$$

A time zero deposit of A_0 will grow to

$$A_t = A_0 e^{\int_0^t r_u\, du}$$

at time t. We think of A_T as the terminal dividend of the bank account. We need to assume that the process $r = (r_t)$ is such that $\int_0^T |r_t|\, dt$ is finite with probability one. Note that the bank account is only *instantaneously* risk-free since future interest rates are generally not known. We refer to r_t as the short-term interest rate or simply the **short rate**. Some authors use the phrase spot rate to distinguish this rate from forward rates. If the zero-coupon yield curve at time t is given by $\tau \mapsto y_t^{t+\tau}$

for $\tau > 0$, we can think of r_t as the limiting value $\lim_{\tau\to 0} y_t^{t+\tau}$, which corresponds to the intercept of the yield curve and the vertical axis in a (τ, y)-diagram.

The short rate is strictly speaking a zero-maturity interest rate. The maturity of the shortest government bond traded in the market may be several months, so that it is impossible to observe the short rate directly from market prices. The short rate in the bond markets can be estimated as the intercept of a yield curve. In the money markets, rates are set for deposits and loans of very short maturities, typically as short as one day. While this is surely a reasonable proxy for the zero-maturity interest rate in the money markets, it is not necessarily a good proxy for the risk-free (government bond) short rate. The reason is that money market rates apply for unsecured loans between financial institutions and hence they reflect the default risk of those investors, see the discussion in Chapter 1. Money market rates are therefore expected to be higher than similar bond market rates.

The prices of the N risky assets are modelled as general Itô processes, see Section 3.5. The price process $P_i = (P_{it})$ of the i'th risky asset is assumed to be of the form

$$dP_{it} = P_{it}\left[\mu_{it}\,dt + \sum_{j=1}^{d} \sigma_{ijt}\,dz_{jt}\right].$$

Here $\mu_i = (\mu_{it})$ denotes the (relative) drift, and $\sigma_{ij} = (\sigma_{ijt})$ reflects the relative sensitivity of the price to the j'th exogenous shock. Note that the price of a given asset may not be sensitive to all the shocks $dz_{1t}, \dots, dz_{dt}$ so that some of the σ_{ijt} may be equal to zero. It can also be that no asset is sensitive to a particular shock. Some shocks may be relevant for investors, but not affect asset prices directly, for example shocks to labour income. If we let $\boldsymbol{\sigma}_{it}$ be the sensitivity vector $(\sigma_{i1t}, \dots, \sigma_{idt})^\top$, the price dynamics of asset i can be rewritten as

$$dP_{it} = P_{it}\left[\mu_{it}\,dt + \boldsymbol{\sigma}_{it}^\top\,d\boldsymbol{z}_t\right]. \tag{4.1}$$

We think of P_{iT} as the terminal dividend of asset i.

We can write the price dynamics of all the N risky assets compactly using vector notation as

$$d\boldsymbol{P}_t = \operatorname{diag}(\boldsymbol{P}_t)\left[\boldsymbol{\mu}_t\,dt + \underline{\underline{\sigma}}_t\,d\boldsymbol{z}_t\right],$$

where

$$\boldsymbol{P}_t = \begin{pmatrix} P_{1t} \\ P_{2t} \\ \vdots \\ P_{Nt} \end{pmatrix}, \qquad \operatorname{diag}(\boldsymbol{P}_t) = \begin{pmatrix} P_{1t} & 0 & \dots & 0 \\ 0 & P_{2t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & P_{Nt} \end{pmatrix},$$

$$\boldsymbol{\mu}_t = \begin{pmatrix} \mu_{1t} \\ \mu_{2t} \\ \vdots \\ \mu_{Nt} \end{pmatrix}, \qquad \underline{\underline{\sigma}}_t = \begin{pmatrix} \sigma_{11t} & \sigma_{12t} & \dots & \sigma_{1dt} \\ \sigma_{21t} & \sigma_{22t} & \dots & \sigma_{2dt} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{N1t} & \sigma_{N2t} & \dots & \sigma_{Ndt} \end{pmatrix}.$$

We assume that the processes μ_i and σ_{ij} are 'well-behaved', for example generating prices with finite variances. The economic interpretation of μ_{it} is the expected rate of return per time period (year) over the next instant. The matrix $\underline{\underline{\sigma}}_t$ captures the sensitivity of the prices to the exogenous shocks and determines the instantaneous variances and covariances (and, hence, also the correlations) of the risky asset prices. In particular, $\underline{\underline{\sigma}}_t \underline{\underline{\sigma}}_t^\top \, dt$ is the $N \times N$ variance-covariance matrix of the rates of return over the next instant $[t, t + dt]$. The volatility of asset i is the standard deviation of the relative price change per time unit over the next instant, that is $\|\boldsymbol{\sigma}_{it}\| = \left(\sum_{j=1}^d \sigma_{ijt}^2\right)^{1/2}$.

4.2.2 Trading strategies

A **trading strategy** is a pair $(\alpha, \boldsymbol{\theta})$, where $\alpha = (\alpha_t)$ is a real-valued process representing the units held of the instantaneously risk-free asset and $\boldsymbol{\theta}$ is an N-dimensional process representing the units held of the N risky assets. To be precise, $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \dots, \boldsymbol{\theta}_N)^\top$, where $\boldsymbol{\theta}_i = (\boldsymbol{\theta}_{it})$ with $\boldsymbol{\theta}_{it}$ representing the units of asset i held at time t. The value of a trading strategy at time t is given by

$$V_t^{\alpha,\boldsymbol{\theta}} = \alpha_t A_t + \boldsymbol{\theta}_t^\top \boldsymbol{P}_t.$$

The gains from holding the portfolio $(\alpha_t, \boldsymbol{\theta}_t)$ over the infinitesimal interval $[t, t + dt]$ is

$$\alpha_t \, dA_t + \boldsymbol{\theta}_t^\top d\boldsymbol{P}_t = \alpha_t r_t e^{\int_0^t r_s \, ds} \, dt + \boldsymbol{\theta}_t^\top d\boldsymbol{P}_t.$$

A trading strategy is called **self-financing** if the future value is equal to the sum of the initial value and the accumulated trading gains so that no money has been added or withdrawn. In mathematical terms, a trading strategy $(\alpha, \boldsymbol{\theta})$ is self-financing if

$$V_t^{\alpha,\boldsymbol{\theta}} = V_0^{\alpha,\boldsymbol{\theta}} + \int_0^t \left(\alpha_s r_s e^{\int_0^s r_u \, du} \, ds + \boldsymbol{\theta}_s^\top d\boldsymbol{P}_s \right)$$

or, in differential terms,

$$\begin{aligned} dV_t^{\alpha,\boldsymbol{\theta}} &= \alpha_t r_t e^{\int_0^t r_u \, du} \, dt + \boldsymbol{\theta}_t^\top d\boldsymbol{P}_t \\ &= \left(\alpha_t r_t A_t + \boldsymbol{\theta}_t^\top \operatorname{diag}(\boldsymbol{P}_t) \boldsymbol{\mu}_t \right) \, dt + \boldsymbol{\theta}_t^\top \operatorname{diag}(\boldsymbol{P}_t) \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t. \end{aligned} \tag{4.2}$$

4.2.3 Redundant assets

An asset is said to be **redundant** if there exists a self-financing trading strategy in other assets which yields the same payoff at time T. In order to be sure to end up with the same payoff or value at time T, the value of the replicating trading strategy must be identical to the price of the asset at any point in time and in any state. Hence, the value process of the strategy and the price process of the asset must be identical. In particular, the value process of the strategy must react to shocks to the economy in the same way as the price process of the asset. Therefore, an asset is redundant whenever the sensitivity vector of its price process is a linear combination of the sensitivity vectors of the price processes of the other assets. This implies that whenever there are redundant assets among the N assets, the rows in the matrix $\underline{\underline{\sigma}}_t$ are linearly dependent.[1]

As the name reflects, a redundant asset does not in any way enhance the opportunities of the agents to move consumption across time and states. The agents can do just as well without the redundant assets. Therefore, we can remove the redundant assets from the set of traded assets. Note that whether an asset is redundant or not depends on the other available assets. Therefore, we should remove redundant assets one by one. First identify one redundant asset and remove that. Then, based on the remaining assets, look for another redundant asset and remove that. Continuing that process until none of the remaining assets are redundant, the number of remaining assets will be equal to the rank[2] of the original sensitivity matrix $\underline{\underline{\sigma}}_t$. Suppose the rank of $\underline{\underline{\sigma}}_t$ equals k for all t. Then there will be k non-redundant assets. We let $\hat{\underline{\underline{\sigma}}}_t$ denote the $k \times d$ matrix obtained from $\underline{\underline{\sigma}}_t$ by removing rows corresponding to redundant assets and let $\hat{\boldsymbol{\mu}}_t$ denote the k-dimensional vector that is left after deleting from $\boldsymbol{\mu}_t$ the elements corresponding to the redundant assets.

4.2.4 Arbitrage

An **arbitrage** is a self-financing trading strategy $(\alpha, \boldsymbol{\theta})$ satisfying one of the following two conditions:

1. $V_0^{\alpha,\boldsymbol{\theta}} < 0$; $V_T^{\alpha,\boldsymbol{\theta}} \geq 0$ with probability 1,
2. $V_0^{\alpha,\boldsymbol{\theta}} \leq 0$; $V_T^{\alpha,\boldsymbol{\theta}} \geq 0$ with probability 1 and $V_T^{\alpha,\boldsymbol{\theta}} > 0$ with strictly positive probability.

A trading strategy $(\alpha, \boldsymbol{\theta})$ satisfying (1) has a negative initial price so the investor receives money when initiating the trading strategy. The terminal payoff of the strategy is non-negative no matter how the world evolves and since the strategy is self-financing there are no intermediate payments. Any rational investor would want to invest in such a trading strategy. Likewise, a trading strategy satisfying (2) will never require the investor to make any payments and it offers a positive probability of a positive terminal payoff. It is like a free lottery ticket.

A straightforward consequence of arbitrage-free pricing is that the price of a redundant asset must be equal to the cost of implementing the self-financing

[1] Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are called linearly independent if $k_1\boldsymbol{a} + k_2\boldsymbol{b} = \boldsymbol{0}$ implies $k_1 = k_2 = 0$, that is $\boldsymbol{a}$ and $\boldsymbol{b}$ cannot be linearly combined into a zero vector. If they are not linearly independent, they are said to be linearly dependent.

[2] The rank of a matrix is defined to be the maximum number of linearly independent rows in the matrix or, equivalently, the maximum number of linearly independent columns. The rank of a $k \times l$ matrix has to be less than or equal to the minimum of k and l. If the rank is equal to the minimum of k and l, the matrix is said to be of *full rank*.

replicating trading strategy. If the redundant asset was cheaper than the replicating trading strategy, an arbitrage can be realized by buying the redundant asset and shorting the replicating trading strategy. Conversely, if the redundant asset was more expensive than the replicating strategy. This observation is the foundation of many models of derivatives pricing including the famous Black–Scholes–Merton model of stock option pricing, see Black and Scholes (1973) and Merton (1973).

Although the definition of arbitrage focuses on payoffs at time T, it does cover shorter term risk-free gains. Suppose, for example, that we can construct a trading strategy with a non-positive initial value (that is a non-positive price), always non-negative values, and a strictly positive value at some time $t < T$. Then this strictly positive value can be invested in the bank account in the period $[t, T]$ generating a strictly positive terminal value.

Any realistic model of equilibrium prices should rule out arbitrage. However, in our continuous-time setting it is in fact possible to construct some strategies that generate something for nothing. These are the so-called doubling strategies. Think of a series of coin tosses enumerated by $n = 1, 2, \dots$. The n'th coin toss takes place at time $1 - 1/(n+1)$. In the n'th toss, you get $\alpha 2^{n-1}$ if heads comes up, and loses $\alpha 2^{n-1}$ otherwise. You stop betting the first time heads comes up. Suppose heads comes up the first time in toss number $(k+1)$. Then in the first k tosses you have lost a total of $\alpha(1 + 2 + \cdots + 2^{k-1}) = \alpha(2^k - 1)$. Since you win $\alpha 2^k$ in toss number $k+1$, your total profit will be $\alpha 2^k - \alpha(2^k - 1) = \alpha$. Since the probability that heads comes up eventually is equal to one, you will gain α with probability one. Similar strategies can be constructed in continuous-time models of financial markets, but are clearly impossible to implement in real life. These strategies are ruled out by requiring that trading strategies have values that are bounded from below, that is that some constant K exists such that $V_t^{\alpha,\boldsymbol{\theta}} \geq -K$ for all t. This is a reasonable restriction since no one can borrow an infinite amount of money. If you have a limited borrowing potential, the doubling strategy described above cannot be implemented.

4.3 STATE-PRICE DEFLATORS, RISK-NEUTRAL PROBABILITIES, AND MARKET PRICES OF RISK

Instead of trying to separately price each of the many, many financial assets traded, it is wiser first to derive a representation of the general pricing mechanism in an arbitrage-free market. In order to price a particular asset the general mechanism can then be combined with the asset-specific payoff. In this section we give three basically equivalent representations of arbitrage-free price systems: state-price deflators, risk-neutral probability measures, and market prices of risk. Once one of these objects has been specified, any payoff stream can be priced.

4.3.1 State-price deflators

A **state-price deflator** is a stochastic process $\zeta = (\zeta_t)$ satisfying

1. $\zeta_0 = 1$,

2. $\zeta_t > 0$ for all $t \in [0, T]$ and all states,
3. $\text{Var}[\zeta_t] < \infty$ for all $t \in [0, T]$,
4. the product of the state-price deflator and the price of any asset is a martingale.

The last condition means that

- $(\zeta_t \exp\{\int_0^t r_u \, du\})$ is a martingale,
- $(\zeta_t P_{it})$ is a martingale for any $i = 1, \dots, N$.

In particular, for all $t < t' \leq T$, we have

$$P_{it}\zeta_t = \mathrm{E}_t \left[P_{it'} \zeta_{t'} \right],$$

or

$$P_{it} = \mathrm{E}_t \left[\frac{\zeta_{t'}}{\zeta_t} P_{it'} \right]. \tag{4.3}$$

Suppose we are given a state-price deflator ζ and hence the distribution of ζ_T/ζ_t. Then the price at time t of an asset with a terminal dividend given by the random variable P_{iT} is equal to $\mathrm{E}_t[(\zeta_T/\zeta_t)P_{iT}]$. Hence, the state-price deflator captures the market-wide pricing information. In particular, for a zero-coupon bond with a unit payment at time T, the time t price is

$$B_t^T = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} \right].$$

Let us write the dynamics of a state-price deflator as

$$d\zeta_t = \zeta_t \left[m_t \, dt + \boldsymbol{v}_t^\top \, d\boldsymbol{z}_t \right] \tag{4.4}$$

for some relative drift m and some 'sensitivity' vector $\boldsymbol{v}$. Define $\zeta_t^* = \zeta_t A_t = \zeta_t \exp\{\int_0^t r_u \, du\}$. By Itô's Lemma,

$$d\zeta_t^* = \zeta_t^* \left[(m_t + r_t) \, dt + \boldsymbol{v}_t^\top \, d\boldsymbol{z}_t \right].$$

Since $\zeta^* = (\zeta_t^*)$ is a martingale, we must have $m_t = -r_t$, that is the relative drift of a state-price deflator is equal to the negative of the short-term interest rate. For any risky asset i, the process $\zeta_t^i = \zeta_t P_{it}$ must be a martingale. From Itô's Lemma and the dynamics of P_i and ζ given in (4.1) and (4.4), we get

$$\begin{aligned} d\zeta_t^i &= \zeta_t \, dP_{it} + P_{it} d\zeta_t + (d\zeta_t)(dP_{it}) \\ &= \zeta_t^i \left[\left(\mu_{it} + m_t + \boldsymbol{\sigma}_{it}^\top \boldsymbol{v}_t \right) \, dt + (\boldsymbol{v}_t + \boldsymbol{\sigma}_{it})^\top \, d\boldsymbol{z}_t \right]. \end{aligned}$$

Hence, for ζ to be a state-price deflator, the equation

$$\mu_{it} + m_t + \boldsymbol{\sigma}_{it}^\top \boldsymbol{v}_t = 0$$

must hold for any asset i, and we can substitute in $m_t = -r_t$. In compact form, the condition on $\boldsymbol{v}$ is then that

$$\boldsymbol{\mu}_t - r_t \mathbf{1} = -\underline{\underline{\sigma}}_t \boldsymbol{v}_t. \tag{4.5}$$

The product of a state-price deflator and the value of a (well-behaved) self-financing trading strategy will also be a martingale so that

$$\zeta_t V_t^{\alpha,\boldsymbol{\theta}} = \mathrm{E}_t\left[\zeta_{t'} V_{t'}^{\alpha,\boldsymbol{\theta}}\right].$$

To see this, first use Itô's Lemma to get

$$d(\zeta_t V_t^{\alpha,\boldsymbol{\theta}}) = \zeta_t\, dV_t^{\alpha,\boldsymbol{\theta}} + V_t^{\alpha,\boldsymbol{\theta}}\, d\zeta_t + (d\zeta_t)(dV_t^{\alpha,\boldsymbol{\theta}}).$$

Substituting in $dV_t^{\alpha,\boldsymbol{\theta}}$ from (4.2) and $d\zeta_t$ from (4.4), we get after some simplification that

$$d(\zeta_t V_t^{\alpha,\boldsymbol{\theta}}) = \zeta_t \boldsymbol{\theta}_t^{\top} \operatorname{diag}(\boldsymbol{P}_t)\left(\boldsymbol{\mu}_t - r_t \mathbf{1} + \underline{\underline{\sigma}}_t \boldsymbol{\nu}_t\right) dt + \zeta_t V_t^{\alpha,\boldsymbol{\theta}} \boldsymbol{\nu}_t^{\top}\, d\boldsymbol{z}_t.$$

From (4.5), we see that the drift is zero so that the process is a martingale.

Given a state-price deflator we can price any asset. But can we be sure that a state-price deflator exists? It turns out that the existence of a state-price deflator is basically equivalent to the absence of arbitrage. Here is the first part of that statement:

Theorem 4.1 *If a state-price deflator exists, prices admit no arbitrage.*

Proof: For simplicity, we will ignore the lower bound on the value processes of trading strategies (the interested reader is referred to Duffie (2001, p. 105) to see how to incorporate the lower bound; this involves local martingales and supermartingales which we will not discuss here). Suppose $(\alpha, \boldsymbol{\theta})$ is a self-financing trading strategy with $V_T^{\alpha,\boldsymbol{\theta}} \geq 0$. Given a state-price deflator $\zeta = (\zeta_t)$ the initial value of the strategy is

$$V_0^{\alpha,\boldsymbol{\theta}} = \mathrm{E}\left[\zeta_T V_T^{\alpha,\boldsymbol{\theta}}\right],$$

which must be non-negative since $\zeta_T > 0$. If, furthermore, there is a positive probability of $V_T^{\alpha,\boldsymbol{\theta}}$ being strictly positive, then $V_0^{\alpha,\boldsymbol{\theta}}$ must be strictly positive. Consequently, arbitrage is ruled out. □

Conversely, under some technical conditions, the absence of arbitrage implies the existence of a state-price deflator. In the absence of arbitrage the optimal consumption strategy of any agent is finite and well-defined and we will now show that the marginal rate of intertemporal substitution of the agent can then be used as a state-price deflator.

In a continuous-time setting it is natural to assume that each agent consumes according to a non-negative continuous-time process $c = (c_t)$. We assume that the life-time utility from a given consumption process is of the time-additive form $\mathrm{E}[\int_0^T e^{-\delta t} u(c_t)\, dt]$. Here $u(\cdot)$ is the utility function and δ the time-preference rate (or subjective discount rate) of this agent. In this case c_t is the consumption rate at time t, that is it is the number of consumption goods consumed per time period. The total number of units of the good consumed over an interval $[t, t+\Delta t]$ is $\int_t^{t+\Delta t} c_s\, ds$ which for small Δt is approximately equal to $c_t \cdot \Delta t$. The agents can shift consumption across time and states by applying appropriate trading strategies.

Suppose $c = (c_t)$ is the optimal consumption process for some agent. Any deviation from this strategy will generate a lower utility. One deviation occurs if the agent at time 0 increases his investment in asset i by ε units. The extra costs of εP_{i0} implies a reduced consumption now. Let us suppose that the agent finances this extra investment by cutting down his consumption rate in the time interval $[0, \Delta t]$ for some small positive Δt by $\varepsilon P_{i0}/\Delta t$. The extra ε units of asset i is resold at time $t < T$, yielding a revenue of εP_{it}. This finances an increase in the consumption rate over $[t, t + \Delta t]$ by $\varepsilon P_{it}/\Delta t$. Since we have assumed so far that the assets pay no dividends before time T, the consumption rates outside the intervals $[0, \Delta t]$ and $[t, t + \Delta t]$ will be unaffected. Given the optimality of $c = (c_t)$, we must have that

$$\begin{aligned} \mathrm{E}\Bigg[\int_0^{\Delta t} & e^{-\delta s}\left(u\left(c_s - \frac{\varepsilon P_{i0}}{\Delta t}\right) - u(c_s)\right) ds \\ & + \int_t^{t+\Delta t} e^{-\delta s}\left(u\left(c_s + \frac{\varepsilon P_{it}}{\Delta t}\right) - u(c_s)\right) ds\Bigg] \le 0. \end{aligned}$$

Dividing by ε and letting $\varepsilon \to 0$, we obtain

$$\mathrm{E}\left[-\frac{P_{i0}}{\Delta t}\int_0^{\Delta t} e^{-\delta s} u'(c_s)\, ds + \frac{P_{it}}{\Delta t}\int_t^{t+\Delta t} e^{-\delta s} u'(c_s)\, ds\right] \le 0.$$

Letting $\Delta t \to 0$, we arrive at

$$\mathrm{E}\left[-P_{i0}u'(c_0) + P_{it}e^{-\delta t}u'(c_t)\right] \le 0,$$

or, equivalently,

$$P_{i0}u'(c_0) \ge \mathrm{E}\left[e^{-\delta t}P_{it}u'(c_t)\right].$$

The reverse inequality can be shown similarly by considering the 'opposite' perturbation, namely a decrease in the investment in asset i by ε units at time 0 over the interval $[0, t]$ leading to higher consumption over $[0, \Delta t]$ and lower consumption over $[t, t + \Delta t]$. Combining the two inequalities, we have that $P_{i0}u'(c_0) = \mathrm{E}[e^{-\delta t}P_{it}u'(c_t)]$ or more generally

$$P_{it} = \mathrm{E}_t\left[e^{-\delta(t'-t)}\frac{u'(c_{t'})}{u'(c_t)}P_{it'}\right], \quad t \le t' \le T. \tag{4.6}$$

With intermediate dividends this relation is slightly different, see Section 4.7.

Comparing (4.3) and (4.6), we see that

$$\zeta_t = e^{-\delta t}\frac{u'(c_t)}{u'(c_0)} \tag{4.7}$$

is a good candidate for a state-price deflator whenever the optimal consumption process c of the agent is well-behaved, as it presumably will be in the absence of arbitrage (the $u'(c_0)$ in the denominator is to ensure that $\zeta_0 = 1$). However, there are some technical subtleties one must consider when going from no arbitrage to the existence of a state-price deflator. Again, we refer the interested reader to Duffie (2001). We summarize in the following theorem:

Theorem 4.2 *If prices admit no arbitrage and technical conditions are satisfied, then a state-price deflator exists.*

The state-price deflator $\zeta_t = e^{-\delta t} u'(c_t)/u'(c_0)$ is the marginal rate of substitution of a particular agent evaluated at her optimal consumption rate. Since the purpose of financial assets is to allow agents to shift consumption across time and states, it is not surprising that the market-wide pricing information can be captured by the marginal rate of substitution. Note that each agent will induce a state-price deflator and since agents have different utility functions, different time preference rates, and different optimal consumption plans, there can potentially be (at least) as many state-price deflators as agents. However, some or all of these state-price deflators may be identical, see the discussion in Section 4.5.

Combining the two previous theorems, we have the following conclusion:

Corollary 4.1 *Under technical conditions, the existence of a state-price deflator is equivalent to the absence of arbitrage.*

4.3.2 Risk-neutral probability measures

For our market with no intermediate dividends, a probability measure $\mathbb{Q}$ is said to be a **risk-neutral probability measure** (or equivalent martingale measure) if the following three conditions are satisfied:

1. $\mathbb{Q}$ is equivalent to $\mathbb{P}$,
2. for any asset i, the discounted price process $\bar{P}_{it} = P_{it} \exp\{-\int_0^t r_s \, ds\}$ is a $\mathbb{Q}$-martingale,
3. the Radon–Nikodym derivative $d\mathbb{Q}/d\mathbb{P}$ has finite variance.

If $\mathbb{Q}$ is a risk-neutral probability measure, then condition (ii) immediately implies that

$$P_{it} = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t'} r_s \, ds} P_{it'} \right] \tag{4.8}$$

for any $t < t' \leq T$. In particular,

$$P_{it} = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{T} r_s \, ds} P_{iT} \right], \tag{4.9}$$

which shows that the price of an asset is equal to the risk-neutral expectation of the discounted terminal dividend, where the discounting is with the accumulated short rate. The joint risk-neutral distribution of the accumulated short rate $\int_t^T r_s \, ds$ and the terminal dividend P_{iT} fully determines the price of an asset. Under some technical conditions on $\boldsymbol{\theta}$, see Duffie (2001, p. 109), the same relations hold for any self-financing trading strategy $(\alpha, \boldsymbol{\theta})$ so that

$$V_t^{\alpha,\boldsymbol{\theta}} = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t'} r_s \, ds} V_{t'}^{\alpha,\boldsymbol{\theta}} \right]. \tag{4.10}$$

These relations show that the risk-neutral probability measure (together with the short-term interest rate process) captures the market-wide pricing information. Note that for the special case of a zero-coupon bond maturing at T, the price at time $t < T$ can be written as

$$B_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s\, ds}\right].$$

Theorem 3.8 implies that the price process of any non-dividend paying asset has a relative drift (that is an expected rate of return) under the risk-neutral measure which is equal to the short-term risk-free interest rate. In a hypothetical world where all investors were risk-neutral, all assets must have the same expected rate of return in equilibrium, because if one asset had a higher return than other assets, all the investors would buy more of that asset, its price would increase, and the expected rate of return decrease. With a risk-free asset available, this implies that all assets must have an expected rate of return equal to the risk-free rate in such a risk-neutral world. This explains the term risk-neutral probability measure: prices in the real world are just as they would have been in an economy in which all investors are risk-neutral and the probability measure is given by $\mathbb{Q}$.

The existence of a risk-neutral probability measure is closely related to absence of arbitrage, as summarized by the following theorem.

Theorem 4.3 *If a risk-neutral probability measure exists, prices admit no arbitrage.*

Proof: Suppose $(\alpha, \boldsymbol{\theta})$ is a self-financing trading strategy satisfying technical conditions ensuring that (4.10) holds. Then

$$V_0^{\alpha,\boldsymbol{\theta}} = \mathrm{E}^{\mathbb{Q}}\left[e^{-\int_0^T r_t\, dt} V_T^{\alpha,\boldsymbol{\theta}}\right].$$

Note that if $V_T^{\alpha,\boldsymbol{\theta}}$ is non-negative with probability one under the real-world probability measure $\mathbb{P}$, then it will also be non-negative with probability one under a risk-neutral probability measure $\mathbb{Q}$ since $\mathbb{Q}$ and $\mathbb{P}$ are equivalent. We see from the equation above that if $V_T^{\alpha,\boldsymbol{\theta}}$ is non-negative, so is $V_0^{\alpha,\boldsymbol{\theta}}$. If, in addition, $V_T^{\alpha,\boldsymbol{\theta}}$ is strictly positive with a strictly positive possibility, then $V_0^{\alpha,\boldsymbol{\theta}}$ must be strictly positive (again using the equivalence of $\mathbb{P}$ and $\mathbb{Q}$). Arbitrage is thus ruled out. □

The next theorem shows that, under technical conditions, there is a one-to-one relation between risk-neutral probability measures and state-price deflators. Hence, they are two equivalent representations of the market-wide pricing mechanism.

Theorem 4.4 *Given a risk-neutral probability measure $\mathbb{Q}$. Let $\xi_t = \mathrm{E}_t[d\mathbb{Q}/d\mathbb{P}]$ and define $\zeta_t = \xi_t \exp\{-\int_0^t r_s\, ds\}$. If ζ_t has finite variance for all $t \leq T$, then $\zeta = (\zeta_t)$ is a state-price deflator.*

Conversely, given a state-price deflator ζ, define $\xi_t = \exp\{\int_0^t r_s\, ds\}\zeta_t$. If ξ_T has finite variance, then a risk-neutral probability measure $\mathbb{Q}$ is defined by $d\mathbb{Q}/d\mathbb{P} = \xi_T$.

Proof: Suppose that $\mathbb{Q}$ is a risk-neutral probability measure. The change of measure implies that

$$\begin{aligned}
\mathrm{E}_t\left[\zeta_s P_{is}\right] &= e^{-\int_0^t r_u\,du}\,\mathrm{E}_t\left[\xi_s P_{is} e^{-\int_t^s r_u\,du}\right] = e^{-\int_0^t r_u\,du}\xi_t\,\mathrm{E}_t^{\mathbb{Q}}\left[P_{is} e^{-\int_t^s r_u\,du}\right] \\
&= e^{-\int_0^t r_u\,du}\xi_t P_{it} = \zeta_t P_{it},
\end{aligned}$$

where the second equality follows from (3.21). Hence, ζ is a state-price deflator (the finite variance condition on ζ_t and the finite variance of prices ensure the existence of the expectations).

Conversely, suppose that ζ is a state-price deflator and define ξ as in the statement of the theorem. Then

$$\mathrm{E}[\xi_T] = \mathrm{E}\left[e^{\int_0^T r_s\,ds}\zeta_T\right] = 1,$$

where the last equality is due to the fact that the product of the state-price deflator and the bank account value is a martingale. Furthermore, ξ_T is strictly positive so $d\mathbb{Q}/d\mathbb{P} = \xi_T$ defines an equivalent probability measure $\mathbb{Q}$. By assumption ξ_T has finite variance. It remains to check that discounted prices are $\mathbb{Q}$-martingales. Again using (3.21), we get

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'} r_s\,ds} P_{it'}\right] = \mathrm{E}_t\left[\frac{\xi_{t'}}{\xi_t} e^{-\int_t^{t'} r_s\,ds} P_{it'}\right] = \mathrm{E}_t\left[\frac{\zeta_{t'}}{\zeta_t} P_{it'}\right] = P_{it},$$

so this condition is also met. Hence, $\mathbb{Q}$ is a risk-neutral probability measure. $\square$

As discussed in the previous subsection, the absence of arbitrage implies the existence of a state-price deflator under some technical conditions, and the above theorem gives a one-to-one relation between state-price deflators and risk-neutral probability measures, also under some technical conditions. Hence, the absence of arbitrage will also imply the existence of a risk-neutral probability measure—again under technical conditions. Let us try to clarify this statement somewhat. The absence of arbitrage by itself does not imply the existence of a risk-neutral probability measure. We must require a little more than absence of arbitrage. As shown by Delbaen and Schachermayer (1994, 1999) the condition that prices admit no 'free lunch with vanishing risk' is equivalent to the existence of a risk-neutral probability measure and hence, by Theorem 4.4, the existence of a state-price deflator. We will not go into the precise and very technical definition of a free lunch with vanishing risk. Just note that while an arbitrage is a free lunch with vanishing risk, there are trading strategies which are not arbitrages but nevertheless are free lunches with vanishing risk. More importantly, we will see below that in markets with sufficiently nice price processes, we can indeed construct a risk-neutral probability measure. So the bottom-line is that absence of arbitrage is virtually equivalent to the existence of a risk-neutral probability measure.

4.3.3 Market prices of risk

If $\mathbb{Q}$ is a risk-neutral probability measure, the discounted prices are $\mathbb{Q}$-martingales. The discounted risky asset prices are given by

$$\bar{\boldsymbol{P}}_t = \boldsymbol{P}_t \, e^{-\int_0^t r_s \, ds}.$$

An application of Itô's Lemma shows that the dynamic of the discounted prices is

$$d\bar{\boldsymbol{P}}_t = \operatorname{diag}(\bar{\boldsymbol{P}}_t) \left[\left(\boldsymbol{\mu}_t - r_t \boldsymbol{1} \right) \, dt + \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t \right]. \tag{4.11}$$

Suppose that $\mathbb{Q}$ is a risk-neutral probability measure. The change of measure from $\mathbb{P}$ to $\mathbb{Q}$ is captured by a random variable, which we denote by $d\mathbb{Q}/d\mathbb{P}$. Define the process $\xi = (\xi_t)$ by $\xi_t = \mathrm{E}_t[d\mathbb{Q}/d\mathbb{P}]$. As explained in Section 3.10, ξ is a martingale. It now follows from the **Martingale Representation Theorem**, see Theorem 3.4, that a d-dimensional process $\psi = (\psi_t)$ exists such that $d\xi_t = \psi_t^{\top} \, d\boldsymbol{z}_t$. By defining the process $\lambda = (\lambda_t)$ by $\lambda_t = -\psi_t/\xi_t$, we have

$$d\xi_t = -\xi_t \boldsymbol{\lambda}_t^{\top} \, d\boldsymbol{z}_t,$$

or, equivalently (using $\xi_0 = \mathrm{E}[d\mathbb{Q}/d\mathbb{P}] = 1$),

$$\xi_t = \exp\left\{ -\frac{1}{2} \int_0^t \|\boldsymbol{\lambda}_s\|^2 \, ds - \int_0^t \boldsymbol{\lambda}_s^{\top} \, d\boldsymbol{z}_s \right\}. \tag{4.12}$$

According to **Girsanov's Theorem**, Theorem 3.10, the process $\boldsymbol{z}^{\mathbb{Q}} = (\boldsymbol{z}_t^{\mathbb{Q}})$ defined by

$$d\boldsymbol{z}_t^{\mathbb{Q}} = d\boldsymbol{z}_t + \boldsymbol{\lambda}_t \, dt, \quad \boldsymbol{z}_0^{\mathbb{Q}} = 0, \tag{4.13}$$

is a standard Brownian motion under the $\mathbb{Q}$-measure. Substituting $d\boldsymbol{z}_t = d\boldsymbol{z}_t^{\mathbb{Q}} - \boldsymbol{\lambda}_t \, dt$ into (4.11), we obtain

$$d\bar{\boldsymbol{P}}_t = \operatorname{diag}(\bar{\boldsymbol{P}}_t) \left[\left(\boldsymbol{\mu}_t - r_t \boldsymbol{1} - \underline{\underline{\sigma}}_t \boldsymbol{\lambda}_t \right) \, dt + \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t^{\mathbb{Q}} \right].$$

If discounted prices are to be $\mathbb{Q}$-martingales, the drift must be zero, so we must have that

$$\underline{\underline{\sigma}}_t \boldsymbol{\lambda}_t = \boldsymbol{\mu}_t - r_t \boldsymbol{1}. \tag{4.14}$$

From these arguments it follows that the existence of a solution $\boldsymbol{\lambda}$ to this system of equations is a necessary condition for the existence of a risk-neutral probability measure. Note that the system has N equations (one for each asset) in d unknowns, $\lambda_1, \ldots, \lambda_d$ (one for each exogenous shock).

On the other hand, if a solution $\boldsymbol{\lambda}$ exists and satisfies certain technical conditions, then a risk-neutral probability measure $\mathbb{Q}$ is defined by $d\mathbb{Q}/d\mathbb{P} = \xi_T$, where ξ_T is obtained by letting $t = T$ in (4.12). Sufficient conditions are that ξ_T has finite variance and that $\exp\left\{\frac{1}{2}\int_0^T \|\boldsymbol{\lambda}_t\|^2\,dt\right\}$ has finite expectation (the latter condition is Novikov's condition which ensures that the process $\xi = (\xi_t)$ is a martingale).

Any process $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$ solving (4.14) is called a **market price of risk process**. To understand this terminology, note that the i'th equation in the system (4.14) can be written as

$$\sum_{j=1}^{d} \sigma_{ijt}\lambda_{jt} = \mu_{it} - r_t.$$

If the price of the i'th asset is only sensitive to the j'th exogenous shock, the equation reduces to

$$\sigma_{ijt}\lambda_{jt} = \mu_{it} - r_t,$$

implying that

$$\lambda_{jt} = \frac{\mu_{it} - r_t}{\sigma_{ijt}}.$$

Therefore, λ_{jt} is the compensation in terms of excess expected return per unit of risk stemming from the j'th exogenous shock.

We summarize our findings as follows:

Theorem 4.5 *Under technical conditions, a risk-neutral probability measure exists if and only if a market price of risk process exists. The risk-neutral probability measure $\mathbb{Q}$ and the market price of risk process $\boldsymbol{\lambda}$ are linked by the relation*

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = \exp\left\{-\frac{1}{2}\int_0^T \|\boldsymbol{\lambda}_t\|^2\,dt - \int_0^T \boldsymbol{\lambda}_t^{\top}\,d\boldsymbol{z}_t\right\}.$$

Combining this theorem with earlier results, we can conclude that the existence of a market price of risk is virtually equivalent to the absence of arbitrage.

With a market price of risk it is easy to see the effects of changing the probability measure from the real-world measure $\mathbb{P}$ to a risk-neutral measure $\mathbb{Q}$. Suppose $\boldsymbol{\lambda}$ is a market price of risk process and let $\mathbb{Q}$ denote the associated risk-neutral probability measure and $\boldsymbol{z}^{\mathbb{Q}}$ the associated standard Brownian motion under $\mathbb{Q}$. Then

$$d\bar{\boldsymbol{P}}_t = \operatorname{diag}(\bar{\boldsymbol{P}}_t)\underline{\underline{\sigma}}_t\,d\boldsymbol{z}_t^{\mathbb{Q}}$$

and

$$d\boldsymbol{P}_t = \operatorname{diag}(\boldsymbol{P}_t)\left[r_t\mathbf{1}\,dt + \underline{\underline{\sigma}}_t\,d\boldsymbol{z}_t^{\mathbb{Q}}\right].$$

So under a risk-neutral probability all asset prices have a drift equal to the short rate. The volatilities are not affected by the change of measure.

Next, let us look at the relation between market prices of risk and state-price deflators. Suppose that $\boldsymbol{\lambda}$ is a market price of risk and ξ_t in (4.12) defines the associated risk-neutral probability measure. From Theorem 4.4 we know that, under a regularity condition, the process ζ defined by

$$\zeta_t = \xi_t e^{-\int_0^t r_s\, ds} = \exp\left\{-\int_0^t r_s\, ds - \frac{1}{2}\int_0^t \|\boldsymbol{\lambda}_s\|^2\, ds - \int_0^t \boldsymbol{\lambda}_s^{\top}\, d\boldsymbol{z}_s\right\}$$

is a state-price deflator. Since $d\xi_t = -\xi_t \boldsymbol{\lambda}_t^{\top} d\boldsymbol{z}_t$, an application of Itô's Lemma implies that

$$d\zeta_t = -\zeta_t \left[r_t\, dt + \boldsymbol{\lambda}_t^{\top} d\boldsymbol{z}_t\right]. \tag{4.15}$$

As we have already seen, the relative drift of a state-price deflator equals the negative of the short-term interest rate. Now, we see that the sensitivity vector of a state-price deflator equals the negative of a market price of risk, consistent with (4.5). Up to technical conditions, there is a one-to-one relation between market prices of risk and state-price deflators.

Let us again consider the key equation (4.14), which is a system of N equations in d unknowns given by the vector $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d)^{\top}$. The number of solutions to this system depends on the rank of the $N \times d$ matrix $\underline{\underline{\sigma}}_t$, which, as discussed in Section 4.2.3, equals the number of non-redundant assets. Let us assume that the rank of $\underline{\underline{\sigma}}_t$ is the same for all t (and all states) and denote the rank by k. We know that $k \leq d$. If $k < d$, there are several solutions to (4.14). We can write one solution as

$$\boldsymbol{\lambda}_t^* = \hat{\underline{\underline{\sigma}}}_t^{\top} \left(\hat{\underline{\underline{\sigma}}}_t \hat{\underline{\underline{\sigma}}}_t^{\top}\right)^{-1} \left(\hat{\boldsymbol{\mu}}_t - r_t \mathbf{1}\right),$$

where $\hat{\underline{\underline{\sigma}}}_t$ and $\hat{\boldsymbol{\mu}}_t$ were defined in Section 4.2.3. In the special case where $k = d$, we have the unique solution

$$\boldsymbol{\lambda}_t^* = \hat{\underline{\underline{\sigma}}}_t^{-1} \left(\hat{\boldsymbol{\mu}}_t - r_t \mathbf{1}\right).$$

4.4 OTHER USEFUL PROBABILITY MEASURES

4.4.1 General martingale measures

Suppose that $\mathbb{Q}$ is a risk-neutral probability measure and let $A_t = \exp\{\int_0^t r_s\, ds\}$ be the time t value of the bank account. According to (4.8) the price P_t of any asset with a single payment date satisfies the relation

$$\frac{P_t}{A_t} = \mathrm{E}_t^{\mathbb{Q}}\left[\frac{P_{t'}}{A_{t'}}\right]$$

for all $t' > t$ before the payment date of the asset, that is the relative price process (P_t/A_t) is a $\mathbb{Q}$-martingale. In a sense, we use the bank account as a numeraire. If the asset pays off P_T at time T, we can compute the time t price as

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[\frac{A_t}{A_T} P_T\right] = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s\, ds} P_T\right].$$

This involves the simultaneous risk-neutral distribution of $\int_t^T r_s \, ds$ and P_T, which might be quite complex.

For some assets we can simplify the computation of the price P_t by using a different, appropriately selected, numeraire asset. Let S_t denote the price process of a particular traded asset or the value process of a dynamic trading strategy. We require that $S_t > 0$. Can we find a probability measure $\mathbb{Q}^S$ so that the relative price process (P_t/S_t) is a $\mathbb{Q}^S$-martingale? Let us write the price dynamics of S_t and P_t as

$$dP_t = P_t \left[\mu_{Pt} \, dt + \boldsymbol{\sigma}_{Pt}^{\top} \, d\boldsymbol{z}_t\right], \qquad dS_t = S_t \left[\mu_{St} \, dt + \boldsymbol{\sigma}_{St}^{\top} \, d\boldsymbol{z}_t\right].$$

Then by Itô's Lemma, see Theorem 3.7 and Example 3.2,

$$d\left(\frac{P_t}{S_t}\right) = \frac{P_t}{S_t} \left[\left(\mu_{Pt} - \mu_{St} + \|\boldsymbol{\sigma}_{St}\|^2 - \boldsymbol{\sigma}_{St}^{\top}\boldsymbol{\sigma}_{Pt}\right) dt + \left(\boldsymbol{\sigma}_{Pt} - \boldsymbol{\sigma}_{St}\right)^{\top} \, d\boldsymbol{z}_t\right]. \tag{4.16}$$

When we change the probability measure, we change the drift rate. In order to obtain a martingale, we need to change the probability measure such that the drift becomes zero. Suppose we can find a well-behaved stochastic process $\boldsymbol{\lambda}_t^S$ such that

$$\left(\boldsymbol{\sigma}_{Pt} - \boldsymbol{\sigma}_{St}\right)^{\top} \boldsymbol{\lambda}_t^S = \mu_{Pt} - \mu_{St} + \|\boldsymbol{\sigma}_{St}\|^2 - \boldsymbol{\sigma}_{St}^{\top}\boldsymbol{\sigma}_{Pt}. \tag{4.17}$$

Then we can define a probability measure $\mathbb{Q}^S$ by the Radon–Nikodym derivative

$$\frac{d\mathbb{Q}^S}{d\mathbb{P}} = \exp\left\{-\frac{1}{2}\int_0^T \|\boldsymbol{\lambda}_t^S\|^2 \, ds - \int_0^T \left(\boldsymbol{\lambda}_t^S\right)^{\top} d\boldsymbol{z}_t\right\}.$$

The process $\boldsymbol{z}^S$ defined by

$$d\boldsymbol{z}_t^S = d\boldsymbol{z}_t + \boldsymbol{\lambda}_t^S \, dt, \quad \boldsymbol{z}_0^S = 0$$

is a standard Brownian motion under $\mathbb{Q}^S$. Substituting $d\boldsymbol{z}_t = d\boldsymbol{z}_t^S - \boldsymbol{\lambda}_t^S \, dt$ into (4.16) we get

$$d\left(\frac{P_t}{S_t}\right) = \frac{P_t}{S_t} \left(\boldsymbol{\sigma}_{Pt} - \boldsymbol{\sigma}_{St}\right)^{\top} \, d\boldsymbol{z}_t^S,$$

so that (P_t/S_t) indeed is a $\mathbb{Q}^S$-martingale.

How can we find a $\boldsymbol{\lambda}^S$ satisfying (4.17)? As we have seen, under weak conditions a market price of risk $\boldsymbol{\lambda}_t$ will exist with the property that $\mu_{Pt} = r_t + \boldsymbol{\sigma}_{Pt}^{\top}\boldsymbol{\lambda}_t$ and $\mu_{St} = r_t + \boldsymbol{\sigma}_{St}^{\top}\boldsymbol{\lambda}_t$. If we substitute in these relations and recall that $\|\boldsymbol{\sigma}_{St}\|^2 = \boldsymbol{\sigma}_{St}^{\top}\boldsymbol{\sigma}_{St}$, the right-hand side of (4.17) simplifies to $\left(\boldsymbol{\sigma}_{Pt} - \boldsymbol{\sigma}_{St}\right)^{\top}\left(\boldsymbol{\lambda}_t - \boldsymbol{\sigma}_{St}\right)$. We can therefore use

$$\boldsymbol{\lambda}_t^S = \boldsymbol{\lambda}_t - \boldsymbol{\sigma}_{St}.$$

In general we refer to such a probability measure $\mathbb{Q}^S$ as a **martingale measure** for the asset with price $S = (S_t)$. In particular, a risk-neutral probability measure $\mathbb{Q}$ is a martingale measure for the bank account. In some cases we might consider using a dividend-paying asset as the numeraire, for example, when valuing a derivative on the asset. This is no problem as long as the dividends of the asset before the maturity date of the derivative are excluded. For example, if you want to value an option on a coupon bond, an appropriate numeraire is the present value of the coupon payments that fall after the option expiration date. See the discussion in Björk (2009, Ch. 26).

Given a martingale measure $\mathbb{Q}^S$ for the asset with price S, the price P_t of an asset with a single payment P_T at time T satisfies

$$P_t = S_t \, \mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{P_T}{S_T}\right]. \tag{4.18}$$

In situations where the distribution of P_T/S_T under the measure $\mathbb{Q}^S$ is relatively simple, this provides a computationally convenient way of stating the price P_t relative to S_t. In the following subsection we look at an important example that will be applied extensively in later chapters. Other specific martingale measures will be introduced and discussed later, for example in Section 6.5.2.

4.4.2 The forward martingale measures

For the pricing of derivative securities that only provide a payoff at a single time T, it is typically convenient to use the zero-coupon bond maturing at time T as the numeraire. Recall that the price at time $t \leq T$ of this bond is denoted by B_t^T and that $B_T^T = 1$. Let $\boldsymbol{\sigma}_t^T$ denote the sensitivity vector of B_t^T so that

$$dB_t^T = B_t^T \left[\left(r_t + (\boldsymbol{\sigma}_t^T)^\top \boldsymbol{\lambda}_t \right) dt + (\boldsymbol{\sigma}_t^T)^\top d\boldsymbol{z}_t \right],$$

assuming the existence of a market price of risk process $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$.

We denote the martingale measure for the zero-coupon bond maturing at T by $\mathbb{Q}^T$ and refer to $\mathbb{Q}^T$ as the ***T*-forward martingale measure**. This type of martingale measure was introduced by Jamshidian (1987) and Geman (1989). The term comes from the fact that under this probability measure the forward price for delivery at time T of any security with no intermediate payments is a martingale, that is the expected change in the forward price is zero. If the price of the underlying asset is P_t, the forward price is P_t/B_t^T, and by definition of the $\mathbb{Q}^T$ measure this relative price is a $\mathbb{Q}^T$-martingale. The expectation under the T-forward martingale measure is sometimes called the expectation in a T-forward risk-neutral world.

The time t price of an asset paying P_T at time T can be computed as

$$P_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}[P_T]. \tag{4.19}$$

Under the probability measure $\mathbb{Q}^T$, the process $\boldsymbol{z}^T$ defined by

$$d\boldsymbol{z}_t^T = d\boldsymbol{z}_t + \left(\boldsymbol{\lambda}_t - \boldsymbol{\sigma}_t^T \right) dt, \qquad \boldsymbol{z}_0^T = 0, \tag{4.20}$$

is a standard Brownian motion according to Girsanov's theorem. In order to compute the price from (4.19) we only have to know (1) the current price of the zero-coupon bond that matures at the payment date of the asset, and (2) the distribution of the random payment of the asset under the T-forward martingale measure $\mathbb{Q}^T$.

We shall apply this pricing technique to derive prices of European options on zero-coupon bonds. The forward martingale measures are also important in the analysis of the so-called market models studied in Chapter 11.

Note that if the yield curve is constant and therefore flat (as in the famous Black–Scholes–Merton model for stock options), the bond price volatility $\boldsymbol{\sigma}_t^T$ is zero and, consequently, there is no difference between the risk-neutral probability measure and the T-forward martingale measure. The two measures differ only when interest rates are stochastic. The general difference is captured by the relation

$$dz_t^T = dz_t^{\mathbb{Q}} - \boldsymbol{\sigma}_t^T \, dt, \tag{4.21}$$

which follows from (4.13) and (4.20). To emphasize the difference between the risk-neutral measure and the forward martingale measures, the risk-neutral probability measure is sometimes referred to as the **spot martingale measure** since it is linked to the short rate or spot rate bank account.

4.5 COMPLETE VS. INCOMPLETE MARKETS

A financial market is said to be (dynamically) **complete** if all relevant risks can be hedged by forming portfolios of the traded financial assets. More formally, let $\mathcal{L}$ denote the set of all random variables (with finite variance) whose outcome can be determined from the exogenous shocks to the economy over the entire period $[0, T]$. In mathematical terms, $\mathcal{L}$ is the set of all random variables that are measurable with respect to the σ-algebra generated by the path of the Brownian motion z over $[0, T]$. On the other hand, let $\mathcal{M}$ denote the set of possible time T values that can be generated by forming self-financing trading strategies in the financial market, that is

$$\mathcal{M} = \left\{ V_T^{\alpha,\boldsymbol{\theta}} \mid (\alpha, \boldsymbol{\theta}) \text{ self-financing with } V_t^{\alpha,\boldsymbol{\theta}} \text{ bounded from below for all } t \in [0, T] \right\}.$$

Of course, for any trading strategy $(\alpha, \boldsymbol{\theta})$ the terminal value $V_T^{\alpha,\boldsymbol{\theta}}$ is a random variable, whose outcome is not determined until time T. Due to the technical conditions imposed on trading strategies, the terminal value will have finite variance, so $\mathcal{M}$ is always a subset of $\mathcal{L}$. If, in fact, $\mathcal{M}$ is equal to $\mathcal{L}$, the financial market is said to be complete. If not, it is said to be incomplete.

In a complete market, any random variable of interest to the investors can be replicated by a trading strategy, that is for any random variable W we can find a self-financing trading strategy with terminal value $V_T^{\alpha,\boldsymbol{\theta}} = W$. Consequently, an investor can obtain exactly her desired exposure to any of the d exogenous shocks.

Intuitively, to have a complete market, sufficiently many financial assets must be traded. However, the assets must also be sufficiently different in terms of their response to the exogenous shocks. After all, we cannot hedge more risk with two perfectly correlated assets than with just one of these assets. Market completeness is therefore closely related to the sensitivity matrix process $\underline{\underline{\sigma}}$ of the traded assets. The following theorem provides the precise relation:

Theorem 4.6 *Suppose that the short-term interest rate r is bounded. Also, suppose that a bounded market price of risk process* $\boldsymbol{\lambda}$ *exists. Then the financial market is complete if and only if the rank of* $\underline{\underline{\sigma}}_t$ *is equal to d (almost everywhere).*

If $N < d$, the matrix $\underline{\underline{\sigma}}_t$ cannot have rank d, so a necessary (but not sufficient) condition for the market to be complete is that at least d risky assets are traded. If $\underline{\underline{\sigma}}_t$ has rank d, then there is exactly one solution to the system of equations (4.14) and, hence, exactly one market price of risk process, namely $\boldsymbol{\lambda}^*$, and (if $\boldsymbol{\lambda}^*$ is sufficiently nice) exactly one risk-neutral probability measure. If the rank of $\underline{\underline{\sigma}}_t$ is strictly less than d, there will be multiple solutions to (4.14) and therefore multiple market prices of risk and multiple risk-neutral probability measures. Combining these observations with the previous theorem, we have the following conclusion:

Theorem 4.7 *Suppose that the short-term interest rate r is bounded and that the market is complete. Then there is a unique market price of risk process* $\boldsymbol{\lambda}$ *and, if* $\boldsymbol{\lambda}$ *satisfies technical conditions, there is a unique risk-neutral probability measure.*

This theorem and Theorem 4.4 together imply that in a complete market, under technical conditions, we have a unique state-price deflator.

A one-period financial market with finitely many, say n, states will be complete when n sufficiently different assets are traded. Just think of the famous binomial model for stock option pricing, where the risk-free asset and the stock make the market complete and leave the stock option redundant and easy to price by absence of arbitrage. When the uncertainty is generated by a d-dimensional standard Brownian motion, there are infinitely many sample paths and thus an infinite state space. It may seem surprising that the market can be complete with only an instantaneously risk-free asset and d risky assets having a price sensitivity matrix $\underline{\underline{\sigma}}_t$ of rank d. While the formal proof of this result is pretty complicated (see Harrison and Pliska (1981, 1983) and Duffie (2001)), the following observations provide the intuition:

- For continuous changes over an instant, only means and variances matter.
- We can approximate the d-dimensional shock dz_t by a random variable that takes on $d+1$ possible values and has the same mean and variance as dz_t.
- For example, a one-dimensional shock dz_t has mean zero and variance dt. This is also true for a random variable ε which equals $\sqrt{dt}$ with a probability of 1/2 and equals $-\sqrt{dt}$ with a probability of 1/2.
- The order one and two moments of a two-dimensional standard Brownian motion $z = (z_1, z_2)$ are $\mathrm{E}[dz_{it}] = 0$, $\mathrm{Var}[dz_{it}] = dt$, $\mathrm{Cov}[dz_{1t}, dz_{2t}] = 0$. These moments are shared by a random variable $(\varepsilon_1, \varepsilon_2)$ with just three states:

$$\begin{aligned}
&\varepsilon_1 = \frac{\sqrt{3\,dt}}{\sqrt{2}} \text{ and } \varepsilon_2 = \frac{\sqrt{dt}}{\sqrt{2}} && \text{with prob. } 1/3\\
&\varepsilon_1 = 0 \text{ and } \varepsilon_2 = -\sqrt{2\,dt} && \text{with prob. } 1/3\\
&\varepsilon_1 = -\frac{\sqrt{3\,dt}}{\sqrt{2}} \text{ and } \varepsilon_2 = \frac{\sqrt{dt}}{\sqrt{2}} && \text{with prob. } 1/3
\end{aligned}$$

- With continuous trading, we can adjust our exposure to the exogenous shocks every instant.

Over each instant we can thus think of the model with uncertainty generated by a d-dimensional standard Brownian motion as a one-period model with $d+1$ states. Therefore it only takes $d+1$ sufficiently different assets to complete the market.

Real financial markets are probably not complete in a broad sense, since most investors face restrictions on the trading strategies they can invest in, for example short-selling and portfolio mix restrictions. Moreover, investors are exposed to risks that cannot be fully hedged by any financial investments, for example labour income risk. An example of an incomplete market is a market where the traded assets are only sensitive to $k < d$ of the d exogenous shocks. Decomposing the d-dimensional standard Brownian motion z into $(Z, \hat{Z})$, where Z is k-dimensional and $\hat{Z}$ is $(d-k)$-dimensional, the dynamics of the traded risky assets can be written as

$$d\boldsymbol{P}_t = \operatorname{diag}(\boldsymbol{P}_t)\left[\boldsymbol{\mu}_t\, dt + \underline{\underline{\sigma}}_t\, d\boldsymbol{Z}_t\right].$$

For example, the dynamics of r_t, $\boldsymbol{\mu}_t$, or $\underline{\underline{\sigma}}_t$ may be affected by the non-traded risks $\hat{Z}$, representing non-hedgeable risk in interest rates, expected returns, and volatilities and correlations, respectively. Or other variables important for the investor, for example his labour income, may be sensitive to $\hat{Z}$. Let us assume for simplicity that $k = N$ and the $k \times k$ matrix $\underline{\underline{\sigma}}_t$ is non-singular. Then we can define a unique market price of risk associated with the traded risks by the k-dimensional vector

$$\Lambda_t = \left(\underline{\underline{\sigma}}_t\right)^{-1}\left(\boldsymbol{\mu}_t - r_t\boldsymbol{1}\right),$$

but for any well-behaved $(d-k)$-dimensional process $\hat{\Lambda}$, the process $\lambda = (\Lambda, \hat{\Lambda})$ will be a market price of risk for all risks. Each choice of $\hat{\Lambda}$ generates a valid market price of risk process and hence a valid risk-neutral probability measure and a valid state-price deflator.

4.6 EQUILIBRIUM AND REPRESENTATIVE AGENTS IN COMPLETE MARKETS

An economy consists of agents and assets. Each agent is characterized by her preferences (utility function) and endowments (initial wealth and future income). An equilibrium for an economy consists of a set of prices for all assets and a feasible trading strategy for each agent such that

1. given the asset prices, each agent has chosen an optimal trading strategy according to her preferences and endowments,
2. markets clear, that is total demand equals total supply for each asset.

To an equilibrium corresponds an equilibrium consumption process for each agent as a result of her endowments and her trading strategy. Clearly, an equilibrium set of prices cannot admit arbitrage.

As shown in Section 4.3, the absence of arbitrage (and some technical conditions) imply that the optimal consumption process for any agent defines a state-price deflator. Assuming time-additive preferences, the state-price deflator associated to agent l is the process $\zeta^l = (\zeta^l_t)$ defined by

$$\zeta^l_t = e^{-\delta^l t} \frac{u'_l(c^l_t)}{u'_l(c^l_0)},$$

where u_l is the utility function, δ^l the time preference rate, and $c^l = (c^l_t)$ the optimal consumption process of agent l.

In general the state-price deflators associated with different agents may differ, but in complete markets there is a unique state-price deflator. Consequently, all the state-price deflators associated with the different agents must be identical. In particular, for any agents k and l and any state ω, we must have that

$$\zeta_t(\omega) = e^{-\delta^k t} \frac{u'_k(c^k_t(\omega))}{u'_k(c^k_0)} = e^{-\delta^l t} \frac{u'_l(c^l_t(\omega))}{u'_l(c^l_0)}.$$

The agents trade until their marginal rates of substitution are perfectly aligned. This is known as **efficient risk-sharing**. In a complete market equilibrium we cannot have $\zeta^k_t(\omega) > \zeta^l_t(\omega)$, because agents k and l will then be able to make a trade that makes both better off. Any such trade is feasible in a complete market, but not necessarily in an incomplete market. In an incomplete market it may thus be impossible to completely align the marginal rates of substitution of the different agents.

Suppose that aggregate consumption at time t is higher in state ω than in state ω'. Then there must be at least one agent, say agent l, who consumes more at time t in state ω than in state ω', $c^l_t(\omega) > c^l_t(\omega')$. Consequently, $u'_l(c^l_t(\omega)) < u'_l(c^l_t(\omega'))$. Let k denote any other agent. If the market is complete we will have that

$$\frac{u'_k(c^k_t(\omega))}{u'_k(c^k_t(\omega'))} = \frac{u'_l(c^l_t(\omega))}{u'_l(c^l_t(\omega'))},$$

for any two states ω, ω'. Consequently, $u'_k(c^k_t(\omega)) < u'_k(c^k_t(\omega'))$ and thus $c^k_t(\omega) > c^k_t(\omega')$ for any agent k. It follows that in a complete market, the optimal consumption of any agent is an increasing function of the aggregate consumption level. Individuals' consumption levels move together.

A consumption allocation is called **Pareto-optimal** if the aggregate endowment cannot be allocated to consumption in another way that leaves all agents at least as well off and some agent strictly better off. An important result is the **First Welfare Theorem**:

Theorem 4.8 *If the financial market is complete, then every equilibrium consumption allocation is Pareto-optimal.*

The intuition is that if it was possible to reallocate consumption so that no agent was worse off and some agent was strictly better off, then the agents would generate such a reallocation by trading the financial assets appropriately. When the market is complete, an appropriate transaction can always be found, which is not necessarily the case in incomplete markets.

Both for theoretical and practical applications it is very cumbersome to deal with the individual utility functions and optimal consumption plans of many different agents. It would be much simpler if we could just consider a single agent. So we want to set up a single-agent economy in which equilibrium asset prices are the same as in the more realistic multi-agent economy. Such a single agent is called a **representative agent**. Like any agent, a representative agent is defined through her preferences and endowments, so the question is under what conditions and how we can construct preferences and endowments for such an agent. Clearly, the endowment of the single agent should be equal to the total endowments of all the individuals in the multi-agent economy. Hence, the main issue is how to define the preferences of the agent so that she is representative. The next theorem states that this can be done whenever the market is complete.

Theorem 4.9 *Suppose all individuals are greedy and risk-averse. If the financial market is complete, the economy has a representative agent.*

When the market is complete, we must look for preferences such that the associated marginal rate of substitution evaluated at the aggregate endowments is equal to the unique state-price deflator. If all agents have identical preferences, then we can use the same preferences for a representative agent. If individual agents have different preferences, the preferences of the representative agent will be some appropriately weighted average of the preferences of the individuals. We will not go into the details here, but refer the interested reader to Duffie (2001) and Munk (2010). Note that in the representative agent economy there can be no trade in the financial assets (who should be the other party in the trade?), and the consumption of the representative agent must equal the aggregate endowment or aggregate consumption in the multi-agent economy. In Chapter 5 we will use these results to link interest rates to aggregate consumption.

4.7 EXTENSION TO INTERMEDIATE DIVIDENDS

Up to now we have assumed that the assets provide a final dividend payment at time T and no dividend payments before. Clearly, we need to extend this to the case of dividends at other dates. We distinguish between lump-sum dividends and continuous dividends. A lump-sum dividend is a payment at a single point in time, whereas a continuous dividend is paid over a period of time.

Suppose $\mathbb{Q}$ is a risk-neutral probability measure. Consider an asset paying only a **lump-sum dividend** of $L_{t'}$ at time $t' < T$. If we invest the dividend in the bank account over the period $[t', T]$, we end up with a value of $L_{t'} \exp\{\int_{t'}^{T} r_u \, du\}$. Thinking of this as a terminal dividend, the value of the asset at time $t < t'$ must be

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(L_{t'}e^{\int_{t'}^T r_u\,du}\right)\right] = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'} r_u\,du}L_{t'}\right].$$

Intermediate lump-sum dividends are therefore valued similarly to terminal dividends and the discounted price process of such an asset will be a $\mathbb{Q}$-martingale over the period $[0,t']$ where the asset 'lives'. An important example is that of a zero-coupon bond paying one at some future date t'. The price at time $t < t'$ of such a bond is given by

$$B_t^{t'} = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'} r_u\,du}\right]. \tag{4.22}$$

In terms of a state-price deflator ζ, we have

$$B_t^{t'} = \mathrm{E}_t\left[\frac{\zeta_{t'}}{\zeta_t}\right].$$

A **continuous dividend** is represented by a dividend rate process $D = (D_t)$, which means that the total dividend paid over any period $[t,t']$ is equal to $\int_t^{t'} D_u\,du$. Over a very short interval $[s, s+ds]$ the total dividend paid is approximately $D_s\,ds$. Investing this in the bank account provides a time T value of $e^{\int_s^T r_u\,du}D_s\,ds$. Integrating up the time T values of all the dividends in the period $[t,T]$, we get a terminal value of $\int_t^T e^{\int_s^T r_u\,du}D_s\,ds$. According to the previous sections the time t value of such a terminal payment is

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(\int_t^T e^{\int_s^T r_u\,du}D_s\,ds\right)\right] = \mathrm{E}_t^{\mathbb{Q}}\left[\int_t^T e^{-\int_t^s r_u\,du}D_s\,ds\right].$$

This implies that for any $t < t' < T$, we have

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'} r_u\,du}P_{t'} + \int_t^{t'} e^{-\int_t^s r_u\,du}D_s\,ds\right] \tag{4.23}$$

and the process with time t value given by $P_t\exp\{-\int_0^t r_u\,du\} + \int_0^t \exp\{-\int_0^s r_u\,du\}D_s\,ds$ is a $\mathbb{Q}$-martingale. In terms of a state-price deflator ζ we have that the process with time t value $\zeta_t P_t + \int_0^t \zeta_s D_s\,ds$ is a $\mathbb{P}$-martingale and

$$P_t = \mathrm{E}_t\left[\frac{\zeta_{t'}}{\zeta_t}P_{t'} + \int_t^{t'}\frac{\zeta_s}{\zeta_t}D_s\,ds\right].$$

In the special case where the payment rate is proportional to the value of the security, that is $D_s = q_s P_s$, it can be shown that

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'}[r_u - q_u]\,du}P_{t'}\right]. \tag{4.24}$$

Pricing expressions for assets that have both continuous and lump-sum dividends can be obtained by combining the expressions above appropriately.

The inclusion of intermediate dividends does not change the link between state-price deflators and the marginal rate of substitution of an agent. We still have the result that $\zeta_t = e^{-\delta t} u'(c_t)/u'(c_0)$ is a valid state-price deflator.

4.8 DIFFUSION MODELS AND THE FUNDAMENTAL PARTIAL DIFFERENTIAL EQUATION

Many financial models assume the existence of one or several so-called **state variables**, that is variables whose current values contain all the relevant information about the economy. Of course, the relevance of information depends on the purpose of the model. Generally, the price of an asset depends on the dynamics of the short-term interest rate, the market prices of relevant risks, and on the distribution of the payoff(s) of the asset. In models with a single state variable we denote the time t value of the state variable by x_t, while in models with several state variables we gather their time t values in the vector $\boldsymbol{x}_t$. By assumption, the current values of the state variables are sufficient information for the pricing and hedging of fixed income securities. In particular, historical values of the state variables, $\boldsymbol{x}_s$ for $s < t$, are irrelevant. It is therefore natural to model the evolution of $\boldsymbol{x}_t$ by a diffusion process since we know that such processes have the Markov property, see Section 3.4. We will refer to models of this type as **diffusion models**. We will first consider diffusion models with a single state variable, which are naturally termed one-factor diffusion models. Afterwards, we shall briefly discuss how the results obtained for one-factor models can be extended to multi-factor models, that is models with several state variables.

4.8.1 One-factor diffusion models

We assume that a single, one-dimensional, state variable contains all the relevant information. The possible values of x_t lie in a set $\mathcal{S} \subseteq \mathbb{R}$. We assume that $x = (x_t)_{t \geq 0}$ is a diffusion process with dynamics given by the stochastic differential equation

$$dx_t = \alpha(x_t, t)\, dt + \beta(x_t, t)\, dz_t,$$

where z is a one-dimensional standard Brownian motion, and α and β are 'well-behaved' functions with values in $\mathbb{R}$. Given a market price of risk $\lambda_t = \lambda(x_t, t)$, we can use (4.13) to write the dynamics of the state variable under the risk-neutral probability measure as

$$dx_t = [\alpha(x_t, t) - \beta(x_t, t)\lambda(x_t, t)]\, dt + \beta(x_t, t)\, dz_t^{\mathbb{Q}}. \tag{4.25}$$

We also assume that the short interest rate depends at most on x and t, that is $r_t = r(x_t, t)$.

Consider a security with a single payment of H_T at time T. We know that the price of the security satisfies $P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} H_T\right]$. Assuming that $H_T = H(x_T, T)$, we can rewrite the price as $P_t = P(x_t, t)$, where

$$P(x,t) = \mathrm{E}_{x,t}^{\mathbb{Q}}\left[e^{-\int_t^T r(x_u,u)\,du} H(x_T, T)\right]$$

and we have exploited the Markov property of (x_t) to write the expectation as a function of the current value of the process. Here $\mathrm{E}_{x,t}^{\mathbb{Q}}$ denotes the expectation given that $x_t = x$. It follows from Itô's Lemma, Theorem 3.6, that the dynamics of $P_t = P(x_t, t)$ is

$$dP_t = P_t\left[\mu(x_t,t)\,dt + \sigma(x_t,t)\,dz_t\right],$$

where the functions μ and σ are defined by

$$\begin{aligned}
\mu(x,t)P(x,t) &= \frac{\partial P}{\partial t}(x,t) + \frac{\partial P}{\partial x}(x,t)\alpha(x,t) + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}(x,t)\beta(x,t)^2, \\
\sigma(x,t)P(x,t) &= \frac{\partial P}{\partial x}(x,t)\beta(x,t).
\end{aligned}$$

We also know that for a market price of risk $\lambda(x_t, t)$, we have

$$\mu(x_t,t) = r(x_t,t) + \sigma(x_t,t)\lambda(x_t,t)$$

for all possible values of x_t and hence

$$\mu(x,t)P(x,t) = r(x,t)P(x,t) + \sigma(x,t)P(x,t)\lambda(x_t,t)$$

for all (x,t). Substituting in μ and σ and rearranging, we arrive at a partial differential equation (PDE) as stated in the following theorem.

Theorem 4.10 *The function P defined by*

$$P(x,t) = \mathrm{E}_{x,t}^{\mathbb{Q}}\left[e^{-\int_t^T r(x_u,u)\,du} H(x_T, T)\right]$$

satisfies the partial differential equation

$$\begin{aligned}
&\frac{\partial P}{\partial t}(x,t) + \left(\alpha(x,t) - \beta(x,t)\lambda(x,t)\right)\frac{\partial P}{\partial x}(x,t) \\
&\quad + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) - r(x,t)P(x,t) = 0, \qquad (x,t) \in \mathcal{S} \times [0,T), \qquad (4.26)
\end{aligned}$$

together with the terminal condition

$$P(x,T) = H(x,T), \qquad x \in \mathcal{S}.$$

The relation between expectations and partial differential equations is generally known as the Feynman–Kac theorem, see (Øksendal, 2003, Thm. 8.2.1). Note that the coefficient of the $\partial P/\partial x$ in the PDE is identical to the risk-neutral drift of the state variable, see (4.25), so that

$$\frac{\partial P}{\partial t} + (\alpha - \beta\lambda)\frac{\partial P}{\partial x} + \frac{1}{2}\beta^2\frac{\partial^2 P}{\partial x^2}$$

is the risk-neutral drift according to Itô's Lemma. The risk-neutral drift has to be identical to rP to avoid arbitrage, and that is exactly what the PDE shows. Also note that the prices of *all* securities with no payments before T solve the same PDE. However, the terminal conditions and thereby also the solutions depend on the payoff characteristics of the securities.

Using the price of a traded asset as the state variable

When the state variable itself is the price of a traded asset, the market price of risk disappears from the pricing PDE. The expected rate of return (corresponding to μ) of this asset is $\alpha(x,t)/x$, and the volatility (corresponding to σ) is $\beta(x,t)/x$. Since Equation (4.14) in particular must hold for this asset, we have that

$$\lambda(x,t) = \frac{\frac{\alpha(x,t)}{x} - r(x,t)}{\frac{\beta(x,t)}{x}} = \frac{\alpha(x,t) - r(x,t)x}{\beta(x,t)}$$
$$\Rightarrow \quad \alpha(x,t) - \beta(x,t)\lambda(x,t) = r(x,t)x.$$

By insertion of this expression, the PDE (4.26) reduces to

$$\frac{\partial P}{\partial t}(x,t) + r(x,t)\left(x\frac{\partial P}{\partial x}(x,t) - P(x,t)\right) + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) = 0,$$
$$(x,t) \in \mathcal{S} \times [0,T).$$

Since no knowledge of the market price of risk is necessary, assets with price of the form $P(x_t,t)$ are in this case priced by pure no-arbitrage arguments. The securities which can be priced in this way are exactly the redundant securities, and the prices will be given relative to the price x_t.

This approach has proved successful in the pricing of stock options with the prime example being the Black–Scholes–Merton model developed by Black and Scholes (1973) and Merton (1973). The model assumes that the risk-free interest rate r (continuously compounded) is constant over time and that the price S_t of the underlying asset follows a continuous stochastic process with a constant relative volatility, that is

$$dS_t = \mu(S_t,t)\,dt + \sigma S_t\,dz_t,$$

where σ is a constant and μ is a 'nice' function.[3] Furthermore, we assume that the underlying asset has no payments in the life of the derivative security. The time t price P_t of a derivative asset is then given by $P_t = P(S_t,t)$ where

[3] It is often assumed that $\mu(S_t,t) = \mu S_t$ for a constant parameter μ, but that is not necessary. However, we must require that the function μ is such that the value space for the price process will be $\mathcal{S} = \mathbb{R}_+$.

$$P(S,t) = \mathrm{E}^{\mathbb{Q}}_{S,t}\left[e^{-\int_t^T r\,du} H(S_T,T)\right] = e^{-r[T-t]}\,\mathrm{E}^{\mathbb{Q}}_{S,t}\left[H(S_T,T)\right]$$

and the risk-neutral dynamics of the underlying asset price is

$$dS_t = rS_t\,dt + \sigma S_t\,dz^{\mathbb{Q}}_t,$$

that is a geometric Brownian motion so that S_T is lognormally distributed. The function $P(S,t)$ solves the PDE

$$\frac{\partial P}{\partial t}(S,t) + rS\frac{\partial P}{\partial S}(S,t) + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 P}{\partial S^2}(S,t) = rP(S,t), \quad (S,t) \in \mathcal{S} \times [0,T), \tag{4.27}$$

with the terminal condition $P(S,T) = H(S,T)$, for all $S \in \mathcal{S}$. For a European call option with an exercise price of K the payoff function is given by $H(S,T) = \max(S-K,0)$. The price $C_t = C(S_t,t)$ can then be found either by solving the PDE (4.27) with the relevant terminal condition or by calculating the discounted risk-neutral expected payoff, that is

$$C(S_t,t) = e^{-r[T-t]}\,\mathrm{E}^{\mathbb{Q}}_{S_t,t}\left[\max(S_T-K,0)\right].$$

Applying Theorem A.4 in Appendix A, the latter approach immediately gives the famous Black–Scholes–Merton formula for the price of a European call option on a stock:[4]

$$C(S_t,t) = S_t N\left(d_1(S_t,t)\right) - Ke^{-r[T-t]} N\left(d_2(S_t,t)\right), \tag{4.28}$$

where

$$d_1(S_t,t) = \frac{\ln(S_t/K) + r[T-t]}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t},$$

$$d_2(S_t,t) = \frac{\ln(S_t/K) + r[T-t]}{\sigma\sqrt{T-t}} - \frac{1}{2}\sigma\sqrt{T-t} = d_1(S_t,t) - \sigma\sqrt{T-t}.$$

[4] According to Abramowitz and Stegun (1972), the cumulative distribution function $N(\cdot)$ of the standard normal distribution can be approximated with six-digit accuracy as follows:

$$N(x) \approx 1 - n(x)\left(a_1 b(x) + a_2 b(x)^2 + a_3 b(x)^3 + a_4 b(x)^4 + a_5 b(x)^5\right), \quad x \geq 0,$$

where $n(x) = e^{-x^2/2}/\sqrt{2\pi}$ is the probability density function, $b(x) = 1/(1+cx)$, and the constants are given by

$$c = 0.2316419, \qquad a_1 = 0.31938153,$$
$$a_2 = -0.356563782, \qquad a_3 = 1.781477937,$$
$$a_4 = -1.821255978, \qquad a_5 = 1.330274429.$$

For $x < 0$, we can use the relation $N(x) = 1 - N(-x)$, where $N(-x)$ can be computed using the approximation above.

It can be verified that the function $C(S, t)$ defined in (4.28) solves the PDE (4.27) with the relevant terminal condition. Similarly, for a European put option the price is

$$\pi(S_t, t) = Ke^{-r[T-t]}N\left(-d_2(S_t, t)\right) - S_t N\left(-d_1(S_t, t)\right).$$

Practitioners often apply slightly modified versions of the Black–Scholes–Merton model and option pricing formula to price derivatives other than stock options, including many fixed income securities. These modifications are often based on Black (1976) who adapted the Black–Scholes–Merton setting to the pricing of European options on commodity futures. However, it is inappropriate to price interest rate derivatives just by modelling the dynamics of the underlying security. Consistent pricing of fixed income securities must be based on the evolution of the entire term structure of interest rates. Broadly speaking, the entire term structure is the 'underlying asset' for all fixed income securities.

Hedging

In a model with a single one-dimensional state variable, a locally risk-free portfolio can be constructed from any two securities. In other words, the bank account can be replicated by a suitable trading strategy of any two securities. Conversely, it is possible to replicate any risky asset by a suitable trading strategy of the bank account and any other risky asset. To replicate asset 1 by a portfolio of the bank account and asset 2, the portfolio must at any point in time consist of

$$\theta_t = \frac{\frac{\partial P_1}{\partial x}(x_t, t)}{\frac{\partial P_2}{\partial x}(x_t, t)} = \frac{\sigma_1(x_t, t)P_1(x_t, t)}{\sigma_2(x_t, t)P_2(x_t, t)}$$

units of asset 2, plus

$$\alpha_t = \left(1 - \frac{\sigma_1(x_t, t)}{\sigma_2(x_t, t)}\right) P_1(x_t, t)$$

invested in the bank account. Then indeed the time t value of the portfolio is

$$\begin{aligned}
\Pi_t &\equiv \alpha_t + \theta_t P_2(x_t, t) \\
&= \left(1 - \frac{\sigma_1(x_t, t)}{\sigma_2(x_t, t)}\right) P_1(x_t, t) + \frac{\sigma_1(x_t, t)}{\sigma_2(x_t, t)} P_1(x_t, t) \\
&= P_1(x_t, t),
\end{aligned}$$

and the dynamics of the portfolio value is

$$
\begin{aligned}
d\Pi_t &= \alpha_t r(x_t, t)\, dt + \theta_t\, dP_2(x_t, t) \\
&= r(x_t, t)\left(1 - \frac{\sigma_1(x_t, t)}{\sigma_2(x_t, t)}\right) P_1(x_t, t)\, dt \\
&\quad + \frac{\sigma_1(x_t, t) P_1(x_t, t)}{\sigma_2(x_t, t) P_2(x_t, t)} \left(\mu_2(x_t, t) P_2(x_t, t)\, dt + \sigma_2(x_t, t) P_2(x_t, t)\, dz_t\right) \\
&= \left(r(x_t, t) + \frac{\sigma_1(x_t, t)}{\sigma_2(x_t, t)} \left(\mu_2(x_t, t) - r(x_t, t)\right)\right) P_1(x_t, t)\, dt \\
&\quad + \sigma_1(x_t, t) P_1(x_t, t)\, dz_t \\
&= \left(r(x_t, t) + \sigma_1(x_t, t)\lambda(x_t, t)\right) P_1(x_t, t)\, dt + \sigma_1(x_t, t) P_1(x_t, t)\, dz_t \\
&= \mu_1(x_t, t) P_1(x_t, t)\, dt + \sigma_1(x_t, t) P_1(x_t, t)\, dz_t \\
&= dP_{1t},
\end{aligned}
$$

so that the trading strategy replicates asset 1. In particular, in one-factor term structure models any fixed income security can be replicated by a portfolio of the bank account and any other fixed income security. We will discuss hedging issues in more detail in Chapter 12.

If the state variable x_t itself is the price of a traded asset, the considerations above imply that any derivative asset can be replicated by a trading strategy that at time t consists of $\frac{\partial P}{\partial x}(x_t, t)$ units of the underlying asset and an appropriate position in the bank account.

Securities with several payment dates

Many financial securities have more than one payment date, for example coupon bonds, swaps, caps, and floors. Theorem 4.10 does not directly apply to such securities. In the extension to securities with several payments, we distinguish again between securities with discrete lump-sum payments and securities with a continuous stream of payments.

First consider a security with discrete lump-sum payments, which are either deterministic or depend on the value of the state variable at the payment date. Suppose that the security provides payments $H_j(x_{T_j})$ at time T_j for $j = 1, \dots, N$ with $T_1 < \dots < T_n$. Clearly, at the time of a payment the value of the security will drop exactly by the payment. The 'ex-payment' value will equal the 'cum-payment' value minus the size of the payment. Letting $t+$ denote 'immediately after time t', we can express this relation as

$$P(x, T_j+) = P(x, T_j) - H_j(x).$$

If the drop in the price $-[P(x, T_j+) - P(x, T_j)]$ was less than the payment $H_j(x)$, an arbitrage profit could be locked in by buying the security immediately before the time of payment and selling it again immediately after the payment was received. Between payment dates, that is in the intervals (T_j, T_{j+1}), the price of the security will satisfy the PDE (4.26). Alternatively, we can apply Theorem 4.10 in order to separately find the current value of each of the payments after which the value of the security follows from a simple summation.

Next consider a security providing continuous payments at the rate $h_t = h(x_t, t)$ throughout $[0, T]$ and a terminal lump-sum payment of $H_T = H(x_T, T)$. From (4.23) we know that the price of such a security in our diffusion setting is given by

$$P(x_t, t) = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r(x_u,u)\,du} H(x_T, T) + \int_t^T e^{-\int_t^s r(x_u,u)\,du} h(x_s, s)\,ds\right].$$

Theorem 4.10 can be extended to show that the function P in this case will solve the PDE

$$\begin{aligned}&\frac{\partial P}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial P}{\partial x}(x,t)\\ &\quad + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) - r(x,t)P(x,t) + h(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T),\end{aligned}$$

with the terminal condition $P(x, T) = H(x, T)$ for all $x \in \mathcal{S}$. The only change in the PDE relative to the case with no intermediate dividends is the addition of the term $h(x, t)$ on the left-hand side of the equation.

In the special case where the payment rate is proportional to the value of the security, that is $h(x, t) = q(x, t)P(x, t)$, we know from (4.24) that the price can be written as

$$P(x_t, t) = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T [r(x_u,u) - q(x_u,u)]\,du} H(x_T, T)\right].$$

The relevant PDE is now

$$\begin{aligned}&\frac{\partial P}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial P}{\partial x}(x,t)\\ &\quad + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) - \big(r(x,t) - q(x,t)\big)P(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T).\end{aligned} \tag{4.29}$$

4.8.2 Multi-factor diffusion models

Assume now that the short-term interest rate, the market prices of risk, and the payoffs we want to price depend on n state variables $x_1, \dots, x_n$ and that the vector $\boldsymbol{x} = (x_1, \dots, x_n)^{\top}$ follows the stochastic process

$$d\boldsymbol{x}_t = \boldsymbol{\alpha}(\boldsymbol{x}_t, t)\,dt + \underline{\underline{\beta}}(\boldsymbol{x}_t, t)\,d\boldsymbol{z}_t, \tag{4.30}$$

where $\boldsymbol{z}$ is an n-dimensional standard Brownian motion. We can write (4.30) componentwise as

$$dx_{it} = \alpha_i(\boldsymbol{x}_t, t)\, dt + \boldsymbol{\beta}_i(\boldsymbol{x}_t, t)^\top\, d\boldsymbol{z}_t = \alpha_i(\boldsymbol{x}_t, t)\, dt + \sum_{j=1}^{n} \beta_{ij}(\boldsymbol{x}_t, t)\, dz_{jt}.$$

The volatility of the i'th state variable is the standard deviation

$$\|\boldsymbol{\beta}_i(\boldsymbol{x}_t, t)\| = \sqrt{\sum_{k=1}^{n} \beta_{ik}(\boldsymbol{x}_t, t)^2},$$

and the instantaneous correlation between changes in the i'th and the j'th state variable is

$$\rho_{ij}(\boldsymbol{x}_t, t) = \frac{\mathrm{Cov}_t[dx_{it}, dx_{jt}]}{\sqrt{\mathrm{Var}_t[dx_{it}]}\sqrt{\mathrm{Var}_t[dx_{jt}]}} = \frac{\sum_{k=1}^{n} \beta_{ik}(\boldsymbol{x}_t, t)\beta_{jk}(\boldsymbol{x}_t, t)}{\|\boldsymbol{\beta}_i(\boldsymbol{x}_t, t)\|\, \|\boldsymbol{\beta}_j(\boldsymbol{x}_t, t)\|}.$$

Consider again a security with a single payment of $H_T = H(\boldsymbol{x}_T, T)$ at time T. Its price is $P_t = P(\boldsymbol{x}_t, t)$, where

$$P(\boldsymbol{x}, t) = \mathrm{E}^{\mathbb{Q}}_{\boldsymbol{x},t}\left[e^{-\int_t^T r(\boldsymbol{x}_u, u)\, du} H(\boldsymbol{x}_T, T)\right].$$

It follows from the multi-dimensional version of Itô's Lemma, Theorem 3.9, that the dynamics of P_t are

$$\frac{dP_t}{P_t} = \mu(\boldsymbol{x}_t, t)\, dt + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\, dz_{jt},$$

where the functions μ and σ_j are defined as

$$\begin{aligned}
\mu(\boldsymbol{x}, t)P(\boldsymbol{x}, t) &= \frac{\partial P}{\partial t}(\boldsymbol{x}, t) + \sum_{j=1}^{n} \frac{\partial P}{\partial x_j}(\boldsymbol{x}, t)\alpha_j(\boldsymbol{x}, t) \\
&\quad + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 P}{\partial x_j \partial x_k}(\boldsymbol{x}, t)\rho_{jk}(\boldsymbol{x}, t)\, \|\boldsymbol{\beta}_j(\boldsymbol{x}, t)\|\, \|\boldsymbol{\beta}_k(\boldsymbol{x}, t)\|, \\
\sigma_j(\boldsymbol{x}, t)P(\boldsymbol{x}, t) &= \sum_{k=1}^{n} \frac{\partial P}{\partial x_k}(\boldsymbol{x}, t)\beta_{kj}(\boldsymbol{x}, t).
\end{aligned}$$

We also know that for a market price of risk $\boldsymbol{\lambda}(\boldsymbol{x}_t, t)$, we have

$$\mu(\boldsymbol{x}_t, t) = r(\boldsymbol{x}_t, t) + \boldsymbol{\sigma}(\boldsymbol{x}_t, t)^\top \boldsymbol{\lambda}(\boldsymbol{x}_t, t) = r(\boldsymbol{x}_t, t) + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\lambda_j(\boldsymbol{x}_t, t). \tag{4.31}$$

Substituting in μ and σ, we arrive at the PDE

$$\begin{aligned}
&\frac{\partial P}{\partial t}(x,t) + \sum_{j=1}^{n}\left(\alpha_j(x,t) - \sum_{k=1}^{n}\beta_{jk}(x,t)\lambda_k(x,t)\right)\frac{\partial P}{\partial x_j}(x,t) \\
&\quad + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\rho_{jk}(x,t)\,\|\boldsymbol{\beta}_j(x,t)\|\,\|\boldsymbol{\beta}_k(x,t)\|\frac{\partial^2 P}{\partial x_j \partial x_k}(x,t) - r(x,t)P(x,t) = 0, \\
&\quad (x,t) \in \mathcal{S} \times [0,T),
\end{aligned}$$

with the obvious terminal condition $P(\boldsymbol{x},T) = H(\boldsymbol{x},T)$, $\boldsymbol{x} \in \mathcal{S}$.

Using matrix notation the PDE can be written more compactly as

$$\begin{aligned}
&\frac{\partial P}{\partial t}(x,t) + \left(\boldsymbol{\alpha}(x,t) - \underline{\underline{\beta}}(x,t)\boldsymbol{\lambda}(x,t)\right)^{\top}\frac{\partial P}{\partial \boldsymbol{x}}(x,t) \\
&\quad + \frac{1}{2}\operatorname{tr}\left(\underline{\underline{\beta}}(x,t)\underline{\underline{\beta}}(x,t)^{\top}\frac{\partial^2 P}{\partial \boldsymbol{x}^2}(x,t)\right) - r(x,t)P(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T),
\end{aligned}$$

where $\partial P/\partial \boldsymbol{x}$ is the vector of first-order derivatives $\partial P/\partial x_j$, $\partial^2 P/\partial \boldsymbol{x}^2$ is the $n \times n$ matrix of second-order derivatives $\partial^2 P/\partial x_i \partial x_j$, and $\operatorname{tr}(\underline{\underline{M}})$ denotes the 'trace' of the matrix $\underline{\underline{M}}$, which is defined as the sum of the diagonal elements, $\operatorname{tr}(\underline{\underline{M}}) = \sum_j M_{jj}$.

In a model with n state variables the bank account can be replicated by a suitably constructed trading strategy in $n+1$ (sufficiently different) securities. Conversely, any security can be replicated by a suitably constructed trading strategy in the bank account and n other (sufficiently different) securities. For securities with more than one payment date the analysis must be modified similarly to the one-dimensional case.

4.9 CONCLUDING REMARKS

This chapter has reviewed the central results of modern asset pricing theory in a continuous-time framework. Ignoring technicalities, we can summarize our main findings as follows:

- The market-wide pricing principles can be represented by three equivalent objects: state-price deflators, risk-neutral probability measures, and market prices of risk. These objects are closely related to individuals' marginal rates of substitution.
- A specification of a state-price deflator, a risk-neutral probability measure, or a market price of risk fixes the prices of all traded assets with given dividends.
- The absence of arbitrage is equivalent to the existence of a state-price deflator, a risk-neutral probability measure, and a market price of risk.

- In a complete and arbitrage-free market, there is a unique state-price deflator, a unique risk-neutral probability measure, and a unique market price of risk.
- In a complete market, a representative agent exists and the unique state-price deflator is that agent's marginal rate of substitution evaluated at the aggregate consumption process.

4.10 EXERCISES

Exercise 4.1 Show that if there is no arbitrage and the short rate can never go negative, then the discount function is non-increasing and all forward rates are non-negative.

Exercise 4.2 Show Equation (4.11).

Exercise 4.3 Consider an asset paying a dividend of D_T at time T and no other dividends. You want to price the asset at time $t < T$. Assume that $D_T = \mathrm{E}_t[D_T]e^X$ for a normally distributed random variable $X \sim N(-\frac{1}{2}\sigma_X^2, \sigma_X^2)$, so that $\mathrm{E}[e^X] = 1$. Also assume that the state-price deflator satisfies $\zeta_T = \zeta_t e^Y$, where X and Y are jointly normally distributed with correlation ρ and $Y \sim N(\mu_Y, \sigma_Y^2)$. Show that the time t price of the asset is

$$P_t = \mathrm{E}_t[D_T] \exp\left\{\mu_Y + \frac{1}{2}\sigma_Y^2 + \rho\sigma_X\sigma_Y\right\}.$$

Exercise 4.4 Suppose the market is complete and $\zeta = (\zeta_t)$ is the unique state-price deflator. Then the present value (the costs) of any consumption process $c = (c_t)_{t\in[0,T]}$ is $\mathrm{E}[\int_0^T \zeta_t c_t \, dt]$. For an agent with time-additive preferences, an initial wealth of W_0 and no future income except from her financial transactions, the utility-maximization problem can then be formulated as

$$\max_{c=(c_t)_{t\in[0,T]}} \mathrm{E}\left[\int_0^T e^{-\delta t} u(c_t)\, dt\right] \qquad \text{s.t. } \mathrm{E}\left[\int_0^T \zeta_t c_t \, dt\right] \leq W_0.$$

Use the Lagrangian technique for constrained optimization to show that the optimal consumption process must satisfy

$$e^{-\delta t} u'(c_t) = \alpha \zeta_t, \quad t \in [0, T],$$

where α is a Lagrange multiplier. Explain why you can conclude that $\zeta_t = e^{-\delta t} u'(c_t)/u'(c_0)$.

5

The Economics of the Term Structure of Interest Rates

5.1 INTRODUCTION

A bond is a standardized and transferable loan agreement between two parties. The issuer of the bond is borrowing money from the holder of the bond and promises to pay back the loan according to a predefined payment scheme. The existence of the bond market allows individuals to trade consumption opportunities at different points in time among each other. An individual who has a clear preference for current capital to finance investments or current consumption can borrow by issuing a bond to an individual who has a clear preference for future consumption opportunities. The price of a bond of a given maturity is, of course, set to align the demand and supply of that bond, and will consequently depend on the attractiveness of the real investment opportunities and on the individuals' preferences for consumption over the maturity of the bond. The term structure of interest rates will reflect these dependencies. In Sections 5.2 and 5.3 we derive relations between equilibrium interest rates and aggregate consumption and production in settings with a representative agent. In Section 5.4 we give some examples of equilibrium term structure models that are derived from the basic relations between interest rates, consumption, and production.

Since agents are (or, at least, should be) concerned with the number of units of goods they consume and not the dollar value of these goods, the relations found in the first part of this chapter apply to real interest rates. However, most traded bonds are nominal, that is they promise the delivery of certain dollar amounts, not the delivery of a certain number of consumption goods. The real value of a nominal bond depends on the evolution of the price of the consumption good. In Section 5.5 we explore the relations between real rates, nominal rates, and inflation. We consider both the case where money has no real effects on the economy and the case where money does affect the real economy.

The development of arbitrage-free dynamic models of the term structure was initiated in the 1970s. Until then, the discussions among economists about the shape of the term structure were based on some relatively imprecise hypotheses. The most well-known is the expectation hypothesis which postulates a close relation between current interest rates or bond yields and expected future interest rates or bond returns. Many economists still seem to rely on the validity of this hypothesis, and a lot of manpower has been spent on testing the hypothesis empirically. In Section 5.6, we review several versions of the expectation hypothesis and discuss the consistency of these versions. We argue that neither of these versions will hold for any reasonable dynamic term structure model. Some alternative traditional hypotheses are briefly reviewed in Section 5.7.

5.2 REAL INTEREST RATES AND AGGREGATE CONSUMPTION

In order to study the link between interest rates and aggregate consumption, we assume the existence of a representative agent with time-additive expected utility $\mathrm{E}[\int_0^T e^{-\delta t} u(C_t)\,dt]$. As discussed in Section 4.6, a representative agent will exist in a complete market. The parameter δ is the subjective time preference rate with higher δ representing a more impatient agent. C_t is the consumption rate of the agent, which is then also the aggregate consumption level in the economy. In terms of the utility and time preference of the representative agent the state price deflator is therefore characterized by

$$\zeta_t = e^{-\delta t} \frac{u'(C_t)}{u'(C_0)},$$

see the discussions leading to (4.7).

Assume that the aggregate consumption process $C = (C_t)$ has dynamics of the form

$$dC_t = C_t \left[\mu_{Ct}\,dt + \boldsymbol{\sigma}_{Ct}^{\top}\,d\boldsymbol{z}_t \right], \tag{5.1}$$

where $\boldsymbol{z} = (\boldsymbol{z}_t)$ is a (possibly multi-dimensional) standard Brownian motion. The dynamics of the state-price deflator will then follow from Itô's Lemma, Theorem 3.6, applied to the function $g(C,t) = e^{-\delta t} u'(C)/u'(C_0)$. Since the relevant derivatives are

$$\frac{\partial g}{\partial t} = -\delta g(C,t), \qquad \frac{\partial g}{\partial C} = e^{-\delta t} \frac{u''(C)}{u'(C_0)} = \frac{u''(C)}{u'(C)} g(C,t),$$

$$\frac{\partial^2 g}{\partial C^2} = e^{-\delta t} \frac{u'''(C)}{u'(C_0)} = \frac{u'''(C)}{u'(C)} g(C,t),$$

the dynamics of $\zeta = (\zeta_t)$ is

$$\begin{aligned} d\zeta_t = -\zeta_t \Bigg[&\left(\delta + \left(\frac{-C_t u''(C_t)}{u'(C_t)} \right) \mu_{Ct} - \frac{1}{2} C_t^2 \frac{u'''(C_t)}{u'(C_t)} \|\boldsymbol{\sigma}_{Ct}\|^2 \right) dt \\ &+ \left(\frac{-C_t u''(C_t)}{u'(C_t)} \right) \boldsymbol{\sigma}_{Ct}^{\top}\,d\boldsymbol{z}_t \Bigg]. \end{aligned}$$

We can compare this with the general dynamics of a state-price deflator in (4.15). The market price of risk is therefore given by

$$\boldsymbol{\lambda}_t = \left(\frac{-C_t u''(C_t)}{u'(C_t)} \right) \boldsymbol{\sigma}_{Ct}, \tag{5.2}$$

and the short rate is

$$r_t = \delta + \frac{-C_t u''(C_t)}{u'(C_t)} \mu_{Ct} - \frac{1}{2} C_t^2 \frac{u'''(C_t)}{u'(C_t)} \|\boldsymbol{\sigma}_{Ct}\|^2. \tag{5.3}$$

This is the interest rate at which the market for short-term borrowing and lending will clear. The equation relates the equilibrium short-term interest rate to the time preference rate and the expected growth rate μ_{Ct} and the variance rate $\|\boldsymbol{\sigma}_{Ct}\|^2$ of aggregate consumption growth over the next instant. We can observe the following relations:

- There is a positive relation between the time preference rate and the equilibrium interest rate. The intuition is that when the agents of the economy are impatient and have a high demand for current consumption, the equilibrium interest rate must be high in order to encourage the agents to save now and postpone consumption.
- The multiplier of μ_{Ct} in (5.3) is the relative risk aversion of the representative agent, which is positive. Hence, there is a positive relation between the expected growth in aggregate consumption and the equilibrium interest rate. This can be explained as follows: we expect higher future consumption and hence lower future marginal utility, so postponed payments due to saving have lower value. Consequently, a higher return on saving is needed to maintain market clearing.
- If u''' is positive, there will be a negative relation between the variance of aggregate consumption and the equilibrium interest rate. If the representative agent has decreasing absolute risk aversion, which is certainly a reasonable assumption, u''' has to be positive. The intuition is that the greater the uncertainty about future consumption, the more the agents will appreciate the sure payments from the risk-free asset and, hence, the lower a return necessary to clear the market for borrowing and lending.

In the special case of constant relative risk aversion, $u(c) = c^{1-\gamma}/(1-\gamma)$, Equation (5.3) simplifies to

$$r_t = \delta + \gamma \mu_{Ct} - \frac{1}{2} \gamma (1+\gamma) \|\boldsymbol{\sigma}_{Ct}\|^2. \tag{5.4}$$

In particular, we see that if the drift and variance rates of aggregate consumption are constant, that is aggregate consumption follows a geometric Brownian motion, then the short-term interest rate will be constant over time. Consequently, the yield curve will be flat and constant over time. This is clearly an unrealistic case. To obtain interesting models we must either allow for variations in the expectation and the variance of aggregate consumption growth or allow for non-constant relative risk aversion (or both).

What can we say about the relation between the equilibrium yield curve and the expectations and uncertainty about future aggregate consumption? Given the consumption dynamics in (5.1), we have

$$\frac{C_T}{C_t} = \exp\left\{ \int_t^T \left(\mu_{Cs} - \frac{1}{2} \|\boldsymbol{\sigma}_{Cs}\|^2 \right) ds + \int_t^T \boldsymbol{\sigma}_{Cs}^\top \, d\boldsymbol{z}_s \right\}.$$

Assuming that the consumption sensitivity $\boldsymbol{\sigma}_{Cs}$ is constant and that the drift rate is such that $\int_t^T \mu_{Cs}\, ds$ is normally distributed, we see that C_T/C_t is lognormally distributed. Assuming time-additive power utility, the state-price deflator $\zeta_T/\zeta_t = e^{-\delta(T-t)}(C_T/C_t)^{-\gamma}$ will then also be lognormally distributed. Consequently, the price of the zero-coupon bond maturing at time T is given by

$$\begin{aligned} B_t^T &= \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\right] = e^{-\delta(T-t)}\,\mathrm{E}_t\left[e^{-\gamma\ln(C_T/C_t)}\right] \\ &= \exp\left\{-\delta(T-t) - \gamma\,\mathrm{E}_t\left[\ln\left(\frac{C_T}{C_t}\right)\right] + \frac{1}{2}\gamma^2\,\mathrm{Var}_t\left[\ln\left(\frac{C_T}{C_t}\right)\right]\right\}, \end{aligned}$$

where we have applied Theorem A.2 in Appendix A. Combining the above expression with (1.3), the continuously compounded zero-coupon rate or yield y_t^T for maturity T is

$$y_t^T = \delta + \gamma\frac{\mathrm{E}_t[\ln(C_T/C_t)]}{T-t} - \frac{1}{2}\gamma^2\frac{\mathrm{Var}_t[\ln(C_T/C_t)]}{T-t}.$$

Since

$$\ln \mathrm{E}_t\left[\frac{C_T}{C_t}\right] = \mathrm{E}_t\left[\ln\left(\frac{C_T}{C_t}\right)\right] + \frac{1}{2}\,\mathrm{Var}_t\left[\ln\left(\frac{C_T}{C_t}\right)\right],$$

the zero-coupon rate can be rewritten as

$$y_t^T = \delta + \gamma\frac{\ln \mathrm{E}_t[C_T/C_t]}{T-t} - \frac{1}{2}\gamma(1+\gamma)\frac{\mathrm{Var}_t[\ln(C_T/C_t)]}{T-t}, \tag{5.5}$$

which is very similar to the short-rate equation (5.4). The yield is increasing in the subjective rate of time preference. The equilibrium yield for the period $[t, T]$ is positively related to the expected growth rate of aggregate consumption over the period and negatively related to the uncertainty about the growth rate of consumption over the period. The intuition for these results is the same as for the short-term interest rate discussed above. We see that the shape of the equilibrium time t yield curve $T \mapsto y_t^T$ is determined by how expectations and variances of consumption growth rates depend on the length of the forecast period. For example, if the economy is expected to enter a short period of high growth rates, real short-term interest rates tend to be high and the yield curve downward-sloping.

Equation (5.5) is based on a lognormal future consumption and power utility. We will discuss such a setting in more detail in Section 7.4. It appears impossible to obtain an exact relation of the same structure as (5.5) for a more general model, where future consumption is not necessarily lognormal and preferences are different from power utility. However, following Breeden (1986), we can derive an approximate relation of a similar form. The equilibrium time t price of a zero-coupon bond paying one consumption unit at time $T \geq t$ is given by

$$B_t^T = \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\right] = e^{-\delta(T-t)}\frac{\mathrm{E}_t\left[u'(C_T)\right]}{u'(C_t)}, \tag{5.6}$$

where C_T is the uncertain future aggregate consumption level. We can write the left-hand side of the equation above in terms of the yield y_t^T of the bond as

$$B_t^T = e^{-y_t^T(T-t)} \approx 1 - y_t^T(T-t),$$

using a first-order Taylor expansion. Turning to the right-hand side of the equation, we will use a second-order Taylor expansion of $u'(C_T)$ around C_t:

$$u'(C_T) \approx u'(C_t) + u''(C_t)(C_T - C_t) + \frac{1}{2}u'''(C_t)(C_T - C_t)^2.$$

This approximation is reasonable when C_T stays relatively close to C_t, which is the case for fairly low and smooth consumption growth and fairly short time horizons. Applying the approximation, the right-hand side of (5.6) becomes

$$\begin{aligned}
e^{-\delta(T-t)} \frac{\mathrm{E}_t\left[u'(C_T)\right]}{u'(C_t)} &\approx e^{-\delta(T-t)} \left(1 + \frac{u''(C_t)}{u'(C_t)}\,\mathrm{E}_t[C_T - C_t]\right. \\
&\qquad \left. + \frac{1}{2}\frac{u'''(C_t)}{u'(C_t)}\,\mathrm{E}_t\left[(C_T - C_t)^2\right]\right) \\
&\approx 1 - \delta(T-t) + e^{-\delta(T-t)} \frac{C_t u''(C_t)}{u'(C_t)}\,\mathrm{E}_t\left[\frac{C_T}{C_t} - 1\right] \\
&\qquad + \frac{1}{2} e^{-\delta(T-t)} C_t^2 \frac{u'''(C_t)}{u'(C_t)}\,\mathrm{E}_t\left[\left(\frac{C_T}{C_t} - 1\right)^2\right],
\end{aligned}$$

where we have used the approximation $e^{-\delta(T-t)} \approx 1 - \delta(T-t)$. Substituting the approximations of both sides into (5.6) and rearranging, we find the following approximate expression for the zero-coupon yield:

$$\begin{aligned}
y_t^T &\approx \delta + e^{-\delta(T-t)} \left(\frac{-C_t u''(C_t)}{u'(C_t)}\right) \frac{\mathrm{E}_t\left[C_T/C_t - 1\right]}{T-t} \\
&\quad - \frac{1}{2} e^{-\delta(T-t)} C_t^2 \frac{u'''(C_t)}{u'(C_t)} \frac{\mathrm{E}_t\left[\left(\frac{C_T}{C_t} - 1\right)^2\right]}{T-t}.
\end{aligned}$$

We can replace $\mathrm{E}_t\left[(C_T/C_t - 1)^2\right]$ by $\mathrm{Var}_t\left[C_T/C_t\right] + (\mathrm{E}_t\left[C_T/C_t - 1\right])^2$ in order to consider the effect of a shift in variance for a fixed expected consumption growth. Again assuming $u' > 0$, $u'' < 0$, and $u''' > 0$, we see that the yield of a given maturity is positively related to the expected growth rate of consumption up to the maturity date and negatively related to the variance of the consumption growth rate up to maturity.

In the simple consumption-based asset pricing model, the representative agent is assumed to have a constant relative risk aversion γ, and aggregate consumption is assumed to have a constant expected growth rate μ_C and a constant variance σ_C^2. The real short rate is then

$$r_t = \delta + \gamma\mu_C - \frac{1}{2}\gamma(1+\gamma)\sigma_C^2, \tag{5.7}$$

which is constant and implies a constant, flat yield curve as already noted. The interest rate level predicted by the model is also unrealistic. In U.S. data over the period 1929–98 the average annual growth rate of aggregate consumption was 1.8% and the volatility of consumption growth was 2.9%, see Bansal and Yaron (2004). Reasonable preference parameters are $\delta = 0.02$ and $\gamma = 4$. Substituting these values into (5.7), the real risk-free rate should have been around 8.36%, far above the historical average of around 1%. The observation that the simple consumption-based asset pricing model predicts too high a risk-free rate is referred to as the *risk-free rate puzzle* and was first pointed out by Weil (1989).[1] Of course, you may disagree about the appropriate values of δ and γ or mistrust the available data on aggregate consumption, see the discussions in Breeden et al. (1989), Grossman et al. (1987), and Wilcox (1992).

Numerous extensions of the simple model have been suggested in the recent asset pricing literature. If you stick to a representative agent setting, you can either generalize the preferences or the aggregate consumption process or both. Suppose we keep constant relative risk aversion, but allow the expected consumption growth rate μ_{Ct} to vary over time. Then we can see the effect on the real short rate from (5.4). If μ_{Ct} is just 1% above [below] the average, the short rate will be γ% above [below] the average. If γ is 4 or higher, this will involve unrealistically large variations in the short rate. This suggests that it is necessary to go beyond the standard assumption of time-additive power utility.

A recent model by Bansal and Yaron (2004) has been very successful in matching stylized asset pricing facts. The representative agent is assumed to have so-called Epstein–Zin preferences. In addition to a time preference rate δ and a relative risk aversion γ, these preferences involve a third parameter ψ, which can be interpreted as an intertemporal elasticity of substitution, that is a measure of the agent's willingness to move consumption over time. The standard time-additive power utility preferences is the special case where $\psi = 1/\gamma$, but there is no reason to believe that the risk aversion and the intertemporal elasticity of substitution should be linked in that specific way. Under the assumption of constant μ_C and σ_C^2, the equilibrium short rate will then be of the form

$$r_t = \delta + \frac{\mu_C}{\psi} - \frac{1}{2}\left(1 - \gamma + \gamma^2 + \frac{\gamma}{\psi}\right)\sigma_C^2.$$

The presence of the new parameter ψ makes it easier to match the observed level of real interest rates. Even with the Epstein–Zin preferences, interest rates will be constant when μ_C and σ_C^2 are constant. Bansal and Yaron (2004) introduce a so-called long-run risk model for aggregate consumption in which the expected growth rate μ_{Ct} is assumed to have a slowly varying (hence, 'long-run') stochastic component, and they also allow consumption variance σ_{Ct}^2 to vary stochastically over time. The risk-free rate will then be of the form

$$r_t = \delta + k_1 + \frac{\mu_{Ct}}{\psi} - \frac{1}{2}k_2\sigma_{Ct}^2,$$

[1] For a reasonable level of risk aversion the simple model predicts an excess expected return on a broad stock index which is much lower than what has been observed historically. This is the so-called *equity premium puzzle* first identified by Mehra and Prescott (1985). For further details, see Cochrane (2005) or Munk (2010).

where k_1 and k_2 are constants determined from the preference parameters and parameters in the dynamics of μ_{Ct} and σ_{Ct}. Consequently, the short rate is no longer constant. The expected growth rate is now multiplied by $1/\psi$, which is probably considerably smaller than γ and, hence, the volatility of the risk-free rate is kept at a reasonable low level. Some implications of the model for the shape and variations of the yield curve are discussed by Wu (2008) and Bansal and Shaliastovich (2009). The combination of Epstein–Zin preferences and the long-run risk dynamics has proven successful in matching historical moments of interest rates and risk premia and thus in explaining various asset pricing puzzles.

As discussed in Chapter 4, the representative agent framework demands market completeness, and real-life financial markets are not complete, for example, individual labour income risk is not fully hedgeable. However, it is extremely difficult to derive closed-form equilibria for incomplete market models. Various numerical studies (for example Telmer (1993) and Aiyagari (1994)) and a few studies obtaining closed-form solutions (for example Constantinides and Duffie (1996) and Christensen et al. (2010)) have shown that when individuals face idiosyncratic income shocks they will self-insure by increasing their precautionary savings in the risk-free asset. Consequently, the equilibrium risk-free rate will be lower.

5.3 REAL INTEREST RATES AND AGGREGATE PRODUCTION

In order to study the relation between interest rates and production, we will look at a slightly simplified version of the general equilibrium model of Cox et al. (1985a).

Consider an economy with a single physical good that can be used either for consumption or investment. All values are expressed in units of this good. The instantaneous rate of return on an investment in the production of the good is

$$\frac{d\eta_t}{\eta_t} = g(x_t)\,dt + \xi(x_t)\,dz_{1t},$$

where z_1 is a standard one-dimensional Brownian motion, and g and ξ are well-behaved real-valued functions (given by Mother Nature) of some state variable x_t. To be more specific, η_0 goods invested in the production process at time 0 will grow to η_t goods at time t if the output of the production process is continuously reinvested in this period. We can interpret g as the expected real growth rate of the economy and the volatility ξ (assumed positive for all x) as a measure of the uncertainty about the growth rate of the economy. The production process has constant returns to scale in the sense that the distribution of the rate of return is independent of the scale of the investment. There is free entry to the production process. We can think of individuals investing in production directly by forming their own firm or indirectly by investing in stocks of production firms. For simplicity we take the first interpretation. All producers, individuals, and firms act competitively so that firms have zero profits and just pass production returns on to their owners. All individuals and firms act as price takers.

We assume that the state variable is one-dimensional and evolves according to the stochastic differential equation

$$dx_t = m(x_t)\,dt + v_1(x_t)\,dz_{1t} + v_2(x_t)\,dz_{2t},$$

where z_2 is another standard one-dimensional Brownian motion independent of z_1, and m, v_1, and v_2 are well-behaved real-valued functions. The instantaneous variance rate of the state variable is $v_1(x)^2 + v_2(x)^2$, the covariance rate of the state variable and the real growth rate is $\xi(x)v_1(x)$ so that the correlation between the state and the growth rate is $v_1(x)/\sqrt{v_1(x)^2 + v_2(x)^2}$. Unless $v_2 \equiv 0$, the state variable is imperfectly correlated with the real production returns. If v_1 is positive [negative], then the state variable is positively [negatively] correlated with the growth rate of the economy (since ξ is assumed positive). Since the state determines the expected returns and the variance of returns on real investments, we may think of x_t as a productivity or technology variable.

In addition to the investment in the production process, we assume that the agents have access to a financial asset with a price P_t with dynamics of the form

$$\frac{dP_t}{P_t} = \mu_t\,dt + \sigma_{1t}\,dz_{1t} + \sigma_{2t}\,dz_{2t}.$$

As a part of the equilibrium we will determine the relation between the expected return μ_t and the volatility coefficients σ_{1t} and σ_{2t}. Finally, the agents can borrow and lend funds at an instantaneously risk-free interest rate r_t, which is also determined in equilibrium. The market is therefore complete. Other financial assets affected by z_1 and z_2 may be traded, but they will be redundant. We will get the same equilibrium relation between expected returns and volatility coefficients for these other assets as for the one modelled explicitly. For simplicity we stick to the case with a single financial asset.

If an agent at each time t consumes at a rate of $c_t \geq 0$, invests a fraction α_t of his wealth in the production process, invests a fraction π_t of wealth in the financial asset, and invests the remaining fraction $1 - \alpha_t - \pi_t$ of wealth in the risk-free asset, his wealth W_t will evolve as

$$\begin{aligned} dW_t = &\left\{r_t W_t + W_t\alpha_t\left(g(x_t) - r_t\right) + W_t\pi_t\left(\mu_t - r_t\right) - c_t\right\}\,dt \\ &+ W_t\alpha_t\xi(x_t)\,dz_{1t} + W_t\pi_t\sigma_{1t}\,dz_{1t} + W_t\pi_t\sigma_{2t}\,dz_{2t}. \end{aligned}$$

Since a negative real investment is physically impossible, we should restrict α_t to the non-negative numbers. However, we will assume that this constraint is not binding. Let us look at an agent maximizing expected utility of future consumption. The indirect utility function is defined as

$$J(W,x,t) = \sup_{(\alpha_s,\pi_s,c_s)_{s\in[t,T]}} \mathrm{E}_t\left[\int_t^T e^{-\delta(s-t)}u(c_s)\,ds\right],$$

where 'sup' is short for supremum, which for most purposes is equivalent to the maximum. The indirect utility is thus the maximal expected utility the agent can obtain given his current wealth and the current value of the state variable. Applying dynamic programming techniques, it can be shown that the optimal choice of α and π satisfies

$$\alpha^* = \frac{-J_W}{WJ_{WW}}\left[(g-r)\frac{\sigma_1^2+\sigma_2^2}{\xi^2\sigma_2^2} - (\mu - r)\frac{\sigma_1}{\xi\sigma_2^2}\right] + \frac{-J_{Wx}}{WJ_{WW}}\frac{\sigma_2\nu_1 - \sigma_1\nu_2}{\xi\sigma_2}, \tag{5.8}$$

$$\pi^* = \frac{-J_W}{WJ_{WW}}\left[-\frac{\sigma_1}{\xi\sigma_2^2}(g-r) + \frac{1}{\sigma_2^2}(\mu - r)\right] + \frac{-J_{Wx}}{WJ_{WW}}\frac{\nu_2}{\sigma_2}. \tag{5.9}$$

In equilibrium, prices and interest rates are such that (a) all agents act optimally and (b) all markets clear. In particular, summing up the positions of all agents in the financial asset we should get zero, and the total amount borrowed by agents on a short-term basis should equal the total amount lent by agents. Since the available production apparatus is to be held by some investors, summing the optimal α's over investors we should get 1. Since we have assumed a complete market, we can construct a representative agent, that is an agent with a given utility function so that the equilibrium interest rates and price processes are the same in the single agent economy as in the larger multi-agent economy. Alternatively, we may think of the case where all agents in the economy are identical so that they will have the same indirect utility function and always make the same consumption and investment choices.

In an equilibrium, we have $\pi^* = 0$ for the representative agent, and hence (5.9) implies that

$$\mu - r = \frac{\sigma_1}{\xi}(g-r) - \frac{\left(\frac{-J_{Wx}}{WJ_{WW}}\right)}{\left(\frac{-J_W}{WJ_{WW}}\right)}\sigma_2\nu_2. \tag{5.10}$$

Substituting this into the expression for α^* and using the fact that $\alpha^* = 1$ in equilibrium, we get that

$$\begin{aligned} 1 &= \left(\frac{-J_W}{WJ_{WW}}\right)\left[(g-r)\frac{\sigma_1^2+\sigma_2^2}{\xi^2\sigma_2^2} - \frac{\sigma_1}{\xi}\frac{\sigma_1}{\xi\sigma_2^2}(g-r) + \frac{\left(\frac{-J_{Wx}}{WJ_{WW}}\right)}{\left(\frac{-J_W}{WJ_{WW}}\right)}\sigma_2\nu_2\frac{\sigma_1}{\xi\sigma_2^2}\right] \\ &\quad + \left(\frac{-J_{Wx}}{WJ_{WW}}\right)\frac{\sigma_2\nu_1 - \sigma_1\nu_2}{\xi\sigma_2} \\ &= \left(\frac{-J_W}{WJ_{WW}}\right)\frac{g-r}{\xi^2} + \left(\frac{-J_{Wx}}{WJ_{WW}}\right)\frac{\nu_1}{\xi}. \end{aligned}$$

Consequently, the equilibrium short-term interest rate can be written as

$$r = g - \left(\frac{-WJ_{WW}}{J_W}\right)\xi^2 + \frac{J_{Wx}}{J_W}\xi\nu_1. \tag{5.11}$$

This equation ties the equilibrium real short-term interest rate to the production side of the economy. Let us address each of the three right-hand side terms:

- The equilibrium real interest rate r is positively related to the expected real growth rate g of the economy. The intuition is that for higher expected growth rates, the productive investments are more attractive relative to the risk-free investment, so to maintain market clearing the interest rate has to be higher.
- The term $-WJ_{WW}/J_W$ is the relative risk aversion of the representative agent's indirect utility, which is assumed to be positive. Hence, we see that the equilibrium real interest rate r is negatively related to the uncertainty about the growth rate of the economy, represented by the instantaneous variance ξ^2. For a higher uncertainty, the safe returns of a risk-free investment is relatively more attractive, so to establish market clearing the interest rate has to decrease.
- The last term in (5.11) is due to the presence of the state variable. The covariance rate of the state variable and the real growth rate of the economy is equal to $\xi\nu_1$. Suppose that high values of the state variable represent good states of the economy so that the marginal utility J_W is decreasing in x, that is $J_{Wx} < 0$. If the state variable and the growth rate of the economy are positively correlated, that is $\nu_1 > 0$, we see from (5.8) that the hedge demand of the productive investment is decreasing, and hence the demand for depositing money at the short rate increasing, in the magnitude of the correlation (both J_{Wx} and J_{WW} are negative). To maintain market clearing, the interest rate must be decreasing in the magnitude of the correlation as reflected by (5.11).

We see from (5.10) that the market prices of risk are given by

$$\lambda_1 = \frac{g-r}{\xi}, \quad \lambda_2 = -\frac{\left(\frac{-J_{Wx}}{WJ_{WW}}\right)}{\left(\frac{-J_W}{WJ_{WW}}\right)}\nu_2 = -\frac{J_{Wx}}{J_W}\nu_2. \tag{5.12}$$

From (5.11) we obtain

$$g - r = \left(\frac{-WJ_{WW}}{J_W}\right)\xi^2 - \frac{J_{Wx}}{J_W}\xi\nu_1,$$

so that we can rewrite λ_1 as

$$\lambda_1 = \left(\frac{-WJ_{WW}}{J_W}\right)\xi - \frac{J_{Wx}}{J_W}\nu_1. \tag{5.13}$$

5.4 EQUILIBRIUM TERM STRUCTURE MODELS

5.4.1 Production-based models

As a special case of their general equilibrium model with production, Cox et al. (1985b) consider a model where the representative agent is assumed to have a logarithmic utility so that the relative risk aversion of the direct utility function is 1. In addition, the agent is assumed to have an infinite time horizon, which implies that the indirect utility function will be independent of time. It can be shown that under these assumptions the indirect utility function of the agent is of the form $J(W,x) = A \ln W + B(x)$. In particular, $J_{Wx} = 0$ and the relative risk aversion of the indirect utility function is also 1. It follows from (5.11) that the equilibrium real short-term interest rate is equal to

$$r(x_t) = g(x_t) - \xi(x_t)^2.$$

The authors further assume that the expected rate of return and the variance rate of the return on the productive investment are both proportional to the state, that is

$$g(x) = k_1 x, \qquad \xi(x)^2 = k_2 x,$$

where $k_1 > k_2$. Then the equilibrium short-rate becomes $r(x) = (k_1 - k_2)x \equiv kx$. Assume now that the state variable follows a square-root process

$$\begin{aligned} dx_t &= \kappa\,(\bar{x} - x_t)\,dt + \rho\sigma_x\sqrt{x_t}\,dz_{1t} + \sqrt{1-\rho^2}\sigma_x\sqrt{x_t}\,dz_{2t} \\ &= \kappa\,(\bar{x} - x_t)\,dt + \sigma_x\sqrt{x_t}\,d\bar{z}_t, \end{aligned}$$

where $\bar{z}$ is a standard Brownian motion with correlation ρ with the standard Brownian motion z_1 and correlation $\sqrt{1-\rho^2}$ with z_2. Then the dynamics of the real short rate is $dr_t = k\,dx_t$, which yields

$$dr_t = \kappa\,(\bar{r} - r_t)\,dt + \sigma_r\sqrt{r_t}\,d\bar{z}_t,$$

where $\bar{r} = k\bar{x}$ and $\sigma_r = \sqrt{k}\sigma_x$. The market prices of risk given in (5.12) and (5.13) simplify to

$$\lambda_1 = \xi(x) = \sqrt{k_2 x} = \sqrt{k_2/k}\,\sqrt{r}, \qquad \lambda_2 = 0.$$

From Chapter 4 we know that the short-rate dynamics and the market prices of risk fully determine prices of all bonds and hence the entire term structure of interest rates. In fact, this is one of the most frequently used dynamic term structure models, the so-called CIR model, which we will discuss in much more detail in Section 7.5.

Longstaff and Schwartz (1992a) study a two-factor version of the production-based equilibrium model. They assume that the production returns are given by

$$\frac{d\eta_t}{\eta_t} = g(x_{1t}, x_{2t})\,dt + \xi(x_{2t})\,dz_{1t},$$

where

$$g(x_1, x_2) = k_1 x_1 + k_2 x_2, \qquad \xi(x_2)^2 = k_3 x_2,$$

so that the state variable x_2 affects both expected returns and uncertainty of production, while the state variable x_1 only affects the expected return. With log utility the short rate is again equal to the expected return minus the variance,

$$r(x_1, x_2) = g(x_1, x_2) - \xi(x_2)^2 = k_1 x_1 + (k_2 - k_3)x_2.$$

The state variables are assumed to follow independent square-root processes,

$$dx_{1t} = (\varphi_1 - \kappa_1 x_{1t})\, dt + \beta_1 \sqrt{x_{1t}}\, dz_{2t},$$
$$dx_{2t} = (\varphi_2 - \kappa_2 x_{2t})\, dt + \beta_2 \sqrt{x_{2t}}\, dz_{3t},$$

where z_2 is independent of z_1 and z_3, but z_1 and z_3 may be correlated. The market prices of risk associated with the Brownian motions are

$$\lambda_1(x_2) = \xi(x_2) = \sqrt{k_2}\sqrt{x_2}, \quad \lambda_2 = \lambda_3 = 0.$$

We will discuss the implications of this model in much more detail in Chapter 8.

5.4.2 Consumption-based models

Other authors take a consumption-based approach for developing models of the term structure of interest rates. For example, Goldstein and Zapatero (1996) present a simple model in which the equilibrium short-term interest rate is consistent with the term structure model of Vasicek (1977). They assume that aggregate consumption evolves as

$$dC_t = C_t\,[\mu_{Ct}\, dt + \sigma_C\, dz_t],$$

where z is a one-dimensional standard Brownian motion, σ_C is a constant, and the expected consumption growth rate μ_{Ct} follows an Ornstein-Uhlenbeck process

$$d\mu_{Ct} = \kappa\,(\bar{\mu}_C - \mu_{Ct})\, dt + \theta\, dz_t.$$

The representative agent is assumed to have a constant relative risk aversion of γ. It follows from (5.4) that the equilibrium real short-term interest rate is

$$r_t = \delta + \gamma\mu_{Ct} - \frac{1}{2}\gamma(1+\gamma)\sigma_C^2$$

with dynamics $dr_t = \gamma\, d\mu_{Ct}$, that is

$$dr_t = \kappa\,(\bar{r} - r_t)\, dt + \sigma_r\, dz_t,$$

where $\sigma_r = \gamma\theta$ and $\bar{r} = \gamma\bar{\mu}_C + \delta - \frac{1}{2}\gamma(1+\gamma)\sigma_C^2$. From (5.2), the market price of risk is given by

$$\lambda = \gamma\sigma_C,$$

which is constant. We will give a thorough treatment of this model in Section 7.4.

In fact, we can generate any of the so-called affine term structure models in this way. Assume that the expected growth rate and the variance rate of aggregate consumption are affine in some state variables, that is

$$\mu_{Ct} = a_0 + \sum_{i=1}^{n} a_i x_{it}, \qquad \|\boldsymbol{\sigma}_{Ct}\|^2 = b_0 + \sum_{i=1}^{n} b_i x_{it},$$

then the equilibrium short rate will be

$$r_t = \left(\delta + \gamma a_0 - \frac{1}{2}\gamma(1+\gamma)b_0\right) + \gamma \sum_{i=1}^{n} \left(a_i - \frac{1}{2}(1+\gamma)b_i\right) x_{it}.$$

Of course, we should have $b_0 + \sum_{i=1}^{n} b_i x_{it} \geq 0$ for all values of the state variables. The market price of risk is $\lambda_t = \gamma\sigma_{Ct}$. If the state variables x_i follow processes of the affine type, we have an affine term structure model. We will return to the affine models both in Chapter 7 and Chapter 8.

For other term structure models developed with the consumption-based approach, see, for example Bakshi and Chen (1997).

5.5 REAL AND NOMINAL INTEREST RATES AND TERM STRUCTURES

In this section we discuss the difference and relation between real interest rates and nominal interest rates. **Nominal interest rates** are related to investments in nominal bonds, which are bonds that promise given payments in a given currency, say dollars. The purchasing power of these payments are uncertain, however, since the future price level of consumer goods is uncertain. **Real interest rates** are related to investments in real bonds, which are bonds whose dollar payments are adjusted by the evolution in the consumer price index and effectively provide a given purchasing power at the payment dates.[2] Although most bond issuers and investors would probably reduce relevant risks by using real bonds rather than nominal bonds, the vast majority of bonds issued and traded at all exchanges are nominal bonds. Surprisingly few real bonds are traded. To the extent that people have preferences for consumption units only (and not for their monetary holdings) they should base their consumption and investment decisions on real interest rates rather than nominal interest rates. The relations between interest rates and consumption and production discussed in the previous sections apply to real interest rates.

In a world where traded bonds are nominal we can quite easily get a good picture of the term structure of nominal interest rates. But what about real interest rates? Traditionally, economists think of nominal rates as the sum of real rates and the expected (consumer price) inflation rate. This relation is often referred to as the Fisher hypothesis or Fisher relation in honor of Fisher (1907). However, neither empirical studies nor modern financial economics theories (as we shall see below) support the Fisher hypothesis.[3]

[2] Since not all consumers will want the same composition of different consumption goods as that reflected by the consumer price index, real bonds will not necessarily provide a perfectly certain purchasing power for each investor.

[3] Of course, at the end of any given period one can compute an ex-post real return by subtracting the realized inflation rate from an ex-post realized nominal return. It is not clear, however, why investors should care about such an ex-post real return.

In the following we shall first derive some generally valid relations between real rates, nominal rates, and inflation and investigate the differences between real and nominal asset prices. Then we will discuss two different types of models in which we can say more about real and nominal rates. The first setting follows the neoclassical tradition in assuming that monetary holdings do not affect the preferences of the agents so that the presence of money has no effects on real rates and real asset returns. Hence, the relations derived earlier in this chapter still apply. However, several empirical findings indicate that the existence of money does have real effects. For example, real stock returns are negatively correlated with inflation and positively correlated with growth in money supply. Also, assets that are positively correlated with inflation have a lower expected return.[4] In the second setting we consider below, money is allowed to have real effects. Economies with this property are called *monetary economies*.

5.5.1 Real and nominal asset pricing

As before, let $\zeta = (\zeta_t)$ denote a state-price deflator, which evolves over time according to

$$d\zeta_t = -\zeta_t \left[r_t \, dt + \boldsymbol{\lambda}_t^{\top} \, d\boldsymbol{z}_t \right],$$

where $r = (r_t)$ is the short-term real interest rate and $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$ is the market price of risk. Then the time t real price of a real zero-coupon bond maturing at time T is given by

$$B_t^T = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} \right].$$

If the real price $P = (P_t)$ of an asset follows the stochastic process

$$dP_t = P_t \left[\mu_{Pt} \, dt + \boldsymbol{\sigma}_{Pt}^{\top} \, d\boldsymbol{z}_t \right],$$

then, in periods where the asset does not pay dividends, the relation

$$\mu_{Pt} - r_t = \boldsymbol{\sigma}_{Pt}^{\top} \boldsymbol{\lambda}_t \tag{5.14}$$

must hold in equilibrium. From Chapter 4 we also know that we can characterize real prices in terms of the risk-neutral probability measure $\mathbb{Q}$, which is formally defined by the change-of-measure process

$$\xi_t \equiv \mathrm{E}_t \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \right] = \exp \left\{ -\frac{1}{2} \int_0^t \|\boldsymbol{\lambda}_s\|^2 \, ds - \int_0^t \boldsymbol{\lambda}_s^{\top} \, d\boldsymbol{z}_s \right\}.$$

The real price of an asset paying no dividends in the time interval (t, T) can then be written as

$$P_t = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} P_T \right] = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^T r_s \, ds} P_T \right].$$

[4] Such results are reported by, for example, Fama (1981), Fama and Gibbons (1982), Chen et al. (1986), and Marshall (1992).

In particular, the time t real price of a real zero-coupon bond maturing at T is

$$B_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s\,ds}\right].$$

In order to study nominal prices and interest rates, we introduce the consumer price index I_t, which is interpreted as the dollar price I_t of a unit of consumption. We write the dynamics of $I = (I_t)$ as

$$dI_t = I_t\left[i_t\,dt + \boldsymbol{\sigma}_{It}^{\top}\,d\boldsymbol{z}_t\right].$$

We can interpret dI_t/I_t as the realized inflation rate over the next instant, i_t as the expected inflation rate, and $\boldsymbol{\sigma}_{It}$ as the percentage sensitivity vector of the inflation rate.

Consider now a nominal bank account which over the next instant promises a risk-free monetary return represented by the nominal short-term interest rate $\tilde{r}_t$. If we let $\tilde{N}_t$ denote the time t dollar value of such an account, we have that

$$d\tilde{N}_t = \tilde{r}_t\tilde{N}_t\,dt.$$

The real price of this account is $N_t = \tilde{N}_t/I_t$, since this is the number of units of the consumption good that have the same value as the account. An application of Itô's Lemma implies a real price dynamics of

$$dN_t = N_t\left[\left(\tilde{r}_t - i_t + \|\boldsymbol{\sigma}_{It}\|^2\right)\,dt - \boldsymbol{\sigma}_{It}^{\top}\,d\boldsymbol{z}_t\right].$$

Note that the real return on this instantaneously nominally risk-free asset, dN_t/N_t, is risky. Since the percentage sensitivity vector is given by $-\boldsymbol{\sigma}_{It}$, it follows from (5.14) that the expected return is given by the real short rate plus $-\boldsymbol{\sigma}_{It}^{\top}\boldsymbol{\lambda}_t$. Comparing this with the drift term in the equation above, we have that

$$\tilde{r}_t - i_t + \|\boldsymbol{\sigma}_{It}\|^2 = r_t - \boldsymbol{\sigma}_{It}^{\top}\boldsymbol{\lambda}_t.$$

Consequently, the nominal short-term interest rate is given by

$$\tilde{r}_t = r_t + i_t - \|\boldsymbol{\sigma}_{It}\|^2 - \boldsymbol{\sigma}_{It}^{\top}\boldsymbol{\lambda}_t, \tag{5.15}$$

that is the nominal short rate is equal to the real short rate plus the expected inflation rate minus the variance of the inflation rate minus a risk premium. The presence of the last two terms invalidates the Fisher relation, which says that the nominal interest rate is equal to the sum of the real interest rate and the expected inflation rate. The Fisher hypothesis will hold if and only if the inflation rate is instantaneously risk-free.

Since most traded assets are nominal, it would be nice to have a relation between expected nominal returns and volatility of nominal prices. For this purpose, let $\tilde{P}_t$ denote the dollar price of a financial asset and assume that the price dynamics can be described by

$$d\tilde{P}_t = \tilde{P}_t\left[\tilde{\mu}_{Pt}\,dt + \tilde{\boldsymbol{\sigma}}_{Pt}^{\top}\,d\boldsymbol{z}_t\right].$$

The real price of this asset is given by $P_t = \tilde{P}_t/I_t$ and by Itô's Lemma

$$dP_t = P_t \left[\left(\tilde{\mu}_{Pt} - i_t - \tilde{\boldsymbol{\sigma}}_{Pt}^{\top} \boldsymbol{\sigma}_{It} + \|\boldsymbol{\sigma}_{It}\|^2 \right) dt + \left(\tilde{\boldsymbol{\sigma}}_{Pt} - \boldsymbol{\sigma}_{It} \right)^{\top} dz_t \right]. \tag{5.16}$$

The expected excess real rate of return on the asset is therefore

$$\begin{aligned} \mu_{Pt} - r_t &= \tilde{\mu}_{Pt} - i_t - \tilde{\boldsymbol{\sigma}}_{Pt}^{\top} \boldsymbol{\sigma}_{It} + \|\boldsymbol{\sigma}_{It}\|^2 - r_t \\ &= \tilde{\mu}_{Pt} - \tilde{r}_t - \tilde{\boldsymbol{\sigma}}_{Pt}^{\top} \boldsymbol{\sigma}_{It} - \boldsymbol{\sigma}_{It}^{\top} \boldsymbol{\lambda}_t, \end{aligned}$$

where we have introduced the nominal short rate $\tilde{r}_t$ by applying (5.15). The volatility vector of the real return on the asset is

$$\boldsymbol{\sigma}_{Pt} = \tilde{\boldsymbol{\sigma}}_{Pt} - \boldsymbol{\sigma}_{It}.$$

Substituting the expressions for $\mu_{Pt} - r_t$ and $\boldsymbol{\sigma}_{Pt}$ into the relation (5.14), we obtain

$$\tilde{\mu}_{Pt} - \tilde{r}_t - \tilde{\boldsymbol{\sigma}}_{Pt}^{\top} \boldsymbol{\sigma}_{It} - \boldsymbol{\sigma}_{It}^{\top} \boldsymbol{\lambda}_t = \left(\tilde{\boldsymbol{\sigma}}_{Pt} - \boldsymbol{\sigma}_{It} \right)^{\top} \boldsymbol{\lambda}_t,$$

and hence

$$\tilde{\mu}_{Pt} - \tilde{r}_t = \tilde{\boldsymbol{\sigma}}_{Pt}^{\top} \tilde{\boldsymbol{\lambda}}_t,$$

where $\tilde{\boldsymbol{\lambda}}_t$ is the nominal market price of risk vector defined by

$$\tilde{\boldsymbol{\lambda}}_t = \boldsymbol{\sigma}_{It} + \boldsymbol{\lambda}_t. \tag{5.17}$$

In terms of expectations, we know that

$$\frac{\tilde{P}_t}{I_t} = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} \frac{\tilde{P}_T}{I_T} \right],$$

from which it follows that

$$\tilde{P}_t = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} \frac{I_t}{I_T} \tilde{P}_T \right] = \mathrm{E}_t \left[\frac{\tilde{\zeta}_T}{\tilde{\zeta}_t} \tilde{P}_T \right],$$

where $\tilde{\zeta}_t = \zeta_t / I_t$ for any t (in particular, $\tilde{\zeta}_0 = 1/I_0$). Since the left-hand side is the current nominal price and the right-hand side involves the future nominal price or payoff, it is reasonable to call $\tilde{\zeta} = (\tilde{\zeta}_t)$ a **nominal state-price deflator**. Its dynamics are given by

$$d\tilde{\zeta}_t = -\tilde{\zeta}_t \left[\tilde{r}_t \, dt + \tilde{\boldsymbol{\lambda}}_t^{\top} \, dz_t \right] \tag{5.18}$$

so the drift rate is (minus) the nominal short rate and the volatility vector is (minus) the nominal market price of risk, completely analogous to the real counterparts.

We can also introduce a **nominal risk-neutral measure** $\tilde{\mathbb{Q}}$ by the change-of-measure process

$$\tilde{\xi}_t \equiv \mathrm{E}_t \left[\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}} \right] = \exp\left\{ -\frac{1}{2} \int_0^t \|\tilde{\boldsymbol{\lambda}}_s\|^2 \, ds - \int_0^t \tilde{\boldsymbol{\lambda}}_s^{\top} \, dz_s \right\}.$$

Then the nominal price of a non-dividend paying asset can be written as

$$\tilde{P}_t = \mathrm{E}_t\left[\frac{\tilde{\zeta}_T}{\tilde{\zeta}_t}\tilde{P}_T\right] = \mathrm{E}_t^{\tilde{\mathbb{Q}}}\left[e^{-\int_t^T \tilde{r}_s\,ds}\tilde{P}_T\right].$$

In particular, the time t nominal price of a nominal zero-coupon bond maturing at T is

$$\tilde{B}_t^T = \mathrm{E}_t\left[\frac{\tilde{\zeta}_T}{\tilde{\zeta}_t}\right] = \mathrm{E}_t^{\tilde{\mathbb{Q}}}\left[e^{-\int_t^T \tilde{r}_s\,ds}\right].$$

To sum up, the prices of nominal bonds are related to the nominal short rate and the nominal market price of risk in exactly the same way as the prices of real bonds are related to the real short rate and the real market price of risk. Models that are based on specific exogenous assumptions about the short rate dynamics and the market price of risk can be applied both to real term structures and to nominal term structures. This is indeed the case for most popular term structure models. However, the equilibrium arguments that some authors offer in support of a particular term structure model, see Section 5.4, typically apply to real interest rates and real market prices of risk. Due to the relations (5.15) and (5.17), the same arguments cannot generally support similar assumptions on nominal rates and market price of risk. Nevertheless, these models are often applied on nominal bonds and term structures.

Above, we derived an equilibrium relation between real and nominal short-term interest rates. What can we say about the relation between longer-term real and nominal interest rates? Applying the well-known relation $\mathrm{Cov}[x, y] = \mathrm{E}[xy] - \mathrm{E}[x]\,\mathrm{E}[y]$, we can write

$$\begin{aligned}
\tilde{B}_t^T &= \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\frac{I_t}{I_T}\right] \\
&= \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\right]\mathrm{E}_t\left[\frac{I_t}{I_T}\right] + \mathrm{Cov}_t\left[\frac{\zeta_T}{\zeta_t}, \frac{I_t}{I_T}\right] \\
&= B_t^T\,\mathrm{E}_t\left[\frac{I_t}{I_T}\right] + \mathrm{Cov}_t\left[\frac{\zeta_T}{\zeta_t}, \frac{I_t}{I_T}\right].
\end{aligned}$$

From the dynamics of the state-price deflator and the price index, we get

$$\begin{aligned}
\frac{\zeta_T}{\zeta_t} &= \exp\left\{-\int_t^T \left(r_s + \frac{1}{2}\|\boldsymbol{\lambda}_s\|^2\right)ds - \int_t^T \boldsymbol{\lambda}_s^{\top}\,d\boldsymbol{z}_s\right\}, \\
\frac{I_t}{I_T} &= \exp\left\{-\int_t^T \left(i_s - \frac{1}{2}\|\boldsymbol{\sigma}_{Is}\|^2\right)ds - \int_t^T \boldsymbol{\sigma}_{Is}^{\top}\,d\boldsymbol{z}_s\right\},
\end{aligned}$$

which can be substituted into the above relation between prices on real and nominal bonds. However, the covariance term on the right-hand side can only be explicitly computed under very special assumptions about the variations over time in r, i, $\boldsymbol{\lambda}$, and $\boldsymbol{\sigma}_I$.

5.5.2 No real effects of inflation

In this subsection we will take as given some process for the consumer price index and assume that monetary holdings do not affect the utility of the agents directly. As before the aggregate consumption level is assumed to follow the process

$$dC_t = C_t\left[\mu_{Ct}\,dt + \boldsymbol{\sigma}_{Ct}^{\top}\,d\boldsymbol{z}_t\right]$$

so that the dynamics of the real state-price density is

$$d\zeta_t = -\zeta_t\left[r_t\,dt + \boldsymbol{\lambda}_t^{\top}\,d\boldsymbol{z}_t\right].$$

The short-term real rate is given by

$$r_t = \delta + \left(-\frac{C_t u''(C_t)}{u'(C_t)}\right)\mu_{Ct} - \frac{1}{2}C_t^2\frac{u'''(C_t)}{u'(C_t)}\|\boldsymbol{\sigma}_{Ct}\|^2$$

and the market price of risk vector is given by

$$\boldsymbol{\lambda}_t = \left(-\frac{C_t u''(C_t)}{u'(C_t)}\right)\boldsymbol{\sigma}_{Ct}. \tag{5.19}$$

By substituting the expression (5.19) for $\boldsymbol{\lambda}_t$ into (5.15), we can write the short-term nominal rate as

$$\tilde{r}_t = r_t + i_t - \|\boldsymbol{\sigma}_{It}\|^2 - \left(-\frac{C_t u''(C_t)}{u'(C_t)}\right)\boldsymbol{\sigma}_{It}^{\top}\boldsymbol{\sigma}_{Ct}.$$

In the special case where the representative agent has constant relative risk aversion, that is $u(C) = C^{1-\gamma}/(1-\gamma)$, and both the aggregate consumption and the price index follow geometric Brownian motions, we get constant rates

$$r = \delta + \gamma\mu_C - \frac{1}{2}\gamma(1+\gamma)\|\boldsymbol{\sigma}_C\|^2, \tag{5.20}$$

$$\tilde{r} = r + i - \|\boldsymbol{\sigma}_I\|^2 - \gamma\boldsymbol{\sigma}_I^{\top}\boldsymbol{\sigma}_C.$$

Breeden (1986) considers the relations between interest rates, inflation, and aggregate consumption and production in an economy with multiple consumption goods. In general the presence of several consumption goods complicates the analysis considerably. Breeden shows that the equilibrium nominal short rate will depend on both an inflation rate computed using the average weights of the different consumption goods and an inflation rate computed using the marginal weights of the different goods, which are determined by the optimal allocation to the different goods of an extra dollar of total consumption expenditure. The average and the marginal consumption weights will generally be different since the representative agent may shift to other consumption goods as his wealth increases. However, in the special and probably unrealistic case of a Cobb–Douglas type utility function, the relative expenditure weights of the different consumption goods will be constant. For that case Breeden obtains results similar to our one-good conclusions.

5.5.3 A model with real effects of money

In the next model we consider, cash holdings enter the direct utility function of the agent(s). This may be rationalized by the fact that cash holdings facilitate frequent consumption transactions. In such a model the price of the consumption good is determined as a part of the equilibrium of the economy, in contrast to the models studied above where we took an exogenous process for the consumer price index. We follow the setup of Bakshi and Chen (1996) closely.

5.5.3.1 The general model

We assume the existence of a representative agent who chooses a consumption process $C = (C_t)$ and a cash process $M = (M_t)$, where M_t is the dollar amount held at time t. As before, let I_t be the unit dollar price of the consumption good. Assume that the representative agent has an infinite time horizon, no endowment stream, and an additively time-separable utility of consumption and the real value of the monetary holdings, that is $\tilde{M}_t = M_t/I_t$. At time t the agent has the opportunity to invest in a nominally risk-free bank account with a nominal rate of return of $\tilde{r}_t$. When the agent chooses to hold M_t dollars in cash over the period $[t, t + dt]$, she therefore gives up a dollar return of $M_t\tilde{r}_t\, dt$, which is equivalent to a consumption of $M_t\tilde{r}_t\, dt/I_t$ units of the good. Given a (real) state-price deflator $\zeta = (\zeta_t)$, the total cost of choosing C and M is thus $\mathrm{E}\left[\int_0^\infty \zeta_t(C_t + M_t\tilde{r}_t/I_t)\, dt\right]$, which must be smaller than or equal to the initial (real) wealth of the agent, W_0. In sum, the optimization problem of the agent can be written as follows:

$$\sup_{(C_t,M_t)} \mathrm{E}\left[\int_0^\infty e^{-\delta t} u\left(C_t, M_t/I_t\right)\, dt\right]$$

$$\text{s.t. } \mathrm{E}\left[\int_0^\infty \zeta_t \left(C_t + \frac{M_t}{I_t}\tilde{r}_t\right)\, dt\right] \leq W_0.$$

The first order conditions are

$$e^{-\delta t} u_C(C_t, M_t/I_t) = \psi\zeta_t, \tag{5.21}$$

$$e^{-\delta t} u_M(C_t, M_t/I_t) = \psi\zeta_t\tilde{r}_t, \tag{5.22}$$

where u_C and u_M are the first-order derivatives of u with respect to the first and second argument, respectively. ψ is a Lagrange multiplier, which is set so that the budget condition holds as an equality. Again, we see that the state-price deflator is given in terms of the marginal utility with respect to consumption. Imposing the initial value $\zeta_0 = 1$ and recalling the definition of $\tilde{M}_t$, we have

$$\zeta_t = e^{-\delta t}\frac{u_C(C_t, \tilde{M}_t)}{u_C(C_0, \tilde{M}_0)}. \tag{5.23}$$

We can apply the state-price deflator to value all payment streams. For example, an investment of one dollar at time t in the nominal bank account generates a continuous payment stream at the rate of $\tilde{r}_s$ dollars to the end of all time. The corresponding real investment at time t is $1/I_t$ and the real dividend at time s is $\tilde{r}_s/I_s$. Hence, we have the relation

$$\frac{1}{I_t} = \mathrm{E}_t\left[\int_t^\infty \frac{\zeta_s}{\zeta_t}\frac{\tilde{r}_s}{I_s}\,ds\right],$$

or, equivalently,

$$\frac{1}{I_t} = \mathrm{E}_t\left[\int_t^\infty e^{-\delta(s-t)}\frac{u_C(C_s,\tilde{M}_s)}{u_C(C_t,\tilde{M}_t)}\frac{\tilde{r}_s}{I_s}\,ds\right]. \tag{5.24}$$

Substituting the first optimality condition (5.21) into the second (5.22), we see that the nominal short rate is given by

$$\tilde{r}_t = \frac{u_M(C_t, M_t/I_t)}{u_C(C_t, M_t/I_t)}. \tag{5.25}$$

The intuition behind this relation can be explained in the following way. If you have an extra dollar now you can either keep it in cash or invest it in the nominally risk-free bank account. If you keep it in cash your utility grows by $u_M(C_t, M_t/I_t)/I_t$. If you invest it in the bank account you will earn a dollar interest of $\tilde{r}_t$ that can be used for consuming $\tilde{r}_t/I_t$ extra units of consumption, which will increase your utility by $u_C(C_t, M_t/I_t)\tilde{r}_t/I_t$. At the optimum, these utility increments must be identical. Combining (5.24) and (5.25), we get that the price index must satisfy the recursive relation

$$\frac{1}{I_t} = \mathrm{E}_t\left[\int_t^\infty e^{-\delta(s-t)}\frac{u_M(C_s,\tilde{M}_s)}{u_C(C_t,\tilde{M}_t)}\frac{1}{I_s}\,ds\right]. \tag{5.26}$$

Let us find expressions for the equilibrium real short rate and the market price of risk in this setting. As always, the real short rate equals minus the percentage drift of the state-price deflator, while the market price of risk equals minus the percentage volatility vector of the state-price deflator. In an equilibrium, the representative agent must consume the aggregate consumption and hold the total money supply in the economy. Suppose that the aggregate consumption and the money supply follow exogenous processes of the form

$$dC_t = C_t\left[\mu_{Ct}\,dt + \boldsymbol{\sigma}_{Ct}^{\top}\,d\boldsymbol{z}_t\right],$$
$$dM_t = M_t\left[\mu_{Mt}\,dt + \boldsymbol{\sigma}_{Mt}^{\top}\,d\boldsymbol{z}_t\right].$$

Assuming that the endogenously determined price index will follow a similar process,

$$dI_t = I_t\left[i_t\,dt + \boldsymbol{\sigma}_{It}^{\top}\,d\boldsymbol{z}_t\right],$$

the dynamics of $\tilde{M}_t = M_t/I_t$ will be

$$d\tilde{M}_t = \tilde{M}_t\left[\tilde{\mu}_{Mt}\,dt + \tilde{\boldsymbol{\sigma}}_{Mt}^{\top}\,d\boldsymbol{z}_t\right],$$

where

$$\tilde{\mu}_{Mt} = \mu_{Mt} - i_t + \|\boldsymbol{\sigma}_{It}\|^2 - \boldsymbol{\sigma}_{Mt}^{\top}\boldsymbol{\sigma}_{It}, \qquad \tilde{\boldsymbol{\sigma}}_{Mt} = \boldsymbol{\sigma}_{Mt} - \boldsymbol{\sigma}_{It}.$$

Given these equations and the relation (5.23), we can find the drift and the volatility vector of the state-price deflator by an application of Itô's Lemma. We find that the equilibrium real short-term interest rate can be written as

$$\begin{aligned} r_t = \delta &+ \left(\frac{-C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\mu_{Ct} + \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\tilde{\mu}_{Mt} \\ &- \frac{1}{2}\frac{C_t^2 u_{CCC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\|\boldsymbol{\sigma}_{Ct}\|^2 - \frac{1}{2}\frac{\tilde{M}_t^2 u_{CMM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\|\tilde{\boldsymbol{\sigma}}_{Mt}\|^2 \\ &- \frac{C_t\tilde{M}_t u_{CCM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\boldsymbol{\sigma}_{Ct}^{\top}\tilde{\boldsymbol{\sigma}}_{Mt}, \end{aligned} \tag{5.27}$$

while the market price of risk vector is

$$\begin{aligned} \boldsymbol{\lambda}_t &= \left(-\frac{C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\boldsymbol{\sigma}_{Ct} + \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\tilde{\boldsymbol{\sigma}}_{Mt} \\ &= \left(-\frac{C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\boldsymbol{\sigma}_{Ct} + \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)(\boldsymbol{\sigma}_{Mt} - \boldsymbol{\sigma}_{It}). \end{aligned} \tag{5.28}$$

With $u_{CM} < 0$, we see that assets that are positively correlated with the inflation rate will have a lower expected real return, other things equal. Intuitively such assets are useful for hedging inflation risk so that they do not have to offer as high an expected return.

The relation (5.15) is also valid in the present setting. Substituting the expression (5.28) for the market price of risk into (5.15), we obtain

$$\begin{aligned} \tilde{r}_t - r_t - i_t + \|\boldsymbol{\sigma}_{It}\|^2 = &- \left(-\frac{C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\boldsymbol{\sigma}_{It}^{\top}\boldsymbol{\sigma}_{Ct} \\ &- \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)}\right)\boldsymbol{\sigma}_{It}^{\top}\tilde{\boldsymbol{\sigma}}_{Mt}. \end{aligned}$$

5.5.3.2 *An example*

To obtain more concrete results, we must specify the utility function and the exogenous processes C and M. Assume a utility function of the Cobb-Douglas type,

$$u(C, \tilde{M}) = \frac{\left(C^{\varphi}\tilde{M}^{1-\varphi}\right)^{1-\gamma}}{1-\gamma},$$

where φ is a constant between zero and one, and γ is a positive constant. The limiting case for $\gamma = 1$ is log utility,

$$u(C, \tilde{M}) = \varphi \ln C + (1 - \varphi) \ln \tilde{M}.$$

By inserting the relevant derivatives into (5.27), we see that the real short rate becomes

$$\begin{aligned} r_t = \delta &+ [1 - \varphi(1-\gamma)]\mu_{Ct} - (1-\varphi)(1-\gamma)\tilde{\mu}_{Mt} \\ &- \frac{1}{2}[1 - \varphi(1-\gamma)][2 - \varphi(1-\gamma)]\|\boldsymbol{\sigma}_{Ct}\|^2 \\ &+ \frac{1}{2}(1-\varphi)(1-\gamma)[1 - (1-\varphi)(1-\gamma)]\|\tilde{\boldsymbol{\sigma}}_{Mt}\|^2 \\ &+ (1-\varphi)(1-\gamma)[1 - \varphi(1-\gamma)]\boldsymbol{\sigma}_{Ct}^{\top}\tilde{\boldsymbol{\sigma}}_{Mt}, \end{aligned} \tag{5.29}$$

which for $\gamma = 1$ simplifies to

$$r_t = \delta + \mu_{Ct} - \|\boldsymbol{\sigma}_{Ct}\|^2.$$

We see that with log utility, the real short rate will be constant if aggregate consumption $C = (C_t)$ follows a geometric Brownian motion. From (5.25), the nominal short rate is

$$\tilde{r}_t = \frac{1-\varphi}{\varphi} \frac{C_t}{\tilde{M}_t}.$$

The ratio $C_t/\tilde{M}_t$ is called the **velocity of money**. If the velocity of money is constant, the nominal short rate will be constant. Since $\tilde{M}_t = M_t/I_t$ and I_t is endogenously determined, the velocity of money will also be endogenously determined.

To identify the nominal short rate for any γ and the real short rate for $\gamma \neq 1$, we have to determine the price level in the economy, which is given recursively in (5.26). This is possible under the assumption that both C and M follow geometric Brownian motions. We conjecture that $I_t = kM_t/C_t$ for some constant k. From (5.26), we get

$$\frac{1}{k} = \frac{1-\varphi}{\varphi} \int_t^{\infty} e^{-\delta(s-t)} \,\mathrm{E}_t\left[\left(\frac{C_s}{C_t}\right)^{1-\gamma} \left(\frac{M_s}{M_t}\right)^{-1}\right] ds.$$

Inserting the relations

$$\frac{C_s}{C_t} = \exp\left\{\left(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2\right)(s-t) + \boldsymbol{\sigma}_C^{\top}(\boldsymbol{z}_s - \boldsymbol{z}_t)\right\},$$

$$\frac{M_s}{M_t} = \exp\left\{\left(\mu_M - \frac{1}{2}\|\boldsymbol{\sigma}_M\|^2\right)(s-t) + \boldsymbol{\sigma}_M^{\top}(\boldsymbol{z}_s - \boldsymbol{z}_t)\right\},$$

and applying a standard rule for expectations of lognormal variables, we get

$$\frac{1}{k} = \frac{1-\varphi}{\varphi} \int_t^\infty \exp\left\{\left(-\delta + (1-\gamma)(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2) - \mu_M + \|\boldsymbol{\sigma}_M\|^2 \right.\right.$$
$$\left.\left. + \frac{1}{2}(1-\gamma)^2\|\boldsymbol{\sigma}_C\|^2 - (1-\gamma)\boldsymbol{\sigma}_C^\top \boldsymbol{\sigma}_M\right)(s-t)\right\} ds,$$

which implies that the conjecture is true with

$$k = \frac{\varphi}{1-\varphi}\left(\delta - (1-\gamma)(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2) + \mu_M - \|\boldsymbol{\sigma}_M\|^2 \right.$$
$$\left. -\frac{1}{2}(1-\gamma)^2\|\boldsymbol{\sigma}_C\|^2 + (1-\gamma)\boldsymbol{\sigma}_C^\top \boldsymbol{\sigma}_M\right).$$

An application of Itô's Lemma shows that the price index also follows a geometric Brownian motion

$$dI_t = I_t\left[i\,dt + \boldsymbol{\sigma}_I^\top\, d\boldsymbol{z}_t\right],$$

where

$$i = \mu_M - \mu_C + \|\boldsymbol{\sigma}_C\|^2 - \boldsymbol{\sigma}_M^\top \boldsymbol{\sigma}_C, \qquad \boldsymbol{\sigma}_I = \boldsymbol{\sigma}_M - \boldsymbol{\sigma}_C.$$

With $I_t = kM_t/C_t$, we have $\tilde{M}_t = C_t/k$, so that the velocity of money $C_t/\tilde{M}_t = k$ is constant, and the nominal short rate becomes

$$\tilde{r}_t = \frac{1-\varphi}{\varphi}k = \delta - (1-\gamma)(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2) + \mu_M - \|\boldsymbol{\sigma}_M\|^2$$
$$- \frac{1}{2}(1-\gamma)^2\|\boldsymbol{\sigma}_C\|^2 + (1-\gamma)\boldsymbol{\sigma}_C^\top \boldsymbol{\sigma}_M,$$

which is also a constant. With log utility, the nominal rate simplifies to $\delta + \mu_M - \|\boldsymbol{\sigma}_M\|^2$. In order to obtain the real short rate in the non-log case, we have to determine $\tilde{\mu}_{Mt}$ and $\tilde{\boldsymbol{\sigma}}_{Mt}$ and plug into (5.29). We get $\tilde{\mu}_{Mt} = \mu_C + \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2 + \boldsymbol{\sigma}_M^\top \boldsymbol{\sigma}_C$ and $\tilde{\boldsymbol{\sigma}}_{Mt} = \boldsymbol{\sigma}_C$ and hence

$$r_t = \delta + \gamma\mu_C - \gamma\|\boldsymbol{\sigma}_C\|^2\left[\frac{1}{2}(1+\gamma) + \varphi(1-\gamma)\right],$$

which is again a constant. In comparison with (5.20) for the case where money has no real effects, the last term in the equation above is new.

5.5.3.3 *Another example*

Bakshi and Chen (1996) also explore another model specification in which both nominal and real short rates are time-varying, but evolve independently of each other. Interest rates will only be stochastic with more general processes for aggregate consumption and money supply than the geometric Brownian motions used above. Bakshi and Chen assume log-utility ($\gamma = 1$) in which case we have already seen that

$$r_t = \delta + \mu_{Ct} - \|\boldsymbol{\sigma}_{Ct}\|^2, \qquad \tilde{r}_t = \frac{1-\varphi}{\varphi}\frac{C_t}{\tilde{M}_t} = \frac{1-\varphi}{\varphi}\frac{C_t I_t}{M_t}.$$

The dynamics of aggregate consumption is assumed to be

$$dC_t = C_t\left[(\alpha_C + \kappa_C x_t)\,dt + \sigma_C\sqrt{x_t}\,dz_{1t}\right],$$

where x can be interpreted as a technology variable and is assumed to follow the process

$$dx_t = \kappa_x(\theta_x - x_t)\,dt + \sigma_x\sqrt{x_t}\,dz_{1t}.$$

The money supply is assumed to be $M_t = M_0 e^{\mu_M^* t} g_t/g_0$, where

$$dg_t = g_t\left[\kappa_g(\theta_g - g_t)\,dt + \sigma_g\sqrt{g_t}\left(\rho_{CM}\,dz_{1t} + \sqrt{1-\rho_{CM}^2}\,dz_{2t}\right)\right],$$

and where z_1 and z_2 are independent one-dimensional Brownian motions. Following the same basic procedure as in the previous model specification, the authors show that the real short rate is

$$r_t = \delta + \alpha_C + (\kappa_C - \sigma_C^2)x_t,$$

while the nominal short rate is

$$\tilde{r}_t = \frac{(\delta + \mu_M^*)(\delta + \mu_M^* + \kappa_g\theta_g)}{\delta + \mu_M^* + (\kappa_g + \sigma_g^2)g_t}.$$

Both rates are time-varying. The real rate is driven by the technology variable x, while the nominal rate is driven by the monetary shock process g. In this set-up, shocks to the real economy have opposite effects of the same magnitude on real rates and inflation so that nominal rates are unaffected.

The real price of a real zero-coupon bond maturing at time T is of the form

$$B_t^T = e^{-a(T-t)-b(T-t)x_t},$$

while the nominal price of a nominal zero-coupon bond maturing at T is

$$\tilde{B}_t^T = \frac{\tilde{a}(T-t) + \tilde{b}(T-t)g_t}{\delta + \mu_M^* + (\kappa_g + \sigma_g^2)g_t},$$

where a, b, $\tilde{a}$, and $\tilde{b}$ are deterministic functions of time for which Bakshi and Chen provide closed-form expressions.

In the very special case where these processes are uncorrelated, that is $\rho_{CM} = 0$, the real and nominal term structures of interest rates are independent of each other! Although this is an extreme result, it does point out that real and nominal term structures in general may have quite different properties.

5.6 THE EXPECTATION HYPOTHESIS

The **expectation hypothesis** relates the current interest rates and yields to expected future interest rates or returns. The basic idea dates back to Fisher (1896) and was further developed and concretized by Hicks (1939) and Lutz (1940). The original motivation of the hypothesis is that when lenders (bond investors) and borrowers (bond issuers) decide between long-term or short-term bonds, they will compare the price or yield of a long-term bond to the expected price or return on a roll-over strategy in short-term bonds. Hence, long-term rates and expected future short-term rates will be linked. Of course, a cornerstone of modern finance theory is that, when comparing different strategies, investors will also take the risks into account. So even before going into the specifics of the hypothesis you should really be quite skeptical, at least when it comes to very strict interpretations of the expectation hypothesis.

The vague idea that current yields and interest rates are linked to expected future rates and returns can be concretized in a number of ways. Below we will present and evaluate a number of versions. This analysis follows Cox et al. (1981a) quite closely. We find that some versions are equivalent, some versions inconsistent. We end up concluding that none of the variants of the expectations hypothesis are consistent with any realistic behaviour of interest rates. Hence, the analysis of the shape of the yield curve and models of term structure dynamics should not be based on this hypothesis. Hence, it is surprising, maybe even disappointing, that empirical tests of the expectation hypothesis have generated such a huge literature in the past and that the hypothesis still seems to be widely accepted among economists.

5.6.1 Versions of the pure expectation hypothesis

The first version of the pure expectation hypothesis that we will discuss says that prices in the bond markets are set so that the expected gross returns on all self-financing trading strategies over a given period are identical. In particular, the expected gross return from buying at time t a zero-coupon bond maturing at time T and reselling it at time $t' \leq T$, which is given by $\mathrm{E}_t[B_{t'}^T / B_t^T]$, will be independent of the maturity date T of the bond (but generally not independent of t'). Let us refer to this as the **gross return** pure expectation hypothesis.

This version of the hypothesis is consistent with pricing in a world of risk-neutral investors. If we have a representative individual with time-additive expected utility, we know that zero-coupon bond prices satisfy

$$B_t^T = \mathrm{E}_t \left[e^{-\delta(t'-t)} \frac{u'(C_{t'})}{u'(C_t)} B_{t'}^T \right],$$

where u is the instantaneous utility function, δ is the time preference rate, and C denotes aggregate consumption. If the representative individual is risk-neutral, his marginal utility is constant, which implies that

$$\mathrm{E}_t \left[\frac{B_{t'}^T}{B_t^T} \right] = e^{\delta(t'-t)},$$

which is clearly independent of T. Clearly, the assumption of risk-neutrality is not very attractive. There is also another serious problem with this hypothesis. As is to be shown in Exercise 5.9, it cannot hold when interest rates are uncertain.

A slight variation of the above is to align all expected continuously compounded returns, that is $\frac{1}{t'-t}\,\mathrm{E}_t[\ln\left(B_{t'}^T/B_t^T\right)]$ for all T. In particular with $T = t'$, the expected continuously compounded rate of return is known to be equal to the zero-coupon yield for maturity t', which we denote by $y_t^{t'} = -\frac{1}{t'-t}\ln B_t^{t'}$. We can therefore formulate the hypothesis as

$$\frac{1}{t'-t}\,\mathrm{E}_t\left[\ln\left(\frac{B_{t'}^T}{B_t^T}\right)\right] = y_t^{t'}, \text{ all } T \geq t'.$$

Let us refer to this as the **rate of return** pure expectation hypothesis. For $t' \to t$, the right-hand side approaches the current short rate r_t, while the left-hand side approaches the absolute drift rate of $\ln B_t^T$.

An alternative specification of the pure expectation hypothesis claims that the expected return over the next time period is the same for all investments in bonds and deposits. In other words there is no difference between expected returns on long-maturity and short-maturity bonds. In the continuous-time limit we consider returns over the next instant. The risk-free return over $[t, t+dt]$ is $r_t\,dt$, so for any zero-coupon bond, the hypothesis claims that

$$\mathrm{E}_t\left[\frac{dB_t^T}{B_t^T}\right] = r_t\,dt, \qquad \text{for all } T > t, \tag{5.30}$$

or, using Theorem 3.8, that

$$B_t^T = \mathrm{E}_t\left[e^{-\int_t^T r_s\,ds}\right], \qquad \text{for all } T > t.$$

This is the **local** pure expectations hypothesis.

Another interpretation says that the return from holding a zero-coupon bond to maturity should equal the expected return from rolling over short-term bonds over the same time period, that is

$$\frac{1}{B_t^T} = \mathrm{E}_t\left[e^{\int_t^T r_s\,ds}\right], \qquad \text{for all } T > t \tag{5.31}$$

or, equivalently,

$$B_t^T = \left(\mathrm{E}_t\left[e^{\int_t^T r_s\,ds}\right]\right)^{-1}, \qquad \text{for all } T > t.$$

This is the **return-to-maturity** pure expectation hypothesis.

A related claim is that the yield on any zero-coupon bond should equal the 'expected yield' on a roll-over strategy in short bonds. Since an investment of one at time t in the bank account generates $e^{\int_t^T r_s\,ds}$ at time T, the ex-post realized yield is $\frac{1}{T-t}\int_t^T r_s\,ds$. Hence, this **yield-to-maturity** pure expectation hypothesis says that

$$y_t^T = -\frac{1}{T-t} \ln B_t^T = \mathrm{E}_t \left[\frac{1}{T-t} \int_t^T r_s \, ds \right], \tag{5.32}$$

or, equivalently,

$$B_t^T = e^{-\mathrm{E}_t \left[\int_t^T r_s \, ds \right]}, \qquad \text{for all } T > t.$$

Finally, the **unbiased** pure expectation hypothesis states that the forward rate for time T prevailing at time $t < T$ is equal to the time t expectation of the short rate at time T, that is that forward rates are unbiased estimates of future spot rates. In symbols,

$$f_t^T = \mathrm{E}_t[r_T], \qquad \text{for all } T > t.$$

This implies that

$$-\ln B_t^T = \int_t^T f_t^s \, ds = \int_t^T \mathrm{E}_t[r_s] \, ds = \mathrm{E}_t \left[\int_t^T r_s \, ds \right],$$

from which we see that the unbiased version of the pure expectation hypothesis is indistinguishable from the yield-to-maturity version.

We will first show that *the different versions are inconsistent* when future rates are uncertain. This follows from an application of Jensen's inequality which states that if X is a random variable and f is a convex function, that is $f'' > 0$, then $\mathrm{E}[f(X)] > f(\mathrm{E}[X])$. Since $f(x) = e^x$ is a convex function, we have $\mathrm{E}[e^X] > e^{\mathrm{E}[X]}$ for any random variable X. In particular for $X = \int_t^T r_s \, ds$, we get

$$\mathrm{E}_t \left[e^{\int_t^T r_s \, ds} \right] > e^{\mathrm{E}_t \left[\int_t^T r_s \, ds \right]} \quad \Rightarrow \quad e^{-\mathrm{E}_t \left[\int_t^T r_s \, ds \right]} > \left(\mathrm{E}_t \left[e^{\int_t^T r_s \, ds} \right] \right)^{-1}.$$

This shows that the bond price according to the yield-to-maturity version is strictly greater than the bond price according to the return-to-maturity version. For $X = -\int_t^T r_s \, ds$, we get

$$\mathrm{E}_t \left[e^{-\int_t^T r_s \, ds} \right] > e^{\mathrm{E}_t \left[-\int_t^T r_s \, ds \right]} = e^{-\mathrm{E}_t \left[\int_t^T r_s \, ds \right]},$$

hence the bond price according to the local version of the hypothesis is strictly greater than the bond price according to the yield-to-maturity version. We can conclude that at most one of the versions of the local, return-to-maturity, and yield-to-maturity pure expectations hypothesis can hold.

5.6.2 The pure expectation hypothesis and equilibrium

Next, let us see whether the different versions can be consistent with any equilibrium. Assume that interest rates and bond prices are generated by a d-dimensional standard Brownian motion z. Assuming absence of arbitrage there exists a market price of risk process $\boldsymbol{\lambda}$ so that for any maturity T, the zero-coupon bond price dynamics is of the form

$$dB_t^T = B_t^T \left[\left(r_t + \left(\boldsymbol{\sigma}_t^T\right)^{\mathsf{T}} \boldsymbol{\lambda}_t \right) dt + \left(\boldsymbol{\sigma}_t^T\right)^{\mathsf{T}} dz_t \right], \tag{5.33}$$

where $\boldsymbol{\sigma}_t^T$ denotes the d-dimensional sensitivity vector of the bond price. Recall that the same $\boldsymbol{\lambda}_t$ applies to all zero-coupon bonds so that $\boldsymbol{\lambda}_t$ is independent of the maturity of the bond. Comparing with (5.30), we see that the local expectation hypothesis will hold if and only if $\left(\boldsymbol{\sigma}_t^T\right)^{\mathsf{T}} \boldsymbol{\lambda}_t = 0$ for all T. This is true if either investors are risk-neutral or interest rate risk is uncorrelated with aggregate consumption. Neither of these conditions hold in real life.

To evaluate the return-to-maturity version, first note that an application of Itô's Lemma on (5.33) shows that

$$d\left(\frac{1}{B_t^T}\right) = \frac{1}{B_t^T} \left[\left(-r_t - \left(\boldsymbol{\sigma}_t^T\right)^{\mathsf{T}} \boldsymbol{\lambda}_t + \|\boldsymbol{\sigma}_t^T\|^2 \right) dt - \left(\boldsymbol{\sigma}_t^T\right)^{\mathsf{T}} dz_t \right].$$

On the other hand, according to the hypothesis (5.31) the relative drift of $1/B_t^T$ equals $-r_t$; compare Theorem 3.8. To match the two expressions for the drift, we must have

$$\left(\boldsymbol{\sigma}_t^T\right)^{\mathsf{T}} \boldsymbol{\lambda}_t = \|\boldsymbol{\sigma}_t^T\|^2, \qquad \text{for all } T. \tag{5.34}$$

Is this possible? Cox et al. (1981a) conclude that it is impossible. If the exogenous shock z and therefore $\boldsymbol{\sigma}_t^T$ and $\boldsymbol{\lambda}_t$ are one-dimensional, they are right, since $\boldsymbol{\lambda}_t$ must then equal $\boldsymbol{\sigma}_t^T$, and this must hold for all T. Since $\boldsymbol{\lambda}_t$ is independent of T and the volatility $\boldsymbol{\sigma}_t^T$ approaches zero for $T \to t$, this can only hold if $\boldsymbol{\lambda}_t \equiv 0$ (risk-neutral investors) or $\boldsymbol{\sigma}_t^T \equiv 0$ (deterministic interest rates). However, as pointed out by McCulloch (1993) and Fisher and Gilles (1998), in multi-dimensional cases the key condition (5.34) may indeed hold, at least in very special cases. Let $\boldsymbol{\varphi}$ be a d-dimensional function with the property that $\|\boldsymbol{\varphi}(\tau)\|^2$ is independent of τ. Define $\boldsymbol{\lambda}_t = 2\boldsymbol{\varphi}(0)$ and $\boldsymbol{\sigma}_t^T = \boldsymbol{\varphi}(0) - \boldsymbol{\varphi}(T-t)$. Then (5.34) is indeed satisfied. However, all such functions $\boldsymbol{\varphi}$ seem to generate very strange bond price dynamics. The examples given in the two papers mentioned above are

$$\boldsymbol{\varphi}(\tau) = k \begin{pmatrix} \sqrt{2e^{-\tau} - e^{-2\tau}} \\ 1 - e^{-\tau} \end{pmatrix}, \qquad \boldsymbol{\varphi}(\tau) = k_1 \begin{pmatrix} \cos(k_2 \tau) \\ \sin(k_2 \tau) \end{pmatrix},$$

where k, k_1, k_2 are constants.

As discussed above, the rate of return version implies that the absolute drift rate of the log-bond price equals the short rate. We can see from (5.32) that the same is true for the yield-to-maturity version and hence the unbiased version.[5] On the other hand Itô's Lemma and (5.33) imply that

$$d\left(\ln B_t^T\right) = \left(r_t + (\boldsymbol{\sigma}_t^T)^{\top}\boldsymbol{\lambda}_t - \frac{1}{2}\|\boldsymbol{\sigma}_t^T\|^2\right)\, dt + \left(\boldsymbol{\sigma}_t^T\right)^{\top}\, d\boldsymbol{z}_t. \tag{5.35}$$

Hence, these versions of the hypothesis will hold if and only if

$$(\boldsymbol{\sigma}_t^T)^{\top}\boldsymbol{\lambda}_t = \frac{1}{2}\|\boldsymbol{\sigma}_t^T\|^2, \quad \text{for all } T.$$

Again, it is possible that the condition holds. Just let φ and $\boldsymbol{\sigma}_t^T$ be as for the return-to-maturity hypothesis and let $\boldsymbol{\lambda}_t = \varphi(0)$. But such specifications are not likely to represent real-life term structures.

The conclusion to be drawn from this analysis is that neither of the different versions of the pure expectation hypothesis seem to be consistent with any reasonable description of the term structure of interest rates.

5.6.3 The weak expectation hypothesis

Above, we looked at versions of the **pure** expectation hypothesis that all align an expected return or yield with a current interest rate or yield. However, as pointed out by Campbell (1986), there is also a **weak** expectation hypothesis that allows for a difference between the relevant expected return/yield and the current rate/yield, but restricts this difference to be constant over time.

The local weak expectation hypothesis says that

$$\mathrm{E}_t\left[\frac{dB_t^T}{B_t^T}\right] = \left(r_t + g(T-t)\right)\, dt$$

for some deterministic function g. In the pure version g is identically zero. For a given time-to-maturity there is a constant 'instantaneous holding term premium'. Comparing with (5.33), we see that this hypothesis will hold when the market price of risk $\boldsymbol{\lambda}_t$ is constant and the bond price sensitivity vector $\boldsymbol{\sigma}_t^T$ is a deterministic function of time-to-maturity. These conditions are satisfied in the Vasicek (1977) model and in other models of the Gaussian class.

[5] According to the yield-to-maturity hypothesis

$$\begin{aligned}\frac{1}{\Delta t}\,\mathrm{E}_t\left[\ln B_{t+\Delta t}^T - \ln B_t^T\right] &= \frac{1}{\Delta t}\,\mathrm{E}_t\left[-\,\mathrm{E}_{t+\Delta t}\left[\int_{t+\Delta t}^T r_s\, ds\right] + \mathrm{E}_t\left[\int_t^T r_s\, ds\right]\right]\\ &= \frac{1}{\Delta t}\,\mathrm{E}_t\left[\int_t^{t+\Delta t} r_s\, ds\right],\end{aligned}$$

which approaches r_t as $\Delta t \to 0$. This means that the absolute drift of $\ln B_t$ equals r_t.

Similarly, the weak yield-to-maturity expectation hypothesis says that

$$f_t^T = \mathrm{E}_t[r_T] + h(T-t)$$

for some deterministic function h with $h(0) = 0$, that is that there is a constant 'instantaneous forward term premium'. The pure version requires h to be identically equal to zero. It can be shown that this condition implies that the drift of $\ln B_t^T$ equals $r_t + h(T-t)$.[6] Comparing with (5.35), we see also that this hypothesis will hold when λ_t is constant and σ_t^T is a deterministic function of $T-t$ as is the case in the Gaussian models.

The class of Gaussian models have several unrealistic properties. For example, such models allow negative interest rates and require bond and interest rate volatilities to be independent of the level of interest rates. So far, the validity of even weak versions of the expectation hypothesis have not been shown in more realistic term structure models.

5.7 LIQUIDITY PREFERENCE, MARKET SEGMENTATION, AND PREFERRED HABITATS

Another traditional explanation of the shape of the yield curve is given by the **liquidity preference hypothesis** introduced by Hicks (1939). He realized that the expectation hypothesis basically ignores investors' aversion towards risk and argued that expected returns on long-term bonds should exceed the expected returns on short-term bonds to compensate for the higher price fluctuations of long-term bonds. According to this view the yield curve should tend to be increasing. Note that the word 'liquidity' in the name of the hypothesis is not

[6] From the weak yield-to-maturity hypothesis, it follows that $-\ln B_t^T = \int_t^T (\mathrm{E}_t[r_s] + h(s-t))\, ds$. Hence,

$$\begin{aligned}
\frac{1}{\Delta t}\mathrm{E}_t\left[\ln B_{t+\Delta t}^T - \ln B_t^T\right] &= \frac{1}{\Delta t}\mathrm{E}_t\left[-\int_{t+\Delta t}^T \left(\mathrm{E}_{t+\Delta t}[r_s] + h(s-(t+\Delta t))\right) ds\right. \\
&\qquad \left. + \int_t^T (\mathrm{E}_t[r_s] + h(s-t))\, ds\right] \\
&= \frac{1}{\Delta t}\mathrm{E}_t\left[\int_t^{t+\Delta t} r_s\, ds\right] \\
&\qquad - \frac{1}{\Delta t}\left(\int_{t+\Delta t}^T h(s-(t+\Delta t))\, ds - \int_t^T h(s-t)\, ds\right).
\end{aligned}$$

The limit of $\frac{1}{\Delta t}\left(\int_{t+\Delta t}^T h(s-(t+\Delta t))\, ds - \int_t^T h(s-t)\, ds\right)$ as $\Delta t \to 0$ is exactly the derivative of $\int_t^T h(s-t)\, ds$ with respect to t. Applying Leibnitz' rule and $h(0) = 0$, this derivative equals $-\int_t^T h'(s-t)\, ds = -h(T-t)$. In sum, the drift rate of $\ln B_t^T$ becomes $r_t + h(T-t)$ according to the hypothesis.

used in the usual sense of the word. Short-term bonds are not necessarily more liquid than long-term bonds. A better name would be 'the maturity preference hypothesis'.

In contrast the **market segmentation hypothesis** introduced by Culbertson (1957) claims that investors will typically prefer to invest in bonds with time-to-maturity in a certain interval, a maturity segment, perhaps in an attempt to match liabilities with similar maturities. For example, a pension fund with liabilities due in 20–30 years can reduce risk by investing in bonds of similar maturity. On the other hand, central banks typically operate in the short end of the market. Hence, separated market segments can exist without any relation between the bond prices and the interest rates in different maturity segments. If this is really the case, we cannot expect to see continuous or smooth yield curves and discount functions across the different segments.

A more realistic version of this hypothesis is the **preferred habitats hypothesis** put forward by Modigliani and Sutch (1966). An investor may prefer bonds with a certain maturity, but should be willing to move away from that maturity if she is sufficiently compensated in terms of a higher yield.[7] The different segments are therefore not completely independent of each other, and yields and discount factors should depend on maturity in a smooth way.

It is really not possible to quantify the market segmentation or the preferred habitats hypothesis without setting up an economy with agents having different favourite maturities. The resulting equilibrium yield curve will depend heavily on the degree of risk aversion of the various agents as illustrated by an analysis of Cox et al. (1981a).

5.8 CONCLUDING REMARKS

In this chapter we have derived links between equilibrium interest rates and aggregate consumption and production that are useful in interpreting and understanding shifts in the level of interest rates and the shape of the yield curve. We have derived relations between nominal rates, real rates, and inflation, and among other things concluded that the term structure of nominal rates can behave very differently to the term structure of real rates. We have shown that some popular term structure models can be supported by equilibrium considerations. Finally, we have discussed and criticized traditional hypotheses about the shape of the yield curve.

The equilibrium models and arguments of this chapter were set in a relatively simple framework, for example assuming the existence of a representative agent with time-additive utility. For models of the equilibrium term structure of interest rates with investor heterogeneity or more general utility functions than studied in this chapter, see, for example, Duffie and Epstein (1992), Wang (1996), Riedel (2000, 2004), Wachter (2006). The effects of central banks on the term structure are discussed and modelled by, for example, Babbs and Webber (1994), Balduzzi et al. (1997), and Piazzesi (2005).

[7] In a sense the liquidity preference hypothesis simply says that all investors prefer short bonds.

5.9 EXERCISES

Exercise 5.1 The term premium at time t for the future period $[t', T]$ is the current forward rate for that period minus the expected spot rate, that is $f_t^{t',T} - \mathrm{E}_t[y_{t'}^T]$. This exercise will give a link between the term premium and a state-price deflator $\zeta = (\zeta_t)$.

(a) Show that

$$B_t^T = B_t^{t'} \, \mathrm{E}_t\left[B_{t'}^T\right] + \mathrm{Cov}_t\left[\frac{\zeta_{t'}}{\zeta_t}, \frac{\zeta_T}{\zeta_{t'}}\right]$$

for any $t \leq t' \leq T$.

(b) Using the above result, show that

$$\mathrm{E}_t\left[e^{-y_{t'}^T(T-t')}\right] - e^{-f_t^{t',T}(T-t')} = -\frac{1}{B_t^{t'}} \mathrm{Cov}_t\left[\frac{\zeta_{t'}}{\zeta_t}, \frac{\zeta_T}{\zeta_{t'}}\right].$$

Using the previous result and the approximation $e^x \approx 1 + x$, show that

$$f_t^{t',T} - \mathrm{E}_t[y_{t'}^T] \approx -\frac{1}{(T-t')B_t^{t'}} \mathrm{Cov}_t\left[\frac{\zeta_{t'}}{\zeta_t}, \frac{\zeta_T}{\zeta_{t'}}\right].$$

Exercise 5.2 Verify (5.16).

Exercise 5.3 Verify (5.18) by applying Itô's Lemma on the relation $\tilde{\zeta}_t = \zeta_t / I_t$.

Exercise 5.4 The purpose of this exercise is to show that the claim of the gross return pure expectation hypothesis is inconsistent with interest rate uncertainty. In the following we consider time points $t_0 < t_1 < t_2$.

(a) Show that if the hypothesis holds, then

$$\frac{1}{B_{t_0}^{t_1}} = \frac{1}{B_{t_0}^{t_2}} \mathrm{E}_{t_0}\left[B_{t_1}^{t_2}\right].$$

Hint: Compare two investment strategies over the period $[t_0, t_1]$. The first strategy is to buy at time t_0 zero-coupon bonds maturing at time t_1. The second strategy is to buy at time t_0 zero-coupon bonds maturing at time t_2 and to sell them again at time t_1.

(b) Show that if the hypothesis holds, then

$$\frac{1}{B_{t_0}^{t_2}} = \frac{1}{B_{t_0}^{t_1}} \mathrm{E}_{t_0}\left[\frac{1}{B_{t_1}^{t_2}}\right].$$

(c) Show from the two previous questions that the hypothesis implies that

$$\mathrm{E}_{t_0}\left[\frac{1}{B_{t_1}^{t_2}}\right] = \frac{1}{\mathrm{E}_{t_0}\left[B_{t_1}^{t_2}\right]}. \tag{*}$$

(d) Show that (*) can only hold under full certainty. *Hint: Use Jensen's inequality.*

Exercise 5.5 Go through the derivations in Section 5.5.3.

Exercise 5.6 Constantinides (1992) develops a theory of the nominal term structure of interest rates by specifying exogenously the nominal state-price deflator $\tilde{\zeta}$. In a slightly simplified version, his assumption is that

$$\tilde{\zeta}_t = e^{-gt+(x_t-\alpha)^2},$$

where g and α are constants, and $x = (x_t)$ follows the Ornstein–Uhlenbeck process

$$dx_t = -\kappa x_t \, dt + \sigma \, dz_t,$$

where κ and σ are positive constants with $\sigma^2 < \kappa$ and $z = (z_t)$ is a standard one-dimensional Brownian motion.

(a) Derive the dynamics of the nominal state-price deflator. Express the nominal short-term interest rate, $\tilde{r}_t$, and the nominal market price of risk, $\tilde{\lambda}_t$, in terms of the variable x_t.

(b) Find the dynamics of the nominal short rate.

(c) Find parameter constraints that ensure that the short rate stays positive. *Hint: The short rate is a quadratic function of x. Find the minimum value of this function.*

(d) What is the distribution of x_T given x_t?

(e) Let Y be a normally distributed random variable with mean μ and variance v^2. Show that

$$\mathrm{E}\left[e^{-\gamma Y^2}\right] = (1+2\gamma v^2)^{-1/2} \exp\left\{-\frac{\gamma \mu^2}{1+2\gamma v^2}\right\}.$$

(f) Use the results of the two previous questions to derive the time t price of a nominal zero-coupon bond with maturity T, that is $\tilde{B}_t^T$. It will be an exponential-quadratic function of x_t. What is the yield on this bond?

(g) Find the percentage volatility σ_t^T of the price of the zero-coupon bond maturing at T.

(h) The instantaneous expected excess rate of return on the zero-coupon bond maturing at T is often called the term premium for maturity T. Explain why the term premium is given by $\sigma_t^T \tilde{\lambda}_t$ and show that the term premium can be written as

$$4\sigma^2\alpha^2 \left(1 - F(T-t)\right) \left(\frac{x_t}{\alpha} - 1\right) \left(\frac{x_t}{\alpha} - \frac{1 - F(T-t)e^{\kappa(T-t)}}{1 - F(T-t)}\right),$$

where

$$F(\tau) = \frac{1}{\frac{\sigma^2}{\kappa} + \left(1 - \frac{\sigma^2}{\kappa}\right) e^{2\kappa\tau}}.$$

For which values of x_t will the term premium for maturity T be positive/negative? For a given state x_t, is it possible that the term premium is positive for some maturities and negative for others?

6

Fixed Income Securities

6.1 INTRODUCTION

Modern financial markets offer a large variety of different fixed income securities and, as shown in Section 1.4, the markets for such securities are of an enormous size. In this chapter we will describe and discuss the main fixed income securities more formally. We will specify the payments of these securities and observe some important links between different derivatives. We will also explore what we can conclude about the prices of the securities without specifying any concrete model of the term structure of interest rates, but only imposing the well-accepted no-arbitrage pricing paradigm. From Chapter 4 we know that the absence of arbitrage implies the existence of a risk-neutral probability measure $\mathbb{Q}$ and other equivalent martingale measures. Accordingly, in this chapter we shall state security prices as an expectation under a risk-adjusted probability measure of the properly discounted future payments from the security.

Among other things, we show:

- a put-call parity for European options on bonds, both zero-coupon bonds and coupon bonds,
- that prices of caps and floors follow from prices of portfolios on certain European options on zero-coupon bonds,
- how swap rates are related to the yield curve,
- that prices of European swaptions follow from prices of certain European options on coupon bonds.

Consequently, we can price many frequently traded securities as long as we price bonds and European call options on bonds. In later chapters we can therefore focus on the pricing of these 'basic' securities.

Section 6.2 deals with forwards and futures, Section 6.3 with options, Section 6.4 with caps and floors, and Section 6.5 with swaps and swaptions. Some features of American-style derivatives are discussed in Section 6.6. Finally, Section 6.7 gives a short overview of the pricing models we are going to study in the following chapters and discusses criteria for choosing between the many different models.

6.2 FORWARDS AND FUTURES

6.2.1 General results on forward prices and futures prices

A forward with maturity date T and delivery price K provides a payoff of $P_T - K$ at time T and no other payments. Here P is the underlying variable, typically the price of an asset or a specific interest rate. From the general asset pricing theory of Chapter 4 and in particular Equation (4.9) it follows that the time $t < T$ value of such a payoff can be written as

$$\begin{aligned} V_t &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(P_T - K\right)\right] \\ &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} P_T\right] - K\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right] \\ &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} P_T\right] - K B_t^T, \end{aligned}$$

where the last equality is due to (4.22). For forwards contracted upon at time t, the delivery price K is set so that the value of the forward at time t is zero. This value of K is called the forward price at time t (for the delivery date T) and is denoted by F_t^T. By definition, the forward price for immediate delivery is identical to the value of the underlying, $F_T^T = P_T$. Solving the equation $V_t = 0$, we get that the forward price is given by

$$F_t^T = \frac{\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} P_T\right]}{B_t^T}. \tag{6.1}$$

If the underlying variable is the price of a traded asset with no payments in the period $[t, T]$, we have

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} P_T\right] = P_t,$$

so that the forward price can be written as

$$F_t^T = \frac{P_t}{B_t^T}.$$

By invoking the covariance definition $\mathrm{Cov}[x, y] = \mathrm{E}[xy] - \mathrm{E}[x]\,\mathrm{E}[y]$, we get that

$$\begin{aligned} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} P_T\right] &= \mathrm{Cov}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}, P_T\right] + \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right] \mathrm{E}_t^{\mathbb{Q}}[P_T] \\ &= \mathrm{Cov}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}, P_T\right] + B_t^T\,\mathrm{E}_t^{\mathbb{Q}}[P_T]. \end{aligned}$$

Upon substitution of this into (6.1) we get the following expression for the forward price:

$$F_t^T = \mathrm{E}_t^{\mathbb{Q}}[P_T] + \frac{\mathrm{Cov}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}, P_T\right]}{B_t^T}. \tag{6.2}$$

We can also characterize the forward price in terms of the T-forward martingale measure introduced in Section 4.4.2. The forward price process for contracts with delivery date T is a martingale under the T-forward martingale measure. This is clear from the following considerations. With B_t^T as the numeraire, we have that the forward price F_t^T is set so that

$$\frac{0}{B_t^T} = \mathrm{E}_t^{\mathbb{Q}^T}\left[\frac{P_T - F_t^T}{B_T^T}\right]$$

and hence

$$F_t^T = \mathrm{E}_t^{\mathbb{Q}^T}[P_T] = \mathrm{E}_t^{Q^T}[F_T^T],$$

which implies that the forward price F_t^T is a $\mathbb{Q}^T$-martingale.

Consider now a futures contract with final settlement at time T. The marking-to-market at a given date involves the payment of the change in the so-called futures price of the contract relative to the previous settlement date. Let Φ_t^T be the futures price at time t. The futures price at the settlement time is by definition equal to the price of the underlying security, $\Phi_T^T = P_T$. At maturity of the contract the futures thus gives a payoff equal to the difference between the price of the underlying asset at that date and the futures price at the previous settlement date. After the last settlement before maturity, the futures is therefore indistinguishable from the corresponding forward contract, so the values of the futures and the forward at that settlement date must be identical. At the next-to-last settlement date before maturity, the futures price is set to that value that ensures that the net present value of the upcoming settlement at the last settlement date before maturity (which depends on this futures price) *and* the final payoff is equal to zero. Similarly at earlier settlement dates. We assume for mathematical convenience that the futures is continuously marked-to-market so that over any infinitesimal interval $[t, t + dt]$ it provides a payment of $d\Phi_t^T$. The following theorem characterizes the futures price:

Theorem 6.1 *The futures price Φ_t^T is a martingale under the risk-neutral probability measure $\mathbb{Q}$, so that in particular*

$$\Phi_t^T = \mathrm{E}_t^{\mathbb{Q}}[P_T]. \tag{6.3}$$

Proof: We will prove the theorem by first considering a discrete-time setting in which positions can be changed and the futures contracts marked-to-market at times $t, t + \Delta t, t + 2\Delta t, \ldots, t + N\Delta t \equiv T$. This proof is originally due to Cox et al. (1981b). A proof based on the same idea, but formulated directly in continuous-time, was given by Duffie and Stanton (1992).

The idea is to set up a self-financing strategy that requires an initial investment at time t equal to the futures price Φ_t^T. Hence, at time t, Φ_t^T is invested in the bank account. In addition, $e^{r_t \Delta t}$ futures contracts are acquired (at a price of zero).

At time $t + \Delta t$, the deposit at the bank account has grown to $e^{r_t \Delta t} \Phi_t^T$. The marking-to-market of the futures position yields a payoff of $e^{r_t \Delta t} \left(\Phi_{t+\Delta t}^T - \Phi_t^T \right)$, which is deposited at the bank account, so that the balance of the account becomes $e^{r_t \Delta t} \Phi_{t+\Delta t}^T$. The position in futures is increased (at no extra costs) to a total of $e^{(r_{t+\Delta t} + r_t)\Delta t}$ contracts.

At time $t + 2\Delta t$, the deposit has grown to $e^{(r_{t+\Delta t}+r_t)\Delta t} \Phi_{t+\Delta t}^T$, which together with the marking-to-market payment of $e^{(r_{t+\Delta t}+r_t)\Delta t} \left(\Phi_{t+2\Delta t}^T - \Phi_{t+\Delta t}^T \right)$ gives a total of $e^{(r_{t+\Delta t}+r_t)\Delta t} \Phi_{t+2\Delta t}^T$.

Continuing this way, the balance of the bank account at time $T = t + N\Delta t$ will be

$$e^{\left(r_{t+(N-1)\Delta t} + \cdots + r_t\right)\Delta t} \Phi_{t+N\Delta t}^T = e^{\left(r_{t+(N-1)\Delta t} + \cdots + r_t\right)\Delta t} \Phi_T^T = e^{\left(r_{t+(N-1)\Delta t} + \cdots + r_t\right)\Delta t} P_T.$$

The continuous-time limit of this is $e^{\int_t^T r_u \, du} P_T$. The time t value of this payment is Φ_t^T, since this is the time t investment required to obtain that terminal payment. On the other hand, we can value the time T payment by discounting by $e^{-\int_t^T r_u \, du}$ and taking the risk-neutral expectation. Hence,

$$\Phi_t^T = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^T r_u \, du} \left(e^{\int_t^T r_u \, du} P_T \right) \right] = \mathrm{E}_t^{\mathbb{Q}} [P_T],$$

as was to be shown. □

Comparing with (4.24), we see that we can think of the futures price as the price of a traded asset with a continuous dividend given by the product of the current price and the short-term interest rate.

From (6.2) and (6.3) we get that the difference between the forward price F_t^T and the futures price Φ_t^T is given by

$$F_t^T - \Phi_t^T = \frac{\mathrm{Cov}_t^{\mathbb{Q}} \left[e^{-\int_t^T r_u \, du}, P_T \right]}{B_t^T}. \tag{6.4}$$

The forward price and the futures price will only be identical if the two random variables P_T and $\exp\left(-\int_t^T r_u \, du \right)$ are uncorrelated under the risk-neutral probability measure. In particular, this is true if the short rate r_t is constant or deterministic.

The forward price is larger [smaller] than the futures price if the variables $\exp\left(-\int_t^T r_u \, du \right)$ and P_T are positively [negatively] correlated under the risk-neutral probability measure. An intuitive, heuristic argument for this goes as follows. If the forward price and the futures price are identical, the total undiscounted payments from the futures contract will be equal to the terminal payment of the forward. Suppose the interest rate and the spot price of the underlying asset are positively correlated, which ought to be the case whenever $\exp\left(-\int_t^T r_u \, du \right)$ and P_T are negatively correlated. Then the marking-to-market payments of the futures tend to be positive when the interest rate is high and negative when the interest rate is low. So positive payments can be reinvested at a high interest rate, whereas negative payments can be financed at a low interest rate. With such a correlation, the futures contract is clearly more attractive than a forward contract when the futures price and the forward price are identical. To maintain a zero initial value of both contracts, the futures price has to be larger than the forward price. Conversely, if the sign of the correlation is reversed.

6.2.2 Forwards on bonds

From (6.1) the unique no-arbitrage forward price on a zero-coupon bond is

$$F_t^{T,S} = \frac{B_t^S}{B_t^T}, \tag{6.5}$$

where T is the delivery date of the futures and $S > T$ is the maturity date of the underlying bond. At the delivery time T the gain or loss from the forward position will be known. The gain from a long position in a forward written on a zero-coupon with face value H and maturity at S is equal to $H(B_T^S - F_t^{T,S})$. If we write the spot bond price B_T^S in terms of the spot LIBOR rate l_T^S and the forward bond price $F_t^{T,S}$ in terms of the forward LIBOR rate $L_t^{T,S}$, it follows from (1.2), (1.6), and (6.5) that the gain is equal to

$$\begin{aligned} H\left(B_T^S - F_t^{T,S}\right) &= H\left(\frac{1}{1+(S-T)l_T^S} - \frac{1}{1+(S-T)L_t^{T,S}}\right) \\ &= \frac{(S-T)(L_t^{T,S} - l_T^S)H}{\left(1+(S-T)l_T^S\right)\left(1+(S-T)L_t^{T,S}\right)}. \end{aligned} \tag{6.6}$$

An investor with a long position in the forward will realize a gain if the spot bond price at delivery turns out to be above the forward price, that is if the spot interest rate at delivery turns out to be below the forward interest rate when the forward position was taken. We can think of a short position in a forward on a zero-coupon bond as a way to lock in the borrowing rate for the period between the delivery date of the forward and the maturity date of the bond.

Next, consider a forward on a coupon bond with payments Y_i at time T_i, $i = 1, \dots, n$. The bond price at time t will be $B_t = \sum_{T_i > t} Y_i B_t^{T_i}$, and the unique no-arbitrage forward price is

$$\begin{aligned} F_t^{T,\text{cpn}} &= \frac{\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} B_T\right]}{B_t^T} \\ &= \frac{\sum_{T_i > T} Y_i\, \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} B_T^{T_i}\right]}{B_t^T} \end{aligned}$$

$$
\begin{aligned}
&= \frac{\sum_{T_i>T} Y_i B_t^{T_i}}{B_t^T} \\
&= \frac{B_t - \sum_{t<T_i<T} Y_i B_t^{T_i}}{B_t^T} \\
&= \sum_{T_i>T} Y_i F_t^{T,T_i}.
\end{aligned}
\tag{6.7}
$$

In particular, this relation implies that forward prices on coupon bonds follow from forward prices of zero-coupon bonds.

6.2.3 Interest rate forwards—forward rate agreements

As discussed in Section 1.2, forward interest rates are rates for a future period relative to the time when the rate is set. Many participants in the financial markets may on occasion be interested in 'locking in' an interest rate for a future period, either in order to hedge risk involved with varying interest rates or to speculate in specific changes in interest rates. In the money markets the agents can lock in an interest rate by entering a forward rate agreement (FRA). Suppose the relevant future period is the time interval between T and S, where $S > T$. In principle, a forward rate agreement with a face value H and a contract rate of K involves two payments: a payment of $-H$ at time T and a payment of $H[1 + (S - T)K]$ at time S (of course, the payments to the other part of the agreement are H at time T and $-H[1 + (S - T)K]$ at time S). In practice, the contract is typically settled at time T, so that the two payments are replaced by a single payment of $B_T^S H[1 + (S - T)K] - H$ at time T.

Usually the contract rate K is set so that the present value of the future payment(s) is zero at the time the contract is made. Suppose the contract is made at time $t < T$. Then the time t value of the two future payments of the contract is equal to $-HB_t^T + H[1 + (S - T)K]B_t^S$. This is zero if and only if

$$
K = \frac{1}{S-T}\left(\frac{B_t^T}{B_t^S} - 1\right) = L_t^{T,S},
$$

compare (1.6), that is when the contract rate equals the forward rate prevailing at time t for the period between T and S. For this contract rate, we can think of the forward rate agreement having a single payment at time T, which is given by

$$
B_T^S H[1 + (S-T)K] - H = H\left(\frac{1 + (S-T)L_t^{T,S}}{1 + (S-T)l_T^S} - 1\right) = \frac{(S-T)(L_t^{T,S} - l_T^S)H}{1 + (S-T)l_T^S}.
\tag{6.8}
$$

The numerator is exactly the interest lost by lending out H from time T to time S at the forward rate given by the FRA rather than the realized spot rate. Of course, this amount may be negative, so that a gain is realized. The division by $1 + (S - T)l_T^S$

corresponds to discounting the gain/loss from time S back to time T. The time T value stated in (6.8) is closely related, but not identical, to the gain/loss on a forward on a zero-coupon bond, compare (6.6).

6.2.4 Futures on bonds

Theorem 6.1 implies that the time t futures price for a futures on a zero-coupon bond maturing at time $S > T$ is given by

$$\Phi_t^{T,S} = \mathrm{E}_t^{\mathbb{Q}}\left[B_T^S\right].$$

For a futures on a coupon bond with payments Y_i at time T_i the final settlement is based on the bond price $B_T = \sum_{T_i > T} Y_i B_T^{T_i}$ and hence the futures price is

$$\Phi_t^{T,\mathrm{cpn}} = \mathrm{E}_t^{\mathbb{Q}}\left[\sum_{T_i>T} Y_i B_T^{T_i}\right] = \sum_{T_i>T} Y_i\, \mathrm{E}_t^{\mathbb{Q}}\left[B_T^{T_i}\right] = \sum_{T_i>T} Y_i \Phi_t^{T,T_i}, \tag{6.9}$$

so that the the futures price on a coupon bond is a payment-weighted average of futures prices of the zero-coupon bonds maturing at the payment dates of the coupon bond. In later chapters we can hence focus on futures on zero-coupon bonds.

Since bond prices generally are negatively correlated with interest rates, we expect that the covariance in (6.4) will be positive and, hence, that forward prices on bonds are higher than the corresponding futures prices.

6.2.5 Interest rate futures—Eurodollar futures

Interest rate futures trade with a very high volume at major international exchanges such as CME (the Chicago Mercantile Exchange) and the NYSE-Euronext LIFFE (London International Financial Futures & Options Exchange). We shall simply refer to all these contracts as Eurodollar futures and refer to the underlying interest rate as the 3-month LIBOR spot rate, whose value at time t we denote by $l_t^{t+0.25}$.

The price quotation of Eurodollar futures is somewhat complicated, since the amounts paid in the marking-to-market settlements are not exactly the changes in the quoted futures price. We must therefore distinguish between the **quoted** futures price, $\tilde{\mathcal{E}}_t^T$, and the **actual** futures price, $\mathcal{E}_t^T$, with the settlements being equal to changes in the actual futures price. At the maturity date of the contract, T, the quoted Eurodollar futures price is defined in terms of the prevailing 3-month LIBOR rate according to the relation

$$\tilde{\mathcal{E}}_T^T = 100\left(1 - l_T^{T+0.25}\right), \tag{6.10}$$

which using (1.2) can be rewritten as

$$\tilde{\mathcal{E}}_T^T = 100\left(1 - 4\left(\frac{1}{B_T^{T+0.25}} - 1\right)\right) = 500 - 400\,\frac{1}{B_T^{T+0.25}}.$$

Traders and analysts typically transform the Eurodollar futures price to an interest rate, the so-called **LIBOR futures rate**, which we denote by φ_t^T and define by

$$\varphi_t^T = 1 - \frac{\tilde{\mathcal{E}}_t^T}{100} \quad \Leftrightarrow \quad \tilde{\mathcal{E}}_t^T = 100\left(1 - \varphi_t^T\right).$$

It follows from (6.10) that the LIBOR futures rate converges to the 3-month LIBOR spot rate, as the maturity of the futures contract approaches.

The actual Eurodollar futures price is given by

$$\mathcal{E}_t^T = 100 - 0.25(100 - \tilde{\mathcal{E}}_t^T) = \frac{1}{4}\left(300 + \tilde{\mathcal{E}}_t^T\right) = 100 - 25\varphi_t^T$$

per 100 dollars of nominal value. It is the change in the actual futures price which is exchanged in the marking-to-market settlements. At the CME the nominal value of the Eurodollar futures is 1 million dollars. A quoted futures price of $\tilde{\mathcal{E}}_t^T = 94.47$ corresponds to a LIBOR futures rate of 5.53% and an actual futures price of

$$\frac{1{,}000{,}000}{100} \cdot [100 - 25 \cdot 0.0553] = 986{,}175.$$

If the quoted futures price increases to 94.48 the next day, corresponding to a drop in the LIBOR futures rate of one basis point (0.01 percentage points), the actual futures price becomes

$$\frac{1{,}000{,}000}{100} \cdot [100 - 25 \cdot 0.0552] = 986{,}200.$$

An investor with a long position will therefore receive $986,200 - 986,175 = 25$ dollars at the settlement at the end of that day.

If we simply sum up the individual settlements without discounting them to the terminal date, the total gain on a long position in a Eurodollar futures contract from t to expiration at T is given by

$$\mathcal{E}_T^T - \mathcal{E}_t^T = \left(100 - 25\varphi_T^T\right) - \left(100 - 25\varphi_t^T\right) = -25\left(\varphi_T^T - \varphi_t^T\right)$$

per 100 dollars of nominal value, that is the total gain on a contract with nominal value H is equal to $-0.25\left(\varphi_T^T - \varphi_t^T\right)H$. The gain will be positive if the three-month spot rate at expiration turns out to be below the futures rate when the position was taken. Conversely for a short position. The gain/loss on a Eurodollar futures contract is closely related to the gain/loss on a forward rate agreement, as can be seen from substituting $S = T + 0.25$ into (6.8). Recall that the rates φ_T^T and $l_T^{T+0.25}$ are identical. However, it should be emphasized that in general the futures rate φ_t^T

and the forward rate $L_t^{T,T+0.25}$ will be different due to the marking-to-market of the futures contract.

The final settlement is based on the terminal actual futures price

$$\begin{aligned}\mathcal{E}_T^T &\equiv 100 - 0.25\left(100 - \tilde{\mathcal{E}}_T^T\right)\\ &= 100 - 0.25\left(400\left[(B_T^{T+0.25})^{-1} - 1\right]\right)\\ &= 100\left[2 - (B_T^{T+0.25})^{-1}\right].\end{aligned}$$

It follows from Theorem 6.1 that the actual futures price at any earlier point in time t can be computed as

$$\mathcal{E}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[\mathcal{E}_T^T\right] = 100\left(2 - \mathrm{E}_t^{\mathbb{Q}}\left[(B_T^{T+0.25})^{-1}\right]\right).$$

The quoted futures price is therefore

$$\tilde{\mathcal{E}}_t^T = 4\mathcal{E}_t^T - 300 = 500 - 400\,\mathrm{E}_t^{\mathbb{Q}}\left[(B_T^{T+0.25})^{-1}\right]. \tag{6.11}$$

6.3 EUROPEAN OPTIONS

In this section, we focus on European options. Some aspects of American options are discussed in Section 6.6.

6.3.1 General pricing results for European options

We can use the idea of changing the numeraire and the probability measure to obtain a general characterization of the price of a European call option. Let T be the expiry date and K the exercise price of the option, so that the option payoff at time T is of the form

$$C_T = \max(P_T - K, 0).$$

For an option on a traded asset, P_T is the price of the underlying asset at the option expiry date. For an option on a given interest rate, P_T denotes the value of this interest rate at the expiry date. According to (4.19) the time t price of the option is

$$C_t = B_t^T\,\mathrm{E}_t^{\mathbb{Q}^T}\left[\max(P_T - K, 0)\right], \tag{6.12}$$

where $\mathbb{Q}^T$ is the T-forward martingale measure, that is the pricing measure associated with using the zero-coupon bond maturing at T as the numeraire. We can rewrite the payoff as

$$C_T = (P_T - K)\,\mathbf{1}_{\{P_T>K\}},$$

where $\mathbf{1}_{\{P_T>K\}}$ is the indicator for the event $P_T > K$. This indicator is a random variable whose value will be 1 if the realized value of P_T turns out to be larger than K and the value is 0 otherwise. Hence, the option price can be rewritten as[1]

$$\begin{aligned}
C_t &= B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[(P_T - K)\mathbf{1}_{\{P_T>K\}} \right] \\
&= B_t^T \left(\mathrm{E}_t^{\mathbb{Q}^T} \left[P_T \mathbf{1}_{\{P_T>K\}} \right] - K \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\mathbf{1}_{\{P_T>K\}} \right] \right) \\
&= B_t^T \left(\mathrm{E}_t^{\mathbb{Q}^T} \left[P_T \mathbf{1}_{\{P_T>K\}} \right] - K \mathbb{Q}_t^T (P_T > K) \right) \\
&= B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[P_T \mathbf{1}_{\{P_T>K\}} \right] - K B_t^T \mathbb{Q}_t^T (P_T > K).
\end{aligned}$$

Here $\mathbb{Q}_t^T(P_T > K)$ denotes the probability (using the probability measure $\mathbb{Q}^T$) of $P_T > K$ given the information known at time t. This can be interpreted as the probability of the option finishing in-the-money, computed in a hypothetical forward-risk-neutral world.

For an option on a traded asset we can rewrite the first term in the above pricing formula, since P_t is then a valid numeraire with a corresponding probability measure $\mathbb{Q}^P$. Applying (4.18) for both the numeraires B_t^T and P_t, we get

$$\begin{aligned}
B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[P_T \mathbf{1}_{\{P_T>K\}} \right] &= P_t \, \mathrm{E}_t^{\mathbb{Q}^P} \left[\mathbf{1}_{\{P_T>K\}} \right] \\
&= P_t \mathbb{Q}_t^P (P_T > K).
\end{aligned}$$

This assumes that the underlying asset pays no dividends in the interval $[t, T]$. The call price is therefore

$$C_t = P_t \mathbb{Q}_t^P (P_T > K) - K B_t^T \mathbb{Q}_t^T (P_T > K). \tag{6.13}$$

Both probabilities in this formula show the probability of the option finishing in-the-money, but under two different probability measures. To compute the price of the European call option in a concrete model we 'just' have to compute these probabilities. In some cases, however, it is easier to work directly on (6.12).

For a put option the analogous result is

$$\pi_t = K B_t^T \mathbb{Q}_t^T (P_T \leq K) - P_t \mathbb{Q}_t^P (P_T \leq K). \tag{6.14}$$

[1] Since the indicator $\mathbf{1}_{\{P_T>K\}}$ takes on the value 1 or 0; its expected value equals 1 times the probability that it will have the value 1 plus 0 times the probability that it will have the value 0, that is the expected value equals the probability that the indicator will have the value 1, which again equals the probability that $P_T > K$.

We can now also derive a general put-call parity for European options. Combining (6.13) and (6.14) we get

$$\begin{aligned} C_t - \pi_t &= P_t\big(\mathbb{Q}_t^P(P_T > K) + \mathbb{Q}_t^P(P_T \leq K)\big) - KB_t^T\left(\mathbb{Q}_t^T(P_T > K) + \mathbb{Q}_t^T(P_T \leq K)\right) \\ &= P_t - KB_t^T \end{aligned}$$

so that

$$C_t + KB_t^T = \pi_t + P_t. \tag{6.15}$$

We note again that this assumes that the underlying asset provides no dividends in the interval $[t, T]$, otherwise the time t value of these intermediate payments must be subtracted from P_t in the above equation. A consequence of the put–call parity is that we can focus on the pricing of European call options. The prices of European put options will then follow immediately.

The put–call parity can also be shown using the following simple replication argument. A portfolio consisting of a call option and K zero-coupon bonds maturing at the same time as the option yields a payoff at time T of

$$\max(P_T - K, 0) + K = \max(P_T, K)$$

and will have a current time t price given by the left-hand side of (6.15). Another portfolio consisting of a put option and one unit of the underlying asset has a time T value of

$$\max(K - P_T, 0) + P_T = \max(K, P_T)$$

and a time t price corresponding to the right-hand side of (6.15). Therefore, there will be an obvious arbitrage opportunity unless (6.15) is satisfied.

6.3.2 Options on bonds

Turning to options on bonds, we will first consider options on zero-coupon bonds although, apparently, no such options are traded at any exchange. However, we shall see later that other, frequently traded, fixed income securities can be considered as portfolios of European options on zero-coupon bonds. This is true for caps and floors, which we turn to in Section 6.4. We will also show later that, under certain assumptions on the dynamics of interest rates, any European option on a coupon bond is equivalent to a portfolio of certain European options on zero-coupon bonds; see Chapter 7. For these reasons, it is important to be able to price European options on zero-coupon bonds.

Let us first fix some notation. The time of maturity of the option is denoted by T. The underlying zero-coupon bond gives a payment of 1 (dollar) at time S, where $S \geq T$. The exercise price of the option is denoted by K. We let $C_t^{K,T,S}$ denote the time t price of such a European call option. At maturity the value of the call equals its payoff:

$$C_T^{K,T,S} = \max\left(B_T^S - K, 0\right).$$

We let $\pi_t^{K,T,S}$ denote the time t price of a similar put option. The value at maturity is equal to

$$\pi_T^{K,T,S} = \max\left(K - B_T^S, 0\right).$$

Note that only options with an exercise price between 0 and 1 are interesting, since the price of the underlying zero-coupon bond at expiry of the option will be in this interval, assuming non-negative interest rates.

From the general option pricing results derived above, we can conclude that the call price can be written as

$$C_t^{K,T,S} = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max\left(B_T^S - K, 0\right)\right] \tag{6.16}$$

and as

$$C_t^{K,T,S} = B_t^S \mathbb{Q}_t^S(B_T^S > K) - K B_t^T \mathbb{Q}_t^T(B_T^S > K), \tag{6.17}$$

where $\mathbb{Q}^S$ denotes the S-forward martingale measure and $\mathbb{Q}^T$, as before, is the T-forward martingale measure. We will use these equations in later chapters to derive closed-form option pricing formulas in specific models of the term structure of interest rates. The probabilities in (6.17) will be determined by the precise assumptions of the model. The put–call parity for European options on zero-coupon bonds is

$$C_t^{K,T,S} + K B_t^T = \pi_t^{K,T,S} + B_t^S.$$

Next, consider options on coupon bonds. Assume that the underlying coupon bond has payments Y_i at time T_i ($i = 1, 2, \dots, n$), where $T_1 < T_2 < \cdots < T_n$. Let B_t denote the time t price of this bond, that is

$$B_t = \sum_{T_i > t} Y_i B_t^{T_i}.$$

Let $C_t^{K,T,\mathrm{cpn}}$ and $\pi_t^{K,T,\mathrm{cpn}}$ denote the time t prices of a European call and a European put, respectively, expiring at time T, having an exercise price of K and the coupon bond above as the underlying asset. Of course, we must have that $T < T_n$. The time T values of the options are given by their payoffs:

$$C_T^{K,T,\mathrm{cpn}} = \max\left(B_T - K, 0\right) = \max\left(\sum_{T_i > T} Y_i B_T^{T_i} - K, 0\right),$$

$$\pi_T^{K,T,\mathrm{cpn}} = \max\left(K - B_T, 0\right) = \max\left(K - \sum_{T_i > T} Y_i B_T^{T_i}, 0\right).$$

Such options are only interesting if the exercise price is positive and less than $\sum_{T_i>T} Y_i$, which is the upper bound for B_T with non-negative forward rates. Note that (i) only the payments of the bonds after maturity of the option are relevant for the payoff and the value of the option;[2] (ii) we have assumed that the payoff of the option is determined by the difference between the exercise price and the **true** bond price rather than the **quoted** bond price. The true bond price is the sum of the quoted bond price and accrued interest.[3] Some aspects of options on the quoted bond price are discussed by Munk (2002).

The general pricing formula for options implies that the price of a European call on a coupon bond can be written as

$$C_t^{K,T,\mathrm{cpn}} = \left(B_t - \sum_{T_i \in (t,T]} Y_i B_t^{T_i} \right) \mathbb{Q}_t^B \left(B_T > K \right) - K B_t^T \mathbb{Q}_t^T \left(B_T > K \right).$$

Here $\mathbb{Q}^B$ indicates the martingale measure corresponding to using as a numeraire the present value of the payments of the underlying coupon bond after the option expiry date. Note that the first term on the right-hand side is the present value of the payments of the underlying bond that comes after the option maturity date. The put–call parity for European options on coupon bonds is as follows:

$$C_t^{K,T,\mathrm{cpn}} + K B_t^T = \pi_t^{K,T,\mathrm{cpn}} + B_t - \sum_{t<T_i\leq T} Y_i B_t^{T_i}. \tag{6.18}$$

In Exercise 6.2 you are asked to give a replication argument supporting (6.18).

We cannot derive unique option prices without making concrete assumptions about the dynamics of the underlying asset and interest rates. But using the no-arbitrage principle only, we can derive bounds on option prices. Merton (1973) derived well-known bounds on the prices of European options on stocks, which are know reproduced in many option pricing textbooks, for example Hull (2009). The bounds that can be obtained for bond options are not just a simple reformulation of the bounds available for stock options due to

- the close relation between the appropriate discount factor and the price of the underlying asset,
- the existence of an upper bound on the price of the underlying bond: under the reasonable assumption that all forward rates are non-negative, the price of a bond will be less than or equal to the sum of its remaining payments.

Although the obtainable bounds for bond options are tighter than those for stock options, they still leave quite a large interval in which the price can lie. For proofs and examples see Munk (2002) and Exercise 6.2.

[2] In particular, we assume that in the case where the expiry date of the option coincides with a payment date of the underlying bond, it is the bond price excluding that payment which determines the payoff of the option.

[3] The quoted price is sometimes referred to as the *clean* price. Similarly, the true price is sometimes called the **dirty** price.

6.3.3 Black's formula for bond options

Practitioners often use Black–Scholes–Merton type formulas for interest rate derivatives. The formulas are based on the Black (1976) variant of the Black–Scholes–Merton model developed for stock option pricing, compare Section 4.8. Black's formula for a European call option on a bond is

$$\begin{aligned} C_t^{K,T,\mathrm{cpn}} &= B_t^T \left[F_t^{T,\mathrm{cpn}} N\left(\hat{d}_1(F_t^{T,\mathrm{cpn}}, t)\right) - KN\left(\hat{d}_2(F_t^{T,\mathrm{cpn}}, t)\right)\right], \\ &= \left(B_t - \sum_{t<T_i<T} Y_i B_t^{T_i} \right) N\left(\hat{d}_1(F_t^{T,\mathrm{cpn}}, t)\right) - KB_t^T N\left(\hat{d}_2(F_t^{T,\mathrm{cpn}}, t)\right), \end{aligned}$$

where $F_t^{T,\mathrm{cpn}}$ is the forward price of the bond, and

$$\hat{d}_1(F,t) = \frac{\ln(F/K)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}, \tag{6.19}$$

$$\hat{d}_2(F,t) = \frac{\ln(F/K)}{\sigma\sqrt{T-t}} - \frac{1}{2}\sigma\sqrt{T-t} = \hat{d}_1(F,t) - \sigma\sqrt{T-t}. \tag{6.20}$$

As discussed briefly in Section 4.8, the use of Black's formula for interest rate derivatives is generally not theoretically supported and may lead to pricing allowing arbitrage. To ensure consistent arbitrage-free pricing of fixed income securities we have to model the dynamics of the entire term structure of interest rates.

6.4 CAPS, FLOORS, AND COLLARS

6.4.1 Caps

An **(interest rate) cap** is designed to protect an investor who has borrowed funds on a floating interest rate basis against the risk of paying very high interest rates. Suppose the loan has a face value of H and payment dates $T_1 < T_2 < \cdots < T_n$, where $T_{i+1} - T_i = \delta$ for all i.[4] The interest rate to be paid at time T_i is determined by the δ-period money market interest rate prevailing at time $T_{i-1} = T_i - \delta$ so that the payment at time T_i is equal to $H\delta l_{T_i-\delta}^{T_i}$. Note that the interest rate is set at the beginning of the period, but paid at the end. Define $T_0 = T_1 - \delta$. The dates $T_0, T_1, \dots, T_{n-1}$ where the rate for the coming period is determined are called the **reset dates** of the loan.

A cap with a face value of H, payment dates T_i ($i = 1, \dots, n$) as above, and a so-called cap rate K yields a time T_i payoff of $H\delta \max(l_{T_i-\delta}^{T_i} - K, 0)$, for $i = 1, 2, \dots, n$. If a borrower buys such a cap, the net payment at time T_i cannot exceed $H\delta K$.

[4] In practice, there will not be exactly the same number of days between successive reset dates, and the calculations below must be slightly adjusted by using the relevant **day count convention**.

The period length δ is often referred to as the **frequency** or the **tenor** of the cap.[5] In practice, the frequency is typically either 3, 6, or 12 months. Note that the time distance between payment dates coincides with the 'maturity' of the floating interest rate. Also note that while a cap is tailored for interest rate hedging, it can also be used for interest rate speculation.

A cap can be seen as a portfolio of n **caplets**, namely one for each payment date of the cap. The i'th caplet yields a payoff at time T_i of

$$C^i_{T_i} = H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right) \tag{6.21}$$

and no other payments. A caplet is a call option on the spot interest rate prevailing at time $T_i - \delta$ for a period of length δ, but where the payment takes place at time T_i although it is already fixed at time $T_i - \delta$.

In the following we will find the value of the i'th caplet before time T_i. Since the payoff becomes known at time $T_i - \delta$, we can obtain its value in the interval between $T_i - \delta$ and T_i by a simple discounting of the payoff, that is

$$C^i_t = B^{T_i}_t H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right), \quad T_i - \delta \le t \le T_i.$$

In particular,

$$C^i_{T_i-\delta} = B^{T_i}_{T_i-\delta} H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right). \tag{6.22}$$

To find the value before the fixing of the payoff, that is for $t < T_i - \delta$, we shall use two strategies. The first is simply to take relevant expectations of the payoff. Since the payoff comes at T_i, we know from Section 4.4.2 that the value of the payoff can be found as the product of the expected payoff computed under the T_i-forward martingale measure and the current discount factor for time T_i payments, that is

$$C^i_t = H\delta B^{T_i}_t \, \mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(l^{T_i}_{T_i-\delta} - K, 0\right)\right], \quad t < T_i - \delta.$$

The price of a cap can therefore be determined as

$$C_t = H\delta \sum_{i=1}^{n} B^{T_i}_t \, \mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(l^{T_i}_{T_i-\delta} - K, 0\right)\right], \quad t < T_0.$$

In Chapter 11 we will look at a class of models that prices caps by directly modelling the dynamics of the forward rates $L^{T_i-\delta,T_i}_t$ under the relevant $\mathbb{Q}^{T_i}$ probability measures, which will imply the $\mathbb{Q}^{T_i}$-distribution of the spot rates $l^{T_i}_{T_i-\delta} = L^{T_i-\delta,T_i}_{T_i-\delta}$ needed to compute the expectations above.

[5] The word tenor is sometimes used for the set of payment dates $T_1, \ldots, T_n$.

The second pricing strategy links caps to bond options. Applying (1.2), we can rewrite (6.22) as

$$\begin{aligned}
\mathcal{C}^i_{T_i-\delta} &= B^{T_i}_{T_i-\delta} H \max\left(1 + \delta l^{T_i}_{T_i-\delta} - [1+\delta K], 0\right) \\
&= B^{T_i}_{T_i-\delta} H \max\left(\frac{1}{B^{T_i}_{T_i-\delta}} - [1+\delta K], 0\right) \\
&= H(1+\delta K) \max\left(\frac{1}{1+\delta K} - B^{T_i}_{T_i-\delta}, 0\right).
\end{aligned}$$

We can now see that the value at time $T_i - \delta$ is identical to the payoff of a European put option expiring at time $T_i - \delta$ that has an exercise price of $1/(1+\delta K)$ and is written on a zero-coupon bond maturing at time T_i. To rule out arbitrage, the value of the i'th caplet at an earlier point in time $t \leq T_i - \delta$ must equal the value of that put option. With the notation used earlier we thus have

$$\mathcal{C}^i_t = H(1+\delta K)\pi_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}.$$

To find the value of the entire cap contract we simply have to add up the values of all the caplets corresponding to the remaining payment dates of the cap. Before the first reset date, T_0, none of the cap payments are known, so the value of the cap is given by

$$\mathcal{C}_t = \sum_{i=1}^{n} \mathcal{C}^i_t = H(1+\delta K) \sum_{i=1}^{n} \pi_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad t < T_0.$$

At all dates after the first reset date, the next payment of the cap will already be known. If we use the notation $T_{i(t)}$ for the nearest following payment date after time t, as in Section 1.2.5, the value of the cap at any time t in $[T_0, T_n]$ (exclusive of any payment received exactly at time t) can be written as

$$\begin{aligned}
\mathcal{C}_t = {} & H B_t^{T_{i(t)}} \delta \max\left(l^{T_{i(t)}}_{T_{i(t)}-\delta} - K, 0\right) \\
& + (1+\delta K) H \sum_{i=i(t)+1}^{n} \pi_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad T_0 \leq t \leq T_n.
\end{aligned}$$

If $T_{n-1} < t < T_n$, we have $i(t) = n$, and there will be no terms in the sum, which is then considered to be equal to zero. In later chapters we will discuss models for pricing bond options. From the results above, cap prices will follow from prices of European puts on zero-coupon bonds.

Note that the interest rates and the discount factors appearing in the expressions above are taken from the money market, not from the government bond market. Also note that since caps and most other contracts related to money market rates trade OTC, one should take the default risk of the two parties into account when valuing the cap. Here, default simply means that the party cannot pay the amounts promised in the contract. Official money market rates and the associated discount function apply to loan and deposit arrangements between large financial institutions, and thus they reflect the default risk of these corporations. If the parties in an OTC transaction have a default risk significantly different from that, the discount rates in the formulas should be adjusted accordingly. However, it is quite complicated to do that in a theoretically correct manner, so we will not discuss this issue any further at this point.

6.4.2 Floors

An **(interest rate) floor** is designed to protect an investor who has lent funds on a floating rate basis against receiving very low interest rates. The contract is constructed just as a cap except that the payoff at time T_i $(i = 1, \dots, n)$ is given by

$$\mathcal{F}^i_{T_i} = H\delta \max\left(K - l^{T_i}_{T_i-\delta}, 0\right), \tag{6.23}$$

where K is called the floor rate. Buying an appropriate floor, an investor who has provided another investor with a floating rate loan will in total at least receive the floor rate. Of course, an investor can also speculate in low future interest rates by buying a floor or speculate in high future interest rates by selling a floor. The (hypothetical) contracts that only yield one of the payments in (6.23) are called **floorlets**. Obviously, we can think of a floorlet as a European put on the floating interest rate with delayed payment of the payoff.

Analogously to the analysis for caps, we can price the floor directly as

$$\mathcal{F}_t = H\delta \sum_{i=1}^{n} B^{T_i}_t \, \mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(K - l^{T_i}_{T_i-\delta}, 0\right)\right], \quad t < T_0,$$

which is the approach taken in the models studied in Chapter 11. Alternatively, we can express the floorlet as a European call on a zero-coupon bond, and hence a floor is equivalent to a portfolio of European calls on zero-coupon bonds. More precisely, the value of the i'th floorlet at time $T_i - \delta$ is

$$\mathcal{F}^i_{T_i-\delta} = H(1+\delta K) \max\left(B^{T_i}_{T_i-\delta} - \frac{1}{1+\delta K}, 0\right).$$

The total value of the floor contract at any time $t < T_0$ is therefore given by

$$\mathcal{F}_t = H(1+\delta K) \sum_{i=1}^{n} C_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad t < T_0,$$

and later the value is

$$\begin{aligned}\mathcal{F}_t = {} & H B_t^{T_{i(t)}} \delta \max\left(K - l^{T_{i(t)}}_{T_{i(t)}-\delta}, 0\right) \\ & + (1+\delta K) H \sum_{i=i(t)+1}^{n} C_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad T_0 \le t \le T_n.\end{aligned}$$

6.4.3 Black's formula for caps and floors

Black's formula for the caplet price is

$$C_t^i = H\delta B_t^{T_i}\left[L_t^{T_i-\delta,T_i} N\left(\hat{d}_1^i(L_t^{T_i-\delta,T_i},t)\right) - KN\left(\hat{d}_2^i(L_t^{T_i-\delta,T_i},t)\right)\right], \quad t < T_i - \delta, \tag{6.24}$$

where the functions $\hat{d}_1^i$ and $\hat{d}_2^i$ are given by

$$\hat{d}_1^i(L_t^{T_i-\delta,T_i},t) = \frac{\ln(L_t^{T_i-\delta,T_i}/K)}{\sigma_i\sqrt{T_i-\delta-t}} + \frac{1}{2}\sigma_i\sqrt{T_i-\delta-t},$$

$$\hat{d}_2^i(L_t^{T_i-\delta,T_i},t) = \hat{d}_1^i(L_t^{T_i-\delta,T_i},t) - \sigma_i\sqrt{T_i-\delta-t}.$$

Again, the price for the entire cap is obtained by summation. For a floor the corresponding formula is

$$\mathcal{F}_t = H\delta\sum_{i=1}^{n} B_t^{T_i}\left[KN\left(-\hat{d}_2^i(L_t^{T_i-\delta,T_i},t)\right) - L_t^{T_i-\delta,T_i} N\left(-\hat{d}_1^i(L_t^{T_i-\delta,T_i},t)\right)\right], \quad t \le T_0.$$

In Chapter 11 we will consider some very special term structure models that indeed support the use of Black's formula at least for some caps and floors.

The prices of stock options are often expressed in terms of implicit volatilities. The implicit volatility for a given European call option on a stock is the value of σ which by substitution into the Black–Scholes–Merton formula (4.28), together with the observable variables S_t, r, K, and $T-t$, yields a price equal to the observed market price. Similarly, prices of caps, floors, and swaptions can be expressed in terms of implicit interest rate volatilities computed with reference to the Black pricing formula. According to (6.24) different σ-values must be applied for each caplet in a cap. For a cap with more than one remaining payment date, many combinations of the σ_i's will result in the same cap price. If we require that all the σ_i's must be equal, only one common value will result in the market price. This value is called the implicit **flat volatility** of the cap. If caps with different maturities, but the same frequency and overlapping payment dates, are traded, a term structure of volatilities, $\sigma_1, \sigma_2, \dots, \sigma_n$, can be derived. For example, if a 1-year and a 2-year cap on the 1-year LIBOR rate are traded, the unique value of σ_1 that makes Black's price equal to the market price of the 1-year cap can be determined. Next, by applying this value of σ_1, a unique value of σ_2 can be determined so that the Black price and the market price of the 2-year cap are identical. The volatilities σ_i determined by this procedure are called implicit **spot volatilities**.

A graph of the spot volatilities as a function of the maturity, that is σ_i as a function of $T_i-\delta$, will usually be a humped curve, that is an increasing curve for maturities up to 2–3 years and then a decreasing curve for longer maturities.[6] A similar, though slightly flatter, curve is obtained by depicting the flat volatilities as a function of the maturity of the cap, since flat volatilities are averages of spot

[6] See the examples and discussions in Brigo and Mercurio (2006).

volatilities. The picture is the same whether implicit or historical forward rate volatilities are used.

6.4.4 Collars

A **collar** is a contract designed to ensure that the interest rate payments on a floating rate borrowing arrangement stay between two pre-specified levels. A collar can be seen as a portfolio of a long position in a cap with a cap rate K_c and a short position in a floor with a floor rate of $K_f < K_c$ (and the same payment dates and underlying floating rate). The payoff of a collar at time T_i, $i = 1, 2, \dots, n$, is thus

$$\begin{aligned}\mathcal{L}^i_{T_i} &= H\delta\left[\max\left(l^{T_i}_{T_i-\delta} - K_c, 0\right) - \max\left(K_f - l^{T_i}_{T_i-\delta}, 0\right)\right] \\ &= \begin{cases} -H\delta\left[K_f - l^{T_i}_{T_i-\delta}\right], & \text{if } l^{T_i}_{T_i-\delta} \le K_f, \\ 0, & \text{if } K_f \le l^{T_i}_{T_i-\delta} \le K_c, \\ H\delta\left[l^{T_i}_{T_i-\delta} - K_c\right], & \text{if } K_c \le l^{T_i}_{T_i-\delta}. \end{cases}\end{aligned}$$

The value of a collar with cap rate K_c and floor rate K_f is of course given by

$$\mathcal{L}_t(K_c, K_f) = \mathcal{C}_t(K_c) - \mathcal{F}_t(K_f),$$

where the expressions for the values of caps and floors derived earlier can be substituted in.

An investor who has borrowed funds on a floating rate basis will, by buying a collar, ensure that the paid interest rate always lies in the interval between K_f and K_c. Clearly, a collar gives cheaper protection against high interest rates than a cap (with the same cap rate K_c), but on the other hand the full benefits of very low interest rates are sacrificed. In practice, K_f and K_c are often set such that the value of the collar is zero at the inception of the contract.

6.4.5 Exotic caps and floors

Above we considered standard, **plain vanilla** caps, floors, and collars. In addition to these instruments, several contracts trade on the international OTC markets with cash flows that are similar to plain vanilla contracts, but deviate in one or more aspects. The deviations complicate the pricing methods considerably. Let us briefly look at a few of these exotic securities. For further details and other examples of non-standard contracts, see Musiela and Rutkowski (1997, Ch. 16) and Brigo and Mercurio (2006, Ch. 13).

- A **bounded cap** is like an ordinary cap except that the cap owner will only receive the scheduled payoff if the sum of the payments received so far from the contract does not exceed a certain pre-specified level. Consequently, the ordinary cap payments in (6.21) are to be multiplied with an indicator function. The payoff at the end of a given period will depend not only on the interest rate in the beginning of the period, but also on previous interest rates. As many other exotic instruments, a bounded cap is therefore a path-dependent asset.

- A **dual strike cap** is similar to a cap with a cap rate of K_1 in periods when the underlying floating rate $l_t^{t+\delta}$ stays below a pre-specified level $\hat{l}$, and similar to a cap with a cap rate of K_2, where $K_2 > K_1$, in periods when the floating rate is above $\hat{l}$.
- A **cumulative cap** ensures that the accumulated interest rate payments do not exceed a given level.
- A **knock-out cap** will at any time T_i give the standard payoff in (6.21) unless the floating rate $l_t^{t+\delta}$ during the period $[T_i - \delta, T_i]$ has exceeded a certain level. In that case the payoff is zero.

Options on caps and floors are also traded. Since caps and floors themselves are (portfolios of) options, the options on caps and floors are so-called **compound options**. An option on a cap is called a **caption** and provides the holder with the right at a future point in time, T_0, to enter into a cap starting at time T_0 (with payment dates $T_1, \dots, T_n$) against paying a given exercise price.

6.5 SWAPS AND SWAPTIONS

6.5.1 Swaps

Many different types of swaps are traded on the OTC markets, for example currency swaps, credit swaps, and asset swaps, but in line with the theme of this chapter we will focus on interest rate swaps. An **(interest rate) swap** is an exchange of two cash flow streams that are determined by certain interest rates. In the simplest and most common interest rate swap, a **plain vanilla swap**, two parties exchange a stream of fixed interest rate payments and a stream of floating interest rate payments. The payments are in the same currency and are computed from the same (hypothetical) face value or notional principal. The floating rate is usually a money market rate, for example a LIBOR rate, possibly augmented or reduced by a fixed margin. The fixed interest rate is usually set so that the swap has zero net present value when the parties agree on the contract. While the two parties can agree upon any maturity, most interest rate swaps have a maturity between 2 and 10 years.

Let us briefly look at the uses of interest rate swaps. An investor can transform a floating rate loan into a fixed rate loan by entering into an appropriate swap, where the investor receives floating rate payments (netting out the payments on the original loan) and pays fixed rate payments. This is called a **liability transformation**. Conversely, an investor who has lent money at a floating rate, that is owns a floating rate bond, can transform this to a fixed rate bond by entering into a swap, where he pays floating rate payments and receives fixed rate payments. This is an **asset transformation**. Hence, interest rate swaps can be used for hedging interest rate risk on both (certain) assets and liabilities. On the other hand, interest rate swaps can also be used for taking advantage of specific expectations of future interest rates, that is for speculation.

Swaps are often said to allow the two parties to exploit their **comparative advantages** in different markets. Concerning interest rate swaps, this argument presumes that one party has a comparative advantage (relative to the other party) in the market for fixed rate loans, while the other party has a comparative advantage (relative to the first party) in the market for floating rate loans. However, these markets are integrated, and the existence of comparative advantages conflicts with modern financial theory and the efficiency of the money markets. Apparent comparative advantages can be due to differences in default risk premia. For details we refer the reader to the discussion in Hull, 2009 (Ch. 7).

Next, we will discuss the valuation of swaps. As for caps and floors, we assume that both parties in the swap have a default risk corresponding to the 'average default risk' of major financial institutions reflected by the money market interest rates. For a description of the impact on the payments and the valuation of swaps between parties with different default risk, see Duffie and Huang (1996) and Huge and Lando (1999). Furthermore, we assume that the fixed rate payments and the floating rate payments occur at exactly the same dates throughout the life of the swap. This is true for most, but not all, traded swaps. For some swaps, the fixed rate payments only occur once a year, whereas the floating rate payments are quarterly or semi-annual. The analysis below can easily be adapted to such swaps.

In a plain vanilla interest rate swap, one party pays a stream of fixed rate payments and receives a stream of floating rate payments. This party is said to have a pay fixed, receive floating swap, or a fixed-for-floating swap, or simply a **payer swap**. The counterpart receives a stream of fixed rate payments and pays a stream of floating rate payments. This party is said to have a pay floating, receive fixed swap, or a floating-for-fixed swap, or simply a **receiver swap**. Note that the names payer swap and receiver swap refer to the fixed rate payments.

We consider a swap with payment dates $T_1, \ldots, T_n$, where $T_{i+1} - T_i = \delta$. The floating interest rate determining the payment at time T_i is the money market (LIBOR) rate $l_{T_i-\delta}^{T_i}$. In the following we assume that there is no fixed extra margin on this floating rate. If there were such an extra charge, the value of the part of the flexible payments that is due to the extra margin could be computed in the same manner as the value of the fixed rate payments of the swap, see below. We refer to $T_0 = T_1 - \delta$ as the starting date of the swap. As for caps and floors, we call $T_0, T_1, \ldots, T_{n-1}$ the reset dates, and δ the frequency or the tenor. Typical swaps have δ equal to 0.25, 0.5, or 1 corresponding to quarterly, semi-annual, or annual payments and interest rates.

We will find the value of an interest rate swap by separately computing the value of the fixed rate payments (V^{fix}) and the value of the floating rate payments (V^{fl}). The fixed rate is denoted by K. This is a nominal, annual interest rate, so that the fixed rate payments equal $HK\delta$, where H is the notional principal or face value (which is not swapped). The value of the remaining fixed payments is simply

$$V_t^{\text{fix}} = \sum_{i=i(t)}^{n} HK\delta B_t^{T_i} = HK\delta \sum_{i=i(t)}^{n} B_t^{T_i}. \tag{6.25}$$

The floating rate payments are exactly the same as the coupon payments on a floating rate bond, which was discussed in Section 1.2.5, that is at time T_i ($i = 1, 2, \dots, n$) the payment is $H\delta l^{T_i}_{T_i-\delta}$. Note that this payment is already known at time $T_i - \delta$. According to (1.13), the value of such a floating bond at any time $t \in [T_0, T_n)$ is given by $H(1 + \delta l^{T_{i(t)}}_{T_{i(t)}-\delta})B_t^{T_{i(t)}}$. Since this is the value of both the coupon payments and the final repayment of face value, the value of the coupon payments only must be

$$\begin{aligned} V_t^{\mathrm{fl}} &= H(1 + \delta l^{T_{i(t)}}_{T_{i(t)}-\delta})B_t^{T_{i(t)}} - HB_t^{T_n} \\ &= H\delta l^{T_{i(t)}}_{T_{i(t)}-\delta} B_t^{T_{i(t)}} + H\left[B_t^{T_{i(t)}} - B_t^{T_n}\right], \quad T_0 \le t < T_n. \end{aligned}$$

At and before time T_0, the first term is not present, so the value of the floating rate payments is simply

$$V_t^{\mathrm{fl}} = H\left[B_t^{T_0} - B_t^{T_n}\right], \quad t \le T_0. \tag{6.26}$$

We will also develop an alternative expression for the value of the floating rate payments of the swap. The time $T_i - \delta$ value of the coupon payment at time T_i is

$$H\delta l^{T_i}_{T_i-\delta} B^{T_i}_{T_i-\delta} = H\delta \frac{l^{T_i}_{T_i-\delta}}{1 + \delta l^{T_i}_{T_i-\delta}},$$

where we have applied (1.2). Consider a strategy of buying a zero-coupon bond with face value H maturing at $T_i - \delta$ and selling a zero-coupon bond with the same face value H but maturing at T_i. The time $T_i - \delta$ value of this position is

$$HB^{T_i-\delta}_{T_i-\delta} - HB^{T_i}_{T_i-\delta} = H - \frac{H}{1 + \delta l^{T_i}_{T_i-\delta}} = H\delta \frac{l^{T_i}_{T_i-\delta}}{1 + \delta l^{T_i}_{T_i-\delta}},$$

which is identical to the value of the floating rate payment of the swap. Therefore, the value of this floating rate payment at any time $t \le T_i - \delta$ must be

$$H\left(B_t^{T_i-\delta} - B_t^{T_i}\right) = H\delta B_t^{T_i} \frac{\frac{B_t^{T_i-\delta}}{B_t^{T_i}} - 1}{\delta} = H\delta B_t^{T_i} L_t^{T_i-\delta, T_i},$$

where we have applied (1.6). Thus, the value at time $t \le T_i - \delta$ of getting $H\delta l^{T_i}_{T_i-\delta}$ at time T_i is equal to $H\delta B_t^{T_i} L_t^{T_i-\delta,T_i}$, that is the unknown future spot rate $l^{T_i}_{T_i-\delta}$ in the payoff is replaced by the current forward rate for $L_t^{T_i-\delta,T_i}$ and then discounted by the current risk-free discount factor $B_t^{T_i}$. The value at time $t > T_0$ of all the remaining floating coupon payments can therefore be written as

$$V_t^{\mathrm{fl}} = H\delta B_t^{T_{i(t)}} l^{T_{i(t)}}_{T_{i(t)}-\delta} + H\delta \sum_{i=i(t)+1}^{n} B_t^{T_i} L_t^{T_i-\delta,T_i}, \quad T_0 \le t < T_n.$$

At or before time T_0, the first term is not present, so we get

$$V_t^{\mathrm{fl}} = H\delta \sum_{i=1}^{n} B_t^{T_i} L_t^{T_i-\delta,T_i}, \quad t \le T_0. \tag{6.27}$$

The value of a payer swap is

$$\mathbf{P}_t = V_t^{\mathrm{fl}} - V_t^{\mathrm{fix}},$$

while the value of a receiver swap is

$$\mathbf{R}_t = V_t^{\mathrm{fix}} - V_t^{\mathrm{fl}}.$$

In particular, the value of a payer swap at or before its starting date T_0 can be written as

$$\mathbf{P}_t = H\delta \sum_{i=1}^{n} B_t^{T_i} \left(L_t^{T_i-\delta,T_i} - K \right), \quad t \le T_0, \tag{6.28}$$

using (6.25) and (6.27), or as

$$\mathbf{P}_t = H \left(\left[B_t^{T_0} - B_t^{T_n} \right] - \sum_{i=1}^{n} K\delta B_t^{T_i} \right), \quad t \le T_0, \tag{6.29}$$

using (6.25) and (6.26). If we let $Y_i = K\delta$ for $i = 1, \dots, n-1$ and $Y_n = 1 + K\delta$, we can rewrite (6.29) as

$$\mathbf{P}_t = H \left(B_t^{T_0} - \sum_{i=1}^{n} Y_i B_t^{T_i} \right), \quad t \le T_0. \tag{6.30}$$

Also note the following relation between a cap, a floor, and a payer swap having the same payment dates and where the cap rate, the floor rate, and the fixed rate in the swap are all identical:

$$\mathcal{C}_t = \mathcal{F}_t + \mathbf{P}_t.$$

This follows from the fact that the payments from a portfolio of a floor and a payer swap exactly match the payments of a cap.

The **swap rate** $\tilde{l}_{T_0}^{\delta}$ prevailing at time T_0 for a swap with frequency δ and payments dates $T_i = T_0 + i\delta$, $i = 1, 2, \dots, n$, is defined as the unique value of the fixed rate that makes the present value of a swap starting at T_0 equal to zero, that is $\mathbf{P}_{T_0} = \mathbf{R}_{T_0} = 0$. The swap rate is sometimes called the equilibrium swap rate or the par swap rate. Applying (6.28), we can write the swap rate as a weighted average of the relevant forward rates:

$$\tilde{l}_{T_0}^{\delta} = \frac{\sum_{i=1}^{n} L_{T_0}^{T_i-\delta,T_i} B_{T_0}^{T_i}}{\sum_{i=1}^{n} B_{T_0}^{T_i}} = \sum_{i=1}^{n} w_i L_{T_0}^{T_i-\delta,T_i}, \tag{6.31}$$

where $w_i = B_{T_0}^{T_i} / \sum_{i=1}^n B_{T_0}^{T_i}$. Alternatively, we can let $t = T_0$ in (6.29) yielding

$$\mathbf{P}_{T_0} = H\left(1 - B_{T_0}^{T_n} - K\delta \sum_{i=1}^n B_{T_0}^{T_i}\right),$$

so that the swap rate can be expressed as

$$\tilde{l}_{T_0}^{\delta} = \frac{1 - B_{T_0}^{T_n}}{\delta \sum_{i=1}^n B_{T_0}^{T_i}}. \tag{6.32}$$

Substituting (6.32) into the expression just above it, the time T_0 value of an agreement to pay a fixed rate K and receive the prevailing market rate at each of the dates $T_1, \dots, T_n$, can be written in terms of the current swap rate as

$$\begin{aligned}\mathbf{P}_{T_0} &= H\left(\tilde{l}_{T_0}^{\delta}\delta\left(\sum_{i=1}^n B_{T_0}^{T_i}\right) - K\delta\left(\sum_{i=1}^n B_{T_0}^{T_i}\right)\right) \\ &= \left(\sum_{i=1}^n B_{T_0}^{T_i}\right) H\delta\left(\tilde{l}_{T_0}^{\delta} - K\right).\end{aligned} \tag{6.33}$$

A **forward swap** (or deferred swap) is an agreement to enter into a swap with a future starting date T_0 and a fixed rate which is already set. Of course, the contract also fixes the frequency, the maturity, and the notional principal of the swap. The value at time $t \le T_0$ of a forward payer swap with fixed rate K is given by the equivalent expressions (6.28)–(6.30). The **forward swap rate** $\tilde{L}_t^{\delta,T_0}$ is defined as the value of the fixed rate that makes the forward swap have zero value at time t. The forward swap rate can be written as

$$\tilde{L}_t^{\delta,T_0} = \frac{B_t^{T_0} - B_t^{T_n}}{\delta \sum_{i=1}^n B_t^{T_i}} = \frac{\sum_{i=1}^n L_t^{T_i-\delta,T_i} B_t^{T_i}}{\sum_{i=1}^n B_t^{T_i}}. \tag{6.34}$$

Note that both the swap rate and the forward swap rate depend on the frequency and the maturity of the underlying swap. To indicate this dependence, let $\tilde{l}_t^{\delta}(n)$ denote the time t swap rate for a swap with payment dates $T_i = t + i\delta$, $i = 1, 2, \dots, n$. If we depict the swap rate as a function of the maturity, that is the function $n \mapsto \tilde{l}_t^{\delta}(n)$ (only defined for $n = 1, 2, \dots$), we get a **term structure of swap rates** for the given frequency. Many financial institutions participating in the swap market will offer swaps of varying maturities under conditions reflected by their posted term structure of swap rates. In Exercise 6.8, the reader is asked to show how the discount factors $B_{T_0}^{T_i}$ can be derived from a term structure of swap rates.

6.5.2 Swaptions

A **European swaption** gives its holder the right, but not the obligation, at the expiry date T_0 to enter into a specific interest rate swap that starts at T_0 and has a given fixed rate K. No exercise price is to be paid if the right is utilized. The rate K is sometimes referred to as the exercise rate of the swaption. We distinguish between a **payer swaption**, which gives the right to enter into a payer swap, and a **receiver swaption**, which gives the right to enter into a receiver swap. As for caps and floors, two different pricing strategies can be taken. One strategy is to link the swaption payoff to the payoff of another well-known derivative. The other strategy is to directly take relevant expectations of the swaption payoff.

Let us first see how we can link swaptions to options on bonds. Let us focus on a European receiver swaption. The receiver swap has payment dates $T_i = T_0 + i\delta$, $i = 1, 2, \dots, n$, and a fixed rate K. According to (6.30), its value at time T_0 is given by

$$\mathbf{R}_{T_0} = H\left(\sum_{i=1}^{n} Y_i B_{T_0}^{T_i} - 1\right),$$

where $Y_i = K\delta$ for $i = 1, \dots, n-1$ and $Y_n = 1 + K\delta$. Hence, the time T_0 payoff of a receiver swaption is

$$\mathcal{R}_{T_0} = \max\left(\mathbf{R}_{T_0} - 0, 0\right) = H \max\left(\sum_{i=1}^{n} Y_i B_{T_0}^{T_i} - 1, 0\right),$$

which is equivalent to the payoff of H European call options on a bullet bond with face value 1, n payment dates, a period of δ between successive payments, and an annualized coupon rate K. The exercise price of each option equals the face value 1. The price of a European receiver swaption must therefore be equal to the price of these call options. In many of the pricing models we develop in later chapters, we can compute such prices quite easily.

Similarly, a European payer swaption yields a payoff of

$$\mathcal{P}_{T_0} = \max\left(\mathbf{P}_{T_0} - 0, 0\right) = \max\left(-\mathbf{R}_{T_0}, 0\right) = H \max\left(1 - \sum_{i=1}^{n} Y_i B_{T_0}^{T_i}, 0\right).$$

This is identical to the payoff from H European put options expiring at T_0 and having an exercise price of 1 with a bond paying Y_i at time T_i ($i = 1, 2, \dots, n$) as its underlying asset.

Alternatively, we can apply (6.33) to express the payoff of a European payer swaption as

$$\mathcal{P}_{T_0} = \left(\sum_{i=1}^{n} B_{T_0}^{T_i}\right) H\delta \max\left(\tilde{l}_{T_0}^{\delta} - K, 0\right), \tag{6.35}$$

where $\tilde{l}_{T_0}^{\delta}$ is the (equilibrium) swap rate prevailing at time T_0. What is an appropriate numeraire for pricing this swaption? If we were to use the zero-coupon bond maturing at T_0 as the numeraire, we would have to find the expectation of the payoff $\mathcal{P}_{T_0}$ under the T_0-forward martingale measure $\mathbb{Q}^{T_0}$. But since the payoff depends on several different bond prices, the distribution of $\mathcal{P}_{T_0}$ under $\mathbb{Q}^{T_0}$ is rather complicated. It is more convenient to use another numeraire, namely the annuity bond, which at each of the dates $T_1, \dots, T_n$ provides a payment of 1 dollar. The value of this annuity at time $t \leq T_0$ equals $G_t = \sum_{i=1}^n B_t^{T_i}$. In particular, the payoff of the swaption can be restated as

$$\mathcal{P}_{T_0} = G_{T_0} H\delta \max\left(\tilde{l}_{T_0}^{\delta} - K, 0\right),$$

and the payoff expressed in units of the annuity bond is simply $H\delta \max\left(\tilde{l}_{T_0}^{\delta} - K, 0\right)$. The martingale measure corresponding to the annuity being the numeraire is called the **swap martingale measure** and will be denoted by $\mathbb{Q}^G$ in the following. The price of the European payer swaption can now be written as

$$\mathcal{P}_t = G_t \, \mathrm{E}_t^{\mathbb{Q}^G}\left[\frac{\mathcal{P}_{T_0}}{G_{T_0}}\right] = G_t H\delta \, \mathrm{E}_t^{\mathbb{Q}^G}\left[\max\left(\tilde{l}_{T_0}^{\delta} - K, 0\right)\right],$$

so we only need to know the distribution of the swap rate $\tilde{l}_{T_0}^{\delta}$ under the swap martingale measure. In Chapter 11 we will look at models of swap rate dynamics under the swap martingale measure that allow us to price swaptions using the above formula.

Similar to the put–call parity for bonds we have the following **payer–receiver parity** for European swaptions having the same underlying swap and the same exercise rate:

$$\mathcal{P}_t - \mathcal{R}_t = \mathbf{P}_t, \quad t \leq T_0, \tag{6.36}$$

compare Exercise 6.8. In words, a payer swaption minus a receiver swaption is indistinguishable from a forward payer swap.

The market standard for pricing European swaptions is Black's formula, which for a payer swaption is

$$\mathcal{P}_t = H\delta\left(\sum_{i=1}^n B_t^{T_i}\right)\left[\tilde{L}_t^{\delta,T_0} N\left(\hat{d}_1(\tilde{L}_t^{\delta,T_0}, t)\right) - KN\left(\hat{d}_2(\tilde{L}_t^{\delta,T_0}, t)\right)\right], \quad t < T_0, \tag{6.37}$$

where the functions $\hat{d}_1$ and $\hat{d}_2$ are as in (6.19) and (6.20) with $T = T_0$. The analogous formula for a European receiver swaption is

$$\mathcal{R}_t = H\delta\left(\sum_{i=1}^n B_t^{T_i}\right)\left[KN\left(-\hat{d}_2(\tilde{L}_t^{\delta,T_0}, t)\right) - \tilde{L}_t^{\delta,T_0} N\left(-\hat{d}_1(\tilde{L}_t^{\delta,T_0}, t)\right)\right], \ t < T_0.$$

Again, the assumptions behind this are generally inappropriate. However, we will see in Chapter 11 that these pricing formulae can be backed by a very special no-arbitrage model of swap rate dynamics.

If we consider formula (6.31) and assume as an approximation that the weights w_i are constant over time, the variance of the future swap rate can be written as

$$\mathrm{Var}_t[\tilde{l}^{\delta}_{T_0}] = \mathrm{Var}_t\left[\sum_{i=1}^{n} w_i L_{T_0}^{T_i-\delta,T_i}\right] = \sum_{i=1}^{n}\sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij},$$

where σ_i denotes the standard deviation of the forward rate $L_{T_0}^{T_i-\delta,T_i}$, and ρ_{ij} denotes the correlation between the forward rates $L_{T_0}^{T_i-\delta,T_i}$ and $L_{T_0}^{T_j-\delta,T_j}$. The prices of swaptions will therefore depend on both the volatilities of the relevant forward rates and their correlations. If implicit forward rate volatilities have already been determined from the market prices of caplets and caps, **implicit forward rate correlations** can be determined from the market prices of swaptions by an application of (6.37).

While a large majority of traded swaptions are European, so-called **Bermuda swaptions** are also traded. A Bermuda swaption can be exercised at a number of pre-specified dates and, therefore, resembles an American option. When the Bermuda swaption is exercised, the holder receives a position in a swap with certain payment dates. Most Bermuda swaptions are constructed such that the underlying swap has some fixed, potential payment dates $T_1, \dots, T_n$. If the Bermuda swaption is exercised at, say, time t', only the remaining swap payments will be effective, that is the payments at date $T_{i(t')}, \dots, T_n$. Later exercise results in a shorter swap. The possible exercise dates will usually coincide with the potential swap payment dates. Exercise of a Bermuda payer (receiver) swaption at date T_l results in a payoff at that date equal to the payoff of a European payer (receiver) swaption expiring at that date with a swap with payment dates $T_{l+1}, \dots, T_n$. Bermuda swaptions are often issued together with a given swap. Such a 'package' is called a **cancellable swap** or a **puttable swap**. Typically, the Bermuda swaption cannot be exercised over a certain period in the beginning of the swap. When practitioners talk of, say, a '10 year non-call 2 year Bermuda swaption', they mean an option on a 10 year swap, where the option at the earliest can be exercised 2 years into the swap and then on all subsequent payment dates of the swap. A less-traded variant is a constant maturity Bermuda swaption, where the optionholder upon exercise receives a swap with the same time to maturity no matter when the option is exercised.

6.5.3 Exotic swap instruments

The following examples of exotic swap market products are adapted from Musiela and Rutkowski (1997) and Hull (2009):

- **Float-for-floating swap:** Two floating interest rates are swapped, for example the 3-month LIBOR rate and the yield on a given government bond.
- **Amortizing swap:** The notional principal is reduced from period to period following a pre-specified scheme, for example so that the notional principle

at any time reflects the outstanding debt on a loan with periodic instalments (as for an annuity or a serial bond).

- **Step-up swap:** The notional principal increases over time in a pre-determined way.
- **Accrual swap:** The scheduled payments of one party are only to be paid as long as the floating rate lies in some interval $\mathcal{I}$. Assume for concreteness that it is the fixed rate payments that have this feature. At the swap payment date T_i the effective fixed rate payment is then $H\delta K N_1/N_2$, where N_1 is the number of days in the period between T_{i-1} and T_i, where the floating rate $l_t^{t+\delta}$ was in the interval $\mathcal{I}$, and N_2 is the total number of days in the period. The interval $\mathcal{I}$ may even differ from period to period either in a deterministic way or depending on the evolution of the floating interest rate so far.
- **Constant maturity swap:** At the payment dates a fixed rate is exchanged for the (equilibrium) swap rate on a swap of a given, constant maturity, that is the floating rate is itself a swap rate.
- **Extendable swap:** One party has the right to extend the life of the swap under certain conditions.
- **Forward swaption:** A forward swaption gives the right to enter into a forward swap, that is the swaption expires at time t^* before the starting date of the swap T_0. The payoff is

$$H\delta \sum_{i=1}^{n} \max\left(\tilde{L}_{T_0}^{\delta,t^*} - K, 0\right) B_{t^*}^{T_i} = \left(\sum_{i=1}^{n} B_{t^*}^{T_i}\right) H\delta \max\left(\tilde{L}_{T_0}^{\delta,t^*} - K, 0\right).$$

- **Swap rate spread option:** The payoff is determined by the difference between (equilibrium) swap rates for two different maturities. Recall that $\tilde{l}_{T_0}^{\delta}(m)$ denotes the swap rate for a swap with payment dates $T_1, \dots, T_m$, where $T_i = T_0 + i\delta$. An (m, n)-period European swap rate spread call option with an exercise rate K yields a payoff at time T_0 of

$$\max\left(\tilde{l}_{T_0}^{\delta}(m) - \tilde{l}_{T_0}^{\delta}(n) - K, 0\right).$$

  The corresponding put has a payoff of

$$\max\left(K - \left[\tilde{l}_{T_0}^{\delta}(m) - \tilde{l}_{T_0}^{\delta}(n)\right], 0\right).$$

- **Yield curve swap:** In a one-period yield curve swap one party receives at a given date T a swap rate $\tilde{l}_T^{\delta}(m)$ and pays a rate $K + \tilde{l}_T^{\delta}(n)$, both computed on the basis of a given notional principal H. A multi-period yield curve swap has, say, L payment dates $T_1, \dots, T_L$. At time T_l one party receives an interest rate of $\tilde{l}_{T_l}^{\delta}(m)$ and pays an interest rate of $K + \tilde{l}_{T_l}^{\delta}(n)$.

In addition, several instruments combine elements of interest rate swaps and currency swaps. For example, in a **differential swap** a domestic floating rate is swapped for a foreign floating rate.

6.6 AMERICAN-STYLE DERIVATIVES

Consider an American-style derivative where the holder can choose to exercise the derivative at the expiration date T or at any time before T. Let P_τ denote the payoff if the derivative is exercised at time $\tau \leq T$. In general, P_τ may depend on the evolution of the economy up to and including time τ, but it is usually a simple function of the time τ price of an underlying security or the time τ value of a particular interest rate. At each point in time the holder of the derivative must decide whether or not he will exercise. Of course, this decision must be based on the available information, so we are seeking an entire **exercise strategy** that tells us exactly in what states of the world we should exercise the derivative. We can represent an exercise strategy by an indicator function $I(\omega, t)$, which for any given state of the economy ω at time t either has the value 1 or 0, where the value 1 indicates exercise and 0 indicates non-exercise. For a given exercise strategy I, the derivative will be exercised the first time $I(\omega, t)$ takes on the value 1. We can write this point in time as

$$\tau(I) = \inf\{s \in [t, T] \mid I(\omega, s) = 1\},$$

where 'inf' is short for infimum (which is roughly equivalent to the minimum, but in some cases there is a difference between the infimum and the minimum of a set). This is called a stopping time in the literature on stochastic processes.

By our earlier analysis, the value of getting the payoff $V_{\tau(I)}$ at time $\tau(I)$ is given by $\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau(I)} r_u\, du} P_{\tau(I)}\right]$. If we let $\mathcal{I}[t, T]$ denote the set of all possible exercise strategies over the time period $[t, T]$, the time t value of the American-style derivative must therefore be

$$V_t = \sup_{I \in \mathcal{I}[t,T]} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau(I)} r_u\, du} P_{\tau(I)}\right].$$

Here, 'sup' is short for supremum (which is roughly equivalent to the maximum). An optimal exercise strategy I^* is such that

$$V_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau(I^*)} r_u\, du} P_{\tau(I^*)}\right].$$

Note that the optimal exercise strategy and the price of the derivative must be solved for simultaneously. This complicates the pricing of American-style derivatives considerably. In fact, in all situations where early exercise may be relevant, we will not be able to compute closed-form pricing formulas for American-style derivatives. We have to resort to numerical techniques.

In a diffusion model with a one-dimensional state variable x, we can write the indicator function representing the exercise strategy of an American-style derivative as $I(x, t)$, so that $I(x, t) = 1$ if and only if the derivative is exercised at time t when $x_t = x$. An exercise strategy divides the space $\mathcal{S} \times [0, T]$ of points (x, t) into an exercise region and a continuation region. The **continuation region** corresponding to a given exercise strategy I is the set

$$\mathcal{C}_I = \{(x, t) \in \mathcal{S} \times [0, T] \mid I(x, t) = 0\}$$

and the **exercise region** is then the remaining part

$$\mathcal{E}_I = \{(x,t) \in \mathcal{S} \times [0,T] \mid I(x,t) = 1\},$$

which can also be written as $\mathcal{E}_I = (\mathcal{S} \times [0,T]) \setminus \mathcal{C}_I$. To an optimal exercise strategy $I^*(x,t)$ corresponds optimal continuation and exercise regions $\mathcal{C}^*$ and $\mathcal{E}^*$. It is intuitively clear that the price function $P(x,t)$ for an American-style derivative must satisfy the fundamental PDE (4.26) in the continuation region corresponding to the optimal exercise strategy, that is for $(x,t) \in \mathcal{C}^*$. Since the continuation region is not known, but is part of the solution, it is impossible to solve such a PDE explicitly except for trivial cases. However, numerical solution techniques for PDEs can, with some modifications, also be applied to the case of American-style derivatives; see Chapter 16.

What can we say about early exercise of American options on bonds? It is well-known that it is never strictly advantageous to exercise an American call option on a non-dividend paying stock before time T; see Merton (1973) and Hull (2009). By analogy, this is also true for American call options on zero-coupon bonds. At first glance, it may appear optimal to exercise an American call on a zero-coupon bond immediately in case the price of the underlying bond is equal to 1, because this will imply a payoff of $1-K$, which is the maximum possible payoff under the assumption of non-negative interest rates. However, the price of the underlying bond will only equal 1 if interest rates are zero and stay at zero for sure. Therefore, exercising the option at time T will also provide a payoff of $1-K$, and since interest rates are zero, the present value of the payoff is also equal to $1-K$. Hence, there is no strict advantage to early exercise. As for stock options, premature exercise of an American put option on a zero-coupon bond will be advantageous for sufficiently low prices of the underlying zero-coupon bond, that is sufficiently high interest rates.

When and under what circumstances should one consider exercising an American call on a coupon bond? This is equivalent to the question of exercising an American call on a dividend-paying stock, which is discussed, for example in (Hull, 2009, Ch. 13). The following conclusions can therefore be stated. The only points in time when it can be optimal to exercise an American call on a bond is just before the payment dates of the bond. Let T_l be the last payment date before expiration of the option. Then it cannot be optimal to exercise the call just before T_l if the payment Y_l is less than $K(1 - B^T_{T_l})$. If the opposite relation holds, it *may* be optimal to exercise just before T_l. Similarly, at any earlier payment date $T_i \in [t, T_l]$, exercise is ruled out if the payment at that date Y_i is less than $K(1 - B^{T_{i+1}}_{T_i})$. Broadly speaking, early exercise of the call will only be relevant if the short-term interest rate is relatively low and the bond payment is relatively high.[7] Regarding early exercise of put options, it can never be optimal to exercise an American put on a bond just before a payment on the bond. At all other points in time early exercise

[7] Some countries have markets with trade in mortgage-backed bonds where the issuer has an American call option on the bond. These bonds are annuity bonds where the payments are considerably higher than for a standard 'bullet' bond with the same face value. Optimality of early exercise of such a call is therefore more likely than exercise of a call on a standard bond. More information is provided in Chapter 14.

may be optimal for sufficiently low bond prices, that is sufficiently high interest rates.

For American options on bonds, it is also possible to find no-arbitrage price bounds, and, as a counterpart to the put–call parity, relatively tight bounds on the difference between the prices of an American call and an American put. Again the reader is referred to Munk (2002).

6.7 AN OVERVIEW OF TERM STRUCTURE MODELS

Economists and financial analysts apply term structure models in order to:

- improve their understanding of the way the term structure of interest rates is set by the market and how it evolves over time,
- price fixed income securities in a consistent way,
- facilitate the management of the interest rate risk that affects the valuation of individual securities, financial investment portfolios, and real investment projects.

As we shall see in the following chapters, a large number of different term structure models have been suggested in the last three decades. All the models have both desirable and undesirable properties so that the choice of model will depend on how one weighs the pros and the cons. Ideally, we seek a model which has as many as possible of the following characteristics:[8]

(a) **flexible:** the model should be able to handle most situations of practical interest, that is it should apply to most fixed income securities and under all likely states of the world;
(b) **simple:** the model should be so simple that it can deliver answers (for example prices and hedge ratios) in a very short time;
(c) **well-specified:** the necessary input for applying the model must be relatively easy to observe or estimate;
(d) **realistic:** the model should not have clearly unreasonable properties;
(e) **empirically acceptable:** the model should be able to describe actual data with sufficient precision;
(f) **theoretically sound:** the model should be consistent with the broadly accepted principles for the behaviour of individual investors and the financial market equilibrium.

No model can completely comply with all these objectives. A realistic, empirically acceptable, and theoretically sound model is bound to be quite complex and will probably not be able to deliver prices and hedge ratios with the speed requested by many practitioners. On the other hand, simpler models will have inappropriate theoretical and/or empirical properties.

[8] The presentation is in part based on Rogers (1995).

We can split the many term structure models into two categories: absolute pricing models and relative pricing models. An **absolute pricing model** of the term structure of interest rates aims at pricing all fixed income securities, both the basic securities, that is bonds and bond-like contracts such as swaps, and the derivative securities such as bond options and swaptions. In contrast, a **relative pricing model** of the term structure takes the currently observed term structure of interest rates, that is the prices of bonds, as given and aims at pricing derivative securities relative to the observed term structure. The same distinction can be used for other asset classes. For example, the Black–Scholes–Merton model is a relative pricing model since it prices stock options relative to the price of the underlying stock, which is taken as given. An absolute stock option pricing model would derive prices of both the underlying stock and the stock option.

Absolute pricing models are sometimes referred to as equilibrium models, while relative pricing models are called pure no-arbitrage models. In this context the term equilibrium model does not necessarily imply that the model is based on explicit assumptions on the preferences and endowments of all market participants (including the bond issuers, for example the government) which in the end determine the supply and demand for bonds and therefore bond prices and interest rates. Indeed, many absolute pricing models of the term structure are based on an assumption on the dynamics of one or several state variables and stipulated relations between the short rate and the state variables and between the market prices of risk and the state variables. These assumptions determine both the current term structure and the dynamics of interest rates and prices of fixed income securities. These models do not explain how these assumptions are produced by the actions of market participants. Nevertheless, it is typically possible to justify the assumptions of these models by some more basic assumptions on preferences, endowments, and so on, so that the model assumptions are compatible with market equilibrium; see the discussion and the examples in Section 5.4. The pure no-arbitrage models offer no explanation as to why the current term structure is as observed.

We can also divide the term structure models into **diffusion models** and **non-diffusion models**. Again, by a diffusion model we mean a model in which all relevant prices and quantities are functions of a state variable of a finite (preferably low) dimension and this state variable follows a Markov diffusion process. A non-diffusion model is a model which does not meet this definition of a diffusion model. While the risk-neutral pricing techniques are valid both in diffusion and non-diffusion models, the PDE approach introduced in Section 4.8 can only be applied in diffusion models. All well-known absolute pricing models of the term structure are diffusion models. We study a number of one-factor and multi-factor diffusion models of the term structure in Chapters 7 and 8. In the diffusion models we derive prices and interest rates as functions of the state variables and relatively few parameters. Consequently, the resulting term structure of interest rates cannot typically fit the currently observed term structure perfectly. If the main application of the model is to price derivative securities, this mismatch is troublesome. If the model is not able to price the underlying securities (that is the zero-coupon bonds) correctly, why trust the model prices for derivative securities? To completely avoid this mismatch one must apply relative pricing models for the derivative securities.

We divide the relative pricing models of the term structure into three subclasses: calibrated diffusion models, Heath–Jarrow–Morton (HJM) models, and market models. The common starting point of all these models is to take the current term structure as given and then model the risk-neutral dynamics of the entire term structure. This is done very directly in the HJM models and the market models. The HJM models are based on assumptions about the dynamics of the entire curve of instantaneous, continuously compounded forward rates, $T \mapsto f_t^T$. It turns out that only the volatility structure of the forward rate curve needs to specified in order to price term structure derivatives. We will discuss the general HJM model and various concrete models in Chapter 10. The market models are closely related to the HJM models, but focus on the pricing of money market products such as caps, floors, and swaptions. These products involve LIBOR rates that are set for specific periods, such as 3 months, 6 months, and 12 months, with a similar compounding period. The market models are all based on assumptions about a number of forward LIBOR rates or swap rates. Again, only the volatility structure of these rates needs to be specified. Market models are studied in Chapter 11. The third subclass of relative pricing models consists of so-called calibrated diffusion models. These models can be seen as extensions of absolute pricing models of the diffusion type. The basic idea is to replace one of the constant parameters in a diffusion model by a suitable deterministic function of time that will make the term structure of the model exactly match the currently observed term structure in the market. These calibrated diffusion models can be reformulated as HJM models, but since they are developed in a special way we treat them separately in Chapter 9.

6.8 EXERCISES

Exercise 6.1 Show that the no-arbitrage price of a European call on a zero-coupon bond will satisfy

$$\max\left(0, B_t^S - KB_t^T\right) \leq C_t^{K,T,S} \leq B_t^S(1-K)$$

provided that all interest rates are non-negative. Here, T is the maturity date of the option, K is the exercise price, and S is the maturity date of the underlying zero-coupon bond. Compare with the corresponding bounds for a European call on a stock, see, for example (Hull, 2009, Ch. 9). Derive similar bounds for a European call on a coupon bond.

Exercise 6.2 Show the put–call parity for options on coupon bonds by a replication argument, that is form two portfolios that have the same payoffs and conclude from their prices that (6.18) must hold.

Exercise 6.3 Let $\tilde{l}_{T_0}^{\delta}(k)$ be the equilibrium swap rate for a swap with payment dates $T_1, T_2, \ldots, T_k$, where $T_i = T_0 + i\delta$ as usual. Suppose that $\tilde{l}_{T_0}^{\delta}(1), \ldots, \tilde{l}_{T_0}^{\delta}(n)$

are known. Find a recursive procedure for deriving the associated discount factors $B_{T_0}^{T_1}, B_{T_0}^{T_2}, \ldots, B_{T_0}^{T_n}$.

Exercise 6.4 Show the parity (6.36). Show that a payer swaption and a receiver swaption (with identical terms) will have identical prices, if the exercise rate of the contracts is equal to the forward swap rate $\tilde{L}_t^{\delta,T_0}$.

Exercise 6.5 Consider a swap with starting date T_0 and a fixed rate K. For $t \leq T_0$, show that $V_t^{\mathrm{fl}}/V_t^{\mathrm{fix}} = \tilde{L}_t^{\delta,T_0}/K$, where $\tilde{L}_t^{\delta,T_0}$ is the forward swap rate.

7

One-Factor Diffusion Models

7.1 INTRODUCTION

This chapter is devoted to the study of one-factor diffusion models of the term structure of interest rates. They all take the short rate as the sole state variable and, hence, implicitly assume that the short rate contains all the information about the term structure that is relevant for pricing and hedging interest rate dependent claims. All the models assume that the short rate is a diffusion process

$$dr_t = \alpha(r_t, t)\, dt + \beta(r_t, t)\, dz_t,$$

where $z = (z_t)_{t \geq 0}$ is a one-dimensional standard Brownian motion under the real-world probability measure $\mathbb{P}$. The market price of risk at time t is of the form $\lambda(r_t, t)$. It then follows from the analysis in Section 4.3 that the short rate dynamics under the risk-neutral probability measure $\mathbb{Q}$ is

$$dr_t = \hat{\alpha}(r_t, t)\, dt + \beta(r_t, t)\, dz_t^{\mathbb{Q}}, \tag{7.1}$$

where $z^{\mathbb{Q}} = (z_t^{\mathbb{Q}})$ is a standard Brownian motion under $\mathbb{Q}$, and

$$\hat{\alpha}(r, t) = \alpha(r, t) - \beta(r, t)\lambda(r, t).$$

We let $\mathcal{S} \subseteq \mathbb{R}$ denote the value space for the short rate, that is the set of values which the short rate can have with strictly positive probability.[1]

A model of the type (7.1) is called **time-homogeneous** if $\hat{\alpha}$ and β are functions of the interest rate only and not of time. Otherwise it is called **time-inhomogeneous**. In the time-homogeneous models the distribution of a given variable at a future date depends only on the current short rate and how far into the future we are looking. For example, the distribution of $r_{t+\tau}$ given $r_t = r$ is the same for all values of t—the distribution depends only on the 'horizon' τ and the initial value r. Similarly, asset prices will only depend on the current short rate and the time to maturity of the asset. For example, the price of a zero-coupon bond $B_t^T = B^T(r_t, t)$ only depends on r_t and the time to maturity $T - t$, compare Theorem 7.1 below. In time-inhomogeneous models, these considerations are not valid, which renders the analysis of such models slightly more complicated. Furthermore, time homogeneity seems to be a realistic property: why should the drift and the volatility

[1] Recall that since the real-world and the risk-neutral probability measures are equivalent, the process can have exactly the same values under the different probability measures.

of the short rate depend on the calendar date? Surely the drift and the volatility change over time, but this is due to changes in fundamental economic variables not just the passage of time. However, time-inhomogeneous models have some practical advantages, which makes them worthwhile looking at. We will do that in Chapter 9. In the present chapter we consider only time-homogeneous models.

We will focus on the pricing of bonds, forwards and futures on bonds, Eurodollar futures, and European options on bonds within the different models. As discussed in Chapter 6, these option prices lead to prices of other important assets such as caps, floors, and European swaptions. The pricing techniques applied are those developed in Chapters 4: solution of a partial differential equation (PDE) or computation of the expected payoff under a suitable martingale measure.

In Section 7.2 we will consider some general aspects of the so-called affine models. Then in Sections 7.3–7.5 we will look at three specific affine models, namely the classic models of Merton (1970), Vasicek (1977), and Cox, Ingersoll, and Ross (1985b). We look at some recent extensions of the models of Vasicek (1977) and Cox, Ingersoll, and Ross (1985b) in Section 7.6. Some non-affine models are outlined and discussed in Section 7.7. Section 7.8 gives a short introduction to the issues of estimating the parameters of the models and testing to what extent the models are supported by the data. Finally, Section 7.9 offers some concluding remarks.

7.2 AFFINE MODELS

In a time-homogeneous one-factor model, the dynamics of the short rate is of the form

$$dr_t = \hat{\alpha}(r_t)\, dt + \beta(r_t)\, dz_t^{\mathbb{Q}}$$

under the risk-neutral (spot martingale) measure $\mathbb{Q}$. The fundamental PDE of Theorem 4.10 is then

$$\frac{\partial P}{\partial t}(r,t)+\hat{\alpha}(r)\frac{\partial P}{\partial r}(r,t)+\frac{1}{2}\beta(r)^2\frac{\partial^2 P}{\partial r^2}(r,t)-rP(r,t)=0, \quad (r,t)\in\mathcal{S}\times[0,T), \tag{7.2}$$

with the terminal condition

$$P(r,T) = H(r), \quad r \in \mathcal{S},$$

where the function H denotes the interest rate dependent payoff of the asset.

In this section we will study a subset of this class of models, namely the so-called affine models. An affine model is a model where the risk-neutral drift rate $\hat{\alpha}(r)$ and the variance rate $\beta(r)^2$ are affine functions of the short rate, that is of the form

$$\hat{\alpha}(r) = \hat{\varphi} - \hat{\kappa} r, \qquad \beta(r)^2 = \delta_1 + \delta_2 r, \tag{7.3}$$

where $\hat{\varphi}$, $\hat{\kappa}$, δ_1, and δ_2 are constants. We require that $\delta_1 + \delta_2 r \geq 0$ for all the values of r which the process for the short rate can have, that is for $r \in \mathcal{S}$, so that the variance is well-defined. The dynamics of the short rate under the risk-neutral probability measure is therefore given by the stochastic differential equation

$$dr_t = \left(\hat{\varphi} - \hat{\kappa} r_t\right) dt + \sqrt{\delta_1 + \delta_2 r_t}\, dz_t^{\mathbb{Q}}. \tag{7.4}$$

This subclass of models is tractable and results in nice, explicit pricing formulas for bonds and forwards on bonds and, in most cases, also for bond futures, Eurodollar futures, and European options on bonds.

7.2.1 Bond prices, zero-coupon rates, and forward rates

As before, B_t^T denotes the price at time t of a zero-coupon bond giving a payment of 1 unit of account with certainty at time T and nothing at all other points in time. We know that in a one-factor model, this price can be written as a function of time and the current short rate, $B_t^T = B^T(r_t, t)$. The following theorem shows that, in a model of the type (7.4), $B^T(r, t)$ is an exponential-affine function of the current short rate. The proof of this result is based only on the fact that $B^T(r, t)$ satisfies the partial differential equation (7.2) with the terminal condition $B^T(r, T) = 1$.

Theorem 7.1 *In the model* (7.4) *the time t price of a zero-coupon bond maturing at time T is given as*

$$B^T(r, t) = e^{-a(T-t)-b(T-t)r}, \tag{7.5}$$

where the functions $a(\tau)$ and $b(\tau)$ satisfy the following system of ordinary differential equations:

$$\frac{1}{2}\delta_2 b(\tau)^2 + \hat{\kappa} b(\tau) + b'(\tau) - 1 = 0, \quad \tau > 0, \tag{7.6}$$

$$a'(\tau) - \hat{\varphi} b(\tau) + \frac{1}{2}\delta_1 b(\tau)^2 = 0, \quad \tau > 0, \tag{7.7}$$

together with the conditions $a(0) = b(0) = 0$.

Proof: We will show that the function $B^T(r, t)$ in (7.5) is a solution to the partial differential equation (7.2). Since $a(0) = b(0) = 0$, the terminal condition $B^T(r, T) = 1$ is satisfied for all $r \in \mathcal{S}$. The relevant derivatives are

$$\begin{aligned}
\frac{\partial B^T}{\partial t}(r, t) &= B^T(r, t) \left(a'(T-t) + b'(T-t)r\right), \\
\frac{\partial B^T}{\partial r}(r, t) &= -B^T(r, t) b(T-t), \\
\frac{\partial^2 B^T}{\partial r^2}(r, t) &= B^T(r, t) b(T-t)^2.
\end{aligned} \tag{7.8}$$

After substituting these derivatives into (7.2) and dividing through by $B^T(r,t)$, we get

$$a'(T-t) + b'(T-t)r - b(T-t)\hat{\alpha}(r) + \frac{1}{2}b(T-t)^2\beta(r)^2 - r = 0,$$
$$(r,t) \in \mathcal{S} \times [0,T). \tag{7.9}$$

Substituting (7.3) into (7.9) and gathering terms involving r, we find that the functions a and b must satisfy the equation

$$\left(a'(T-t) - \hat{\varphi}b(T-t) + \frac{1}{2}\delta_1 b(T-t)^2\right)$$
$$+ \left(\frac{1}{2}\delta_2 b(T-t)^2 + \hat{\kappa} b(T-t) + b'(T-t) - 1\right) r = 0, \quad (r,t) \in \mathcal{S} \times [0,T).$$

This can only be true if (7.6) and (7.7) hold.[2] □

Conversely, it can be shown that the zero-coupon bond prices $B^T(r,t)$ are only of the exponential-affine form (7.5), if the drift rate and the variance rate are affine functions of the short rate as in (7.3).[3]

The differential equations (7.6)–(7.7) are called **Ricatti equations**. The functions a and b are determined by first solving (7.6) with the condition $b(0) = 0$ to obtain the b-function. The solution to (7.7) with the condition $a(0) = 0$ can be written in terms of the b-function as

$$a(\tau) = \hat{\varphi} \int_0^\tau b(u)\, du - \frac{1}{2}\delta_1 \int_0^\tau b(u)^2\, du, \tag{7.10}$$

since $a(\tau) = a(\tau) - a(0) = \int_0^\tau a'(u)\, du$. For many frequently applied specifications of $\hat{\varphi}, \hat{\kappa}, \delta_1$, and δ_2, explicit expressions for a and b can be obtained in this way. For other specifications the Ricatti equations can be solved numerically by very efficient methods. In all the models we will consider, the function $b(\tau)$ is positive for all τ. Consequently, bond prices will be decreasing in the short rate consistent with the traditional relation between bond prices and interest rates.

Next, we study the yield curves in the affine models (7.4). The zero-coupon rate at time t for the period up to time T is denoted by y_t^T and is also a function of the current short rate, $y_t^T = y^T(r_t, t)$. With continuous compounding we have

$$B^T(r,t) = e^{-y^T(r,t)(T-t)},$$

compare (1.3). It follows from (7.5) that

$$y^T(r,t) = -\frac{\ln B^T(r,t)}{T-t} = \frac{a(T-t)}{T-t} + \frac{b(T-t)}{T-t} r, \tag{7.11}$$

[2] Suppose $A + Br = 0$ for all $r \in \mathcal{S}$. Given $r_1, r_2 \in \mathcal{S}$, where $r_1 \neq r_2$. Then, $A + Br_1 = 0$ and $A + Br_2 = 0$. Subtracting one of these equations from the other, we get $B[r_1 - r_2] = 0$, which implies that $B = 0$. It follows immediately that A must also equal zero.

[3] For details see (Duffie, 2001, Sec. 7E).

that is any zero-coupon yield is an affine function of the short rate. If b is positive, all zero-coupon yields are increasing in the short rate. An increase in the short rate will induce an upward shift of the entire yield curve $T \mapsto y^T(r,t)$. However, unless $b(\tau)$ is proportional to τ, the shift is not a parallel shift since the coefficient $b(T-t)/(T-t)$ is maturity dependent. In the important models, this coefficient is decreasing in maturity T, so that a shift in the short rate affects zero-coupon yields of short maturities more than zero-coupon yields of long maturities, which seems to be a reasonable property. Note that the zero-coupon yield for a fixed time to maturity of τ can be written as

$$y^{t+\tau}(r,t) = \frac{a(\tau)}{\tau} + \frac{b(\tau)}{\tau} r, \tag{7.12}$$

which is independent of t, which again stems from the time homogeneity of the model. In Exercise 7.1, you are asked to show that the slope of the yield curve at zero maturity, $\lim_{\tau \to 0} \frac{\partial y^{t+\tau}(r,t)}{\partial \tau}$, is equal to $(\hat{\varphi} - \hat{\kappa} r)/2$, that is half the risk-neutral drift of the short rate. In particular, if the risk-neutral drift of the short rate is positive [negative], the very short end of the yield curve is sloping upwards [downwards].

The forward rate f_t^T at time t for a loan over an infinitesimally short period beginning at time T is also given by a function of the current short rate, $f_t^T = f^T(r_t,t)$. With continuous compounding we have

$$f^T(r,t) = -\frac{\frac{\partial B^T}{\partial T}(r,t)}{B^T(r,t)},$$

compare (1.9). From (7.5) we get that

$$f^T(r,t) = a'(T-t) + b'(T-t)r. \tag{7.13}$$

Hence, the forward rates are also affine in the short rate r. For a fixed time to maturity τ, the forward rate is

$$f^{t+\tau}(r,t) = a'(\tau) + b'(\tau)r.$$

Let us consider the dynamics of the price $B_t^T = B^T(r,t)$ of a zero-coupon bond with a fixed maturity date T. Note that we are mostly interested in the evolution of prices and interest rates in the real world; the risk-neutral measure is only used for deriving the pricing formulas. From the general asset pricing theory of Chapter 4, we know that the dynamics will be of the form

$$dB_t^T = B_t^T \left[\left(r_t + \sigma^T(r_t,t)\lambda(r_t,t) \right) dt + \sigma^T(r_t,t)\, dz_t \right], \tag{7.14}$$

and from Itô's Lemma the sensitivity term of the zero-coupon bond price is given by

$$\sigma^T(r,t) = \frac{\frac{\partial B^T}{\partial r}(r,t)}{B^T(r,t)} \beta(r,t).$$

In the time-homogeneous affine models it follows from (7.8) that the sensitivity term is

$$\sigma^T(r,t) = -b(T-t)\beta(r). \tag{7.15}$$

With $b(T-t)$ and $\beta(r)$ being positive, $\sigma^T(r,t)$ will be negative. If there is a positive [negative] shock to the short rate, there will a negative [positive] shock to the zero-coupon bond price. The volatility of the zero-coupon bond price is the absolute value of $\sigma^T(r,t)$, that is $b(T-t)\beta(r)$. In equilibrium, risky assets will *normally* have an expected rate of return that exceeds the locally risk-free interest rate. This can only be the case if the market price of risk $\lambda(r,t)$ is negative.

When we look at the dynamics of zero-coupon rates, we are often more interested in the evolution of a rate with a fixed time to maturity $\tau = T-t$ (say, the 5 year interest rate) rather than a rate with a fixed maturity date T. Hence, we study the dynamics of $\bar{y}_t^\tau = y_t^{t+\tau} = y^{t+\tau}(r_t,t)$ for a fixed τ. Itô's Lemma and (7.12) imply that

$$d\bar{y}_t^\tau = \frac{b(\tau)}{\tau}\alpha(r_t)\,dt + \frac{b(\tau)}{\tau}\beta(r_t)\,dz_t \tag{7.16}$$

under the real-world probability measure. Here we have used that $\partial^2 y/\partial r^2 = 0$ and assumed that the market price of risk and, therefore, the drift of the short rate under the real-world measure $\alpha(r_t) = \hat{\alpha}(r_t) + \lambda(r_t)\beta(r_t)$ are time-homogeneous. Similarly, the forward rate with fixed time to maturity τ is $\bar{f}_t^\tau = f_t^{t+\tau} = f^{t+\tau}(r_t,t)$, which by Itô's Lemma and (7.13) evolves as

$$d\bar{f}_t^\tau = b'(\tau)\alpha(r_t)\,dt + b'(\tau)\beta(r_t)\,dz_t.$$

7.2.2 Forwards and futures

Equation (6.5) offers a general characterization of forward prices on zero-coupon bonds. Letting $F^{T,S}(r,t)$ denote the forward price at time t with a current short rate of r for delivery at time T of a zero-coupon bond maturing at time S, we have that $F^{T,S}(r,t) = B^S(r,t)/B^T(r,t)$. In the affine models where the zero-coupon price is given by (7.5), the forward price becomes

$$F^{T,S}(r,t) = \exp\left\{-\left[a(S-t) - a(T-t)\right] - \left[b(S-t) - b(T-t)\right]r\right\}, \tag{7.17}$$

where the functions a and b are the same as in Theorem 7.1.

For a futures on a zero-coupon bond we let $\Phi^{T,S}(r,t)$ denote the futures price. From Section 6.2 we have that the futures price is given by

$$\Phi^{T,S}(r,t) = \mathrm{E}_{r,t}^{\mathbb{Q}}\left[B^S(r_T,T)\right].$$

And from Section 4.8 we know that the price can be found by solving a partial differential equation, which can be obtained by letting $q = r$ in (4.29), that is

$$\frac{\partial \Phi^{T,S}}{\partial t}(r,t) + \hat{\alpha}(r)\frac{\partial \Phi^{T,S}}{\partial r}(r,t) + \frac{1}{2}\beta(r)^2\frac{\partial^2 \Phi^{T,S}}{\partial r^2}(r,t) = 0,$$
$$\forall (r,t) \in \mathcal{S} \times [0,T), \tag{7.18}$$

together with the terminal condition $\Phi^{T,S}(r,T) = B^S(r,T)$. The following theorem characterizes the solution.

Theorem 7.2 *Assume an affine model of the type* (7.4). *For a futures contract with final settlement date T and a zero-coupon bond maturing at time S as the underlying asset, the futures price at time t with a short rate of r is given by*

$$\Phi^{T,S}(r,t) = e^{-\tilde{a}(T-t)-\tilde{b}(T-t)r}, \tag{7.19}$$

where the functions $\tilde{a}(\tau)$ and $\tilde{b}(\tau)$ satisfy the following system of ordinary differential equations:

$$\frac{1}{2}\delta_2\tilde{b}(\tau)^2 + \hat{\kappa}\tilde{b}(\tau) + \tilde{b}'(\tau) = 0, \quad \tau \in (0,T), \tag{7.20}$$

$$\tilde{a}'(\tau) - \hat{\varphi}\tilde{b}(\tau) + \frac{1}{2}\delta_1\tilde{b}(\tau)^2 = 0, \quad \tau \in (0,T), \tag{7.21}$$

with the conditions $\tilde{a}(0) = a(S-T)$ and $\tilde{b}(0) = b(S-T)$, where a and b are as in Theorem 7.1.

If $\delta_2 = 0$, we have $\tilde{b}(\tau) = b(\tau + S - T) - b(\tau)$.

The solution to (7.21) with $\tilde{a}(0) = a(S-T)$ can generally be written as

$$\tilde{a}(\tau) = a(S-T) + \hat{\varphi}\int_0^\tau \tilde{b}(u)\,du - \frac{1}{2}\delta_1\int_0^\tau \tilde{b}(u)^2\,du. \tag{7.22}$$

The proof of this theorem is analogous to the proof of Theorem 7.1, since the PDE (7.18) is almost identical to the PDE (7.2) satisfied by the zero-coupon bond price. The last claim in the theorem above is left for the reader as Exercise 7.10. The claim implies that, for $\delta_2 = 0$, the futures price becomes

$$\Phi^{T,S}(r,t) = e^{-\tilde{a}(T-t)-[b(S-t)-b(T-t)]r}.$$

Comparing with the forward price expression (7.17), we see that, for $\delta_2 = 0$, we have

$$\frac{\frac{\partial F^{T,S}}{\partial r}(r,t)}{F^{T,S}(r,t)} = \frac{\frac{\partial \Phi^{T,S}}{\partial r}(r,t)}{\Phi^{T,S}(r,t)},$$

that is, any change in the term structure of interest rates will generate identical percentage changes in forward prices and futures prices with similar terms.

If the underlying bond is a coupon bond with payments Y_i at time T_i, it follows from (6.7) that the forward price at time t for delivery at time T is given by

$$F^{T,\mathrm{cpn}}(r,t) = \sum_{T_i > T} Y_i F^{T,T_i}(r,t), \tag{7.23}$$

into which we can insert (7.17) on the right-hand side. From (6.9) we get that the same relation holds for futures prices:

$$\Phi^{T,\text{cpn}}(r,t) = \sum_{T_i > T} Y_i \Phi^{T,T_i}(r,t), \tag{7.24}$$

into which we can insert (7.19) on the right-hand side.

For Eurodollar futures we have from (6.11) that the quoted futures price is

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400\, \mathrm{E}^{\mathbb{Q}}_{r,t}\left[(B^{T+0.25}(r,T))^{-1} \right],$$

which in an affine model becomes

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400\, \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{a(0.25)+b(0.25)r_T} \right],$$

where a and b are as in Theorem 7.1. Above, we concluded that for a futures on a zero-coupon bond the futures price is given by

$$\Phi^{T,S}(r,t) = \mathrm{E}^{\mathbb{Q}}_{r,t}\left[B^S(r,T) \right] = \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-a(S-T)-b(S-T)r_T} \right] = e^{-\tilde{a}(T-t)-\tilde{b}(T-t)r},$$

where $\tilde{a}$ and $\tilde{b}$ solve the differential equations (7.20)–(7.21) with $\tilde{a}(0) = a(S-T)$, $\tilde{b}(0) = b(S-T)$. Analogously, we get that

$$\mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{a(0.25)+b(0.25)r_T} \right] = e^{-\hat{a}(T-t)-\hat{b}(T-t)r},$$

where $\hat{a}$ and $\hat{b}$ solve the same differential equations, but with the conditions $\hat{a}(0) = -a(0.25)$, $\hat{b}(0) = -b(0.25)$. In particular, $\hat{a}$ is given as

$$\hat{a}(\tau) = -a(0.25) + \hat{\varphi} \int_0^\tau \hat{b}(u)\, du - \frac{1}{2}\delta_1 \int_0^\tau \hat{b}(u)^2\, du. \tag{7.25}$$

The quoted Eurodollar futures price is therefore

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400 e^{-\hat{a}(T-t)-\hat{b}(T-t)r}. \tag{7.26}$$

If $\delta_2 = 0$, we have $\hat{b}(\tau) = b(\tau) - b(\tau + 0.25)$.

7.2.3 European options on bonds

In Chapter 6 we obtained general pricing formulas for a European call option on a zero-coupon bond. When we explicitly indicate the dependence of prices on the short-term interest rate, the formula (6.16) becomes

$$C^{K,T,S}(r_t,t) = B^T(r_t,t)\, \mathrm{E}^{\mathbb{Q}^T}_t\left[\max\left(B^S(r_T,T) - K, 0 \right) \right].$$

In an affine model where r_T is normally distributed under the T-forward martingale measure $\mathbb{Q}^T$, it follows from the bond pricing formula (7.5) that the bond price $B^S(r_T, T)$ will be lognormally distributed and we will end up with a Black–Scholes–Merton type pricing formula (see Section 4.8). This will be the case in an affine model with a constant volatility, that is $\beta(r) = \beta$ corresponding to $\delta_2 = 0, \delta_1 = \beta^2$. To obtain the precise formula, we need to know the expectation and variance of $B^S(r_T, T)$. It is computationally convenient to use the fact that we can replace the bond price at the maturity of the option, $B^S(r_T, T)$, by the forward price of the underlying bond with delivery at the maturity of the option, $F^{T,S}(r_T, T)$. When $B^S(r_T, T) = F^{T,S}(r_T, T)$ is lognormally distributed under $\mathbb{Q}^T$, it follows from an application of Theorem A.4 in Appendix A that the call price is

$$
\begin{aligned}
C^{K,T,S}(r,t) &= B^T(r,t)\, \mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[\max\left(F^{T,S}(r_T,T) - K, 0\right)\right] \\
&= B^T(r,t)\left\{\mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[F^{T,S}(r_T,T)\right] N(d_1) - KN(d_2)\right\},
\end{aligned}
$$

where d_1 and d_2 are given by

$$
\begin{aligned}
d_1 &= \frac{\ln\left(\mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[F^{T,S}(r_T,T)\right]/K\right)}{\sqrt{\mathrm{Var}^{\mathbb{Q}^T}_{r,t}\left[\ln F^{T,S}(r_T,T)\right]}} + \frac{1}{2}\sqrt{\mathrm{Var}^{\mathbb{Q}^T}_{r,t}\left[\ln F^{T,S}(r_T,T)\right]}, \\
d_2 &= d_1 - \sqrt{\mathrm{Var}^{\mathbb{Q}^T}_{r,t}\left[\ln F^{T,S}(r_T,T)\right]}.
\end{aligned}
$$

We also know that forward prices for delivery at time T are martingales under the T-forward martingale measure so that

$$
\mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[F^{T,S}(r_T,T)\right] = F^{T,S}(r,t) = \frac{B^S(r,t)}{B^T(r,t)}.
$$

Hence, the option price can be written as

$$
C^{K,T,S}(r,t) = B^S(r,t)N(d_1) - KB^T(r,t)N(d_2),
$$

with

$$
\begin{aligned}
d_1 &= \frac{1}{v(t,T,S)} \ln\left(\frac{B^S(r,t)}{KB^T(r,t)}\right) + \frac{1}{2}v(t,T,S), \\
d_2 &= d_1 - v(t,T,S)
\end{aligned}
$$

and it only remains to compute

$$
v(t,T,S) \equiv \sqrt{\mathrm{Var}^{\mathbb{Q}^T}_{r,t}\left[\ln F^{T,S}(r_T,T)\right]}.
$$

In order to compute this, recall that the forward price is given as a function of the interest rate in (7.17) and that we are working in the case with constant interest rate volatility, β. We can then apply Itô's Lemma to find the $\mathbb{Q}^T$-dynamics of the forward price. Since the forward price is a $\mathbb{Q}^T$-martingale, the drift will be zero. We get

$$dF^{T,S}(r_t, t) = -F^{T,S}(r_t, t)\beta[b(S-t) - b(T-t)]\, dz_t^T.$$

It follows that

$$\begin{aligned}\ln F^{T,S}(r_T, T) = \ln F^{T,S}(r_t, t) - \frac{1}{2}\beta^2 \int_t^T [b(S-u) - b(T-u)]^2 du \\ - \beta \int_t^T [b(S-u) - b(T-u)]\, dz_u^T\end{aligned}$$

and applying Theorem 3.3 we obtain

$$v(t, T, S)^2 = \mathrm{Var}_{r,t}^{\mathbb{Q}^T}\left[\ln F^{T,S}(r_T, T)\right] = \beta^2 \int_t^T [b(S-u) - b(T-u)]^2\, du.$$

We still need to identify the b function in the specific model. We emphasize that this procedure only works when future values of the short rate are normally distributed.

An alternative to the above procedure is to start with Equation (6.17). When we explicitly indicate the dependence of prices on the short-term interest rate, the formula looks as follows:

$$C^{K,T,S}(r_t, t) = B^S(r_t, t)\mathbb{Q}_t^S(B^S(r_T, T) > K) - KB^T(r_t, t)\mathbb{Q}_t^T(B^S(r_T, T) > K).$$

In affine models we can use the general bond price expression (7.5) to get

$$B_T^S > K \quad \Leftrightarrow \quad r_T < -\frac{a(S-T)}{b(S-T)} - \frac{\ln K}{b(S-T)}.$$

We need to compute the probability of this event under the two forward martingale measures $\mathbb{Q}^S$ and $\mathbb{Q}^T$ conditional on the current interest rate, r_t. From Equation (4.21) we know that the link between the forward martingale measure $\mathbb{Q}^S$ and the risk-neutral probability measure is captured by the sensitivity of the price of the zero-coupon bond maturing at S, which in a homogeneous affine model is known from (7.15). Consequently, we have

$$dz_t^{\mathbb{Q}} = dz_t^S - b(S-t)\beta(r_t)\, dt,$$

so that the $\mathbb{Q}^S$-dynamics of the short rate becomes

$$\begin{aligned}dr_t &= \hat{\alpha}(r_t)\, dt + \beta(r_t)\left(dz_t^S - b(S-t)\beta(r_t)\, dt\right) \\ &= \left(\hat{\alpha}(r_t) - \beta(r_t)^2 b(S-t)\right)\, dt + \beta(r_t)\, dz_t^S \\ &= \left([\hat{\varphi} - \delta_1 b(S-t)] - [\hat{\kappa} + \delta_2 b(S-t)]r_t\right)\, dt + \sqrt{\delta_1 + \delta_2 r_t}\, dz_t^S.\end{aligned} \tag{7.27}$$

The $\mathbb{Q}^T$-dynamics is similar, just replace S by T. Note that also under both these measures, the short-rate has an 'affine' dynamics although with a time-dependent drift. This will facilitate the computation of the probabilities entering the option pricing formula above.

A reasonable affine one-factor model must have the property that bond prices are decreasing in the short rate, which is the case if the function $b(\tau)$ is positive. This is true in the specific models studied later in this chapter. This property can be used to show that a European call option on a coupon bond can be seen as a portfolio of European call options on zero-coupon bonds. Since this result was first derived by Jamshidian (1989), we shall refer to it as **Jamshidian's trick**. As always, the underlying coupon bond is assumed to pay Y_i at time T_i $(i = 1, 2, \ldots, n)$, where $T_1 < T_2 < \cdots < T_n$, so that the price of the bond is

$$B(r, t) = \sum_{T_i > t} Y_i B^{T_i}(r, t),$$

where we sum over all the future payment dates.

Theorem 7.3 *In an affine one-factor model, where the zero-coupon bond prices are given by* (7.5) *with* $b(\tau) > 0$ *for all* τ*, the price of a European call on a coupon bond is*

$$C^{K,T,cpn}(r, t) = \sum_{T_i > T} Y_i C^{K_i,T,T_i}(r, t), \tag{7.28}$$

where $K_i = B^{T_i}(r^*, T)$*, and* r^* *is defined as the solution to the equation*

$$B(r^*, T) = K.$$

Proof: The payoff of the option on the coupon bond is

$$\max(B(r_T, T) - K, 0) = \max\left(\sum_{T_i > T} Y_i B^{T_i}(r_T, T) - K, 0\right).$$

Since the zero-coupon bond price $B^{T_i}(r_T, T)$ is a monotonically decreasing function of the interest rate r_T, the whole sum $\sum_{T_i > T} Y_i B^{T_i}(r_T, T)$ is monotonically decreasing in r_T. Therefore, exactly one value r^* of r_T will make the option finish **at the money**, that is

$$B(r^*, T) = \sum_{T_i > T} Y_i B^{T_i}(r^*, T) = K.$$

Letting $K_i = B^{T_i}(r^*, T)$, we have that $\sum_{T_i > T} Y_i K_i = K$.

For $r_T < r^*$,

$$\sum_{T_i > T} Y_i B^{T_i}(r_T, T) > \sum_{T_i > T} Y_i B^{T_i}(r^*, T) = K,$$

and

$$B^{T_i}(r_T, T) > B^{T_i}(r^*, T) = K_i,$$

so that

$$\max\left(\sum_{T_i>T} Y_i B^{T_i}(r_T,T) - K, 0\right) = \sum_{T_i>T} Y_i B^{T_i}(r_T,T) - K$$
$$= \sum_{T_i>T} Y_i \left(B^{T_i}(r_T,T) - K_i\right)$$
$$= \sum_{T_i>T} Y_i \max\left(B^{T_i}(r_T,T) - K_i, 0\right).$$

For $r_T \geq r^*$,

$$\sum_{T_i>T} Y_i B^{T_i}(r_T,T) \leq \sum_{T_i>T} Y_i B^{T_i}(r^*,T) = K,$$

and

$$B^{T_i}(r_T,T) \leq B^{T_i}(r^*,T) = K_i,$$

so that

$$\max\left(\sum_{T_i>T} Y_i B^{T_i}(r_T,T) - K, 0\right) = 0 = \sum_{T_i>T} Y_i \max\left(B^{T_i}(r_T,T) - K_i, 0\right).$$

Hence, for *all* possible values of r_T we may conclude that

$$\max\left(\sum_{T_i>T} Y_i B^{T_i}(r_T,T) - K, 0\right) = \sum_{T_i>T} Y_i \max\left(B^{T_i}(r_T,T) - K_i, 0\right).$$

The payoff of the option on the coupon bond is thus identical to the payoff of a portfolio of options on zero-coupon bonds, namely a portfolio consisting (for each i with $T_i > T$) of Y_i options on a zero-coupon bond maturing at time T_i and an exercise price of K_i. Consequently, to rule out arbitrage, the value of the option on the coupon bond at time $t \leq T$ equals the value of that portfolio of options on zero-coupon bonds. The formal derivation is as follows:

$$C^{K,T,\mathrm{cpn}}(r,t) = \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-\int_t^T r_u\,du} \max\left(B(r_T,T) - K, 0\right)\right]$$
$$= \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-\int_t^T r_u\,du} \sum_{T_i>T} Y_i \max\left(B^{T_i}(r_T,T) - K_i, 0\right)\right]$$
$$= \sum_{T_i>T} Y_i\, \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-\int_t^T r_u\,du} \max\left(B^{T_i}(r_T,T) - K_i, 0\right)\right]$$
$$= \sum_{T_i>T} Y_i C^{K_i,T,T_i}(r,t),$$

which completes the proof. □

To compute the price of a European call option on a coupon bond we must numerically solve one equation in one unknown (to find r^*) and calculate n' prices of European call options on zero-coupon bonds, where n' is the number of payment dates of the coupon bond after expiration of the option. In the following sections we shall go through three different time-homogeneous, affine models in which the price of a European call option on a zero-coupon bond is given by relatively simple Black–Scholes type expressions.[4]

The price of a European call with expiration date T and an exercise price of K_i which is written on a zero-coupon bond maturing at T_i is given by

$$C^{K_i,T,T_i}(r,t) = B^{T_i}(r,t)\,\mathbb{Q}^{T_i}_{r,t}\left(B^{T_i}(r_T,T) > K_i\right) \\ - K_i B^{T}(r,t)\,\mathbb{Q}^{T}_{r,t}\left(B^{T_i}(r_T,T) > K_i\right).$$

In the proof of Theorem 7.3 we found that

$$B^{T_i}(r_T,T) > K_i \quad \Leftrightarrow \quad r_T < r^*$$

for all i. Together with Theorem 7.3 these expressions imply that the price of a European call on a coupon bond can be written as

$$\begin{aligned} C^{K,T,\mathrm{cpn}}(r,t) &= \sum_{T_i>T} Y_i \left\{ B^{T_i}(r,t)\,\mathbb{Q}^{T_i}_{r,t}(r_T < r^*) - K_i B^{T}(r,t)\,\mathbb{Q}^{T}_{r,t}(r_T < r^*)\right\} \\ &= \sum_{T_i>T} Y_i B^{T_i}(r,t)\,\mathbb{Q}^{T_i}_{r,t}(r_T < r^*) - K B^{T}(r,t)\,\mathbb{Q}^{T}_{r,t}(r_T < r^*). \end{aligned}$$

Note that the probabilities involved are probabilities of the option finishing in the money under different probability measures. The precise model specifications will determine these probabilities and, hence, the option price.

7.3 MERTON'S MODEL

7.3.1 The short rate process

Apparently, the first continuous-time model of the term structure of interest rates was introduced by Merton (1970). In his model the short rate follows a generalized Brownian motion under the risk-neutral probability measure,

$$dr_t = \hat{\varphi}\,dt + \beta\,dz^{\mathbb{Q}}_t,$$

[4] As discussed by Wei (1997), a very precise approximation of the price can be obtained by computing the price of just one European call option on a particular zero-coupon bond. However, since the exact price can be computed very quickly by Jamshidian's trick, the approximation is not that useful in these one-factor models, but more appropriate in multi-factor models. We will examine the approximation more closely in Chapter 12.

where $\hat{\varphi}$ and β are constants. This is a very simple time-homogeneous affine model with a constant drift rate and volatility, which contradicts empirical observations. This assumption implies that

$$r_T = r_t + \hat{\varphi}[T - t] + \beta[z_T^{\mathbb{Q}} - z_t^{\mathbb{Q}}], \quad t < T.$$

Since $z_T^{\mathbb{Q}} - z_t^{\mathbb{Q}} \sim N(0, T - t)$, we see that, given the short rate $r_t = r$ at time t, the future short rate r_T is normally distributed under the risk-neutral measure with mean and variance

$$\mathrm{E}_{r,t}^{\mathbb{Q}}[r_T] = r + \hat{\varphi}[T - t], \quad \mathrm{Var}_{r,t}^{\mathbb{Q}}[r_T] = \beta^2[T - t].$$

If the market price of risk $\lambda(r_t, t)$ is constant, the drift rate of the short rate under the real-world probability measure will also be a constant $\varphi = \hat{\varphi} + \beta\lambda$. In this case the future short rate is also normally distributed under the real-world probability measure with mean $r + \varphi[T - t]$ and variance $\beta^2[T - t]$.

A model (like Merton's) where the future short rate is normally distributed is called a **Gaussian model**. A normally distributed random variable can take on any real valued number, so the value space $\mathcal{S}$ for the interest rate in a Gaussian model is $\mathcal{S} = \mathbb{R}$.[5] In particular, the short rate in a Gaussian model can be negative with strictly positive probability, which conflicts with both economic theory and empirical observations. If the interest rate is negative, a loan is to be repaid with a lower amount than the original proceeds. This allows so-called **mattress arbitrage**: borrow money and put it into your mattress until the loan is due. The difference between the proceeds and the repayment is a risk-free profit. Note, however, that in a deflation period the smaller amount to be repaid may represent a higher purchasing power than the original proceeds, so in such an economic environment borrowing at negative nominal rates is not an arbitrage. On the other hand, who would lend money at a negative nominal rate? It is certainly advantageous to keep the money in the pocket where it earns a zero interest rate. Hence, nominal interest rates should stay non-negative.[6]

7.3.2 Bond pricing

Merton's model is of the affine form (7.4) with $\hat{\kappa} = 0$, $\delta_1 = \beta^2$, and $\delta_2 = 0$. Theorem 7.1 implies that the prices of zero-coupon bonds in Merton's model are exponentially affine,

$$B^T(r, t) = e^{-a(T-t)-b(T-t)r}.$$

According to (7.6), the function $b(\tau)$ solves the simple ordinary differential equation $b'(\tau) = 1$ with $b(0) = 0$, which implies that

$$b(\tau) = \tau. \tag{7.29}$$

[5] Future interest rates may not have the same distribution under the real-world probability measure and the martingale measures, but we know that the measures are equivalent so that the value space is measure-independent.

[6] Real-life bank accounts often provide some services valuable to the customer, so that their deposit rates (net of fees) may be slightly negative.

The function $a(\tau)$ can then be determined from (7.10):

$$a(\tau) = \hat{\varphi} \int_0^\tau u\, du - \frac{1}{2}\beta^2 \int_0^\tau u^2\, du = \frac{1}{2}\hat{\varphi}\tau^2 - \frac{1}{6}\beta^2\tau^3. \tag{7.30}$$

Since the future short rate is normally distributed, the future zero-coupon bond prices are lognormally distributed in Merton's model.

7.3.3 The yield curve

Let us see which shapes the yield curve can have in Merton's model. The Equations (7.12), (7.29), and (7.30) imply that the τ-maturity zero-coupon yield is

$$y_t^{t+\tau} = r + \frac{1}{2}\hat{\varphi}\tau - \frac{1}{6}\beta^2\tau^2.$$

Hence, for all values of $\hat{\varphi}$ and β, the yield curve is a parabola with downward-sloping branches. The maximum zero-coupon yield is obtained for a time to maturity of $\tau = 3\hat{\varphi}/(2\beta^2)$ and equals $r + 3\hat{\varphi}^2/(8\beta^2)$. Moreover, $y_t^{t+\tau}$ is negative for $\tau > \tau^*$, where

$$\tau^* = \frac{3}{\beta^2}\left(\frac{\hat{\varphi}}{2} + \sqrt{\frac{\hat{\varphi}^2}{4} + \frac{2\beta^2 r}{3}}\right).$$

From (7.16) we see that in Merton's model the τ-maturity zero-coupon rate evolves as

$$d\bar{y}_t^\tau = \alpha(r_t)\, dt + \beta\, dz_t$$

under the real-world probability measure, where $\alpha(r_t) = \hat{\varphi} + \beta\lambda(r_t)$ is the real-world drift rate of the short-term interest rate. Since $d\bar{y}_t^\tau$ is obviously independent of τ, the change in all zero-coupon rates will be identical. In other words, the yield curve will only change by parallel shifts (see also Exercise 7.10). We can therefore conclude that Merton's model can only generate a completely unrealistic form and dynamics of the yield curve. Nevertheless, we will still derive forward prices, futures prices, and European option prices, since this illustrates the general procedure in a simple setting.

7.3.4 Forwards and futures

By substituting the expressions (7.29) and (7.30) into (7.17), we get that the forward price on a zero-coupon bond under Merton's assumptions is

$$\begin{aligned} F^{T,S}(r,t) = \exp\Big\{ & -\frac{1}{2}\left[(S-t)^2 - (T-t)^2\right] \\ & + \frac{1}{6}\beta^2\left[(S-t)^3 - (T-t)^3\right] - (S-T)r \Big\}. \end{aligned}$$

In Merton's model δ_2 equals 0, so by Theorem 7.2 the $\tilde{b}$ function in the futures price on a zero-coupon bond is given by $\tilde{b}(\tau) = b(\tau + S - T) - b(\tau) = S - T$. Applying (7.22), the futures price can be written as

$$\Phi^{T,S}(r,t) = \exp\left\{\frac{1}{2}\hat{\varphi}(S-T)(S+T-2t) \right.$$
$$\left. - \frac{1}{6}\beta^2(S-T)^2(2T+S-3t) - (S-T)r\right\}.$$

Forward and futures prices on coupon bonds can be found by inserting the expressions above into (7.23) and (7.24).

In Equation (7.26), we get $\hat{b}(\tau) = b(\tau) - b(\tau + 0.25) = -0.25$ and from (7.25) we conclude that

$$\begin{aligned}\hat{a}(\tau) &= -a(0.25) - 0.25\hat{\varphi}\tau - \frac{1}{2}(0.25)^2\beta^2\tau \\ &= -\frac{1}{2}(0.25)^2\hat{\varphi} + \frac{1}{6}(0.25)^3\beta^2 - 0.25\hat{\varphi}\tau - \frac{1}{2}(0.25)^2\beta^2\tau.\end{aligned}$$

The quoted Eurodollar futures price in Merton's model is therefore

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400e^{-\hat{a}(\tau)+0.25r}.$$

7.3.5 Option pricing

Since the future values of the short rate are normally distributed in Merton's setting, we conclude from the analysis in Section 7.2.3 that the price of a European call option on a zero-coupon bond is given by

$$C^{K,T,S}(r,t) = B^S(r,t)N(d_1) - KB^T(r,t)N(d_2), \tag{7.31}$$

with

$$\begin{aligned}d_1 &= \frac{1}{v(t,T,S)}\ln\left(\frac{B^S(r,t)}{KB^T(r,t)}\right) + \frac{1}{2}v(t,T,S), \\ d_2 &= d_1 - v(t,T,S)\end{aligned}$$

and, since $b(\tau) = \tau$, we have

$$v(t,T,S)^2 = \beta^2\int_t^T [S-u-(T-u)]^2\,du = \beta^2(S-T)^2(T-t).$$

The price of a European call option on a coupon bond can be found by combining the pricing formula (7.31) and Jamshidian's trick of Theorem 7.3:

$$\begin{aligned}
C^{K,T,\mathrm{cpn}}(r,t) &= \sum_{T_i>T} Y_i \left\{ B^{T_i}(r,t) N(d_1^i) - K_i B^T(r,t) N(d_2^i) \right\} \\
&= \sum_{T_i>T} Y_i B^{T_i}(r,t) N(d_1^i) - B^T(r,t) \sum_{T_i>T} Y_i K_i N(d_2^i) \\
&= \sum_{T_i>T} Y_i B^{T_i}(r,t) N(d_1^i) - K B^T(r,t) N(d_2^i),
\end{aligned}$$

where

$$\begin{aligned}
d_1^i &= \frac{1}{v(t,T,T_i)} \ln\left(\frac{B^{T_i}(r,t)}{K_i B^T(r,t)} \right) + \frac{1}{2} v(t,T,T_i), \\
d_2^i &= d_1^i - v(t,T,T_i), \\
v(t,T,T_i) &= \beta [T_i - T] \sqrt{T-t},
\end{aligned}$$

and we have used the fact that the d_2^i's are identical, compare the discussion at the end of Section 7.2.3.

7.4 VASICEK'S MODEL

7.4.1 The short rate process

One of the inappropriate properties of Merton's model is the constant drift of the short rate. With a constant positive [negative] drift the short rate is expected to increase [decrease] in all the futures which is certainly not realistic. Many empirical studies find that interest rates exhibit **mean reversion** in the sense that if an interest rate is high by historical standards, it will typically fall in the near future. Conversely if the current interest rate is low. Vasicek (1977) assumes that the short rate follows an Ornstein–Uhlenbeck process:

$$dr_t = \kappa[\theta - r_t]\, dt + \beta\, dz_t, \tag{7.32}$$

where κ, θ, and β are positive constants. Note that this is the dynamics under the real-world probability measure. As we saw in Section 3.8.2, this process is mean reverting. If $r_t > \theta$, the drift of the interest rate is negative so that the short rate tends to fall towards θ. If $r_t < \theta$, the drift is positive so that the short rate tends to increase towards θ. The short rate is therefore always drawn towards θ, which we call the long-term level of the short rate. However, shocks may pull the short rate further away from its long-term level. The parameter κ determines the speed of adjustment. As in Merton's model the volatility of the short rate is constant, which conflicts with empirical studies of interest rates. See Section 3.8.2 for simulated paths illustrating the impact of the various parameters on the process.

It follows from Section 3.8.2 that Vasicek's model is a Gaussian model. More precisely, the future short rate in Vasicek's model is normally distributed with mean and variance given by

$$\mathrm{E}_{r,t}\left[r_T\right] = \theta + (r - \theta)e^{-\kappa[T-t]}, \tag{7.33}$$

$$\mathrm{Var}_{r,t}\left[r_T\right] = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa[T-t]}\right) \tag{7.34}$$

under the real-world probability measure $\mathbb{P}$. As $T \to \infty$, the mean approaches θ and the variance approaches $\beta^2/(2\kappa)$. As $\kappa \to \infty$, the mean approaches the long-term level θ and the variance approaches zero. As $\kappa \to 0$, the mean approaches the current short rate r_t and the variance approaches $\beta^2[T-t]$. The current difference between the short rate and its long-term level is expected to be halved over a time period of length $T - t = (\ln 2)/\kappa$.

Like other Gaussian models, Vasicek's model assigns a positive probability to negative values of the future short rate (and all other future rates), despite the inappropriateness of this property. Figure 7.1 illustrates the distribution of the short

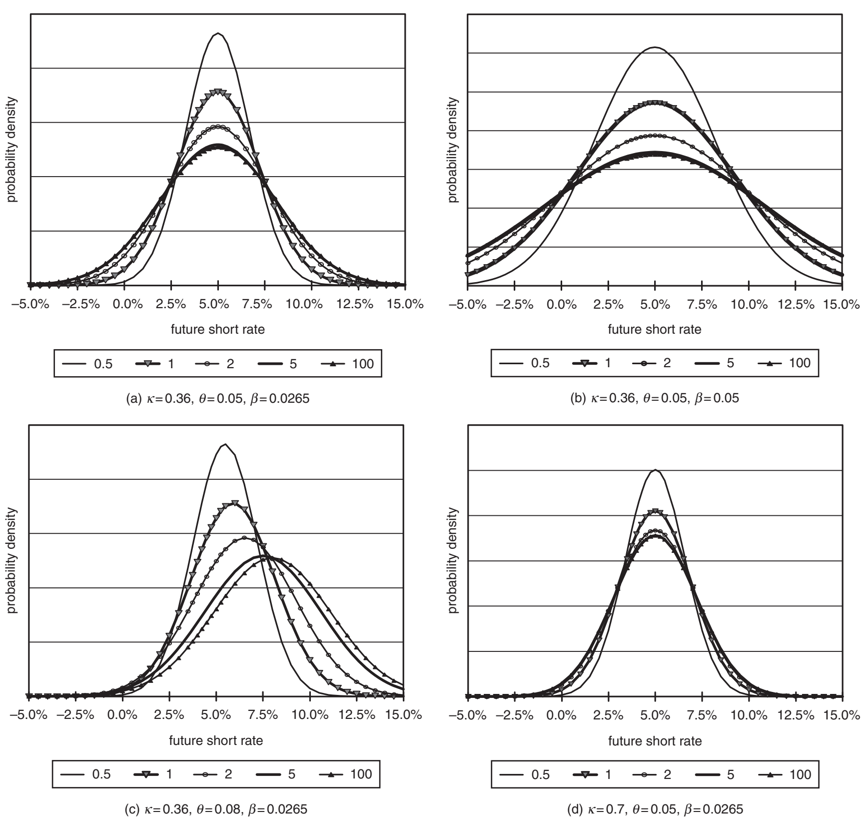

Fig. 7.1: The distribution of r_T for $T - t = 0.5, 1, 2, 5, 100$ years given a current short rate of $r_t = 0.05$.

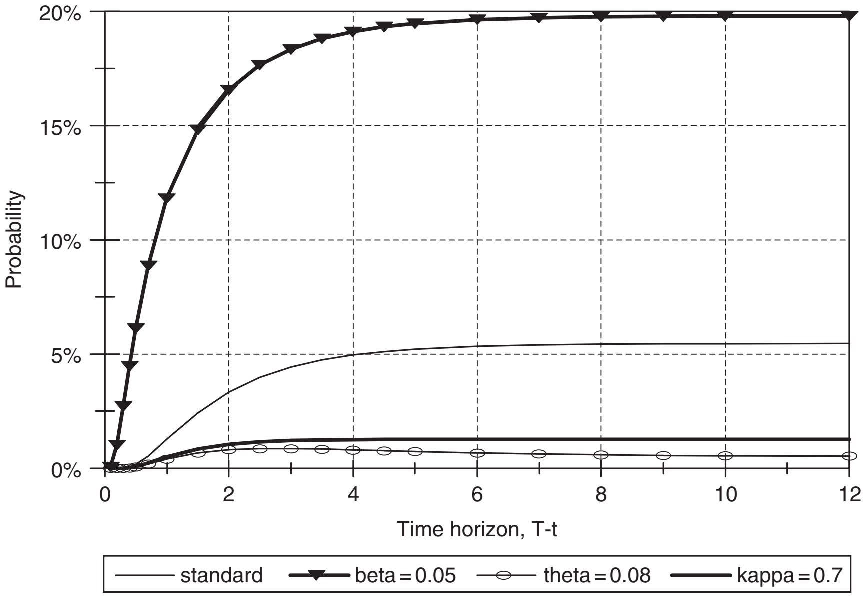

Fig. 7.2: The probability that r_T is negative given $r_t = 0.05$ as a function of the horizon $T - t$. The benchmark parameter values are $\kappa = 0.36$, $\theta = 0.05$, and $\beta = 0.0265$.

rate $\frac{1}{2}, 1, 2, 5$, and 100 years into the future at a current short rate of $r_t = 0.05$ and four different parameter combinations. Figure 7.2 shows the real-world probability of the future short rate r_T being negative (given r_t) for the same four parameter constellations. Since $(r_T - \mathrm{E}_{r,t}[r_T])/\sqrt{\mathrm{Var}_{r,t}[r_T]}$ is standard normally distributed, this probability is easily computed as

$$\mathbb{P}_{r,t}(r_T < 0) = \mathbb{P}_{r,t}\left(\frac{r_T - \mathrm{E}_{r,t}[r_T]}{\sqrt{\mathrm{Var}_{r,t}[r_T]}} < -\frac{\mathrm{E}_{r,t}[r_T]}{\sqrt{\mathrm{Var}_{r,t}[r_T]}}\right) = N\left(-\frac{\mathrm{E}_{r,t}[r_T]}{\sqrt{\mathrm{Var}_{r,t}[r_T]}}\right),$$

into which we can insert (7.33) and (7.34). Clearly, this probability is increasing in the interest rate volatility β and decreasing in the speed of adjustment κ, in the long-term level θ, and in the current level of the short rate r.

For pricing purposes we are interested in the dynamics of the short rate under the risk-neutral (spot martingale) measure and other relevant martingale measures. Vasicek assumed without any explanation that the market price of r-risk is constant, $\lambda(r, t) = \lambda$. As discussed in Section 5.4, it is possible to construct an equilibrium model resulting in Vasicek's assumptions. Since absence of arbitrage is necessary for an equilibrium to exist, it may seem odd that a model allowing negative interest rates is consistent with equilibrium. The reason is that the model does not allow agents to hold cash, so that the 'mattress arbitrage' strategy cannot be implemented. Therefore, the equilibrium model supporting the Vasicek model does not eliminate the critique of the lack of realism of Vasicek's model.

With $\lambda(r,t) = \lambda$, the dynamics of the short rate under the risk-neutral measure $\mathbb{Q}$ becomes

$$\begin{aligned} dr_t &= \kappa[\theta - r_t]\,dt + \beta\left(dz_t^{\mathbb{Q}} - \lambda\,dt\right) \\ &= \kappa[\hat{\theta} - r_t]\,dt + \beta\,dz_t^{\mathbb{Q}}, \end{aligned} \tag{7.35}$$

where $\hat{\theta} = \theta - \lambda\beta/\kappa$. Relative to the real-world dynamics, the only difference is that the parameter θ is replaced by $\hat{\theta}$. Hence, the process has the same qualitative properties under the two probability measures.

7.4.2 Bond pricing

Vasicek's model is an affine model since (7.35) is of the form (7.4) with $\hat{\kappa} = \kappa$, $\hat{\varphi} = \kappa\hat{\theta}$, $\delta_1 = \beta^2$, and $\delta_2 = 0$. It follows from Theorem 7.1 that the price of a zero-coupon bond is

$$B^T(r,t) = e^{-a(T-t)-b(T-t)r}, \tag{7.36}$$

where $b(\tau)$ satisfies the ordinary differential equation

$$\kappa b(\tau) + b'(\tau) - 1 = 0, \quad b(0) = 0,$$

which has the solution

$$b(\tau) = \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right), \tag{7.37}$$

and from (7.10) we get

$$a(\tau) = \kappa\hat{\theta}\int_0^\tau b(u)\,du - \frac{1}{2}\beta^2\int_0^\tau b(u)^2\,du = y_\infty[\tau - b(\tau)] + \frac{\beta^2}{4\kappa}b(\tau)^2. \tag{7.38}$$

Here we have introduced the auxiliary parameter

$$y_\infty = \hat{\theta} - \frac{\beta^2}{2\kappa^2} = \theta - \frac{\lambda\beta}{\kappa} - \frac{\beta^2}{2\kappa^2}$$

and used that

$$\int_0^\tau b(u)\,du = \frac{1}{\kappa}(\tau - b(\tau)), \qquad \int_0^\tau b(u)^2\,du = \frac{1}{\kappa^2}(\tau - b(\tau)) - \frac{1}{2\kappa}b(\tau)^2.$$

In Section 7.4.3 we shall see that y_∞ is the 'long rate', that is the limit of the zero-coupon yields as the maturity goes to infinity.

Let us look at some of the properties of the zero-coupon bond price. Simple differentiation yields

$$\frac{\partial B^T}{\partial r}(r,t) = -b(T-t)B^T(r,t), \qquad \frac{\partial^2 B^T}{\partial r^2}(r,t) = b(T-t)^2 B^T(r,t).$$

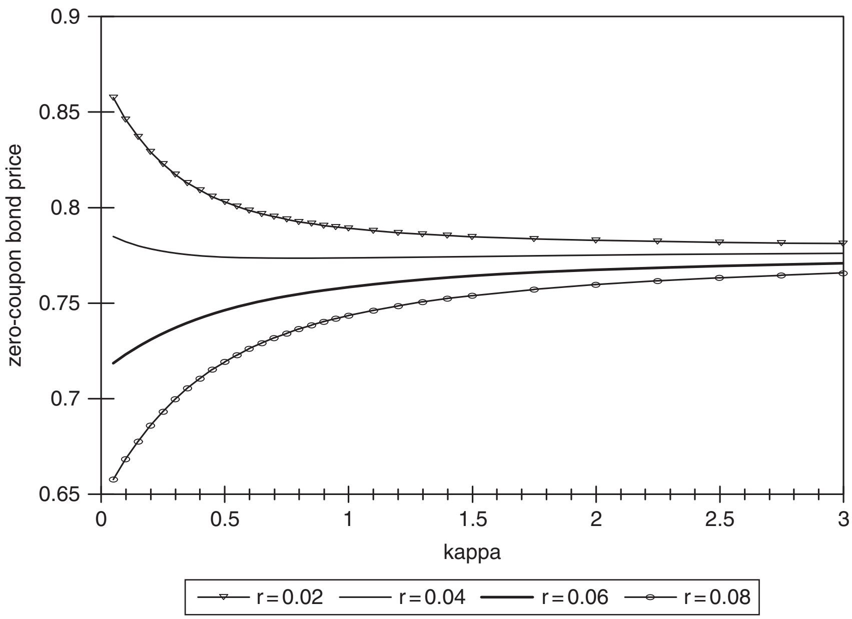

Fig. 7.3: The price of a 5-year zero-coupon bond as a function of the speed of adjustment parameter κ for different values of the current short rate r. The other parameter values are $\theta = 0.05$, $\beta = 0.03$, and $\lambda = -0.15$.

Since $b(\tau) > 0$, the zero-coupon price is a convex, decreasing function of the short rate.

The dependence of the zero-coupon bond price on the parameter κ is illustrated in Fig. 7.3. A high value of κ implies that the future short rate is very likely to be close to θ, and hence the zero-coupon bond price will be relatively insensitive to the current short rate. For $\kappa \to \infty$, the zero-coupon bond price approaches $\exp\{-\theta[T-t]\}$, which is 0.7788 for $\theta = 0.05$ and $T - t = 5$ as in the figure.[7] Conversely, the zero-coupon bond price is highly dependent on the short rate for low values of κ. If the current short rate is below the long-term level, a high κ will imply that $\int_t^T r_u\,du$ is expected to be larger (and $\exp\{-\int_t^T r_u\,du\}$ smaller) than for a low value of κ. In this case, the zero-coupon bond price $B^T(r,t) = \mathrm{E}^{\mathbb{Q}}_{r,t}\left[\exp\left(-\int_t^T r_u\,du\right)\right]$ is thus decreasing in κ. The converse relation holds whenever the current short rate exceeds the long-term level.

Clearly, the zero-coupon price is decreasing in θ as shown in Fig. 7.4 since with higher θ we expect higher future rates and, consequently, a higher value of $\int_t^T r_u\,du$. The prices of long maturity bonds are more sensitive to changes in θ since in the long run θ is more important than the current short rate.

Figure 7.5 shows the relation between zero-coupon bond prices and the interest rate volatility β. Obviously, the price is not a monotonic function of β. For low

[7] Note that $\hat{\theta}$ approaches θ for $\kappa \to \infty$.

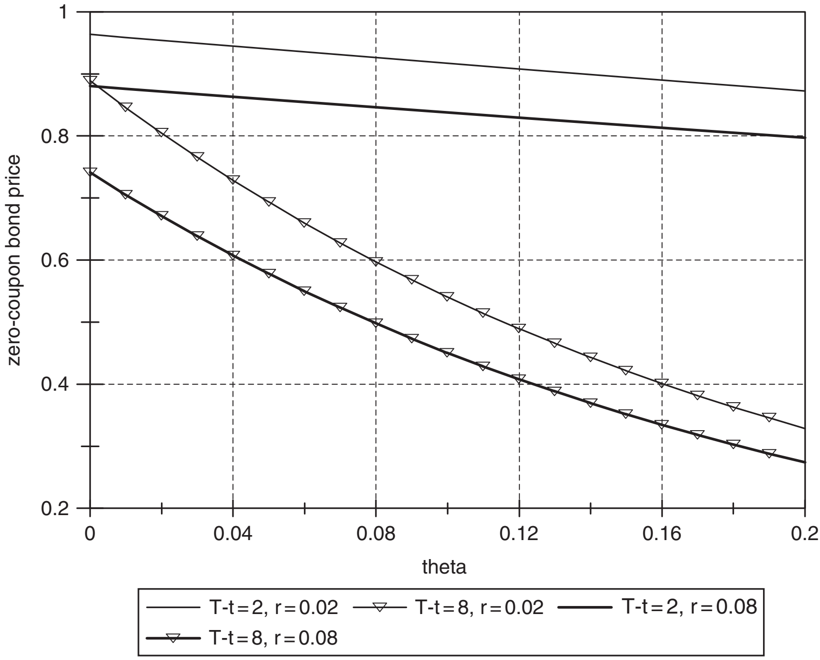

Fig. 7.4: The price of a zero-coupon bond $B^T(r,t)$ as a function of the long-term level θ for different combinations of the time to maturity and the current short rate. The other parameter values are $\kappa = 0.3$, $\beta = 0.03$, and $\lambda = -0.15$.

values of β the prices decrease in β, while the opposite is the case for high β-values. Long-term bonds are more sensitive to β than short-term bonds.

Figure 7.6 illustrates how the zero-coupon bond price depends on the market price of risk parameter λ. Formula (7.14) implies that the dynamics of the zero-coupon bond price $B_t^T = B^T(r,t)$ can be written as

$$dB_t^T = B_t^T \left[\left(r_t + \lambda \sigma^T(r_t,t) \right) \, dt + \sigma^T(r_t,t) \, dz_t \right],$$

where $\sigma^T(r_t,t) = -b(T-t)\beta$ is negative. The more negative λ is, the higher is the excess expected return on the bond demanded by the market participants, and hence the lower the current price. Again the dependence is most pronounced for long-term bonds.

We can also see that the price volatility $|\sigma^T(r_t,t)| = b(T-t)\beta$ is independent of the interest rate level and is concavely increasing in the time to maturity. Also note that the price volatility depends on the parameters κ and β, but not on θ or λ.

Finally, Fig. 7.7 depicts the discount function, that is the zero-coupon bond price as a function of the time to maturity. Note that with a negative short rate, the discount function is not necessarily decreasing. For $\tau \to \infty$, $b(\tau)$ will approach $1/\kappa$, whereas $a(\tau) \to -\infty$ if $y_\infty < 0$, and $a(\tau) \to +\infty$ if $y_\infty > 0$. Consequently, if $y_\infty > 0$, the discount function approaches zero for $T \to \infty$, which is a reasonable

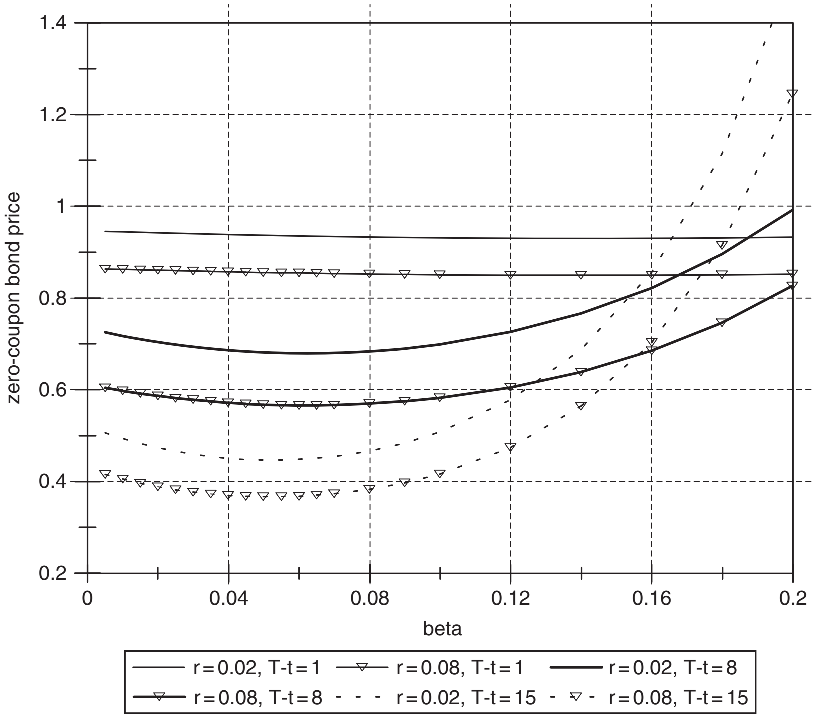

Fig. 7.5: The price of a zero-coupon bond $B^T(r,t)$ as a function of the volatility parameter β for different combinations of the time to maturity $T-t$ and the current short rate r. The values of the fixed parameters are $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = -0.15$.

property. On the other hand, if $y_\infty < 0$, the discount function will diverge to infinity, which is clearly inappropriate. The long rate y_∞ can be negative if the ratio β/κ is sufficiently large.

7.4.3 The yield curve

From (7.11) the zero-coupon yield $y^T(r,t)$ at time t for maturity T is

$$y^T(r,t) = \frac{a(T-t)}{T-t} + \frac{b(T-t)}{T-t} r.$$

Straightforward differentiation results in

$$a'(\tau) = y_\infty[1 - b'(\tau)] + \frac{\beta^2}{2\kappa} b(\tau) b'(\tau), \tag{7.39}$$

$$b'(\tau) = e^{-\kappa\tau}. \tag{7.40}$$

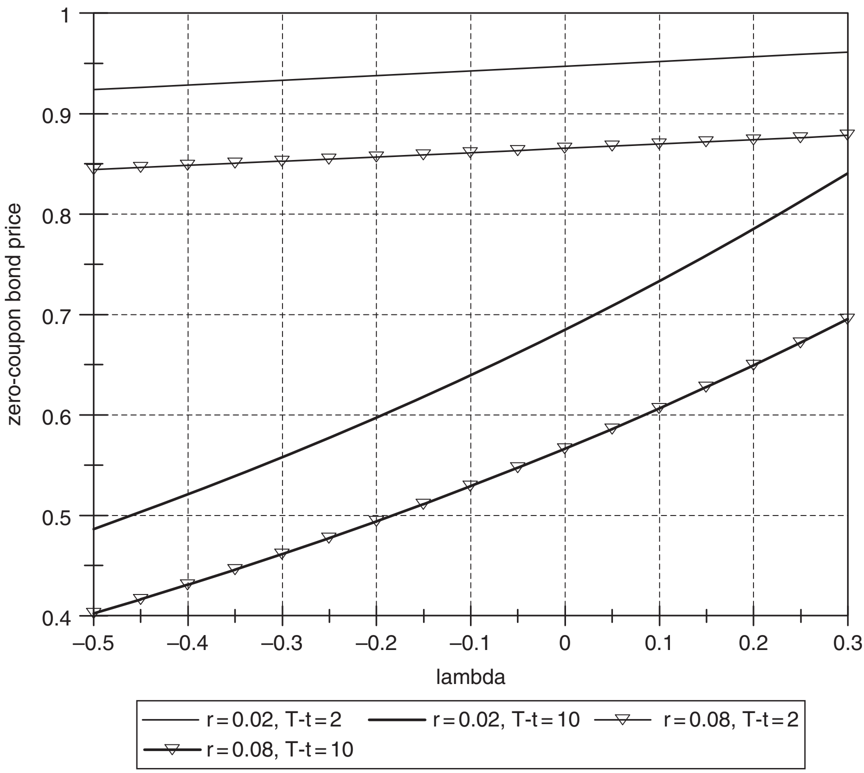

Fig. 7.6: The price of zero-coupon bonds $B^T(r,t)$ as a function of λ for different combinations of the time to maturity $T-t$ and the current short rate r. The values of the fixed parameters are $\kappa = 0.3$, $\theta = 0.05$, and $\beta = 0.03$.

An application of l'Hôpital's rule implies that

$$\lim_{\tau \to 0} \frac{b(\tau)}{\tau} = \lim_{\tau \to 0} \frac{b'(\tau)}{1} = 1 \qquad \text{and} \qquad \lim_{\tau \to 0} \frac{a(\tau)}{\tau} = \lim_{\tau \to 0} \frac{a'(\tau)}{1} = 0,$$

and thus

$$\lim_{T \to t} y^T(r,t) = r,$$

that is the short rate is exactly the intercept of the yield curve as it should be. Similarly, it can be shown that

$$\lim_{\tau \to \infty} \frac{b(\tau)}{\tau} = 0 \qquad \text{and} \qquad \lim_{\tau \to \infty} \frac{a(\tau)}{\tau} = y_\infty,$$

so that

$$\lim_{T \to \infty} y^T(r,t) = y_\infty.$$

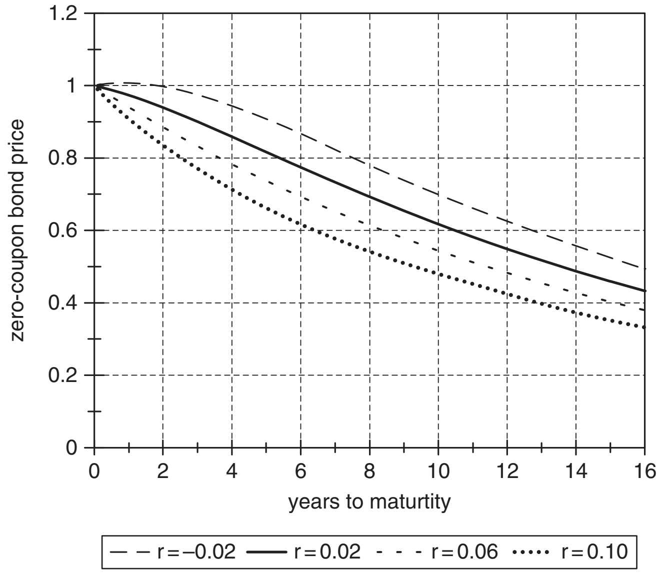

Fig. 7.7: The price of zero-coupon bonds $B^T(r,t)$ as a function of the time to maturity $T-t$. The parameter values are $\kappa = 0.3$, $\theta = 0.05$, $\beta = 0.03$, and $\lambda = -0.15$.

The 'long rate' y_∞ is therefore constant and, in particular, not affected by changes in the short rate. The following theorem lists the possible shapes of the zero-coupon yield curve $T \mapsto y^T(r,t)$ under the assumptions of Vasicek's model.

Theorem 7.4 *In the Vasicek model the zero-coupon yield curve $T \mapsto y^T(r,t)$ will have one of three shapes depending on the parameter values and the current short rate:*

(i) If $r < y_\infty - \frac{\beta^2}{4\kappa^2}$, the yield curve is increasing;
(ii) if $r > y_\infty + \frac{\beta^2}{2\kappa^2}$, the yield curve is decreasing;
(iii) for intermediate values of r, the yield curve is humped, that is increasing in T up to some maturity T^* and then decreasing for longer maturities.

Proof: The zero-coupon yield $y^T(r,t)$ is given by

$$\begin{aligned} y^T(r,t) &= \frac{a(T-t)}{T-t} + \frac{b(T-t)}{T-t} r \\ &= y_\infty + \frac{b(T-t)}{T-t}\left(\frac{\beta^2}{4\kappa} b(T-t) + r - y_\infty\right), \end{aligned}$$

where we have inserted (7.38). We are interested in the relation between the zero-coupon yield and the time to maturity $T-t$, that is the function $Y(\tau) = y^{t+\tau}(r,t)$. Defining $h(\tau) = b(\tau)/\tau$, we have

$$Y(\tau) = y_\infty + h(\tau)\left(\frac{\beta^2}{4\kappa}b(\tau) + r - y_\infty\right).$$

A straightforward computation leads to the derivative

$$Y'(\tau) = h'(\tau)\left(\frac{\beta^2}{4\kappa}b(\tau) + r - y_\infty\right) + h(\tau)e^{-\kappa\tau}\frac{\beta^2}{4\kappa},$$

where we have applied that $b'(\tau) = e^{-\kappa\tau}$. Introducing the auxiliary function

$$g(\tau) = b(\tau) + \frac{h(\tau)e^{-\kappa\tau}}{h'(\tau)}$$

we can rewrite $Y'(\tau)$ as

$$Y'(\tau) = h'(\tau)\left(r - y_\infty + \frac{\beta^2}{4\kappa}g(\tau)\right). \tag{7.41}$$

Below we will argue that $h'(\tau) < 0$ for all τ and that $g(\tau)$ is a monotonically increasing function with $g(0) = -2/\kappa$ and $g(\tau) \to 1/\kappa$ for $\tau \to \infty$. This will imply the claims of the theorem as can be seen from the following arguments. If $r - y_\infty + \beta^2/(4\kappa^2) < 0$, then the parenthesis on the right-hand side of (7.41) is negative for all τ. In this case $Y'(\tau) > 0$ for all τ, and hence the yield curve will be monotonically increasing in the maturity. Similarly, the yield curve will be monotonically decreasing in maturity, that is $Y'(\tau) < 0$ for all τ, if $r - y_\infty - \beta^2/(2\kappa^2) > 0$. For the remaining values of r the expression in the parenthesis on the right-hand side of (7.41) will be negative for $\tau \in [0, \tau^*)$ and positive for $\tau > \tau^*$, where τ^* is uniquely determined by the equation

$$r - y_\infty + \frac{\beta^2}{4\kappa}g(\tau^*) = 0.$$

In that case the yield curve is 'humped'.

Now let us show that $h'(\tau) < 0$ for all τ. Simple differentiation yields $h'(\tau) = (e^{-\kappa\tau}\tau - b(\tau))/\tau^2$, which is negative if $e^{-\kappa\tau}\tau < b(\tau)$ or, equivalently, if $1 + \kappa\tau < e^{\kappa\tau}$, which is clearly satisfied (compare the graphs of the functions $1 + x$ and e^x).

Finally, by application of l'Hôpital's rule, it can be shown that $g(0) = -2/\kappa$ and $g(\tau) \to 1/\kappa$ for $\tau \to \infty$. By differentiation and tedious manipulations it can be shown that g is monotonically increasing. □

Figure 7.8 illustrates the possible shapes of the yield curve. For any maturity the zero-coupon rate is an increasing affine function of the short rate. An increase [decrease] in the short rate will therefore shift the whole yield curve upwards [downwards]. The change in the zero-coupon rate will be decreasing in the maturity, so that shifts are not parallel. *Twists* of the yield curve where short rates and long rates move in opposite directions are not possible.

According to (7.13) the instantaneous forward rate $f^T(r,t)$ prevailing at time t is given by

$$f^T(r,t) = a'(T-t) + b'(T-t)r.$$

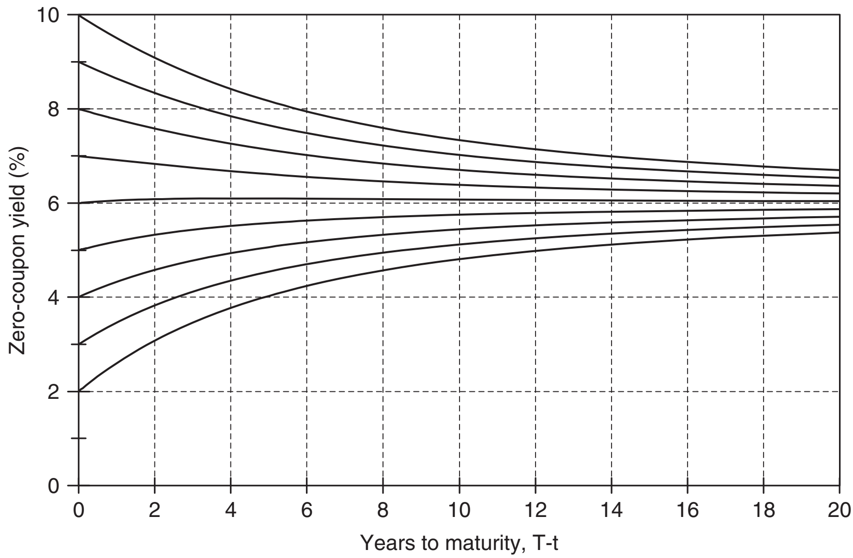

Fig. 7.8: The yield curve for different values of the short rate. The parameter values are $\kappa = 0.3$, $\theta = 0.05$, $\beta = 0.03$, and $\lambda = -0.15$. The long rate is then $y_\infty = 6\%$. The yield curve is increasing for $r < 5.75\%$, decreasing for $r > 6.5\%$, and *humped* for intermediate values of r. The curve for $r = 6\%$ exhibits a very small *hump* with a maximum yield for a time to maturity of approximately 5 years.

Applying (7.39) and (7.40) this expression can be rewritten as

$$\begin{aligned} f^T(r,t) &= -\left(1 - e^{-\kappa[T-t]}\right)\left(\frac{\beta^2}{2\kappa^2}\left(1 - e^{-\kappa[T-t]}\right) - \hat{\theta}\right) + e^{-\kappa[T-t]} r \\ &= \left(1 - e^{-\kappa[T-t]}\right)\left(y_\infty + \frac{\beta^2}{2\kappa^2} e^{-\kappa[T-t]}\right) + e^{-\kappa[T-t]} r. \end{aligned} \tag{7.42}$$

Because the short rate can be negative, so can the forward rates.

7.4.4 Forwards and futures

The forward price on a zero-coupon bond in Vasicek's model is obtained by substituting the functions b and a from (7.37) and (7.38) into the general expression

$$F^{T,S}(r,t) = \exp\{-[a(S-t) - a(T-t)] - [b(S-t) - b(T-t)]\, r\},$$

compare (7.17).

In Vasicek's model the δ_2 parameter in the general dynamics (7.4) is zero, so that the $\tilde{b}$ function involved in the futures price on a zero-coupon bond, $\Phi^{T,S}(r,t) = e^{-\tilde{a}(T-t)-\tilde{b}(T-t)r}$, according to Theorem 7.2 is

$$\tilde{b}(\tau) = b(\tau + S - T) - b(\tau) = e^{-\kappa\tau} b(S - T).$$

Substituting this into (7.22) we get that the $\tilde{a}$ function in the futures price expression is

$$\begin{aligned}\tilde{a}(\tau) &= a(S - T) + \kappa\hat{\theta} b(S - T) \int_0^\tau e^{-\kappa u}\, du - \frac{1}{2}\beta^2 b(S - T)^2 \int_0^\tau e^{-2\kappa u}\, du \\ &= a(S - T) + \kappa\hat{\theta} b(S - T) b(\tau) - \frac{1}{2}\beta^2 b(S - T)^2 \left(b(\tau) - \frac{1}{2}\kappa b(\tau)^2 \right).\end{aligned}$$

Forward and futures prices on coupon bonds are found by inserting the formulas above into (7.23) and (7.24).

For Eurodollar futures, (7.26) implies that the quoted price is given by

$$\tilde{\mathcal{E}}^T(r, t) = 500 - 400 e^{-\hat{a}(T-t) - \hat{b}(T-t) r},$$

and since $\delta_2 = 0$, we have $\hat{b}(\tau) = b(\tau) - b(\tau + 0.25) = -b(0.25) e^{-\kappa\tau}$. From (7.25) we get that

$$\begin{aligned}\hat{a}(\tau) &= -a(0.25) - \kappa\hat{\theta} b(0.25) \int_0^\tau e^{-\kappa u}\, du - \frac{1}{2}\beta^2 b(0.25)^2 \int_0^\tau e^{-2\kappa u}\, du \\ &= -a(0.25) - \kappa\hat{\theta} b(0.25) b(\tau) - \frac{1}{2}\beta^2 b(0.25)^2 \left(b(\tau) - \frac{1}{2}\kappa b(\tau)^2 \right).\end{aligned}$$

7.4.5 Option pricing

The future values of the short rate are normally distributed, so we know from Section 7.2.3 that the price of a European call option on a zero-coupon bond is given by

$$C^{K,T,S}(r, t) = B^S(r, t) N(d_1) - K B^T(r, t) N(d_2), \tag{7.43}$$

with

$$\begin{aligned}d_1 &= \frac{1}{v(t, T, S)} \ln\left(\frac{B^S(r, t)}{K B^T(r, t)} \right) + \frac{1}{2} v(t, T, S), \\ d_2 &= d_1 - v(t, T, S)\end{aligned}$$

and

$$\begin{aligned}v(t, T, S)^2 &= \beta^2 \int_t^T [b(S - u) - b(T - u)]^2\, du \\ &= \frac{\beta^2}{2\kappa^3} \left(1 - e^{-\kappa[S-T]} \right)^2 \left(1 - e^{-2\kappa[T-t]} \right).\end{aligned}$$

This option pricing formula was first derived by Jamshidian (1989).

Figure 7.9 illustrates how the call price depends on the current short rate. An increase in the short rate has the effect that the present value of the exercise price decreases, which leaves the call option more valuable. This effect is known from the Black–Scholes–Merton stock option formula. For bond options there is an additional effect. When the short rate increases, the price of the underlying bond decreases, which will lower the call option value. According to the figure, the latter effect dominates at least for the parameters used when generating the graph. See Exercise 7.3 for more on the relation between the call price and the short rate.

The relation between the call price and the interest rate volatility β is shown in Fig. 7.10. An increase in β yields a higher volatility on the underlying bond, which makes the option more valuable. However, the price of the underlying bond also depends on β. As shown in Fig. 7.5, the bond price will decrease with β for low values of β, and this effect can be so strong that the option price can decrease with β.

Because the function $b(\tau)$ is strictly positive in Vasicek's model, we can apply Jamshidian's trick of Theorem 7.3 for the pricing of a European call option on a coupon bond. This leads to the pricing formula

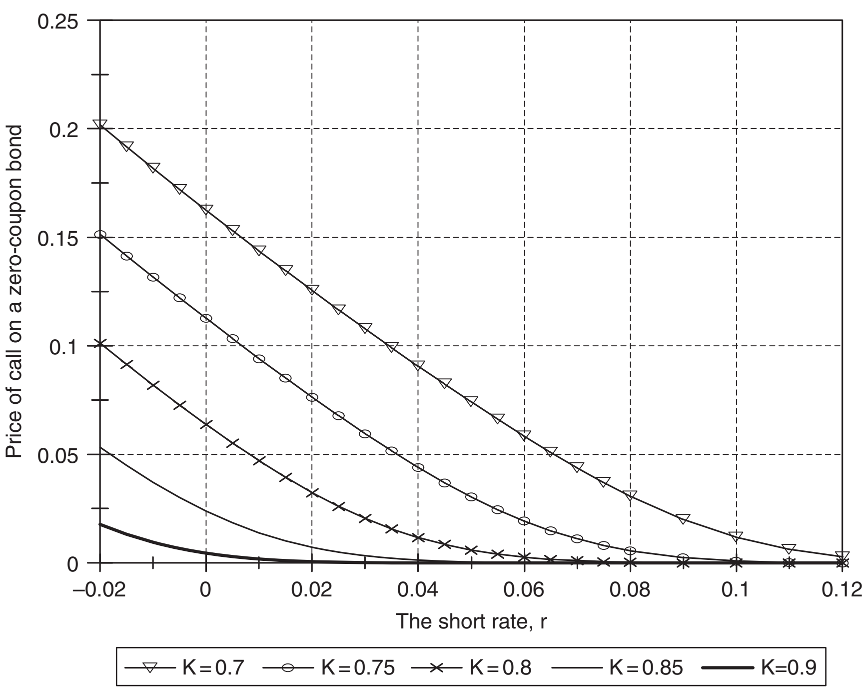

Fig. 7.9: The price of a European call option on a zero-coupon bond as a function of the current short rate r. The option expires in $T - t = 0.5$ years, while the bond matures in $S - t = 5$ years. The prices are computed using Vasicek's model with parameter values $\beta = 0.03$, $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = -0.15$.

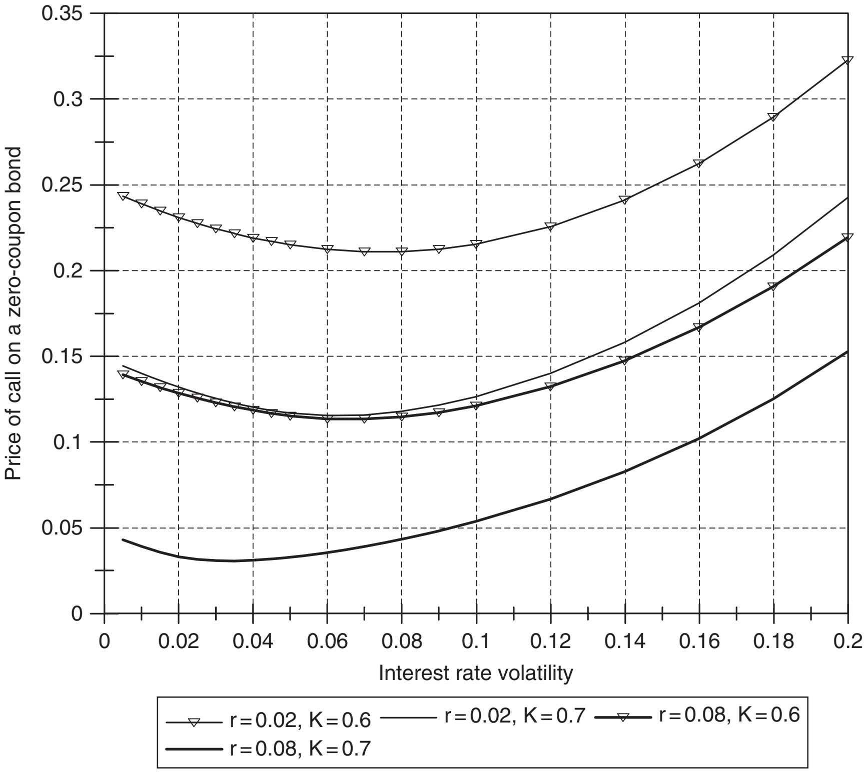

Fig. 7.10: The price of a European call option on a zero-coupon bond as a function of the interest rate volatility β. The option expires in $T - t = 0.5$ years, while the bond matures in $S - t = 5$ years. The prices are computed using Vasicek's model with the parameter values $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = -0.15$.

$$\begin{aligned} C^{K,T,\text{cpn}}(r,t) &= \sum_{T_i > T} Y_i \left\{ B^{T_i}(r,t) N(d_1^i) - K_i B^T(r,t) N(d_2^i) \right\} \\ &= \sum_{T_i > T} Y_i B^{T_i}(r,t) N(d_1^i) - K B^T(r,t) N(d_2^i), \end{aligned}$$

where K_i is defined as $K_i = B^{T_i}(r^*, T)$, r^* is given as the solution to the equation $B(r^*, T) = K$, and

$$\begin{aligned} d_1^i &= \frac{1}{v(t,T,T_i)} \ln\left(\frac{B^{T_i}(r,t)}{K_i B^T(r,t)} \right) + \frac{1}{2} v(t,T,T_i), \\ d_2^i &= d_1^i - v(t,T,T_i), \\ v(t,T,T_i) &= \frac{\beta}{\sqrt{2\kappa^3}} \left(1 - e^{-\kappa[T_i - T]}\right) \left(1 - e^{-2\kappa[T-t]}\right)^{1/2}. \end{aligned}$$

Here we have utilized the knowledge that all the d_2^i's are identical.

7.5 THE COX–INGERSOLL–ROSS MODEL

7.5.1 The short rate process

The one-factor diffusion model suggested by Cox, Ingersoll, and Ross (1985b) is popular both among academics and practitioners. The model assumes that the short rate follows a square-root process

$$dr_t = \kappa\,[\theta - r_t]\;dt + \beta\sqrt{r_t}\,dz_t, \tag{7.44}$$

where κ, θ, and β are positive constants. We will refer to the model as the CIR model. Some of the key properties of square-root processes were discussed in Section 3.8.3. Just as the Vasicek model, the CIR model for the short rate exhibits mean reversion around a long-term level θ. A difference relative to Vasicek's short rate process is the specification of the volatility. In the CIR model the volatility is not constant but an increasing function of the interest rate, so that the short rate is less volatile for low levels than for high levels of the rate. This property seems to be consistent with observed interest rate behaviour, although the square-root relation is not so clear in the data, compare the discussion in Section 7.8. The short rate in the CIR model cannot become negative, which is a major advantage relative to Vasicek's model. As explained in Section 3.8.3, the value space of the short rate in the CIR model is either $\mathcal{S} = (0, \infty)$ or $\mathcal{S} = (0, \infty)$ depending on the parameter values.

In Section 5.4 the CIR model was shown to be a special case of a comprehensive general equilibrium model of the financial markets developed by the same authors in another article, Cox et al. (1985a). The short rate process (7.44) *and* an expression for the market price of interest rate risk, $\lambda(r, t)$, is the output of the general model under specific assumptions on preferences, endowments, and the underlying technology.[8] According to the model the market price of risk is

$$\lambda(r, t) = \frac{\lambda\sqrt{r}}{\beta},$$

where λ on the right-hand side is a constant. The drift of the short rate under the risk-neutral measure is therefore

$$\alpha(r, t) - \beta(r, t)\lambda(r, t) = \kappa[\theta - r] - \frac{\lambda\sqrt{r}}{\beta}\beta\sqrt{r} = \kappa\theta - (\kappa + \lambda)r.$$

Defining $\hat{\kappa} = \kappa + \lambda$ and $\hat{\varphi} = \kappa\theta$, the process for the short rate under the risk-neutral measure can be written as

$$dr_t = (\hat{\varphi} - \hat{\kappa} r_t)\,dt + \beta\sqrt{r_t}\,dz_t^{\mathbb{Q}}. \tag{7.45}$$

Since this is of the form (7.4) with $\delta_1 = 0$ and $\delta_2 = \beta^2$, we see that the CIR model is also an affine model. We can rewrite the dynamics as

$$dr_t = \hat{\kappa}\left[\hat{\theta} - r_t\right]\,dt + \beta\sqrt{r_t}\,dz_t^{\mathbb{Q}},$$

[8] In their general model r is in fact the **real** short-term interest rate, but in practice the model is often used for the **nominal** interest rates.

where $\hat{\theta} = \kappa\theta/(\kappa + \lambda)$. Hence, the short rate also exhibits mean reversion under the risk-neutral probability measure, but both the speed of adjustment and the long-term level are different than under the real-world probability measure. In Vasicek's model only the long-term level was affected by the change of measure.

In the CIR model the distribution of the future short rate r_T (conditional on the current short rate r_t) is given by the non-central χ^2-distribution. The precise density function follows from the analysis of the square-root process in Section 3.8.3. The mean and variance of r_T given $r_t = r$ are

$$\mathrm{E}_{r,t}[r_T] = \theta + (r - \theta)e^{-\kappa[T-t]},$$

$$\mathrm{Var}_{r,t}[r_T] = \frac{\beta^2 r}{\kappa}\left(e^{-\kappa[T-t]} - e^{-2\kappa[T-t]}\right) + \frac{\beta^2\theta}{2\kappa}\left(1 - e^{-\kappa[T-t]}\right)^2.$$

Note that the mean is just as in Vasicek's model, compare (7.33), while the expression for the variance is slightly more complicated than in the Vasicek model, compare (7.34). For $T \to \infty$, the mean approaches θ and the variance approaches $\theta\beta^2/(2\kappa)$. For $\kappa \to \infty$, the mean goes to θ and the variance goes to 0. For $\kappa \to 0$, the mean approaches the current rate r and the variance approaches $\beta^2 r[T - t]$. The future short rate is also non-centrally χ^2-distributed under the risk-neutral measure but, relative to the expressions above, κ is to be replaced by $\hat{\kappa} = \kappa + \lambda$ and θ by $\hat{\theta} = \kappa\theta/(\kappa + \lambda)$.

7.5.2 Bond pricing

The CIR model is affine, so Theorem 7.1 implies that the price of a zero-coupon bond maturing at time T is

$$B^T(r,t) = e^{-a(T-t)-b(T-t)r}, \tag{7.46}$$

where the functions $a(\tau)$ and $b(\tau)$ solve the ordinary differential equations (7.6) and (7.7). For the CIR model, these equations are

$$\frac{1}{2}\beta^2 b(\tau)^2 + \hat{\kappa}b(\tau) + b'(\tau) - 1 = 0, \tag{7.47}$$

$$a'(\tau) - \kappa\theta b(\tau) = 0 \tag{7.48}$$

with the conditions $a(0) = b(0) = 0$. The solution is

$$b(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \hat{\kappa})(e^{\gamma\tau} - 1) + 2\gamma}, \tag{7.49}$$

$$a(\tau) = -\frac{2\hat{\kappa}\hat{\theta}}{\beta^2}\left(\ln(2\gamma) + \frac{1}{2}(\hat{\kappa} + \gamma)\tau - \ln\left[(\gamma + \hat{\kappa})(e^{\gamma\tau} - 1) + 2\gamma\right]\right), \tag{7.50}$$

where $\gamma = \sqrt{\hat{\kappa}^2 + 2\beta^2}$ (you are asked to verify the solution in Exercise 7.5).

Since

$$\frac{\partial B^T}{\partial r}(r,t) = -b(T-t)B^T(r,t), \qquad \frac{\partial^2 B^T}{\partial r^2}(r,t) = b(T-t)^2 B^T(r,t)$$

and $b(\tau) > 0$, the zero-coupon bond price is a convex decreasing function of the short rate. Furthermore, the price is a decreasing function of the time to maturity, a concave increasing function of β^2, a concave increasing function of λ, and a convex decreasing function of θ. The dependence on κ is determined by the relation between the current short rate r and the long-term level θ. If $r > \theta$, the bond price is a concave increasing function of κ. If $r < \theta$, the price is a convex decreasing function of κ.

Manipulating (7.14) slightly, we get that the dynamics of the zero-coupon price $B_t^T = B^T(r,t)$ is

$$dB_t^T = B_t^T \left[r_t \left(1 - \lambda b(T-t)\right)\, dt + \sigma^T(r_t,t)\, dz_t \right],$$

where $\sigma^T(r,t) = -b(T-t)\beta\sqrt{r}$. The volatility $|\sigma^T(r,t)| = b(T-t)\beta\sqrt{r}$ of the zero-coupon bond price is thus an increasing function of the interest rate level and an increasing function of the time to maturity, since $b'(\tau) > 0$ for all τ. Note that the volatility depends on $\hat{\kappa} = \kappa + \lambda$ and β, but not on θ (similar to the Vasicek model).

7.5.3 The yield curve

Next we study the zero-coupon yield curve $T \mapsto y^T(r,t)$. From (7.11) we have that

$$y^T(r,t) = \frac{a(T-t)}{T-t} + \frac{b(T-t)}{T-t} r.$$

It can be shown that $y^t(r,t) = r$ and that

$$y_\infty \equiv \lim_{T\to\infty} y^T(r,t) = \frac{2\hat{\kappa}\hat{\theta}}{\hat{\kappa}+\gamma}.$$

Concerning the shape of the yield curve, Kan (1992) has shown the following result:

Theorem 7.5 *In the CIR model the shape of the yield curve depends on the parameter values and the current short rate as follows:*

(1) *If $\hat{\kappa} > 0$, the yield curve is decreasing for $r \geq \hat{\varphi}/\hat{\kappa} = \kappa\theta/(\kappa+\lambda)$ and increasing for $0 \leq r \leq \hat{\varphi}/\gamma$. For $\hat{\varphi}/\gamma < r < \hat{\varphi}/\hat{\kappa}$, the yield curve is humped, that is first increasing, then decreasing.*

(2) *If $\hat{\kappa} \leq 0$, the yield curve is increasing for $0 \leq r \leq \hat{\varphi}/\gamma$ and humped for $r > \hat{\varphi}/\gamma$.*

The proof of this theorem is rather complicated and is therefore omitted. Estimations of the model typically give $\hat{\kappa} > 0$, so that the first case applies (see references to estimations in Section 7.8).

The term structure of forward rates $T \mapsto f^T(r,t)$ is given by

$$f^T(r,t) = a'(T-t) + b'(T-t)r,$$

which by (7.47) and (7.48) can be rewritten as

$$f^T(r,t) = r + \hat{\kappa}\left[\hat{\theta} - r\right] b(T-t) - \frac{1}{2}\beta^2 r b(T-t)^2. \tag{7.51}$$

7.5.4 Forwards and futures

The forward price on a zero-coupon bond in the CIR model is found by substituting the functions b and a from (7.49) and (7.50) into the general expression

$$F^{T,S}(r,t) = \exp\left\{-\left[a(S-t) - a(T-t)\right] - \left[b(S-t) - b(T-t)\right] r\right\},$$

which is known from (7.17). The forward price on a coupon bond follows from (7.23).

It is more complicated to determine the futures price on a zero-coupon bond in the CIR model than it was in the models of Merton and Vasicek, due to the fact that the parameter δ_2 is non-zero in the CIR model. From Theorem 7.2 we have that the futures price is of the form

$$\Phi^{T,S}(r,t) = e^{-\tilde{a}(T-t)-\tilde{b}(T-t)r},$$

where the function $\tilde{b}$ is a solution to the ordinary differential equation (7.20), which in the CIR model becomes

$$\frac{1}{2}\beta^2\tilde{b}(\tau)^2 + \hat{\kappa}\tilde{b}(\tau) + \tilde{b}'(\tau) = 0, \quad \tau \in (0,T),$$

with the condition $\tilde{b}(0) = b(S-T)$. Then the function $\tilde{a}$ can be determined from (7.22), which in the present case is

$$\tilde{a}(\tau) = a(S-T) + \kappa\theta \int_0^\tau \tilde{b}(u)\, du.$$

The solution is

$$\tilde{b}(\tau) = \frac{2\hat{\kappa}\, b(S-T)}{\beta^2 b(S-T)\left(e^{\hat{\kappa}\tau} - 1\right) + 2\hat{\kappa}\, e^{\hat{\kappa}\tau}},$$

$$\tilde{a}(\tau) = a(S-T) - \frac{2\kappa\theta}{\beta^2} \ln\left(\frac{\tilde{b}(\tau)e^{\hat{\kappa}\tau}}{b(S-T)}\right).$$

The futures price on a coupon bond follows from (7.24).

According to (7.26), the quoted Eurodollar futures price is

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400e^{-\hat{a}(T-t)-\hat{b}(T-t)r},$$

where the functions $\hat{a}$ and $\hat{b}$ in the CIR model can be computed to be

$$\hat{b}(\tau) = \frac{-2\hat{\kappa} b(0.25)}{-\beta^2 b(0.25)\left(e^{\hat{\kappa}\tau} - 1\right) + 2\hat{\kappa} e^{\hat{\kappa}\tau}},$$

$$\hat{a}(\tau) = -a(0.25) - \frac{2\kappa\theta}{\beta^2} \ln\left(-\frac{\hat{b}(\tau) e^{\hat{\kappa}\tau}}{b(0.25)}\right).$$

7.5.5 Option pricing

To price European options on zero-coupon bonds we can try to compute

$$C^{K,T,S}(r,t) = B^T(r,t)\, \mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[\max\left(B^S(r_T,T) - K, 0\right)\right]$$

using the distribution of r_T under the T-forward martingale measure $\mathbb{Q}^T$ (given $r_t = r$). In the models of Merton and Vasicek, this approach was relatively straightforward since r_T was normally distributed and, hence, $B^S(r_T,T)$ lognormally distributed so that we were basically back to the Black–Scholes–Merton case. However, in the CIR model the distribution of r_T and, hence, $B^S(r_T,T)$ is much more complicated.

Instead, we recall from Section 7.2.3 that the option price can alternatively be written as

$$\begin{aligned} C^{K,T,S}(r,t) &= B^S(r,t)\, \mathbb{Q}^S_t\left(r_T < \frac{-\ln K - a(S-T)}{b(S-T)}\right) \\ &\quad - KB^T(r,t)\, \mathbb{Q}^T_t\left(r_T < \frac{-\ln K - a(S-T)}{b(S-T)}\right). \end{aligned} \tag{7.52}$$

Equation (7.27) for the $\mathbb{Q}^S$-dynamics of r specializes to

$$dr_t = \left(\hat{\varphi} - \left[\hat{\kappa} + \beta^2 b(S-t)\right] r_t\right) dt + \beta\sqrt{r_t}\, dz^S_t.$$

In the drift term the coefficient on r_t is now a deterministic function of time, but nevertheless future values will still be non-centrally χ^2-distributed. The probabilities in (7.52) can therefore be computed from the cumulative distribution function of the non-central χ^2-distribution with appropriate parameters. We will skip the details and simply state the resulting pricing formula first derived by Cox et al.:

$$C^{K,T,S}(r,t) = B^S(r,t)\chi^2(h_1; f, g_1) - KB^T(r,t)\chi^2(h_2; f, g_2), \tag{7.53}$$

where $\chi^2(\cdot; f, g)$ is the cumulative distribution function for a non-centrally χ^2-distributed random variable with f degrees of freedom and non-centrality parameter g. The formula has the same structure as in the models of Merton and

Vasicek, but the relevant distribution function is no longer the normal distribution function. The parameters f, g_i, and h_i are given by

$$f = \frac{4\kappa\theta}{\beta^2}, \qquad h_1 = 2\bar{r}\,(\xi + \psi + b(S-T)), \qquad h_2 = 2\bar{r}(\xi + \psi),$$

$$g_1 = \frac{2\xi^2 r e^{\gamma[T-t]}}{\xi + \psi + b(S-T)}, \qquad g_2 = \frac{2\xi^2 r e^{\gamma[T-t]}}{\xi + \psi},$$

and we have introduced the auxiliary parameters

$$\xi = \frac{2\gamma}{\beta^2\left[e^{\gamma[T-t]} - 1\right]}, \qquad \psi = \frac{\hat{\kappa} + \gamma}{\beta^2}, \qquad \bar{r} = -\frac{a(S-T) + \ln K}{b(S-T)}.$$

Note that $\bar{r}$ is exactly the critical interest rate, that is the value of the short rate for which the option finishes at the money, since $B^S(\bar{r}, T) = K$.

To implement the formula, the following approximation of the χ^2-distribution function is useful:

$$\chi^2(h; f, g) \approx N(d),$$

where

$$\begin{aligned}
d &= k\left(\left(\frac{h}{f+g}\right)^m - l\right),\\
m &= 1 - \frac{2}{3}\frac{(f+g)(f+3g)}{(f+2g)^2},\\
k &= \left(2m^2 p\left[1 - p(1-m)(1-3m)\right]\right)^{-1/2},\\
l &= 1 + m(m-1)p - \frac{1}{2}m(m-1)(2-m)(1-3m)p^2,\\
p &= \frac{f+2g}{(f+g)^2}.
\end{aligned}$$

This approximation was originally suggested by Sankaran (1963) and has subsequently been applied in the CIR model by Longstaff (1993). For a more precise approximation, see Ding (1992).

Because of the complexity of the formula it is difficult to evaluate how the call price depends on the parameters and variables involved. Of course, the call price is an increasing function of the time to maturity of the option[9] and a decreasing function of the exercise price. An increase in the short rate r has two effects on the call price: the present value of the exercise price decreases, but the value of the underlying bond also decreases. According to Cox et al. (1985b), numerical computations indicate that the latter effect dominates the first so that the call price is a decreasing function of the interest rate, as we saw it in Vasicek's model.

[9] There are no payments on the underlying asset in the life of the option, so a European call is equivalent to an American call, which clearly increases in value as time to maturity is increased.

For the pricing of European options on coupon bonds we can again apply Jamshidian's trick of Theorem 7.3:

$$C^{K,T,\mathrm{cpn}}(r,t) = \sum_{T_i > T} Y_i B^{T_i}(r,t)\chi^2(h_{1i}; f, g_{1i}) - KB^T(r,t)\chi^2(h_2; f, g_2),$$

where

$$h_{1i} = 2r^* \left(\xi + \psi + b(T_i - T)\right),\ h_2 = 2r^*(\xi + \psi),\ g_{1i} = \frac{2\xi^2 re^{\gamma[T-t]}}{\xi + \psi + b(T_i - T)},$$

and f, g_2, ξ, and ψ are defined just below (7.53). This result was first derived by Longstaff (1993).

7.6 GENERALIZED AFFINE MODELS

In the original Vasicek model of Section 7.4, the market price of risk is specified as a constant, $\lambda(r,t) = \lambda$. Consider the more general specification

$$\lambda(r,t) = \lambda_1 + \lambda_2 r,$$

where λ_1 and λ_2 are constants. If the short rate dynamics under the real-world measure remains unchanged as in (7.32), the risk-neutral drift now becomes

$$\kappa[\theta - r] - \beta[\lambda_1 + \lambda_2 r] = (\kappa\theta - \beta\lambda_1) - (\kappa + \beta\lambda_2)r = \tilde{\kappa}\left[\tilde{\theta} - r\right],$$

where

$$\tilde{\kappa} = \kappa + \beta\lambda_2, \quad \tilde{\theta} = \frac{1}{\tilde{\kappa}}(\kappa\theta - \beta\lambda_1) = \frac{\kappa}{\tilde{\kappa}}\hat{\theta}.$$

The risk-neutral drift and variance are still affine functions of the short rate, so we can apply the same pricing formulas as in the original Vasicek model except that we have to replace κ by $\tilde{\kappa}$ and $\hat{\theta}$ by $\tilde{\theta}$.

The generalization of the Vasicek model is an example of the so-called **essentially affine** models introduced by Duffee (2002). In contrast, Duffee refers to models where the market price of risk is restricted to be proportional to the short rate volatility—as the original Vasicek model—as **completely affine** models. In the completely affine models, the square of the market price of risk is affine in the short rate, while this is not true in the essentially affine models. The essentially affine Vasicek model involves an extra parameter in the market price of risk, which opens up for more flexibility in the link between the risk-neutral dynamics (determining the cross-section of bond prices and thus the shape of the yield curve) and real-world dynamics (determining the time-series behaviour of rates and yields).

In the CIR model of Section 7.5, the market price of risk can also be generalized without leaving the affine framework. If the market price of risk is

$$\lambda(r,t) = \frac{\lambda_1}{\beta\sqrt{r_t}} + \frac{\lambda_2\sqrt{r_t}}{\beta}$$

and the real-world dynamics of the short rate is still given by (7.44), then the risk-neutral drift of the short rate becomes

$$\begin{aligned}\kappa[\theta - r] - \beta\sqrt{r}\left(\frac{\lambda_1}{\beta\sqrt{r}} + \frac{\lambda_2\sqrt{r}}{\beta}\right) &= \kappa[\theta - r] - \lambda_1 - \lambda_2 r \\ &= (\kappa\theta - \lambda_1) - (\kappa + \lambda_2)r = \hat{\kappa}\left[\tilde{\theta} - r\right],\end{aligned}$$

where $\hat{\kappa} = \kappa + \lambda_2$ and $\tilde{\theta} = (\kappa\theta - \lambda_1)/\hat{\kappa}$. Compared to the original CIR model, $\hat{\theta}$ is replaced by $\tilde{\theta}$. This type of generalization was introduced by Cheridito et al. (2007) who call it an **extended affine** model. As the short rate approaches zero, it is clear that the extended market price of risk will diverge to infinity. To maintain the key link between the absence of arbitrage and the existence of a risk-neutral probability measure, Cheridito et al. (2007) show that it is sufficient to require that zero is an unattainable boundary under both the real-world and the risk-neutral measure. This is satisfied if $2\kappa\theta \geq \beta^2$ and $2(\kappa\theta - \lambda_1) \geq \beta^2$, compare the remarks in Section 3.8.3. This extension of the CIR model is thus only possible for some parameters. Again the extra parameter added enhances the flexibility of the model and makes it easier to fit both cross-sectional and time-series bond-price data.

Duarte (2004) introduces another extension of the CIR model. The risk-neutral dynamics is still required to be of an affine form

$$dr_t = \kappa^{\mathbb{Q}}\left(\theta^{\mathbb{Q}} - r_t\right) dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}},$$

so that the usual pricing formulas apply. However, Duarte allows the market price of risk to have the more general form

$$\lambda(r, t) = \frac{\lambda_0}{\beta} + \frac{\lambda_1}{\beta\sqrt{r_t}} + \frac{\lambda_2\sqrt{r_t}}{\beta},$$

where the term with λ_0 is new relative to the extended affine version. The real-world dynamics is therefore

$$dr_t = \left([\kappa^{\mathbb{Q}}\theta^{\mathbb{Q}} + \lambda_1] + \lambda_0\sqrt{r_t} - [\kappa^{\mathbb{Q}} - \lambda_2]r_t\right) dt + \beta\sqrt{r_t}\, dz_t.$$

Note that the real-world drift is no longer affine in the short rate. For this reason, Duarte calls it a **semi-affine square-root** model. For certain parameter values, the market price of risk can now change sign depending on the current level of the short rate. This is a valuable flexibility when matching actual bond-price dynamics.

7.7 NON-AFFINE MODELS

The financial literature contains many other one-factor models than those that fit into the affine framework studied in the previous sections. In this section we will go through the non-affine models that have attracted most attention, which are models where the future values of the short rate are lognormally distributed.

An apparently popular model among practitioners is the model suggested by Black and Karasinski (1991) and, in particular, the special case considered by Black et al. (1990), the so-called BDT model. The general time-homogeneous version of the Black–Karasinski model is

$$d(\ln r_t) = \kappa[\theta - \ln r_t]\,dt + \beta\,dz_t^{\mathbb{Q}}, \tag{7.54}$$

where κ, θ, and β are constants. Typically, practitioners replace the parameters κ, θ, and β by deterministic functions of time which are chosen to ensure that the model prices of bonds and caps are consistent with current market prices. We will discuss this idea in Chapter 9 and stick to the model with constant parameters in this section. Relative to the Vasicek model, r_t is replaced by $\ln r_t$ in the stochastic differential equation. Since r_T (given r_t) is normally distributed in the Vasicek model, it follows that $\ln r_T$ (given r_t) is normally distributed in the Black–Karasinski model, that is r_T is lognormally distributed. A pleasant consequence of this property is that the interest rate stays positive. Also this model exhibits a form of mean reversion. Assume that $\kappa > 0$. If $r_t < e^{\theta}$, the drift rate of $\ln r_t$ is positive so that r_t is expected to increase. Conversely if $r_t > e^{\theta}$. The parameter κ measures the speed at which $\ln r_t$ is drawn towards θ. An application of Itô's Lemma leads to

$$dr_t = \left(\left[\kappa\theta + \frac{1}{2}\beta^2\right] r_t - \kappa r_t \ln r_t\right) dt + \beta r_t\,dz_t^{\mathbb{Q}}.$$

There are no closed-form pricing expressions for bonds nor forwards, futures, nor options within this framework. Black and Karasinski implement their model in a binomial tree in which prices can be computed by the well-known backward iteration procedure.

In the Black–Karasinski model (7.54), the future short rate is lognormally distributed. Another model with this property is the one where the short rate follows a geometric Brownian motion

$$dr_t = r_t\left[\alpha\,dt + \beta\,dz_t^{\mathbb{Q}}\right], \tag{7.55}$$

where α and β are constants. Such a model was applied by Rendleman and Bartter (1980). However, as the Black–Karasinski model, this lognormal model does not allow simple closed-form expressions for the prices we are interested in. Dothan (1978) and Hogan and Weintraub (1993) state some very complicated pricing formulas for zero-coupon bonds which involve complex numbers, Bessel functions, and hyperbolic trigonometric functions! A seemingly fast and accurate recursive procedure for the computation of bond prices in the model (7.55) is described by Hansen and Jørgensen (2000).

In addition to the lack of nice pricing formulas, the lognormal models (7.54) and (7.55) have another very inappropriate property. As shown by Hogan and Weintraub (1993) these models imply that, for all t, T, S with $t < T < S$,

$$\mathrm{E}_{r,t}^{\mathbb{Q}}\left[\left(B^S(r_T, T)\right)^{-1}\right] = \infty. \tag{7.56}$$

As noted by Sandmann and Sondermann (1997), this result has two inexpedient consequences stated in the following theorem.

Theorem 7.6 *In the lognormal one-factor models* (7.54) *and* (7.55) *the following holds:*

(a) *An investment in the bank account over any period of time of strictly positive length is expected to give an infinite return, that is for $t \leq T < S$*

$$\mathrm{E}^{\mathbb{Q}}_{r,t}\left[\exp\left\{\int_T^S r_u\,du\right\}\right] = \infty.$$

(b) *The quoted Eurodollar futures price is $\tilde{\mathcal{E}}^T(r,t) = -\infty$.*

Proof: The first part of the theorem follows by Jensen's inequality, which implies that[10]

$$\begin{aligned} B^S(r,T) &= \mathrm{E}^{\mathbb{Q}}_{r,T}\left[\exp\left\{-\int_T^S r_u\,du\right\}\right] \\ &= \mathrm{E}^{\mathbb{Q}}_{r,T}\left[\left(\exp\left\{\int_T^S r_u\,du\right\}\right)^{-1}\right] > \left(\mathrm{E}^{\mathbb{Q}}_{r,T}\left[\exp\left\{\int_T^S r_u\,du\right\}\right]\right)^{-1} \end{aligned}$$

and hence

$$B^S(r,T)^{-1} < \mathrm{E}^{\mathbb{Q}}_{r,T}\left[\exp\left\{\int_T^S r_u\,du\right\}\right].$$

Taking expectations $\mathrm{E}^{\mathbb{Q}}_{r,t}[\,\cdot\,]$ on both sides, we get[11]

$$\mathrm{E}^{\mathbb{Q}}_{r,t}\left[\exp\left\{\int_T^S r_u\,du\right\}\right] > \mathrm{E}^{\mathbb{Q}}_{r,t}\left[(B^S(r_T,T))^{-1}\right] = \infty,$$

where the equality comes from (7.56).

From (6.11) we have that the quoted Eurodollar futures price is

$$\tilde{\mathcal{E}}^T(r,T) = 500 - 400\,\mathrm{E}^{\mathbb{Q}}_{r,t}\left[(B^{T+0.25}(r_T,T))^{-1}\right].$$

Inserting (7.56) with $S = T + 0.25$ into the expression above, we obtain the second part of the theorem. □

Since Eurodollar futures is a highly liquid product on the international financial markets, it is very inappropriate to use a model which clearly misprices these contracts.

It can be shown that the problematic relation (7.56) is avoided by assuming that either the effective annual interest rates or the LIBOR interest rates are lognormally distributed instead of the continuously compounded interest rates. Models with lognormal LIBOR rates have become very popular in recent years, primarily because they are (at least to some extent) consistent with practitioners' use of Black's pricing formula. We will study such models closely in Chapter 11.

[10] Jensen's inequality says that if X is a random variable and $f(x)$ is a convex function, then $\mathrm{E}[f(X)] > f(\mathrm{E}[X])$. In our application, $X = \exp\left\{\int_T^S r_u\,du\right\}$ and $f(X) = 1/X$.

[11] Here we apply the Law of Iterated Expectations, Theorem 3.1.

In summary, the lognormal models (7.54) and (7.55) have the nice property that negative rates are precluded, but they do not allow simple pricing formulas and they clearly misprice an important class of assets. For these reasons it is difficult to see why they have gained such popularity. The CIR model, for example, also precludes negative interest rates, is analytically tractable, and does not lead to obvious mispricing of any contracts. Furthermore, the model is consistent with a general equilibrium of the economy. Of course, these arguments do not imply that the CIR model provides the best description of the movements of interest rates over time, compare the discussion in the next section.

Finally, let us mention some models that are neither affine nor lognormal. The model

$$dr_t = \kappa \left[\theta - r_t\right] \, dt + \beta r_t \, dz_t^{\mathbb{Q}} \tag{7.57}$$

was suggested by Brennan and Schwartz (1980) and Courtadon (1982). Despite the relatively simple dynamics, no explicit pricing formulas have been derived for bonds nor derivative assets.

Longstaff (1989), Beaglehole and Tenney (1991, 1992), and Leippold and Wu (2002) consider so-called quadratic models where the short rate is given as $r_t = x_t^2$, and x_t follows an Ornstein–Uhlenbeck process (like the r-process in Vasicek's model). This specification ensures non-negative interest rates. The price of a zero-coupon bond is of the form

$$B^T(x,t) = e^{-a(T-t)-b(T-t)x-c(T-t)x^2},$$

where the functions a, b, and c solve ordinary differential equations. Relative to the affine models, a quadratic term has been added. The quadratic models thus give a more flexible relation between zero-coupon bond prices and the short rate. Leippold and Wu (2002) and Jamshidian (1996) obtain some rather complex expressions for the prices of European bond options and other derivatives.

7.8 PARAMETER ESTIMATION AND EMPIRICAL TESTS

To implement a model, one must assume some values of the parameters. In practice the true values of the parameters are unknown, but values can be estimated from observed interest rates and prices. For concreteness we will take the estimation of the Vasicek model as an example, but similar considerations apply to other models. The parameters of Vasicek's model are κ, θ, β, and λ. The estimation methods can be divided into three classes: time-series estimation, cross-section estimation, and panel-data estimation. Below we give a short introduction to these methods and mention some important studies. More details on the estimation and test of dynamic term structure models can be found in textbooks such as Campbell et al. (1997) and James and Webber (2000).

Time-series estimation: With this approach the parameters of the process for the short rate are estimated from a time series of historical observations of a short-term interest rate. The estimation itself can be carried out by means of different statistical methods, for example maximum likelihood (as in Marsh and Rosenfeld (1983) and Ogden (1987)) or various moment matching methods (as in Andersen and Lund (1997), Chan et al. (1992), and Dell' Aquila et al. (2003)). An essential practical problem is that no interest rates of zero maturity are observable so that some proxy must be applied. Interest rates of very short maturities are set at the money market, but due to, for example, the credit risk of the parties involved these rates are not perfect substitutes for the truly risk-free interest rate of zero maturity, which is represented by r_t in the models. Most authors use yields on government bonds with short maturities, such as 1 or 3 months, as an approximation to the short rate. However, Knez et al. (1994) and Duffee (1996) argue that special features of the trading in 1-month U.S. Treasury bills affect the yields on these bonds making them a questionable proxy for the short rate in our term structure models. Chapman et al. (1999) and Honoré (1998) study the sensitivity of model estimation results to various proxies for the short rate.

Another problem in applying the time-series approach is that not all parameters of the model can be identified. In Vasicek's model only the parameters κ, θ, and β that enter the real-world dynamics of the short rate in (7.32) can be estimated. The missing parameter λ only affects the process for r_t under the risk adjusted martingale measures, but the time series of interest rates is of course observed in the real world, that is under the real-world probability measure.

A third problem is that a large number of observations are required to give reasonably certain parameter estimates. However, the longer the observation period is, the less likely it is that the short rate has followed the same process (with constant parameters) during the entire period.

Furthermore, the time-series approach ignores the fact that the interest rate models not only describe the dynamics of the short rate but also describe the entire yield curve and its dynamics.

Cross-section estimation: Alternatively, the parameters of the model can be estimated as the values that will lead to model prices of a cross-section of liquid bonds (and possibly other fixed income securities) that are as close as possible to the current market prices of these assets. Then the estimated model can be applied to price less liquid assets in a way which is consistent with the market prices of the liquid assets. Typically, the parameter values are chosen to minimize the sum of squared deviations of model prices from market prices where the sum runs over all the assets in the chosen cross-section. Such a procedure is simple to implement.

A cross-section estimation cannot identify all parameters of the model either. The current prices only depend on the parameters that affect the short rate dynamics under the risk adjusted martingale measures. For Vasicek's model this is the case for $\hat{\theta}$, β, and κ. The parameters θ and λ cannot be estimated separately. However, if the only use of the model is to derive current prices of other assets, we only need the values of $\hat{\theta}$, β, and κ. In view of the problems connected with observing the short rate, the value of the short rate is often estimated in line with the parameters of the model.

A cross-section estimation completely ignores the time-series dimension of the data. The estimation procedure does not in any way ensure that the parameter values estimated at different dates are of similar magnitudes.[12] The model's results concerning the dynamics of interest rates and asset prices are not used at all in the estimation.

Panel-data estimation: This estimation approach combines the two approaches described above by using both the time-series and the cross-section dimension of the data and the models. Typically, the data used are time series of selected yields of different maturities. With a panel-data approach all the parameters can be estimated. For example, Gibbons and Ramaswamy (1993) and Daves and Ehrhardt (1993) apply this procedure to estimate the CIR model. If we want to apply a model both for pricing certain assets and for assessing and managing the changes in interest rates and prices over time, we should also base our estimation on both cross-sectional and time-series information.

Two relatively simple versions of the panel-data approach are obtained by emphasizing either the time-series dimension or the cross-section dimension and only applying the other dimension to get all the parameters identified. As discussed above, the parameters κ, θ, and β of Vasicek's model can be estimated from a time series of observations of (approximations of) the short rate. The remaining parameter λ can then be estimated as the value that leads to model prices (using the already fixed estimates of κ, θ, and β) that are as close as possible to the current prices on selected, liquid assets. On the other hand, one can estimate κ, $\hat{\theta}$, and β from a cross-section and then estimate θ from a time series of interest rates (using the already fixed estimates of κ and β). In this way an estimate of λ can be determined such that the relation $\hat{\theta} = \theta - \lambda\beta/\kappa$ holds for the estimated parameter values.

In any estimation the parameter values are chosen such that the model fits the data to the best possible extent according to some specified criterion. Typically, an estimation procedure will also generate information on how well the model fits the data. Therefore, most papers referred to above also contain a test of one or several models.

Probably the most frequently cited reference on the estimation, comparison, and test of one-factor diffusion models of the term structure is Chan et al. (1992) (henceforth abbreviated CKLS), who consider time-homogeneous models of the type

$$dr_t = (\theta - \kappa r_t)\, dt + \beta r_t^{\gamma}\, dz_t. \tag{7.58}$$

By restricting the values of the parameters θ, κ, and γ, many of the models studied earlier are obtained as special cases, for example, the models of Merton ($\kappa = \gamma = 0$), Vasicek ($\gamma = 0$), CIR ($\gamma = 1/2$), and the lognormal model (7.55) ($\gamma = 1, \theta = 0$). CKLS use the one-month yield on government bonds as an approximation to the short-term interest rate and apply a time-series approach on U.S. data over the period 1964–89. They estimate eight different restricted models and

[12] For example, Brown and Dybvig (1986) find that the parameter estimates of the CIR model fluctuate considerably over time.

the unrestricted model and test how well they perform in describing the evolution in the short rate over the given period. Their results indicate that it is primarily the value that the model assigns to the parameter γ which determines whether the model is rejected or accepted. The unrestricted estimate of γ is approximately 1.5, and models having a much lower γ-value are rejected in their test, including the Vasicek model and the CIR model. On the other hand, the lognormal model (7.55) is accepted.

Subsequently the CKLS analysis has been criticized on several counts. Firstly, as mentioned above, the one month yield may be a poor approximation to the zero-maturity short rate. This critique can be met in a one-factor model by using the one-to-one relation between the zero-coupon rate of any given maturity and the true short rate. For the affine models this relation is given by (7.11) where the functions a and b are known in closed-form for some affine models (Merton, Vasicek, and CIR), and for the other affine models they can be computed quickly and accurately by solving the Ricatti differential equations (7.6) and (7.7) numerically. For non-affine models the relation can be found by numerically solving the PDE (7.2) for a zero-coupon bond with the given time to maturity (one month in the CKLS case) and transforming the price to a zero-coupon yield. In this way Honoré (1998) transforms a time series of zero-coupon rates with a given maturity to a time series of implicit zero-maturity short rates. Based on the transformed time series of short rates he finds estimates of the parameter γ in the interval between 0.8 and 1.0, which is much lower than the CKLS-estimate.

Another criticism, advanced by Bliss and Smith (1997), is that the data set used by CKLS includes the period between October 1979 and September 1982, when the Federal Reserve, that is the U.S. central bank, followed a highly unusual monetary policy ('the Fed Experiment') resulting in a non-representative dynamics in interest rates, in particular the short rates. Hence, Bliss and Smith allow the parameters to have different values in this sub-period than in the rest of the period used by CKLS (1964–89). Outside the experimental period the unrestricted estimate of γ is 1.0, which is again considerably smaller than the CKLS estimate. The only models that are not rejected on a 5% test level are the CIR model and the Brennan–Schwartz model (7.57).

Finally, applying a different estimation method and a different data set (weekly observations of three month U.S. government bond yields over the period 1954–95), Andersen and Lund (1997) estimate γ to 0.676, which is much lower than the estimate of CKLS. Christensen et al. (2001) discuss some general problems in estimating a process like (7.58), and using a maximum likelihood estimation procedure and 1982–95 data they obtain a γ-estimate of 0.78.

The tests mentioned above are based on a time series of (approximations of) the short rate. Similar tests of the CIR model using other time series are performed by Brown and Dybvig (1986) and Brown and Schaefer (1994). On the other hand, Gibbons and Ramaswamy (1993) test the ability of the CIR model to simultaneously describe the evolution in four zero-coupon rates, namely the 1, 3, 6, and 12 month rates (a panel-data test). With data covering the same period as the CKLS study, they accept the CIR model.

By now it should be clear that the extensive empirical literature cannot give a clear-cut answer to the question of which one-factor model fits the data best. The answer depends on the data and the estimation technique applied. In most tests models with constant interest rate volatility, such as the models of Merton and Vasicek, and typically also all models without mean reversion are rejected. The CIR model is accepted in most tests, and since it both has nice theoretical properties and allows relatively simple closed-form pricing formulas, it is widely used both by academics and practitioners.

7.9 CONCLUDING REMARKS

In this chapter we have studied time-homogeneous one-factor diffusion models of the term structure of interest rates. They are all based on specific assumptions on the evolution of the short rate and on the market price of interest rate risk. The models of Vasicek and Cox et al. are frequently applied both by practitioners for pricing and risk management and by academics for studying the effects of interest rate uncertainty on various financial issues. Both models are consistent with a general economic equilibrium model, although this equilibrium is based on many simplifying and unrealistic assumptions on the economy and its agents. Both models are analytically tractable and generate relatively simple pricing formulas for many fixed income securities. The CIR model has the economically most appealing properties and performs better than the Vasicek model in explaining the empirical bond market data.

The assumption of the models of this chapter, that the short rate contains all relevant information about the yield curve, is very restrictive and not empirically acceptable. Several empirical studies show that at least two and possibly three or four state variables are needed to explain the observed variations in yield curves. As we shall see in the next chapter, many of the multi-factor models suggested in the literature are generalizations of the one-factor models of Vasicek and CIR.

In all time-homogeneous one-factor models the current yield curve is determined by the current short rate and the relevant model parameters. No matter how the parameter values are chosen it is highly unlikely that the yield curve derived from the model can be completely aligned with the yield curve observed in the market. If the model is to be applied for the pricing of derivatives such as futures and options on bonds and caps, floors, and swaptions, it is somewhat disturbing that the model cannot price the underlying zero-coupon bonds correctly. As we will see in Chapter 9, a perfect model fit of the current yield curve can be obtained in a one-factor model by replacing one or more parameters by deterministic functions of time. While these time-inhomogeneous versions of the one-factor models may provide a better basis for derivative pricing, they are not however, unproblematic. Also note that typically the current yield curve is not directly observable in the market, but has to be estimated from prices of coupon bonds. For this purpose practitioners often use a cubic spline or a Nelson–Siegel parametrization as outlined in Chapter 2. If one instead applies the parametrization of the discount function $T \mapsto B^T(r, t)$ that comes out of an economically better founded model, such as the CIR model, the problems of time-inhomogeneous models can be avoided.

7.10 EXERCISES

Exercise 7.1 (Slope of the yield curve at zero maturity) Consider a time-homogeneous affine model in which the yield curve at time t is given by

$$\bar{y}_t^\tau = \frac{a(\tau)}{\tau} + \frac{b(\tau)}{\tau} r,$$

compare (7.12). The slope of the yield curve at zero maturity is the limit $\lim_{\tau \to 0} \frac{\partial \bar{y}_t^\tau}{\partial \tau}$.

(a) Show by differentiation and an application of l'Hôpital's rule that the slope is

$$\lim_{\tau \to 0} \frac{\partial \bar{y}_t^\tau}{\partial \tau} = \frac{1}{2} a''(0) + \frac{1}{2} b''(0) r.$$

(b) Differentiate the ordinary differential equations (7.6) and (7.7) to find expressions for $b''(\tau)$ and $a''(\tau)$.
(c) Conclude that the slope is given by $(\hat{\varphi} - \hat{\kappa} r)/2$.

Exercise 7.2 (Parallel shifts of the yield curve) The purpose of this exercise is to find out under which assumptions the only possible shifts of the yield curve are parallel, that is such that $d\bar{y}_t^\tau$ is independent of τ where $\bar{y}_t^\tau = y_t^{t+\tau}$.

(a) Argue that if the yield curve only changes in the form of parallel shifts, then the zero-coupon yields at time t must have the form

$$y_t^T = y^T(r_t, t) = r_t + h(T - t)$$

for some function h with $h(0) = 0$ and that the prices of zero-coupon bonds are thereby given as

$$B^T(r, t) = e^{-r[T-t] - h(T-t)[T-t]}.$$

(b) Use the partial differential equation (7.2) to show that

$$(^*) \frac{1}{2} \beta(r)^2 (T-t)^2 - \hat{\alpha}(r)(T-t) + h'(T-t)(T-t) + h(T-t) = 0$$

for all (r, t) (with $t \leq T$, of course).
(c) Using (*), show that $\frac{1}{2}\beta(r)^2(T-t)^2 - \hat{\alpha}(r)(T-t)$ must be independent of r. Conclude that both $\hat{\alpha}(r)$ and $\beta(r)$ have to be constants, so that the model is indeed Merton's model.
(d) Describe the possible shapes of the yield curve in an arbitrage-free model in which the yield curve only moves in terms of parallel shifts. Is it possible for the yield curve to be flat in such a model?

Exercise 7.3 (Call on zero-coupon bonds in Vasicek's model) Figure 7.9 shows an example of the relation between the price of a European call on a zero-coupon bond and the current short rate r in the Vasicek model, compare (7.43). The purpose of this exercise is to derive an explicit expression for $\partial C / \partial r$.

(a) Show that

$$B^S(r,t)e^{-\frac{1}{2}d_1(r,t)^2} = KB^T(r,t)e^{-\frac{1}{2}d_2(r,t)^2}.$$

(b) Show that

$$B^S(r,t)n\left(d_1(r,t)\right) - KB^T(r,t)n\left(d_2(r,t)\right) = 0,$$

where $n(y) = \exp(-y^2/2)/\sqrt{2\pi}$ is the probability density function for a standard normally distributed random variable.

(c) Show that

$$\frac{\partial C^{K,T,S}}{\partial r}(r,t) = -B^S(r,t)b(S-t)N\left(d_1(r,t)\right) + KB^T(r,t)b(T-t)N\left(d_2(r,t)\right).$$

Exercise 7.4 (Futures on bonds) Show the last claim in Theorem 7.2.

Exercise 7.5 (CIR zero-coupon bond price) Show that the functions b and a given by (7.49) and (7.50) solve the ordinary differential equations (7.47) and (7.48).

Exercise 7.6 (Comparison of prices in the models of Vasicek and CIR) Compare the prices according to Vasicek's model (7.32) and the CIR-model (7.44) of the following securities:

(a) 1-year and 10-year zero-coupon bonds;
(b) 3-month European call options on a 5-year zero-coupon bond with exercise prices of 0.7, 0.75, and 0.8, respectively;
(c) a 10-year 8% bullet bond with annual payments;
(d) 3-month European call options on a 10-year 8% bullet bond with annual payments for three different exercise prices chosen to represent an in-the-money option, a near-the-money option, and an out-of-the-money option.

In the comparisons use $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = 0$ for both models. The current short rate is $r = 0.05$, and the current volatility on the short rate is 0.03 so that $\beta = 0.03$ in Vasicek's model and $\beta\sqrt{0.05} = 0.03$ in the CIR model.

Exercise 7.7 (Expectation hypothesis in Vasicek) Verify that the local weak and the weak yield-to-maturity versions of the expectation hypothesis hold in the Vasicek model.

8

Multi-Factor Diffusion Models

8.1 INTRODUCTION

The preceding chapter gave an overview of one-factor diffusion models of the term structure of interest rates. All those models are based on an assumed dynamics in the continuously compounded short rate, r. In several of these models we were able to derive relatively simple, explicit pricing formulas for both bonds and European options on bonds and hence also for caps, floors, swaps, and European swaptions, see Chapter 6. The models can generate yield curves of various realistic forms, and the parameters of the models can be estimated quite easily from market data. Several of the empirical tests described in the literature have failed to reject selected one-factor models. Furthermore, the CIR model particularly is theoretically well-founded and based on short rate dynamics with many realistic properties.

However, all the one-factor models also have obviously unrealistic properties. First, they are not able to generate all the yield curve shapes observed in practice. For example, the Vasicek and CIR models can only produce an increasing curve, a decreasing curve, and a curve with a small hump. While the zero-coupon yield curve typically has one of these shapes, it does occasionally have a different shape, for example the yield curve is sometimes decreasing for short maturities and then increasing for longer maturities. Moreover, even when the market yield curve is, say, increasing in maturity, it might be a different increasing relation than that which a one-factor model can generate.

Second, the one-factor models are not able to generate all the types of yield curve *changes* that have been observed. In the affine one-factor models the zero-coupon yield $\bar{y}_t^\tau = y_t^{t+\tau}$ for any maturity τ is of the form

$$\bar{y}_t^\tau = \frac{a(\tau)}{\tau} + \frac{b(\tau)}{\tau} r_t,$$

compare (7.11). If $b(\tau) > 0$, the change in the yield of any maturity will have the same sign as the change in the short rate. Therefore, these models do not allow so-called *twists* of the term structure of interest rates, that is yield curve changes where short-maturity yields and long-maturity yields move in opposite directions.

A third critical point, which is related to the second point above, is that the changes over infinitesimal time periods of any two interest rate dependent variables will be perfectly correlated. This is, for example, the case for any two bond prices or any two yields. This is due to the fact that all unexpected changes are proportional to the shock to the short rate, dz_t. For example, the dynamics of the τ_i-maturity zero-coupon yield in any time-homogeneous one-factor model is of the form

$$d\bar{y}_t^{\tau_i} = \mu_y(r_t, \tau_i)\, dt + \sigma_y(r_t, \tau_i)\, dz_t,$$

where the drift rate μ_y and the volatility σ_y are model-specific functions. The variance of the change in the yield over an infinitesimal time period is therefore

$$\mathrm{Var}_t\left[d\bar{y}_t^{\tau_i}\right] = \sigma_y(r_t, \tau_i)^2\, dt.$$

The covariance between changes in two different zero-coupon yields is

$$\mathrm{Cov}_t\left[d\bar{y}_t^{\tau_1}, d\bar{y}_t^{\tau_2}\right] = \sigma_y(r_t, \tau_1)\sigma_y(r_t, \tau_2)\, dt.$$

Hence, the correlation between the yield changes is

$$\mathrm{Corr}_t\left[d\bar{y}_t^{\tau_1}, d\bar{y}_t^{\tau_2}\right] = \frac{\mathrm{Cov}_t\left[d\bar{y}_t^{\tau_1}, d\bar{y}_t^{\tau_2}\right]}{\sqrt{\mathrm{Var}_t\left[d\bar{y}_t^{\tau_1}\right]}\sqrt{\mathrm{Var}_t\left[d\bar{y}_t^{\tau_2}\right]}} = 1.$$

This conflicts with empirical studies which demonstrate that the actual correlation between changes in zero-coupon yields of different maturities is far from one. Table 8.1 shows correlations between weekly changes in par yields on U.S. government bonds. The correlations are estimated by Canabarro (1995) from data over the period from January 1986 to December 1991. A similar pattern has been documented by other authors, such as Rebonato (1996, Ch. 2) who uses data from the U.K. bond market.

It is thus clear that the one-factor diffusion models may very well be too simple to provide a reasonable fit of both the cross-section and time-series dynamics of bond prices. Intuitively, multi-factor models are more flexible and should be able to generate additional yield curve shapes and yield curve movements relative to the one-factor models. Furthermore, multi-factor models allow non-perfect correlations between different interest rate dependent variables, see the discussion in Section 8.4 below.

Table 8.1: Estimated correlation matrix of weekly changes in par yields on U.S. government bonds. The matrix is extracted from Exhibit 1 in Canabarro (1995).

Maturity (years)	Maturity (years)							
	0.25	0.5	1	2	5	10	20	30
0.25	1.00	0.85	0.80	0.72	0.61	0.52	0.46	0.46
0.50	0.85	1.00	0.90	0.85	0.76	0.68	0.63	0.62
1	0.80	0.90	1.00	0.94	0.87	0.79	0.73	0.73
2	0.72	0.85	0.94	1.00	0.95	0.88	0.82	0.82
5	0.61	0.76	0.87	0.95	1.00	0.96	0.92	0.91
10	0.52	0.68	0.79	0.88	0.96	1.00	0.97	0.96
20	0.46	0.63	0.73	0.82	0.92	0.97	1.00	0.97
30	0.46	0.62	0.73	0.82	0.91	0.96	0.97	1.00

Several empirical studies have investigated how many factors are necessary in order to obtain a sufficiently precise description of the actual evolution of the term structure of interest rates and the correlations between yields of different maturities. Of course, to some extent the result of such an investigation will depend on the chosen data set, the observation period, and the estimation procedure. However, all studies seem to indicate that two or more factors are needed. One way to address this question is to perform a so-called principal component analysis of the variance-covariance matrix of changes in zero-coupon yields of selected maturities. Canabarro (1995) finds that a single factor can describe at most 85.0% of the total variation in his data from the period 1986–91 on the U.S. bond market. The second-most important factor describes an additional 10.3% of the variation, while the third-most and the fourth-most important factors provide additional contributions of 1.9% and 1.2%, respectively. Additional factors contribute in total less than 1.6%. Similar results are reported by Litterman and Scheinkman (1991), who also use U.S. bond market data, and by Rebonato (1996, Ch. 2), who applies U.K. data over the period 1989–92.

A principal component analysis does not provide a precise identification of *which* factors best describe the evolution of the term structure, but it can give some indication of the factors. The studies mentioned above give remarkably similar indications. They all find that the most important factor affects yields of all maturities similarly and hence can be interpreted as a **level factor**. The second-most important factor affects short-maturity yields and long-maturity yields in opposite directions and can therefore be interpreted as a **slope factor**. Finally, the third-most important factor affects yields of very short and long maturities in the same direction, but yields of intermediate maturities (approximately 2–5 years) in the opposite direction. We can interpret this factor as a **curvature factor**. Litterman and Scheinkman argue that the third factor can alternatively be interpreted as a factor representing the term structure of yield volatilities, that is the volatilities of the zero-coupon yields of different maturities.

Other empirical papers have studied how well specific multi-factor models can fit selected bond market data. Empirical tests performed by Stambaugh (1988), Pearson and Sun (1991), Chen and Scott (1993), Brenner et al. (1996), Andersen and Lund (1997), Vetzal (1997), Balduzzi et al. (1998), Boudoukh et al. (1999), Dai and Singleton (2000), and Cheridito et al. (2007) all conclude that different multi-factor models provide a much better description of the shape and movements of the term structure of interest rates than the one-factor special cases of the models. While there seems to be some consensus in the recent literature that a three-factor model can provide a reasonable fit to the data, there is less agreement about exactly which three-factor model to use.

The chapter proceeds in the following way. Section 8.2 introduces the notation and main pricing tools in the general multi-factor setting, after which Section 8.3 specializes to the affine modelling class. Section 8.4 focuses on two-factor affine models and provides a detailed description of two-factor versions of the models of Vasicek and Cox, Ingersoll, and Ross. Specific three-factor models are presented in Section 8.5, whereas Section 8.6 explains and discusses how the market price of risk specification of the standard affine models can be generalized. Section 8.7 briefly introduces some multi-factor models not included in the other sections. Finally, Section 8.8 summarizes the chapter.

8.2 THE GENERAL MULTI-FACTOR SETTING

In this section we review the notation and the general results in multi-factor diffusion models, which were first discussed in Section 4.8. In a general n-factor diffusion model of the term structure of interest rates, the fundamental assumption is that the state of the economy can be represented by an n-dimensional vector process $\boldsymbol{x} = (x_1, \dots, x_n)^\top$ of state variables. In particular, the process $\boldsymbol{x}$ follows a Markov diffusion process,

$$d\boldsymbol{x}_t = \boldsymbol{\alpha}(\boldsymbol{x}_t, t)\, dt + \underline{\underline{\beta}}(\boldsymbol{x}_t, t)\, d\boldsymbol{z}_t, \tag{8.1}$$

where $\boldsymbol{z} = (z_1, \dots, z_n)^\top$ is an n-dimensional standard Brownian motion. Denote by $\mathcal{S} \subseteq \mathbb{R}^n$ the value space of the process, that is the set of possible states. In the expression (8.1), $\boldsymbol{\alpha}$ is a function from $\mathcal{S} \times \mathbb{R}_+$ into $\mathbb{R}^n$, and $\underline{\underline{\beta}}$ is a function from $\mathcal{S} \times \mathbb{R}_+$ into the set of $n \times n$ matrices of real numbers, that is $\underline{\underline{\beta}}(\boldsymbol{x}_t, t)$ is an $n \times n$ matrix. The functions $\boldsymbol{\alpha}$ and $\underline{\underline{\beta}}$ must satisfy certain regularity conditions to ensure that the stochastic differential equation (8.1) has a unique solution, compare Øksendal (2003). We can write (8.1) componentwise as

$$dx_{it} = \alpha_i(\boldsymbol{x}_t, t)\, dt + \sum_{j=1}^{n} \beta_{ij}(\boldsymbol{x}_t, t)\, dz_{jt} = \alpha_i(\boldsymbol{x}_t, t)\, dt + \boldsymbol{\beta}_i(\boldsymbol{x}_t, t)^\top\, d\boldsymbol{z}_t.$$

As discussed in Chapter 4, the absence of arbitrage will imply the existence of a vector process $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_n)^\top$ of market prices of risk, so that for any traded asset we have the relation

$$\mu(\boldsymbol{x}_t, t) = r(\boldsymbol{x}_t, t) + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\lambda_j(\boldsymbol{x}_t, t).$$

Here μ denotes the expected rate of return on the asset, and $\sigma_1, \dots, \sigma_n$ are the volatility terms, that is the price process is

$$dP_t = P_t \left[\mu(\boldsymbol{x}_t, t)\, dt + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\, dz_{jt} \right].$$

See, for example, Equation (4.31).

We also know that the n-dimensional process $\boldsymbol{z}^{\mathbb{Q}} = (z_1^{\mathbb{Q}}, \dots, z_n^{\mathbb{Q}})^\top$ defined by

$$dz_{jt}^{\mathbb{Q}} = dz_{jt} + \lambda_j(\boldsymbol{x}_t, t)\, dt, \quad j = 1, \dots, n,$$

is a standard Brownian motion under the risk-neutral probability measure $\mathbb{Q}$. With the notation

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{x}, t) = \boldsymbol{\alpha}(\boldsymbol{x}, t) - \underline{\underline{\beta}}(\boldsymbol{x}, t)\boldsymbol{\lambda}(\boldsymbol{x}, t),$$

that is

$$\hat{\alpha}_i(\boldsymbol{x}, t) = \alpha_i(\boldsymbol{x}, t) - \sum_{j=1}^{n} \beta_{ij}(\boldsymbol{x}, t)\lambda_j(\boldsymbol{x}, t),$$

we can write the dynamics of the state variables under $\mathbb{Q}$ as

$$d\boldsymbol{x}_t = \hat{\boldsymbol{\alpha}}(\boldsymbol{x}_t, t)\, dt + \underline{\underline{\beta}}(\boldsymbol{x}_t, t)\, d\boldsymbol{z}_t^{\mathbb{Q}}$$

or componentwise as

$$dx_{it} = \hat{\alpha}_i(\boldsymbol{x}_t, t)\, dt + \sum_{j=1}^{n} \beta_{ij}(\boldsymbol{x}_t, t)\, dz_{jt}^{\mathbb{Q}}.$$

From the analysis in Section 4.8, we know that the price $P_t = P(\boldsymbol{x}_t, t)$ of a traded asset of the European type can be found as the solution to the partial differential equation

$$\begin{aligned}
&\frac{\partial P}{\partial t}(\boldsymbol{x}, t) + \sum_{i=1}^{n} \hat{\alpha}_i(\boldsymbol{x}, t)\frac{\partial P}{\partial x_i}(\boldsymbol{x}, t) \\
&\quad + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij}(\boldsymbol{x}, t)\frac{\partial^2 P}{\partial x_i \partial x_j}(\boldsymbol{x}, t) - r(\boldsymbol{x}, t)P(\boldsymbol{x}, t) = 0, \quad (\boldsymbol{x}, t) \in \mathcal{S} \times [0, T),
\end{aligned} \tag{8.2}$$

with the terminal condition

$$P(\boldsymbol{x}, T) = H(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{S},$$

reflecting the payoff at maturity. Here, $\gamma_{ij} = \sum_{k=1}^{n} \beta_{ik}\beta_{jk}$ is the (i, j)'th element of the variance-covariance matrix $\underline{\underline{\beta}}\,\underline{\underline{\beta}}^{\mathsf{T}}$. If ρ_{ij} denotes the correlation between changes in the i'th and the j'th state variables, we have that $\gamma_{ij} = \rho_{ij}\, \|\boldsymbol{\beta}_i\|\, \|\boldsymbol{\beta}_j\|$. Alternatively, the price can be computed from an expectation under the risk-neutral (spot martingale) measure,

$$P(\boldsymbol{x}, t) = \mathrm{E}_{\boldsymbol{x},t}^{\mathbb{Q}}\left[e^{-\int_t^T r(\boldsymbol{x}_u, u)\, du} H(\boldsymbol{x}_T)\right],$$

or from an expectation under the T-forward martingale measure

$$P(\boldsymbol{x}, t) = B^T(\boldsymbol{x}, t)\, \mathrm{E}_{\boldsymbol{x},t}^{\mathbb{Q}^T}\left[H(\boldsymbol{x}_T)\right],$$

or as an expectation under another convenient martingale measure.

Just as in the analysis of one-factor models in Chapter 7, we focus on the time-homogeneous models in which the functions $\hat{\boldsymbol{\alpha}}$ and $\underline{\underline{\beta}}$, and also the short rate r, do not depend on time, but only depend on the state variables. In particular,

$$d\boldsymbol{x}_t = \hat{\boldsymbol{\alpha}}(\boldsymbol{x}_t)\, dt + \underline{\underline{\beta}}(\boldsymbol{x}_t)\, d\boldsymbol{z}_t^{\mathbb{Q}}.$$

8.3 AFFINE MULTI-FACTOR MODELS

We focus first on so-called affine models, which in a multi-factor setting were introduced by Duffie and Kan (1996) and further analysed by Dai and Singleton (2000). In the affine multi-factor models the risk-neutral dynamics of the vector of state variables is of the form

$$d\boldsymbol{x}_t = \left(\hat{\boldsymbol{\varphi}} - \underline{\underline{\hat{\kappa}}}\boldsymbol{x}_t\right) dt + \underline{\underline{\Gamma}}\sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\, d\boldsymbol{z}_t^{\mathbb{Q}}, \tag{8.3}$$

where $\underline{\underline{V}}(\boldsymbol{x}_t)$ is the diagonal $n \times n$ matrix

$$\underline{\underline{V}}(\boldsymbol{x}_t) = \begin{pmatrix} \nu_1 + \boldsymbol{v}_1^{\top}\boldsymbol{x}_t & 0 & \dots & 0 \\ 0 & \nu_2 + \boldsymbol{v}_2^{\top}\boldsymbol{x}_t & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \nu_n + \boldsymbol{v}_n^{\top}\boldsymbol{x}_t \end{pmatrix}$$

and $\sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}$ is the diagonal $n \times n$ matrix in which the entry (i, i) is given by $\sqrt{\nu_i + \boldsymbol{v}_i^{\top}\boldsymbol{x}_t}$. Moreover, $\hat{\boldsymbol{\varphi}} = (\hat{\varphi}_1, \dots, \hat{\varphi}_n)^{\top}$ is a constant vector and $\underline{\underline{\hat{\kappa}}}$ and $\underline{\underline{\Gamma}}$ are constant $n \times n$ matrices. In particular, we can write the dynamics of the i'th state variable as

$$dx_{it} = \left(\hat{\varphi}_i - \sum_{j=1}^{n} \hat{\kappa}_{ij} x_{jt}\right) dt + \sum_{j=1}^{n} \Gamma_{ij}\sqrt{\nu_j + \boldsymbol{v}_j^{\top}\boldsymbol{x}_t}\, dz_{jt}^{\mathbb{Q}}.$$

Obviously, we have to make sure that $\nu_j + \boldsymbol{v}_j^{\top}\boldsymbol{x}_t \geq 0$ for each $j = 1, \dots, n$ and for all possible values of $\boldsymbol{x}_t$. This is no problem if $\boldsymbol{v}_j = \boldsymbol{0}$ for all j, but in other cases we will have to impose some parameter restrictions as will be discussed below.

A time-homogeneous multi-factor diffusion model is said to be affine if the dynamics of the state variables under the risk-neutral probability measure is of the form (8.3) *and* the short rate $r_t = r(\boldsymbol{x}_t)$ is an affine function of $\boldsymbol{x}$, that is a constant scalar ξ_0 and a constant n-dimensional vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^{\top}$ exist such that

$$r(\boldsymbol{x}) = \xi_0 + \boldsymbol{\xi}^{\top}\boldsymbol{x} = \xi_0 + \sum_{i=1}^{n} \xi_i x_i. \tag{8.4}$$

The condition on the short rate is trivially satisfied in the one-factor models in Chapter 7 since they all take the short rate itself as the state variable. Similarly, the condition is satisfied in the multi-factor models in which the short rate is one of the state variables.

If the vector $\boldsymbol{\lambda}(\boldsymbol{x})$ of market prices of risk is of the form

$$\boldsymbol{\lambda}(\boldsymbol{x}) = \sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\bar{\boldsymbol{\lambda}} \tag{8.5}$$

for a constant vector $\bar{\boldsymbol{\lambda}}$, then the real-world dynamics of the state variable vector will be

$$dx_t = \left(\varphi - \underline{\underline{\kappa}} x_t\right) dt + \underline{\underline{\Gamma}} \sqrt{\underline{\underline{V}}(x_t)}\, dz_t, \tag{8.6}$$

where

$$\varphi = \hat{\varphi} + \underline{\underline{\Gamma}} \psi, \qquad \underline{\underline{\kappa}} = \underline{\underline{\hat{\kappa}}} - \underline{\underline{\Gamma}}\, \underline{\underline{\Psi}},$$

and here ψ is a vector with elements $\psi_i = \bar{\lambda}_i \nu_i$, whereas $\underline{\underline{\Psi}}$ is a matrix in which any row i is given by $\bar{\lambda}_i \boldsymbol{\nu}_i$ (you are asked to verify (8.6) in Exercise 8.1). Hence, the real-world dynamics is also in the affine class.

8.3.1 Pricing in multi-factor affine models

The zero-coupon bond prices in the multi-factor affine model are characterized in the following theorem.

Theorem 8.1 *In an admissible affine model, the price of a zero-coupon bond has the exponential-affine form*

$$\begin{aligned} B_t^T = B^T(x_t, t) &= \exp\left\{ -a(T-t) - \boldsymbol{b}(T-t)^\top x_t \right\} \\ &= \exp\left\{ -a(T-t) - \sum_{j=1}^n b_j(T-t) x_{jt} \right\} \end{aligned} \tag{8.7}$$

where the deterministic functions a and $\boldsymbol{b}$ satisfy the system of ordinary differential equations (ODEs)

$$\boldsymbol{b}'(\tau) = -\underline{\underline{\hat{\kappa}}}^\top \boldsymbol{b}(\tau) - \frac{1}{2} \sum_{i=1}^n \left[\underline{\underline{\Gamma}}^\top \boldsymbol{b}(\tau)\right]_i^2 \boldsymbol{\nu}_i + \boldsymbol{\xi}, \tag{8.8}$$

$$a'(\tau) = \hat{\varphi}^\top \boldsymbol{b}(\tau) - \frac{1}{2} \sum_{i=1}^n \left[\underline{\underline{\Gamma}}^\top \boldsymbol{b}(\tau)\right]_i^2 \nu_i + \xi_0$$

with the initial conditions $a(0) = 0$, $\boldsymbol{b}(0) = \mathbf{0}$. Here $\left[\underline{\underline{\Gamma}}^\top \boldsymbol{b}(\tau)\right]_i$ denotes element i of the vector $\underline{\underline{\Gamma}}^\top \boldsymbol{b}(\tau)$.

This can be seen by substitution into the partial differential equation (8.2) imposing the conditions of the affine model (you are asked to show this in Exercise 8.2). We can write the ODEs without matrix and vector notation as

$$b_i'(\tau) = -\sum_{j=1}^n \hat{\kappa}_{ji} b_j(\tau) - \frac{1}{2} \sum_{k=1}^n \nu_{ki} \left(\sum_{j=1}^n \Gamma_{jk} b_j(\tau) \right)^2 + \xi_i, \quad i = 1, \dots, n, \tag{8.9}$$

$$a'(\tau) = \sum_{j=1}^n \hat{\varphi}_j b_j(\tau) - \frac{1}{2} \sum_{k=1}^n \nu_k \left(\sum_{j=1}^n \Gamma_{jk} b_j(\tau) \right)^2 + \xi_0 \tag{8.10}$$

with the initial conditions $a(0) = b_1(0) = \cdots = b_n(0) = 0$. As in the one-factor models, $a(\tau)$ can be found by integration once all $b_j(\tau)$ have been determined.

There is also a converse to the above theorem. Under certain regularity conditions, the zero-coupon bond prices are only of the exponential-affine form if $\hat{\boldsymbol{\alpha}}$, $\underline{\underline{\beta}}\,\underline{\underline{\beta}}^{\mathsf{T}}$, and r are affine functions of $\boldsymbol{x}$, see Duffie and Kan (1996).

In an affine n-factor model the zero-coupon yields $\bar{y}_t^\tau = -(\ln B_t^{t+\tau})/\tau$ are

$$\bar{y}^\tau(\boldsymbol{x}) = \frac{a(\tau)}{\tau} + \sum_{j=1}^{n} \frac{b_j(\tau)}{\tau} x_j, \tag{8.11}$$

and the forward rates $\bar{f}_t^\tau = f_t^{t+\tau}$ are

$$\bar{f}^\tau(\boldsymbol{x}) = a'(\tau) + \sum_{j=1}^{n} b'(\tau) x_j.$$

The dynamics of the zero-coupon bond price B_t^T is

$$\frac{dB_t^T}{B_t^T} = r(\boldsymbol{x}_t)\,dt + \sum_{j=1}^{n} \sigma_j^T(\boldsymbol{x}_t, t)\,dz_{jt}^{\mathbb{Q}},$$

where the sensitivities σ_j^T are given by

$$\sigma_j^T(\boldsymbol{x}, t) = -\sqrt{\nu_j + \boldsymbol{\nu}_j^{\mathsf{T}}\boldsymbol{x}} \sum_{k=1}^{n} \Gamma_{kj} b_k(T-t). \tag{8.12}$$

In Chapter 6 we discussed general methods for pricing a European call option on a zero-coupon bond. One approach is based on the formula

$$C^{K,T,S}(\boldsymbol{x}, t) = B^T(\boldsymbol{x}, t)\,\mathrm{E}_{\boldsymbol{x},t}^{\mathbb{Q}^T}\left[\max\left(B^S(\boldsymbol{x}_T, T) - K, 0\right)\right].$$

In a Gaussian model, $\boldsymbol{x}_T$ is normally distributed, so it follows from (8.7) that the bond price at the maturity of the option is lognormally distributed and it is relatively easy to compute the above expectation and hence the price of the call option. This is similar to the Black–Scholes–Merton model for stock options and to the one-factor term structure models of Merton and Vasicek studied in Chapter 7. We shall apply this approach below in a two-factor version of the Vasicek model. Alternatively, we can compute the price as

$$C^{K,T,S}(\boldsymbol{x}_t, t) = B^S(\boldsymbol{x}_t, t)\mathbb{Q}_t^S(B^S(\boldsymbol{x}_T, T) > K) - KB^T(\boldsymbol{x}_t, t)\mathbb{Q}_t^T(B^S(\boldsymbol{x}_T, T) > K), \tag{8.13}$$

compare Equation (6.17). Using (8.7), we see that

$$B^S(\boldsymbol{x}_T, T) > K \quad \Leftrightarrow \quad \sum_{j=1}^{n} b_j(S-T)x_j < -a(S-T) - \ln K,$$

so to apply (8.13) we need to know the distribution of a linear combination of the state variables under the appropriate probability measures. It is also possible to derive closed-form pricing formulas for options on zero-coupon bonds in some non-Gaussian affine models, for example in the two-factor version of the Cox–Ingersoll–Ross model that we study in Section 8.4.2.

As demonstrated in Section 7.2.3, the price of a European call option on a coupon bond is in the one-factor affine models given as the price of a portfolio of European call options on zero-coupon bonds, compare (7.28). This is the case whenever the price of any zero-coupon bond is a monotonic function of the short rate. The same trick cannot be applied in multi-factor models so that the prices of options on coupon bonds (and consequently also swaptions, see Section 6.5.2) must be computed using numerical methods. However, it is possible to approximate very accurately the price of a European option on a coupon bond by the price of a single European option on a carefully selected zero-coupon bond. For details on the approximation, see Chapter 12 and Munk (1999). As we shall see below, several of the multi-factor models provide a closed-form expression for the price of a European option on a zero-coupon bond so that the approximate price of the coupon bond option is easily obtainable.

Other techniques for approximating prices of European options on coupon bonds have been suggested in the literature. For example, in the framework of affine models Collin-Dufresne and Goldstein (2002b) and Singleton and Umantsev (2002) introduce two approximations that may dominate (with respect to accuracy and computational speed) the approximation outlined above, but these approximations are much harder to understand. Another promising and relatively simple approach has been proposed by Schrager and Pelsser (2006).

8.3.2 Categorization of affine models

Dai and Singleton (2000) provide a very useful categorization of affine models. They categorize the class of n-factor affine models into $n+1$ subclasses denoted by $\mathbb{A}_0(n), \mathbb{A}_1(n), \dots, \mathbb{A}_n(n)$, where $\mathbb{A}_m(n)$ is the subclass of n-factor models in which m of the state variables affect $\underline{\underline{V}}(x)$ and thus the instantaneous variance-covariance matrix $\mathrm{Var}_t[d\boldsymbol{x}_t]$. Without loss of generality, it can be assumed that the m variance-influencing state variables are the first m, that is $x_1, \dots, x_m$. Dai and Singleton discuss conditions for the model being admissible in the sense that $\nu_i + \boldsymbol{\nu}_i^{\mathsf{T}} \boldsymbol{x}_t > 0$ for all possible values of $\boldsymbol{x}_t$. This involves various parameter restrictions. For example, any two variance-influencing state variables must have zero instantaneous correlation. For any variance-influencing state variables x_i, the drift rate must have $\varphi_i \geq 0$, $\kappa_{ij} = 0$ for $j = m+1, \dots, n$ and $\kappa_{ij} \leq 0$ for $j = 1, \dots, m$ with $j \neq i$, and x_i is only sensitive to dz_{it} with a volatility proportional to $\sqrt{x_{it}}$. These conditions will ensure that all the variance-influencing state variables stay non-negative. In loose terms, the variance-influencing variables are instantaneously uncorrelated square-root processes, where the drift of each variable may be positively related to the values of the other square-root processes and cannot depend on the non-variance-influencing state variables. These non-variance-influencing state variables can have drift rates involving all state variables. The sensitivity of a non-variance-influencing state variable x_i for $i = m+1, \dots, n$ towards any of the shocks dz_{jt} $(j = 1, \dots, m)$

that affect the variance-influencing variables is restricted to being proportional to $\sqrt{x_{jt}}$, whereas the sensitivity towards the remaining shocks dz_{jt} $(j = m+1, \dots, n)$ must be of the form $\Gamma_{ij}\sqrt{\nu_j + \sum_{k=1}^{m} \nu_{jk} x_{kt}}$ with $\nu_j, \nu_{jk} \geq 0$.

Dai and Singleton formulate a *canonical* representation of each $\mathbb{A}_m(n)$ model subclass, but the natural formulations of interesting multi-factor models typically do not fit into that formulation, so we will not pursue it further here. The specific affine multi-factor models studied in the literature can easily be assigned to a given model subclass $\mathbb{A}_m(n)$ by observing the total number of factors and the number of factors appearing in the volatility terms. An important contribution of the analysis of Dai and Singleton is to show that the various models studied in the literature are often unnecessarily restrictive in the sense that they can be generalized without leaving the subclass they belong to. The possible generalizations can improve the empirical fit of the model.

The simplest subclasses are the extremes, $\mathbb{A}_0(n)$ and $\mathbb{A}_n(n)$. Below we provide some general results for these two subclasses. After that, we will take a more detailed look at two- and three-factor affine models.

8.3.2.1 Multi-factor Gaussian models

For the $\mathbb{A}_0(n)$ subclass, there are no state variables affecting any volatility terms, that is in the general affine model (8.3) we have $\boldsymbol{\nu}_i = 0$ for all i. Without loss of generality we can assume that $\nu_i = 1$ for all i, so that the dynamics of the state variable is of the form

$$d\boldsymbol{x}_t = \left(\boldsymbol{\varphi} - \underline{\underline{\kappa}}\boldsymbol{x}_t\right) dt + \underline{\underline{\Gamma}}\, d\boldsymbol{z}_t.$$

Consequently, the state variable vector $\boldsymbol{x} = (\boldsymbol{x}_t)$ follows an n-dimensional Ornstein–Uhlenbeck process. If the market price of risk is constant in line with (8.5), the risk-neutral dynamics will be of the form

$$d\boldsymbol{x}_t = \left(\hat{\boldsymbol{\varphi}} - \underline{\underline{\kappa}}\boldsymbol{x}_t\right) dt + \underline{\underline{\Gamma}}\, d\boldsymbol{z}_t^{\mathbb{Q}}. \tag{8.14}$$

Future values of the state variables are normally distributed. Since the short rate is an affine function of the state variables, future values of the short rate are also normally distributed. Hence, the models in $\mathbb{A}_0(n)$ are Gaussian n-factor models. The expressions for means, variances, and covariances of the state variables (and hence of the short rate) will be simple only when the matrix $\underline{\underline{\hat{\kappa}}}$ is diagonal.[1]

Gaussian models are very tractable and allow analytical expressions for both bond prices and prices of European options on zero-coupon bonds. The bond prices follow from (8.7), where the ordinary differential equation (8.9) simplifies to

$$\boldsymbol{b}'(\tau) = -\underline{\underline{\hat{\kappa}}}^{\top}\boldsymbol{b}(\tau) + \boldsymbol{\xi}, \quad \boldsymbol{b}(0) = \boldsymbol{0}.$$

[1] Generally the moments depend on the eigenvalues and the eigenvectors of the matrix $\underline{\underline{\hat{\kappa}}}$, compare the discussion in Langetieg (1980).

The sensitivities of the zero-coupon bond prices σ_i^T defined in (8.12) depend only on the time to maturity of the bond,

$$\sigma_i^T(t) = -\sum_{k=1}^{n} \Gamma_{ki} b_k(T-t).$$

Since the vector of state variables $\boldsymbol{x}_T$ given $\boldsymbol{x}_t = \boldsymbol{x}$ is normally distributed, the zero-coupon bond price is lognormally distributed. We have from Section 7.2.3 that the price of a European call option on a zero-coupon bond is given by

$$C^{K,T,S}(\boldsymbol{x},t) = B^S(\boldsymbol{x},t) N(d_1) - K B^T(\boldsymbol{x},t) N(d_2),$$

where

$$d_1 = \frac{1}{v(t,T,S)} \ln\left(\frac{B^S(\boldsymbol{x},t)}{K B^T(\boldsymbol{x},t)}\right) + \frac{1}{2} v(t,T,S),$$
$$d_2 = d_1 - v(t,T,S),$$

and

$$\begin{aligned} v(t,T,S)^2 = \mathrm{Var}_t^{\mathbb{Q}^T}\left[\ln F_T^{T,S}\right] &= \sum_{i=1}^{n} \int_t^T \left(\sigma_i^S(u) - \sigma_i^T(u)\right)^2 du \\ &= \sum_{i=1}^{n} \int_t^T \left(\sum_{k=1}^{n} \Gamma_{ki}\left[b_k(S-u) - b_k(T-u)\right]\right)^2 du. \end{aligned} \tag{8.15}$$

A specific two-factor Gaussian model is considered in the following section.

8.3.2.2 Multi-factor CIR models

In the $\mathbb{A}_n(n)$ subclass of affine models, all n state variables affect conditional variances. The conditions ensuring admissibility imply that the dynamics of the state variables must be of the form

$$dx_{it} = \left(\varphi_i - \sum_{j=1}^{n} \kappa_{ij} x_{jt}\right) dt + \Gamma_i \sqrt{x_{it}}\, dz_{it},$$

with $\varphi_i > 0$ and $\kappa_{ij} \leq 0$ for $j \neq i$. To be consistent with (8.5), the market prices of risk must be of the form $\lambda_i(\boldsymbol{x}) = \bar{\lambda}_i \sqrt{x_{it}}$. Hence, the risk-neutral dynamics is

$$dx_{it} = \left(\varphi_i - \sum_{j=1}^{n} \hat{\kappa}_{ij} x_{jt}\right) dt + \Gamma_i \sqrt{x_{it}}\, dz_{it}^{\mathbb{Q}},$$

where $\hat{\kappa}_{ii} = \kappa_{ii} + \Gamma_i \bar{\lambda}_i$ and $\hat{\kappa}_{ij} = \kappa_{ij}$ for $j \neq i$. It seems natural to refer to such a model as a multi-factor CIR model. Note that the state variables are instantaneously uncorrelated, but in general they are not independent processes, since the drift rate of one variable can be positively related to the values of the other state variables. Consequently, future values of the state variables can indeed be positively correlated, but they cannot be negatively correlated.

The ODEs (8.9) simplify to

$$b_i'(\tau) = -\sum_{j=1}^{n} \hat{\kappa}_{ji} b_j(\tau) - \frac{1}{2}\Gamma_i^2 b_i(\tau)^2 + \xi_i, \quad i = 1, \dots, n,$$

but are in general still too complicated to solve in closed form. An exception is the special case where $\kappa_{ij} = 0$ for $j \neq i$, the different x_i processes are completely independent of each other. The ODEs are then disentangled to

$$b_i'(\tau) = -\hat{\kappa}_i b_j(\tau) - \frac{1}{2}\Gamma_i^2 b_i(\tau)^2 + \xi_i, \quad i = 1, \dots, n,$$

where $\hat{\kappa}_i$ is short for $\hat{\kappa}_{ii}$. The solution is analogous to the function $b(\tau)$ in the one-factor CIR model, compare (7.49). Alternatively, we can compute the zero-coupon bond price as follows. First, without loss of generality we can assume $\xi_1 = \dots = \xi_n = 1$, because we can scale any state variable up or down by multiplying φ_i and Γ_i by the same constant. Hence, we can assume that the short rate is $r_t = \xi_0 + \sum_{i=1}^n x_{it}$, and the zero-coupon bond price becomes

$$\begin{aligned}
B^T(\boldsymbol{x}, t) &= \mathrm{E}^{\mathbb{Q}}\left[e^{-\int_t^T r_u\, du} \,\middle|\, \boldsymbol{x}_t = \boldsymbol{x} \right] = \mathrm{E}^{\mathbb{Q}}\left[e^{-\xi_0[T-t] - \sum_{i=1}^n \int_t^T x_{iu}\, du} \,\middle|\, \boldsymbol{x}_t = \boldsymbol{x} \right] \\
&= e^{-\xi_0[T-t]}\, \mathrm{E}^{\mathbb{Q}}\left[\prod_{i=1}^n e^{-\int_t^T x_{iu}\, du} \,\middle|\, \boldsymbol{x}_t = \boldsymbol{x} \right] \\
&= e^{-\xi_0[T-t]} \prod_{i=1}^n \mathrm{E}^{\mathbb{Q}}\left[e^{-\int_t^T x_{iu}\, du} \,\middle|\, x_{it} = x_i \right],
\end{aligned}$$

where we have used the independence of the state variables. Because each state variable x_i follows a process of the same type as the short rate in the one-factor CIR model, we get

$$\mathrm{E}^{\mathbb{Q}}\left[e^{-\int_t^T x_{iu}\, du} \,\middle|\, x_{it} = x_i \right] = e^{-a_i(T-t) - b_i(T-t)x_i},$$

where

$$\begin{aligned}
b_i(\tau) &= \frac{2(e^{\gamma_i \tau} - 1)}{(\gamma_i + \hat{\kappa}_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i}, \\
a_j(\tau) &= -\frac{2\hat{\varphi}_i}{\Gamma_i^2}\left(\ln(2\gamma_i) + \frac{1}{2}(\hat{\kappa}_i + \gamma_i)\tau - \ln\left[(\gamma_i + \hat{\kappa}_i)(e^{\gamma_i \tau} - 1) + 2\gamma_i \right] \right),
\end{aligned}$$

and $\gamma_i = \sqrt{\hat{\kappa}_i^2 + 2\Gamma_i^2}$, cf. (7.46), (7.49), and (7.50). Consequently, the zero-coupon bond price is

$$B^T(\boldsymbol{x}, t) = e^{-\xi_0[T-t]} \prod_{i=1}^{n} e^{-a_i(T-t)-b_i(T-t)x_i} = e^{-a(T-t)-\sum_{i=1}^{n} b_i(T-t)x_i},$$

where $a(\tau) = \xi_0 \tau + \sum_{i=1}^{n} a_i(\tau)$. The risk-neutral dynamics of the zero-coupon bond price is

$$\frac{dB_t^T}{B_t^T} = r(\boldsymbol{x}_t)\, dt + \sum_{i=1}^{n} \sigma_i^T(\boldsymbol{x}_t, t)\, dz_{it}^{\mathbb{Q}},$$

where the sensitivities σ_i^T are given by

$$\sigma_i^T(\boldsymbol{x}, t) = -\Gamma_i \sqrt{x_i} b_i(T-t). \tag{8.16}$$

In the independent multi-factor CIR model, future values of the state variables will be non-centrally χ^2-distributed and it is possible to derive a closed-form pricing formula for European options on zero-coupon bonds involving an n-dimensional non-central χ^2 cumulative distribution function. We will consider the independent two-factor CIR model more closely in the following section.

8.4 TWO-FACTOR AFFINE DIFFUSION MODELS

8.4.1 The two-factor Vasicek model

Beaglehole and Tenney (1991) and Hull and White (1994b) have suggested a Gaussian two-factor model, which is a relatively simple extension of the one-factor Vasicek model. The Vasicek model has short rate dynamics

$$dr_t = \kappa \left[\hat{\theta} - r_t\right] dt + \beta\, dz_t^{\mathbb{Q}} = \left(\hat{\varphi} - \kappa r_t\right) dt + \beta\, dz_t^{\mathbb{Q}},$$

compare Section 7.4. The extension is to let the long-term level $\hat{\theta}$ follow a similar stochastic process. Hull and White formulate the generalized model as follows:

$$\begin{aligned} dr_t &= \left(\hat{\varphi} + \varepsilon_t - \kappa_r r_t\right) dt + \beta_r\, dz_{1t}^{\mathbb{Q}}, \\ d\varepsilon_t &= -\kappa_\varepsilon \varepsilon_t\, dt + \beta_\varepsilon \rho\, dz_{1t}^{\mathbb{Q}} + \beta_\varepsilon \sqrt{1-\rho^2}\, dz_{2t}^{\mathbb{Q}}. \end{aligned}$$

The process $\varepsilon = (\varepsilon_t)$ exhibits mean reversion around zero and represents the deviation of the current view on the long-term level of the short rate from the average view (under the risk-neutral measure). Here, β_ε is the volatility of the ε-process, and ρ is the correlation between changes in the short rate and changes in ε. All constant parameters are assumed to be positive except ρ, which can have any value in the interval $[-1, 1]$. This two-factor model is a special case of the general Gaussian multi-factor model (8.14), namely the special case given by the parameter restrictions

$$\begin{aligned}
\hat{\varphi}_1 &= \hat{\varphi}, & \hat{\kappa}_{11} &= \kappa_r, \\
\hat{\kappa}_{12} &= -1, & \Gamma_{11} &= \beta_r, \\
\Gamma_{12} &= 0, & \hat{\varphi}_2 &= 0, \\
\hat{\kappa}_{21} &= 0, & \hat{\kappa}_{22} &= \kappa_\varepsilon, \\
\Gamma_{21} &= \beta_\varepsilon \rho, & \Gamma_{22} &= \beta_\varepsilon \sqrt{1-\rho^2}.
\end{aligned}$$

Since the short rate is itself the first state variable, we must in addition put $\xi_1 = 1$ and $\xi_0 = \xi_2 = 0$.

After these substitutions the ordinary differential equations for b_1 and b_2 become

$$\begin{aligned}
b_1'(\tau) &= -\kappa_r b_1(\tau) + 1, & b_1(0) &= 0, \\
b_2'(\tau) &= b_1(\tau) - \kappa_\varepsilon b_2(\tau), & b_2(0) &= 0.
\end{aligned} \tag{8.17}$$

The first of these is identical to the differential equation solved in the original one-factor Vasicek model, see Section 7.4.2, so we know that the solution is

$$\mathcal{B}_\kappa(\tau) = \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right).$$

In particular, $b_1(\tau) = \mathcal{B}_{\kappa_r}(\tau)$. It can be verified (see Exercise 8.3) that the solution to the equation for b_2 is given by

$$b_2(\tau) = \frac{1}{\kappa_r - \kappa_\varepsilon]}(\mathcal{B}_{\kappa_\varepsilon}(\tau) - \mathcal{B}_{\kappa_r}(\tau)) \tag{8.18}$$

Finally, the equation for the function a can be rewritten as

$$a'(\tau) = \hat{\varphi} b_1(\tau) - \frac{1}{2}\beta_r^2 b_1(\tau)^2 - \frac{1}{2}\beta_\varepsilon^2 b_2(\tau)^2 - \rho\beta_r\beta_\varepsilon b_1(\tau) b_2(\tau), \quad a(0) = 0,$$

from which it follows that

$$\begin{aligned}
a(\tau) = a(\tau) - a(0) &= \int_0^\tau a'(u)\,du \\
&= \hat{\varphi}\int_0^\tau \mathcal{B}_{\kappa_r}(u)\,du - \frac{1}{2}\beta_r^2 \int_0^\tau \mathcal{B}_{\kappa_r}(u)^2\,du \\
&\quad - \frac{1}{2}\frac{\beta_\varepsilon^2}{(\kappa_r - \kappa_\varepsilon)^2}\int_0^\tau (\mathcal{B}_{\kappa_\varepsilon}(u) - \mathcal{B}_{\kappa_r}(u))^2 du \\
&\quad - \frac{\rho\beta_r\beta_\varepsilon}{\kappa_r - \kappa_\varepsilon}\int_0^\tau \mathcal{B}_{\kappa_r}(u)(\mathcal{B}_{\kappa_\varepsilon}(u) - \mathcal{B}_{\kappa_r}(u))du.
\end{aligned}$$

These integrals are easily computed (as in the one-factor Vasicek model), and we end up with

$$\begin{aligned}
a(\tau) &= \frac{\hat{\varphi}}{\kappa_r}\left(\tau - \mathcal{B}_{\kappa_r}(\tau)\right) \\
&\quad - \frac{1}{2\kappa_r^2}\left(\beta_r^2 - \frac{2\rho\beta_r\beta_\varepsilon}{\kappa_r - \kappa_\varepsilon} + \frac{\beta_\varepsilon^2}{(\kappa_r - \kappa_\varepsilon)^2}\right)\left(\tau - \mathcal{B}_{\kappa_r}(\tau) - \frac{\kappa_r}{2}\mathcal{B}_{\kappa_r}(\tau)^2\right)
\end{aligned}$$

$$
\begin{aligned}
&- \frac{1}{2} \frac{\beta_\varepsilon^2}{\kappa_\varepsilon^2 (\kappa_r - \kappa_\varepsilon)^2} \left(\tau - \mathcal{B}_{\kappa_\varepsilon}(\tau) - \frac{\kappa_\varepsilon}{2} \mathcal{B}_{\kappa_\varepsilon}(\tau)^2 \right) \\
&+ \frac{1}{\kappa_r \kappa_\varepsilon (\kappa_r - \kappa_\varepsilon)} \left(\frac{\beta_\varepsilon^2}{\kappa_r - \kappa_\varepsilon} - \rho \beta_r \beta_\varepsilon \right) \left(\tau - \mathcal{B}_{\kappa_r}(\tau) - \mathcal{B}_{\kappa_\varepsilon}(\tau) + \mathcal{B}_{\kappa_r + \kappa_\varepsilon}(\tau) \right).
\end{aligned}
$$

The zero-coupon yield for a time-to-maturity of τ is given by

$$
\bar{y}^\tau(r, \varepsilon) = \frac{a(\tau)}{\tau} + \frac{b_1(\tau)}{\tau} r + \frac{b_2(\tau)}{\tau} \varepsilon. \tag{8.19}
$$

In Exercise 8.4 you are asked to investigate experimentally what shapes the yield curve might have depending on the current values of the state variables and the parameters of the model. You should find that the model can generate a wider variety of yield curve shapes than the one-factor Vasicek model. For some parameter combinations it is also possible to obtain twists of the yield curve, so that short-maturity and long-maturity yields move in different directions. Other aspects of the two-factor Vasicek model are discussed in Exercises 8.5 and 8.6.

The relevant variance $v(t, T, S)^2$ entering the price of an option on a zero-coupon bond follows from (8.15):

$$
\begin{aligned}
v(t, T, S)^2 &= \beta_r^2 \int_t^T \left[b_1(S - u) - b_1(T - u) \right]^2 \, du \\
&\quad + \beta_\varepsilon^2 \int_t^T \left[b_2(S - u) - b_2(T - u) \right]^2 \, du \\
&\quad + 2\rho\beta_r\beta_\varepsilon \int_t^T \left[b_1(S-u) - b_1(T-u) \right] \left[b_2(S - u) - b_2(T - u) \right] \, du,
\end{aligned}
$$

where the integrals can be computed explicitly.

Hull and White (1994b) show further how to obtain a perfect fit of the model to an observed yield curve by replacing the constant $\hat{\varphi}$ by a suitable time-dependent function.

8.4.2 The independent two-factor CIR model aka the Longstaff–Schwartz model

8.4.2.1 Model description

An interesting example of a multi-factor CIR model is the two-factor model of Longstaff and Schwartz (1992a). As the one-factor CIR model, the Longstaff–Schwartz model is a special case of the general equilibrium model studied by Cox et al. (1985a). With some empirical support, Longstaff and Schwartz assume that the economy has one state variable, x_1, affecting only expected returns on productive investments and another state variable, x_2, affecting both expected returns and the uncertainty about the returns on productive investments. The two state variables x_1 and x_2 are assumed to follow the independent square-root processes

$$dx_{1t} = (\varphi_1 - \kappa_1 x_{1t})\, dt + \beta_1 \sqrt{x_{1t}}\, dz_{1t},$$
$$dx_{2t} = (\varphi_2 - \kappa_2 x_{2t})\, dt + \beta_2 \sqrt{x_{2t}}\, dz_{2t}$$

under the real-world probability measure. All the constants are positive.

Under certain specifications of preferences, endowments, technology, and so on, and an appropriate scaling of the two state variables, the equilibrium short rate is exactly the sum of the two state variables,

$$r_t = x_{1t} + x_{2t}. \tag{8.20}$$

Furthermore, the market price of risk associated with x_1, that is $\lambda_1(\boldsymbol{x}, t)$, is equal to zero, while the market price of risk associated with x_2 is of the form $\lambda_2(\boldsymbol{x}, t) = \lambda\sqrt{x_2}/\beta_2$, where λ is a constant. Hence, the standard Brownian motions under the risk-neutral probability measure $\mathbb{Q}$ are given by

$$dz^{\mathbb{Q}}_{1t} = dz_{1t}, \qquad dz^{\mathbb{Q}}_{2t} = dz_{2t} + \frac{\lambda}{\beta_2}\sqrt{x_{2t}}\, dt. \tag{8.21}$$

The dynamics of the state variables under the risk-neutral measure becomes

$$dx_{1t} = \left(\hat{\varphi}_1 - \hat{\kappa}_1 x_{1t}\right)\, dt + \beta_1 \sqrt{x_{1t}}\, dz^{\mathbb{Q}}_{1t},$$
$$dx_{2t} = \left(\hat{\varphi}_2 - \hat{\kappa}_2 x_{2t}\right)\, dt + \beta_2 \sqrt{x_{2t}}\, dz^{\mathbb{Q}}_{2t},$$

where $\hat{\varphi}_1 = \varphi_1$, $\hat{\kappa}_1 = \kappa_1$, $\hat{\varphi}_2 = \varphi_2$, and $\hat{\kappa}_2 = \kappa_2 + \lambda$.

8.4.2.2 The yield curve

According to the analysis for general multi-factor CIR models, the zero-coupon bond price $B^T(x_1, x_2, t)$ can be written as

$$B^T(x_1, x_2, t) = e^{-a(T-t) - b_1(T-t)x_1 - b_2(T-t)x_2}, \tag{8.22}$$

where $a(\tau) = a_1(\tau) + a_2(\tau)$,

$$b_j(\tau) = \frac{2(e^{\gamma_j \tau} - 1)}{(\gamma_j + \hat{\kappa}_j)(e^{\gamma_j \tau} - 1) + 2\gamma_j}, \quad j = 1, 2,$$

$$a_j(\tau) = -\frac{2\hat{\varphi}_j}{\beta_j^2}\left(\ln(2\gamma_j) + \frac{1}{2}(\hat{\kappa}_j + \gamma_j)\tau - \ln\left[(\gamma_j + \hat{\kappa}_j)(e^{\gamma_j \tau} - 1) + 2\gamma_j\right]\right),$$

and $\gamma_j = \sqrt{\hat{\kappa}_j^2 + 2\beta_j^2}$ for $j = 1, 2$.

The state variables x_1 and x_2 are abstract variables that are not directly observable. Longstaff and Schwartz perform a change of variables to the short rate, r_t, and the instantaneous variance rate of the short rate, v_t. Strictly speaking, these variables cannot be directly observed either, but they can be estimated from bond market data. In addition, the new variables seem important for the pricing of bonds and interest rate derivatives, and it is easier to relate to prices as functions of r and v instead of functions of x_1 and x_2. Since r_t is given by (8.20), we get $dr_t = dx_{1t} + dx_{2t}$, that is

$$dr_t = (\varphi_1 + \varphi_2 - \kappa_1 x_{1t} - \kappa_2 x_{2t})\, dt + \beta_1 \sqrt{x_{1t}}\, dz_{1t} + \beta_2 \sqrt{x_{2t}}\, dz_{2t}.$$

The instantaneous variance is $\mathrm{Var}_t(dr_t) = v_t\, dt$, where

$$v_t = \beta_1^2 x_{1t} + \beta_2^2 x_{2t}, \tag{8.23}$$

so that the dynamics of v_t is

$$dv_t = \left(\beta_1^2 \varphi_1 + \beta_2^2 \varphi_2 - \beta_1^2 \kappa_1 x_{1t} - \beta_2^2 \kappa_2 x_{2t}\right) dt + \beta_1^3 \sqrt{x_{1t}}\, dz_{1t} + \beta_2^3 \sqrt{x_{2t}}\, dz_{2t}.$$

If $\beta_1 \neq \beta_2$, the Equations (8.20) and (8.23) imply that

$$x_{1t} = \frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}, \qquad x_{2t} = \frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}. \tag{8.24}$$

The dynamics of r and v can then be rewritten as

$$\begin{aligned} dr_t = &\left(\varphi_1 + \varphi_2 - \frac{\kappa_1 \beta_2^2 - \kappa_2 \beta_1^2}{\beta_2^2 - \beta_1^2} r_t - \frac{\kappa_2 - \kappa_1}{\beta_2^2 - \beta_1^2} v_t\right) dt \\ &+ \beta_1 \sqrt{\frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}}\, dz_{1t} + \beta_2 \sqrt{\frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}}\, dz_{2t}, \end{aligned} \tag{8.25}$$

$$\begin{aligned} dv_t = &\left(\beta_1^2 \varphi_1 + \beta_2^2 \varphi_2 - \beta_1^2 \beta_2^2 \frac{\kappa_1 - \kappa_2}{\beta_2^2 - \beta_1^2} r_t - \frac{\beta_2^2 \kappa_2 - \beta_1^2 \kappa_1}{\beta_2^2 - \beta_1^2} v_t\right) dt \\ &+ \beta_1^3 \sqrt{\frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}}\, dz_{1t} + \beta_2^3 \sqrt{\frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}}\, dz_{2t}. \end{aligned} \tag{8.26}$$

Since both x_1 and x_2 stay non-negative, it follows from (8.24) that v_t at any point in time will lie between $\beta_1^2 r_t$ and $\beta_2^2 r_t$. It can be shown that (8.25) and (8.26) imply that changes in r_t and v_t are positively correlated, which is in accordance with empirical observations of the relation between the level and the volatility of interest rates.

Substituting (8.24) into (8.22), we can write the zero-coupon bond price as a function of r and v:

$$B^T(r, v, t) = e^{-a(T-t) - \tilde{b}_1(T-t) r - \tilde{b}_2(T-t) v},$$

where

$$\tilde{b}_1(\tau) = \frac{\beta_2^2 b_1(\tau) - \beta_1^2 b_2(\tau)}{\beta_2^2 - \beta_1^2}, \qquad \tilde{b}_2(\tau) = \frac{b_2(\tau) - b_1(\tau)}{\beta_2^2 - \beta_1^2}.$$

Note that the zero-coupon bond price involves six parameters, namely β_1, β_2, $\hat{\kappa}_1$, $\hat{\kappa}_2$, $\hat{\varphi}_1$, and $\hat{\varphi}_2$. The partial derivatives $\partial B^T/\partial r$ and $\partial B^T/\partial v$ can be either positive or negative so, in contrast to the one-factor models in Chapter 7, the zero-coupon bond price is not a monotonically decreasing function of the short rate. According to Longstaff and Schwartz, the derivative $\partial B^T/\partial r$ is typically negative for short-term bonds, but it can be positive for long-term bonds. The derivative $\partial B^T/\partial v$ approaches zero when time-to-maturity goes to zero, so that very short-term bonds are affected primarily by the short rate and only to a small extent by the variance of the short rate. If the short rate r_t at some point in time is zero (in which case v_t is also zero), it will become strictly positive immediately afterwards and, hence, $B^T(0,0,t) < 1$ for $t < T$. Finally, $B^T(r,v,t) \to 0$ for $r \to \infty$ (in which case also $v \to \infty$).

The zero-coupon yield $\bar{y}_t^\tau = y_t^{t+\tau}$ is given by $\bar{y}_t^\tau = \bar{y}^\tau(r_t, v_t)$, where

$$\bar{y}^\tau(r,v) = \frac{a(\tau)}{\tau} + \frac{\tilde{b}_1(\tau)}{\tau} r + \frac{\tilde{b}_2(\tau)}{\tau} v,$$

which is an affine function of r and v. It can be shown that $\bar{y}^\tau(r,v) \to r$ for $\tau \to 0$ and that the asymptotic long rate is constant since

$$\bar{y}^\tau(r,v) \to \frac{\hat{\varphi}_1}{\beta_1^2}(\gamma_1 - \hat{\kappa}_1) + \frac{\hat{\varphi}_2}{\beta_2^2}(\gamma_2 - \hat{\kappa}_2) \qquad \text{for } \tau \to \infty.$$

According to Longstaff and Schwartz, the yield curve $\tau \mapsto \bar{y}^\tau(r,v)$ can have many different shapes. For example, it can be monotonically increasing or decreasing, humped (first increasing, then decreasing), it can have a trough (first decreasing, then increasing), or both a hump and a trough. We can see most of these shapes in Fig. 8.1. Note that for a given short rate the shape of the yield curve may depend on the variance factor. Partial changes in r and v may imply a significant change of the shape of the yield curve, for example, a twist so that different maturity segments of the yield curve move in opposite directions. The Longstaff–Schwartz model is therefore much more flexible than the one-factor CIR model.

The forward rate $\bar{f}_t^\tau = f_t^{t+\tau}$ is given by $\bar{f}_t^\tau = \bar{f}^\tau(r_t, v_t)$, where

$$\bar{f}^\tau(r,v) = a'(\tau) + \tilde{b}_1'(\tau) r + \tilde{b}_2'(\tau) v.$$

All zero-coupon yields and forward rates are non-negative in this model.

The dynamics of the zero-coupon bond price is of the form

$$\begin{aligned}
\frac{dB_t^T}{B_t^T} &= r_t\,dt - \beta_1\sqrt{x_{1t}}b_1(T-t)\,dz_{1t}^{\mathbb{Q}} - \beta_2\sqrt{x_{2t}}b_2(T-t)\,dz_{2t}^{\mathbb{Q}} \\
&= (r_t - \lambda x_{2t} b_2(T-t))\,dt - \beta_1\sqrt{x_{1t}}b_1(T-t)\,dz_{1t} - \beta_2\sqrt{x_{2t}}b_2(T-t)\,dz_{2t} \\
&= \left(r_t + \frac{\lambda}{\beta_2^2 - \beta_1^2} b_2(T-t)(\beta_1^2 r_t - v_t)\right) dt \\
&\quad - \beta_1\sqrt{\frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}}\, b_1(T-t)\,dz_{1t} - \beta_2\sqrt{\frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}}\, b_2(T-t)\,dz_{2t},
\end{aligned}$$

where we have applied (8.16), (8.21), and (8.24). The so-called term premium, that is the expected rate of return on a zero-coupon bond in excess of the short rate, is

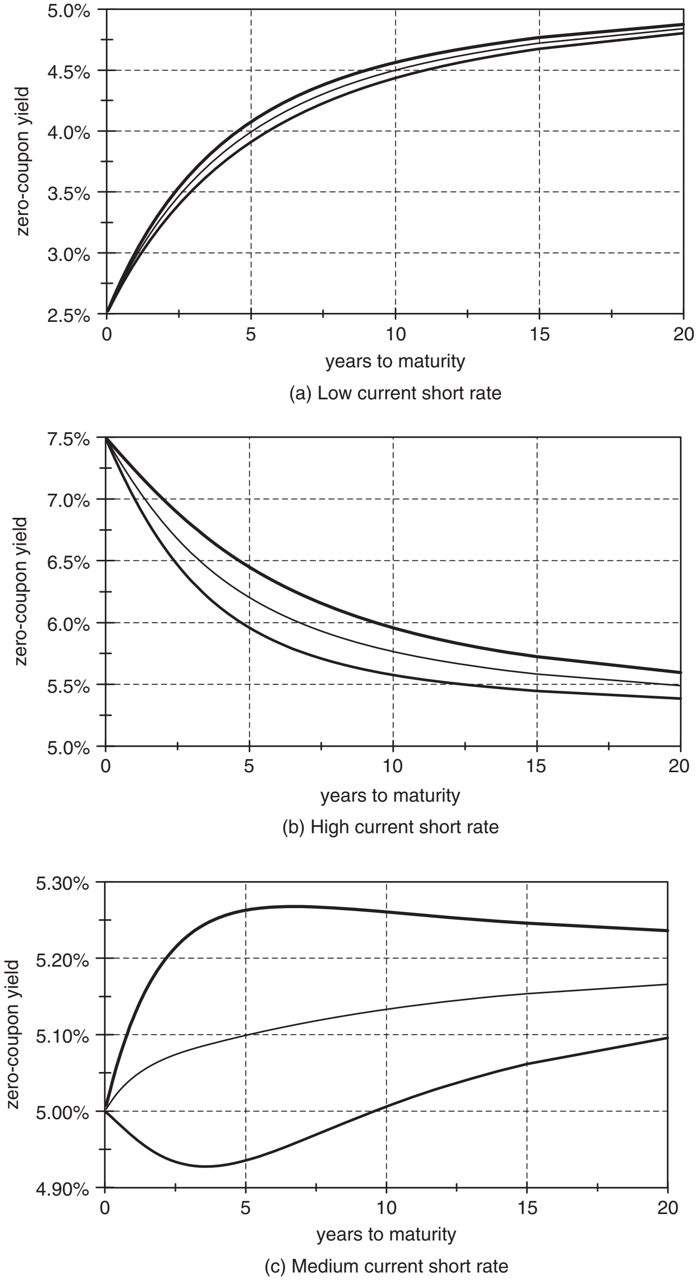

Fig. 8.1: Zero-coupon yield curves in the Longstaff–Schwartz model. The parameter values are $\beta_1 = 0.1$, $\beta_2 = 0.2$, $\kappa_1 = 0.3$, $\kappa_2 = 0.45$, $\varphi_1 = \varphi_2 = 0.01$, and $\lambda = 0$. The asymptotic long rate is 5.20%. The very thick lines are for a high value of v, namely $(0.75\beta_2^2 + 0.25\beta_1^2)r$; the thin lines are for a medium value of v, namely $(0.5\beta_2^2 + 0.5\beta_1^2)r$; and the medium thick lines are for a low value of v, namely $(0.25\beta_2^2 + 0.75\beta_1^2)r$.

$\lambda b_2(T-t)(\beta_1^2 r_t - v_t)/(\beta_2^2-\beta_1^2)$, which is positive if $\lambda < 0$. It is consistent with empirical studies that the term premium is affected by two stochastic factors (r and v) and depends on the interest rate volatility. The volatility $\|\sigma^T(r,v,t)\|$ of the zero-coupon bond price B_t^T is in the Longstaff–Schwartz model (you are asked to verify this in Exercise 8.7) given by

$$\begin{aligned}\|\sigma^T(r,v,t)\|^2 &= \frac{\beta_1^2\beta_2^2}{\beta_2^2-\beta_1^2}\left(b_1(T-t)^2 - b_2(T-t)^2\right) r_t \\ &\quad + \frac{\beta_2^2 b_2(T-t)^2 - \beta_1^2 b_1(T-t)^2}{\beta_2^2-\beta_1^2} v_t. \end{aligned} \tag{8.27}$$

Since the function $T \mapsto \|\sigma^T(r,v,t)\|$ depends on both of r and v, the two-factor model is able to generate more flexible term structures of volatilities than the one-factor models. It can be shown that the volatility $\|\sigma^T(r,v,t)\|$ is an increasing function of the time to maturity $T-t$.

8.4.2.3 Options and other derivatives

The price of a European call on a zero-coupon bond can be computed using (8.13). In the Longstaff–Schwartz model the two state variables are non-centrally χ^2-distributed so, not surprisingly, the relevant probabilities are taken from the two-dimensional non-central χ^2-distribution. Longstaff and Schwartz state the precise formula, which in our notation looks as follows:

$$\begin{aligned} C^{K,T,S}(r,v,t) &= B^S(r,v,t)\chi^2\left(\theta_1,\theta_2; 4\varphi_1/\beta_1^2, 4\varphi_2/\beta_2^2, \omega_1[\beta_2^2 r - v], \omega_2[v - \beta_1^2 r]\right) \\ &\quad - KB^T(r,v,t)\chi^2\left(\hat{\theta}_1,\hat{\theta}_2; 4\varphi_1/\beta_1^2, 4\varphi_2/\beta_2^2, \hat{\omega}_1[\beta_2^2 r - v], \hat{\omega}_2[v - \beta_1^2 r]\right), \end{aligned} \tag{8.28}$$

where, for $i = 1,2$, $\hat{b}_i(\tau) = \gamma_i b_i(\tau)/(e^{\gamma_i\tau}-1)$ and

$$\theta_i = \frac{-4\gamma_i^2[a(S-T)+\ln K]}{\beta_i^2(e^{\gamma_i[T-t]}-1)^2\hat{b}_i(S-t)},$$

$$\hat{\theta}_i = \frac{-4\gamma_i^2[a(S-T)+\ln K]}{\beta_i^2(e^{\gamma_i[T-t]}-1)^2\hat{b}_i(T-t)\hat{b}_i(S-T)},$$

$$\omega_i = \frac{4\gamma_i e^{\gamma_i[T-t]}\hat{b}_i(S-t)}{\beta_i^2(\beta_2^2-\beta_1^2)(e^{\gamma_i[T-t]}-1)\hat{b}_i(S-T)},$$

$$\hat{\omega}_i = \frac{4\gamma_i e^{\gamma_i[T-t]}\hat{b}_i(T-t)}{\beta_i^2(\beta_2^2-\beta_1^2)(e^{\gamma_i[T-t]}-1)}.$$

Here $\chi^2(\cdot,\cdot)$ is the cumulative distribution function for a two-dimensional non-central χ^2-distribution. To be more precise, the value of the cumulative distribution function is

$$\chi^2(\theta_1, \theta_2; c_1, c_2, d_1, d_2) = \int_0^{\theta_1} f_{\chi^2(c_1,d_1)}(u) \left[\int_0^{\theta_2 - u\theta_2/\theta_1} f_{\chi^2(c_2,d_2)}(s)\, ds \right] du,$$

where $f_{\chi^2(c,d)}$ is the probability density function for a one-dimensional random variable which is non-centrally χ^2-distributed with c degrees of freedom and non-centrality parameter d. Note that the inner integral can be written as the cumulative distribution function for a one-dimensional non-central χ^2-distribution evaluated in the point $\theta_2 - u\theta_2/\theta_1$. As discussed in the context of the one-factor CIR model, this one-dimensional cumulative distribution function can be approximated by the cumulative distribution function of a standard one-dimensional normal distribution. The value of the two-dimensional χ^2-distribution function can then be obtained by a numerical integration. Chen and Scott (1992) provide a detailed analysis of the computation of the two-dimensional χ^2-distribution function. They conclude that, despite the necessary numerical integration, the option price can be computed much faster using (8.28) than using Monte Carlo simulation or numerical solution of the fundamental partial differential equation. Longstaff and Schwartz state that the partial derivatives $\partial C/\partial r$ and $\partial C/\partial v$ can be either positive or negative, which is not surprising considering the fact that the price of the underlying bond can also be either positively or negatively related to r and v.

In the Longstaff–Schwartz model the prices of many derivative securities can only be computed using numerical techniques. One approach is to solve numerically the fundamental partial differential equation with the appropriate terminal conditions, see Chapter 16. For that purpose the formulation of the model in terms of the original state variables, x_1 and x_2, is preferable. The PDE to be solved is

$$\begin{aligned}
&\frac{\partial P}{\partial t}(x_1, x_2, t) + \left(\hat{\varphi}_1 - \hat{\kappa}_1 x_1\right) \frac{\partial P}{\partial x_1}(x_1, x_2, t) + \left(\hat{\varphi}_2 - \hat{\kappa}_2 x_2\right) \frac{\partial P}{\partial x_2}(x_1, x_2, t) \\
&\quad + \frac{1}{2}\beta_1^2 x_1 \frac{\partial^2 P}{\partial x_1^2}(x_1, x_2, t) + \frac{1}{2}\beta_2^2 x_2 \frac{\partial^2 P}{\partial x_2^2}(x_1, x_2, t) \\
&\quad - (x_1 + x_2) P(x_1, x_2, t) = 0, \quad (x_1, x_2, t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times [0, T).
\end{aligned}$$

Note that since x_1 and x_2 are independent, there is no term with the mixed second-order derivative $\frac{\partial^2 P}{\partial x_1 \partial x_2}$. This fact simplifies the numerical solution. The PDE for the price function in terms of the variables r and v will involve a mixed second-order derivative since r and v are not independent. Furthermore, the value space for the variables x_1 and x_2 is simpler than the value space for r and v since the possible values of v depend on the value of r. This will complicate the numerical solution of the PDE involving r and v even further.

8.4.2.4 Additional remarks

To implement the Longstaff–Schwartz model, the current values of the short rate and the current variance rate of the short rate must be determined, and parameter values have to be estimated. Longstaff and Schwartz discuss the estimation procedure both in the original article and in other articles, see Longstaff and Schwartz 1993a, 1994. Clewlow and Strickland (1994) and Rebonato (1996, Ch. 12) discuss several practical problems in the parameter estimation. Longstaff and Schwartz (1993a) explain how to obtain a perfect fit of the model yield curve to the observed yield curve by replacing the parameter $\hat{\kappa}_2$ with a suitable time-dependent function. However, this extended version of the model exhibits time-inhomogeneous volatilities, which is problematic as will be discussed in Section 9.6. In Longstaff and Schwartz (1992b) the authors consider the pricing of caps and swaptions within their two-factor model, whereas in Longstaff and Schwartz (1993b) they discuss the importance of taking stochastic interest rate volatility into account when measuring the interest rate risk of bonds.

8.4.3 Other two-factor models

Cox et al. (1985b) introduce several multi-factor versions of their famous one-factor model. The short rate in their one-factor model is really the *real* short rate, the bonds they price are real bonds promising delivery of certain prespecified consumption units, and the prices are also stated in consumption units. To derive prices in monetary units (for example dollars) of nominal securities, that is securities with payoff specified in monetary units, they focus on including the consumer price index as an additional state variable. In their extensions they continue to assume that the real short rate follows the process

$$dr_t = \kappa[\theta - r_t]\,dt + \beta\sqrt{r_t}\,dz_{1t},$$

as in the one-factor model. The first extension is to let the consumer price index I_t follow a geometric Brownian motion

$$dI_t = I_t\left[\pi\,dt + \beta_I\,dz_{2t}\right],$$

where π denotes the expected inflation rate, and where the market price of risk associated with the consumer price uncertainty is zero. A well-known application of Itô's Lemma implies that

$$d(\ln I_t) = \left(\pi - \frac{1}{2}\beta_I^2\right)dt + \beta_I\,dz_{2t}$$

so that the extended model is affine in r_t and $\ln I_t$. The price in monetary units of a nominal zero-coupon bond maturing at time T is

$$B^T(I, r, t) = I^{-1} e^{-a(T-t)-\left(\pi-\frac{1}{2}\beta_I^2\right)(T-t)-b(T-t)r},$$

where the functions $a(\tau)$ and $b(\tau)$ are exactly as in the one-factor CIR model, see (7.49) and (7.50).

8.5 THREE-FACTOR AFFINE MODELS

Balduzzi et al. (1996) suggest a three-factor affine model in which the three state variables are the short rate r_t, the long-term level θ_t of the short rate, and the instantaneous variance v_t of the short rate. They assume that the real-world dynamics is of the form

$$dr_t = \kappa_r[\theta_t - r_t]\,dt + \sqrt{v_t}\,dz_{1t},$$
$$d\theta_t = \kappa_\theta[\bar{\theta} - \theta_t]\,dt + \beta_\theta\,dz_{2t},$$
$$dv_t = \kappa_v[\bar{v} - v_t]\,dt + \rho\beta_v\sqrt{v_t}\,dz_{1t} + \sqrt{1-\rho^2}\beta_v\sqrt{v_t}\,dz_{3t}.$$

Here, ρ is the correlation between changes in the short rate level and the variance of the short rate. Furthermore, the market prices of risk are assumed to have a form that implies that the dynamics under the risk-neutral measure is

$$dr_t = (\kappa_r[\theta_t - r_t] - \lambda_r v_t)\;dt + \sqrt{v_t}\,dz^{\mathbb{Q}}_{1t},$$
$$d\theta_t = \left(\kappa_\theta[\bar{\theta} - \theta_t] - \lambda_\theta\beta_\theta\right)\;dt + \beta_\theta\,dz^{\mathbb{Q}}_{2t},$$
$$dv_t = (\kappa_v[\bar{v} - v_t] - \lambda_v v_t)\;dt + \rho\beta_v\sqrt{v_t}\,dz^{\mathbb{Q}}_{1t} + \sqrt{1-\rho^2}\beta_v\sqrt{v_t}\,dz^{\mathbb{Q}}_{3t},$$

where λ_r, λ_θ, and λ_v are constants. The model is in the subclass $\mathbb{A}_1(3)$ as exactly one of the three variables, namely v_t, affects the instantaneous variance-covariance matrix. The zero-coupon bond prices are

$$B^T(r,\theta,v,t) = e^{-a(T-t)-b_1(T-t)r-b_2(T-t)\theta-b_3(T-t)v}.$$

The authors find explicit expressions for b_1 and b_2, but a and b_3 must be found by numerical solution of the appropriate ordinary differential equations, see (8.9) and (8.10). The model can produce a wide variety of interesting yield curve shapes. The estimation of the model is also discussed by the authors.

The model by Balduzzi et al. falls into the $\mathbb{A}_1(3)$ model class (it is no problem to reformulate the model so that the variance-influencing variable v is the first state variable, as assumed in the discussion in Section 8.3.2). Dai and Singleton (2000) show that a more general $\mathbb{A}_1(3)$ exists with six additional parameters. They demonstrate in an empirical study that the more general version fits a certain data set of U.S. swap rates of different maturities much better than the original model by Balduzzi et al. Some, but not all, of the six added parameters can be set to zero without significantly reducing the fit of the model. The main virtue of the extended model is that it allows a more flexible correlation structure of the three state variables and, in particular, a negative correlation between some of the state variables. No closed-form solution for the functions a, b_1, b_2, and b_3 can be given in the extended versions of the model, but the relevant ordinary differential equations can be solved numerically.

Chen (1996) studies a three-factor model with the same three state variables r_t, θ_t, and v_t. In the simplest version of the model, the dynamics under the real-world probability measure is given as

$$dr_t = \kappa_r[\theta_t - r_t]\,dt + \sqrt{v_t}\,dz_{1t},$$
$$d\theta_t = \kappa_\theta[\bar{\theta} - \theta_t]\,dt + \beta_\theta\sqrt{\theta_t}\,dz_{2t},$$
$$dv_t = \kappa_v[\bar{v} - v_t]\,dt + \beta_v\sqrt{v_t}\,dz_{3t},$$

and the market prices of risk are such that the dynamics of the state variables have the same structure under the risk-neutral measure, but with different constants κ_r, κ_θ, $\bar{\theta}$, κ_v, and $\bar{v}$. The model is in the $\mathbb{A}_2(3)$ subclass as two of the three state variables are influencing the instantaneous variances and covariances. The zero-coupon bond prices are of the form

$$B^T(r, \theta, v, t) = e^{-a(T-t)-b_1(T-t)r-b_2(T-t)\theta-b_3(T-t)v}.$$

Chen is able to find explicit, but complicated, expressions for the functions a, b_1, b_2, and b_3. In addition, Chen considers a more general three-factor model, which is not included in the affine class of models.

Dai and Singleton (2000) show that the Chen model can be generalized within the $\mathbb{A}_2(3)$ class. The maximal generalization adds eight constants, but even a generalized model with just two extra constants performs much better empirically than Chen's original model. Again, the gain from the generalization seems to be mainly due to allowing a negative correlation between some of the state variables. In these extended models, the functions a, b_1, b_2, and b_3 have to be found by solving the relevant ordinary differential equations numerically.

Note that negative correlations between the state variables are not possible in the multi-factor CIR models and thus models in the $\mathbb{A}_3(3)$ subclass. Models in that subclass cannot therefore fit the data very well. According to Dai and Singleton (2000), the most general $\mathbb{A}_1(3)$ model offers the most flexible instantaneous correlation structure, whereas the most general $\mathbb{A}_2(3)$ model offers the highest flexibility in specifying time-varying volatilities. At least in the study of Dai and Singleton (2000), the flexible correlations are more important than the flexible volatilities, so they prefer a model in the $\mathbb{A}_1(3)$ subclass. Another interesting conclusion of the study is that the short rate tends to mean-revert relatively fast to a factor that itself is mean-reverting to a constant at a slower speed.

Section 8.4.3 referred to a two-factor version of the CIR model in which the second factor is the (log) consumer price index. The model is affine and leads to a closed-form solution for nominal bond prices. Cox et al. (1985b) also considered a three-factor version of the model in which the expected inflation rate π is also assumed to be stochastic so that

$$\begin{aligned} dI_t &= I_t \left[\pi_t \, dt + \beta_I \sqrt{\pi_t} \, dz_{2t} \right], \\ d\pi_t &= \kappa_\pi [\theta_\pi - \pi_t] \, dt + \rho \beta_\pi \sqrt{\pi_t} \, dz_{2t} + \sqrt{1-\rho^2} \beta_\pi \sqrt{\pi_t} \, dz_{3t}. \end{aligned}$$

The resulting model is a three-factor affine model with the state variables r_t, $\ln I_t$, and π_t. For the precise expression of the price of a nominal zero-coupon bond we refer the reader to the article, Cox et al. (1985b).[2] Other affine models of this type have been studied by Chen and Scott (1993).

[2] In addition, Cox et al. (1985b) state an explicit, but very complicated, pricing formula for the nominal bond in the case where the expected inflation rate follows the process

$$d\pi_t = \kappa_\pi [\theta_\pi - \pi_t] \, dt + \rho \beta_\pi \pi_t^{3/2} \, dz_{2t} + \sqrt{1-\rho^2} \beta_\pi \pi_t^{3/2} \, dz_{3t},$$

and the dynamics of r_t and I_t are as before. This model does not belong to the affine class.

8.6 GENERALIZED AFFINE MODELS

In the framework of one-factor models, we discussed in Section 7.6 how the market price of risk can be generalized within the affine framework. The main advantage is to allow a larger difference between the real-world and the risk-neutral dynamics of the short rate, which can make it easier for the model to match both the cross-section of bond prices and yields at a given date as well as the time-series behaviour of interest rates. The multi-factor models can be generalized in a similar way.

Following the terminology of Duffee (2002), the models in which the short rate is given by (8.4), the risk-neutral dynamics of the state variables by (8.3), and the market price of risk by (8.5) are referred to as the **completely affine** models. The generalizations are considering a more flexible specification of the market price of risk, whereas the specification of the short rate and the risk-neutral dynamics of the state variables remain the same. Consequently, the real-world dynamics of the state variables is generalized.

In the **essentially affine** models introduced by Duffee (2002), the market price of risk is assumed to be of the form

$$\boldsymbol{\lambda}_t = \sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\bar{\boldsymbol{\lambda}}_1 + \sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)^-}\,\underline{\underline{\bar{\lambda}}}_2\boldsymbol{x}_t, \tag{8.29}$$

where $\underline{\underline{\bar{\lambda}}}_2$ is an $n \times n$ matrix of constants, and $\underline{\underline{V}}(\boldsymbol{x}_t)^-$ is a diagonal $n \times n$ matrix with elements

$$\left[\underline{\underline{V}}(\boldsymbol{x}_t)^-\right]_{i,i} = \begin{cases} \left(\nu_i + \boldsymbol{v}_i^\top \boldsymbol{x}_t\right)^{-1}, & \text{if } \inf(\nu_i + \boldsymbol{v}_i^\top \boldsymbol{x}_t) > 0, \\ 0, & \text{otherwise.} \end{cases}$$

The square root of the diagonal matrix $\underline{\underline{V}}(\boldsymbol{x}_t)^-$ is simply the diagonal matrix in which each entry is the square root of the corresponding entry in $\underline{\underline{V}}(\boldsymbol{x}_t)^-$. The second term in (8.29) is new compared to the completely affine models. The change of measure to the risk-neutral probability measure is captured by a drift adjustment involving

$$\sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\boldsymbol{\lambda}_t = \underline{\underline{V}}(\boldsymbol{x}_t)\bar{\boldsymbol{\lambda}}_1 + \underline{\underline{I}}^-\underline{\underline{\bar{\lambda}}}_2\boldsymbol{x}_t,$$

where $\underline{\underline{I}}^-$ is a diagonal matrix within which element (i, i) equals 1 if $\inf(\nu_i + \boldsymbol{v}_i^\top \boldsymbol{x}_t) > 0$ and zero otherwise. It can now be verified that the real-world dynamics of the state variables is still affine. The essentially affine model allows for more flexibility in the market prices of risk associated with the state variables that do not affect instantaneous variances and covariances. Hence, the generalization loosens up the tight link between the market price of risk and the volatility structure. Also, some of the individual elements of $\boldsymbol{\lambda}_t$ can now change sign, that is be positive for some values of the state variable vector and negative for others.

Duffee provides an interesting example illustrating some of the strengths of the essentially affine framework. Suppose the real-world dynamics of the short rate r_t follows the Ornstein–Uhlenbeck process assumed in the one-factor Vasicek model, and let f_t denote some other variable so that the real-world dynamics is

$$df_t = \kappa_f\left(\bar{f} - f_t\right)\,dt + \sigma_f\sqrt{f_t}\,dz_{1t},$$
$$dr_t = \kappa_r\left(\bar{r} - r_t\right)\,dt + \sigma_r\,dz_{2t}.$$

In order to comply with the restrictions of the completely affine model, the market price of risk vector $\boldsymbol{\lambda}_t$ associated with (z_1, z_2) must be of the form

$$\boldsymbol{\lambda}_t = \begin{pmatrix}\lambda_{1t}\\ \lambda_{2t}\end{pmatrix} = \begin{pmatrix}\sqrt{f_t} & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\bar{\lambda}_{11}\\ \bar{\lambda}_{12}\end{pmatrix} = \begin{pmatrix}\bar{\lambda}_{11}\sqrt{f_t}\\ \bar{\lambda}_{12}\end{pmatrix}$$

so that, in particular, the market price of risk associated with the short rate shock is constant. With the essentially affine framework we can allow the market price of risk to be of the form

$$\boldsymbol{\lambda}_t = \begin{pmatrix}\bar{\lambda}_{11}\sqrt{f_t}\\ \bar{\lambda}_{12}\end{pmatrix} + \begin{pmatrix}0 & 0\\ 0 & 1\end{pmatrix}\begin{pmatrix}\bar{\lambda}_{2,11} & \bar{\lambda}_{2,12}\\ \bar{\lambda}_{2,21} & \bar{\lambda}_{2,22}\end{pmatrix}\begin{pmatrix}f_t\\ r_t\end{pmatrix} = \begin{pmatrix}\bar{\lambda}_{11}\sqrt{f_t}\\ \bar{\lambda}_{12} + \bar{\lambda}_{2,21}f_t + \bar{\lambda}_{2,22}r_t\end{pmatrix}.$$

Here, the market price of risk associated with the short rate shock can depend on the short rate itself and also on a second variable, f_t. The risk-neutral short rate drift will thus also depend on f_t. Therefore, the variable f_t can affect bond prices and the yield curve, even though it does not affect the real-world dynamics of the short rate.

Cheridito et al. (2007) introduce the **extended affine** framework in which the market price of risk satisfies

$$\sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\boldsymbol{\lambda}_t = \bar{\boldsymbol{\lambda}}_1 + \underline{\underline{\bar{\lambda}}}_2\boldsymbol{x}_t,$$

where the matrix $\underline{\underline{\bar{\lambda}}}_2$ is possibly restricted to ensure that the process $\boldsymbol{x}_t$ is well-defined under both the real-world and the risk-neutral measure. Basically, the extended affine model adds flexibility to the market prices of risk related to the variance-influencing state variables. If the square-root type state variables could take on the value zero, the associated market prices of risk would be undefined, so 'Feller conditions' ensuring strict positive of these state variables must be imposed (see the related discussion for the extended one-factor CIR model in Section 7.6). Then the volatility matrix $\underline{\underline{V}}(\boldsymbol{x}_t)$ stays strictly positive, and the model is arbitrage-free. Still the extended affine models have the questionable property that some market prices of risk are unbounded from above.

For pure Gaussian affine models in the class $\mathbb{A}_0(n)$ for any n, the extended affine model collapses to the essential affine model, and both are more general than the completely affine model. For models in the class $\mathbb{A}_n(n)$, the essentially affine model collapses to the completely affine model, whereas the extended affine model is more general. However, due to the additional parameter restrictions of the extended affine model, it would be incorrect to say that the extended affine framework nests the essential affine framework or even the completely affine framework.

Duarte (2004) adds a constant to the market price of risk specification of the essentially affine models so that

$$\boldsymbol{\lambda}_t = \boldsymbol{\lambda}_0 + \sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\bar{\boldsymbol{\lambda}}_1 + \sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)^-}\,\underline{\underline{\bar{\lambda}}}_2\boldsymbol{x}_t.$$

The drift adjustment associated with the measure change now involves the term $\sqrt{\underline{\underline{V}}(\boldsymbol{x}_t)}\boldsymbol{\lambda}_0$, so that the drift of the state variables cannot be affine under both the real-world and the risk-neutral measure. Assuming that the risk-neutral drift is affine, our general exponential-affine bond pricing formula remains valid. The real-world drift is then non-affine, which slightly complicates empirical work. Such a model is called **semi-affine**.

Feldhütter (2008a) performs a comprehensive empirical investigation of various three-factor models and their ability to explain the dynamics of the U.S. Treasury yield curve over the period 1952–2004. He concludes that extended affine models match the historical risk premia better and the time-varying volatility slightly better than the essentially affine models. The distribution of yields is better matched by extended affine models, whereas the essentially affine models are better in matching the shape of the yield curve. Finally, Feldhütter finds that the semi-affine models have a superior cross-sectional and time-series fit relative to both essentially and extended affine models.

8.7 OTHER MULTI-FACTOR DIFFUSION MODELS

8.7.1 Unspanned stochastic volatility

It has long been recognized that the volatility of interest rates varies over time in a non-deterministic way. This is a key motivation behind the construction of models in which one or more of the state variables affect the variance-covariance structure of the state variables and thus the variance-covariance structure of the short rate and yields of different maturities. In the stochastic volatility models traditionally applied in practice and discussed in this book until this point, the zero-coupon bond pricing function $B^T(\boldsymbol{x}, t)$ is depending non-trivially on all state variables and thus in particular on the volatility-determining state variables. Because the first-order partial derivatives of the zero-coupon bond price with respect to these volatility factors are generally non-zero, it is possible to set up a trading strategy in bonds of different maturities which is completely hedged against volatility shocks, that is the stochastic volatility is spanned by the traded bonds.

However, some recent empirical studies document unspanned stochastic volatility in the sense that a part of the stochastic volatility in the yield curve cannot be hedged away using only bonds. Simple fixed-income derivatives like caps and swaptions, which obviously depend on the volatility of interest rates, cannot be perfectly replicated by trading even a larger number of bonds. Bond markets are incomplete. For example, based on 1995–2000 data from the U.S., the U.K., and Japan, Collin-Dufresne and Goldstein (2002a) find that only a (small) part of the returns on at-the-money straddles can be explained by changes in the underlying swap rates in a regression analysis. An at-the-money straddle is a portfolio consisting of an at-the-money cap and an at-the-money floor. By construction, such a straddle is neutral to small changes in the interest rate level, but very sensitive to changes in volatility. The results thus show that variations in interest rate volatility are only partly due to variations in the level of interest rates. Note that this is model-independent evidence of unspanned stochastic volatility: no model is assumed for the pricing of the caps and floors involved. For further empirical support of unspanned stochastic volatility, see Heidari and Wu (2003), Li and Zhao (2006), Jarrow et al. (2007), and Trolle and Schwartz (2009).

Collin-Dufresne and Goldstein (2002a) show that no two-factor diffusion model can exhibit unspanned stochastic volatility. Consider for example a two-factor diffusion model with the short rate r_t and a volatility factor v_t as the state variables. Bond prices are surely depending on r_t, so they have to be independent of v_t to produce incomplete markets, that is zero-coupon bond prices are of the form $B^T(r_t, t)$. Let $\sigma_r(r_t, v_t)$ be the volatility and $\mu_r(r_t, v_t)$ be the risk-neutral drift of the short rate. Then the bond pricing function must satisfy the partial differential equation

$$\mu_r(r, v)\frac{\partial B^T}{\partial r}(r, t) + \frac{1}{2}\sigma_r(r, v)^2\frac{\partial^2 B^T}{\partial r^2}(r, t) = rB^T(r, t) - \frac{\partial B^T}{\partial t}(r, t), \quad T > t$$

compare (7.2). Since the right-hand side is independent of v, this must also be true for the left-hand side. The derivatives of the bond price must be such that any v-dependence in the risk-neutral drift is balanced by a v-dependence in the variance, which puts a tight restriction on the coefficients of the two terms. However, that restriction cannot hold due to the fact that the ratio of the first-order derivative to the second-order derivative of the bond price (roughly the ratio of the duration to the convexity, see the terminology introduced in Chapter 12) will depend on the maturity T of the bond. The left-hand side will only be independent of v, if both μ_r and σ_r are independent of v, that is we are back to a one-factor complete-market model.

Collin-Dufresne and Goldstein (2002a) also show that the set of three-factor models with unspanned stochastic volatility is very limited. For example, no Gaussian three-factor model can exhibit this feature. The three-factor models of Balduzzi et al. (1996) and Chen (1996) described in Section 8.5 cannot display unspanned stochastic volatility, but the most general models in the model classes $\mathbb{A}_1(3)$, $\mathbb{A}_2(3)$, and $\mathbb{A}_3(3)$ have the necessary flexibility. Both Collin-Dufresne and Goldstein (2002a) and Casassus et al. (2005) suggest and study examples of three-factor models featuring unspanned stochastic volatility.

A particularly simple and illustrative case is the model in which the risk-neutral dynamics is

$$\begin{aligned}
dr_t &= \kappa_r(\theta_t - r_t)\, dt + \sqrt{v_t}\, dz^{\mathbb{Q}}_{1t}, \\
d\theta_t &= \left(\gamma_\theta - 2\kappa_r\theta_t + \kappa_r^{-1}v_t\right)\, dt \\
dv_t &= \mu_v(v_t)\, dt + \sigma_v(v_t)\, dz^{\mathbb{Q}}_{2t},
\end{aligned}$$

where μ_v and σ_v can be any well-behaved functions. The model nests the original one-factor Vasicek model in which v_t and θ_t are constants. Although there are only two exogenous shocks (standard Brownian motions), all three variables (r_t, θ_t, v_t) are necessary to have a Markov diffusion process. Casassus et al. (2005) show that bond prices in this model are of the form

$$B^T_t = e^{-a(T-t) - b_1(T-t)r_t - b_2(T-t)\theta_t},$$

where

$$b_1(\tau) = \frac{1}{\kappa_r}\left(1 - e^{-\kappa_r \tau}\right), \quad b_2(\tau) = \frac{1}{2\kappa_r}\left(1 - e^{-\kappa_r \tau}\right)^2, \quad a(\tau) = \gamma_\theta \int_0^\tau b_2(u)\, du.$$

Obviously, bond prices are independent of the variable v_t, the instantaneous variance of the short rate, so the model exhibits unspanned stochastic volatility.

8.7.2 A model with a short and a long rate

One of the very first two-factor term structure models was suggested by Brennan and Schwartz (1979). They take the short rate r_t and the long rate l_t to be the state variables. The long rate is the yield on a consol (an infinite maturity bond) which provides a continuous payment at a constant rate c. The idea of the model is in line with empirical studies since the short rate can be seen as an indicator for the level of the yield curve, while the difference between the long and the short rate is a measure of the slope of the yield curve. The specific long rate dynamics assumed by Brennan and Schwartz is unacceptable, however. The problem is that the long rate is given by $l_t = c/L_t$, where L_t is the price of the consol, and we know that this price follows from the short rate process and the pricing formula

$$L_t = \mathrm{E}_t^{\mathbb{Q}}\left[\int_t^\infty e^{-\int_t^s r_u\, du} c\, ds\right].$$

The drift and the volatility of L_t, and hence of l_t, are therefore closely related to the short rate r_t. Brennan and Schwartz assume, for example, that the volatility of the long rate is proportional to the long rate and independent of the short rate. For a more detailed discussion of this issue, see Hogan (1993) and Duffie et al. (1995). In addition to the model formulation problems, the Brennan-Schwartz model does not allow closed-form pricing formulas for bonds or derivatives. Although it should be possible to construct a theoretically acceptable model with the short and the long rate as the state variables, no such model has apparently been suggested in the finance literature.

8.7.3 Key rate models

In their analysis of affine multi-factor models Duffie and Kan (1996) focus on models in which the state variables are zero-coupon yields of selected maturities, for example the 1-year, the 5-year, the 10-year, and the 30-year zero-coupon yield. We will refer to the selected interest rates as **key rates**. A clear advantage of such a model is that it is easy to observe (or at least estimate) the state variables from market data, much easier than the short rate volatility, for example. The yield curve obtained in these models automatically matches the market yields for the selected maturities. Other multi-factor models tend to have difficulties in matching the long end of the yield curve, which is problematic for the pricing and hedging of long-term bonds and options on long-term bonds. Many practitioners measure the sensitivity of different securities towards changes in different maturity segments of the yield curve. For that purpose it is clearly convenient to use a model with a direct relation between security prices and representative yields for the different maturity segments.

As shown in (8.11), the zero-coupon yields $\bar{y}_t^\tau = y_t^{t+\tau}$ in a general n-factor affine diffusion model are given by $\bar{y}_t^\tau = \bar{y}^\tau(\boldsymbol{x}_t)$, where

$$\bar{y}^\tau(\boldsymbol{x}) = \frac{a(\tau)}{\tau} + \sum_{j=1}^n \frac{b_j(\tau)}{\tau} x_j.$$

Here, the functions a, $b_1, \dots, b_n$ solve the ordinary differential equations (8.9) and (8.10) with the initial conditions $a(0) = b_j(0) = 0$. If each state variable x_j is the zero-coupon yield for a given time to maturity τ_j, that is,

$$\bar{y}^{\tau_j}(\boldsymbol{x}) = x_j,$$

we must have that

$$b_j(\tau_j) = \tau_j, \qquad a(\tau_j) = b_i(\tau_j) = 0, \quad i \neq j. \tag{8.30}$$

These conditions impose very complicated restrictions on the parameters in the drift and the volatility terms in the dynamics of the state variables, that is the key rates. Explicit expressions for the functions a and $b_1, \dots, b_n$ can only be found in the Gaussian models. In general, the Ricatti equations with the extra conditions (8.30) have to be solved numerically.

An alternative procedure is to start with an affine model with other state variables so that the conditions (8.30) do not have to be imposed in the solution of the Ricatti equations. Subsequently, the variables can be changed to the desired key rates. Since the zero-coupon yields are affine functions of the original state variables, the model with the transformed state variables (that is the key rates) is also an affine model.

8.7.4 Quadratic models

In Section 7.7 we gave a short introduction to quadratic one-factor models, that is models in which the short rate is the square of a state variable which follows an Ornstein–Uhlenbeck process. There are also multi-factor quadratic term structure models. The vector of state variables $\boldsymbol{x}$ follows a multi-dimensional Ornstein–Uhlenbeck process

$$d\boldsymbol{x}_t = (\hat{\boldsymbol{\varphi}} - \underline{\underline{\hat{\kappa}}}\boldsymbol{x}_t)\,dt + \underline{\underline{\Gamma}}\,d\boldsymbol{z}_t^{\mathbb{Q}},$$

and the short rate is a quadratic function of the state variables, that is

$$r_t = \xi + \boldsymbol{\psi}^\top \boldsymbol{x}_t + \boldsymbol{x}_t^\top \underline{\underline{\Theta}}\,\boldsymbol{x}_t = \xi + \sum_{i=1}^n \psi_i x_{it} + \sum_{i=1}^n \sum_{j=1}^n \Theta_{ij} x_{it} x_{jt}.$$

The zero-coupon bond prices are then of the form

$$\begin{aligned}B^T(\boldsymbol{x},t) &= \exp\left\{-a(T-t) - \boldsymbol{b}(T-t)^{\top}\boldsymbol{x} - \boldsymbol{x}^{\top}\underline{\underline{c}}(T-t)\boldsymbol{x}\right\}\\ &= \exp\left\{-a(T-t) - \sum_{i=1}^{n} b_i(T-t)x_i - \sum_{i=1}^{n}\sum_{j=1}^{n} c_{ij}(T-t)x_i x_j\right\},\end{aligned}$$

where the functions a, b_i, and c_{ij} can be found by solving a system of ordinary differential equations. These equations have explicit solutions only in very simple cases, but efficient numerical solution techniques exist. Special cases of this model class have been studied by Beaglehole and Tenney (1992) and Jamshidian (1996), whereas Leippold and Wu (2002) provide a general characterization of the quadratic models.

8.8 FINAL REMARKS

To give a precise description of the evolution of the term structure of interest rates over time, it seems to be necessary to use models with more than one state variable. However, it is more complicated to estimate and apply multi-factor models than one-factor models. Is the additional effort worthwhile? Do multi-factor models generate prices and hedge ratios that are significantly different from those generated by one-factor models? Of course, the answer will depend on the precise results we want from the model.

Buser et al. (1990) compare the prices on selected options on long-term bonds computed with different time-homogeneous models. They conclude that when the model parameters are chosen so that the current short rate, the slope of the yield curve, and some interest rate volatility measure are the same in all the models, the model prices are very close, except when the interest rate volatility is large. However, they only consider specific derivatives and do not compare hedge strategies, only prices.

For a comparison of derivative prices in different models to be fair, the models should produce identical prices of the underlying assets, which in the case of interest rate derivatives are the zero-coupon bonds of all maturities. As we will discuss in more detail in Chapter 9, the models studied so far can be generalized in such a way that the term structure of interest rates produced by the model and the observed term structure match exactly. The model is said to be calibrated to the observed term structure. Basically, one of the model parameters has to be replaced by a carefully chosen time-dependent function, which results in a time-inhomogeneous version of the model. The models can also be calibrated to match prices of derivative securities. Several authors assume a presumably reasonable two-factor model and calibrate a simpler one-factor model to the yield curve of the two-factor model using the extension technique described above. They compare the prices on various derivatives and the efficiency of hedging strategies for the two-factor and the calibrated one-factor model. We will take a closer a look at these studies in Section 9.9. The overall conclusion is that the calibrated one-factor models should be used only for the pricing of securities that resemble the securities to which the model is calibrated. For the pricing of other securities and, in particular, for the construction of hedging strategies it is important to apply multi-factor models that provide a good description of the actual evolution of the term structure of interest rates. And with the well-developed multi-factor models studied in this chapter and the efficient modern computational methods, realistic multi-factor models can easily be used in all practical applications.

Another conclusion of Chapter 9 is that the calibrated factor models should be used with caution. They have some unrealistic properties that may affect the prices of derivative securities. In Chapters 10 and 11 we will consider models that from the outset are developed to match the observed yield curve.

Just as at the end of Chapter 7 we come to the defence of time-homogeneous diffusion models. In practice, the zero-coupon yield curve is not directly observable, but has to be estimated, typically by using observed prices on coupon bonds. Frequently, the estimation procedure is based on a relatively simple parametrization of the discount function as, for example a cubic spline or a Nelson–Siegel parametrization described in Chapter 2. Probably, an equally good fit to the observed market prices and an economically more appropriate yield curve estimate can be obtained by applying the parametrization of the discount function $T \mapsto B_t^T$ that comes from an economically founded model such as those discussed in this chapter. Hence, it may be better to use the time-homogeneous version of the model than to calibrate a time-inhomogeneous version perfectly to an estimate of the current yield curve.

8.9 EXERCISES

Exercise 8.1 ($\mathbb{P}$ and $\mathbb{Q}$ dynamics in the general affine model) Verify (8.6).

Exercise 8.2 (Bond prices in the general affine model) Show Theorem 8.1.

Exercise 8.3 (Solution of ODE in two-factor Vasicek) Show that (8.18) is a solution to (8.17).

Exercise 8.4 (Yield curves in two-factor Vasicek) In the two-factor Vasicek model, the zero-coupon yields are given by (8.19). Develop a spreadsheet in which you can compute yields for a large number of maturities and illustrate the yield curve graphically. What shapes can the yield curve have? How does the shape depend on the current values of the state variables and the parameters of the model? You can use the following benchmark parameters: $\hat{\varphi} = 0.0225$, $\kappa_r = 0.36$, $\beta_r = 0.03$, $\kappa_\varepsilon = 0.1$, $\beta_\varepsilon = 0.02$, and $\rho = 0$. Start with initial values $r = 0.05$ and $\varepsilon = 0$.

Exercise 8.5 (Negative short rate in two-factor Vasicek) In the two-factor Vasicek model, derive an expression for the (risk-neutral) probability that the short rate is negative at time T given that $r_t = r$ and $\varepsilon_t = \varepsilon$ at time $t < T$. Illustrate graphically how this probability depends on T using the parameter values listed in Exercise 8.4. Start with initial values $r = 0.05$ and $\varepsilon = 0$.

Exercise 8.6 (Yield correlations in two-factor Vasicek) In the two-factor Vasicek model, derive an expression for the instantaneous correlation between changes in two different zero-coupon yields, $\bar{y}_t^{\tau_1}$ and $\bar{y}_t^{\tau_2}$. Which parameters affect the yield correlation? Compute a matrix of these probabilities similar to Table 8.1 using the parameter values listed in Exercise 8.4. Investigate the effects of varying the parameter ρ.

Exercise 8.7 (Volatilities in Longstaff-Schwartz) In the Longstaff-Schwartz model, show that the instantaneous variance of the zero-coupon bond price is given by (8.27).

9

Calibration of Diffusion Models

9.1 INTRODUCTION

In Chapters 7 and 8 we have studied diffusion models in which the drift rates, the variance rates, and the covariance rates of the state variables do not explicitly depend on time, but only on the current value of the state variables. Such diffusion processes are called time-homogeneous. The drift rates, variances, and covariances are simple functions of the state variables and a small set of parameters. The derived prices and interest rates are also functions of the state variables and these few parameters. Consequently, the resulting term structure of interest rates will typically not fit the currently observable term structure perfectly. It is generally impossible to find values of a small number of parameters so that the model can perfectly match the infinitely many values that a term structure consists of. This property appears to be inappropriate when the models are to be applied to the pricing of derivative securities. If the model is not able to price the underlying securities (that is the zero-coupon bonds) correctly, why trust the model prices for derivative securities?

In order to be able to fit the observable term structure we need more parameters. This can be obtained by replacing one of the model parameters with a carefully chosen time-dependent function so that the resulting model is time-inhomogeneous. The model is said to be calibrated to the market term structure. The calibrated model is consistent with the observed term structure and is therefore a relative pricing model (or pure no-arbitrage model), see the classification introduced in Section 6.7. The model can also be calibrated to other market information such as the term structure of interest rate volatilities. This requires that an additional parameter is allowed to depend on time.

For time-homogeneous diffusion models the current prices and yields and the distribution of future prices and yields do not depend directly on the calendar date, only on the time to maturity. For example, the zero-coupon yield $y_t^{t+\tau}$ does not depend directly on t, but is determined by the maturity τ and value of the state variables $\boldsymbol{x}_t$. Consequently, if the state variables have the same values at two different points in time, the yield curve will also be the same. Time homogeneity seems to be a reasonable property of a term structure model. When interest rates and prices change over time, it is due to changes in the economic environment (the state variables) rather than the simple passing of time. In contrast, the time-inhomogeneous models discussed in this chapter involve a direct dependence on calendar time. We have to be careful not to introduce unrealistic time dependencies that are likely to affect the prices of the derivative securities we are interested in.

In this chapter we consider the calibration of the one-factor models discussed in Chapter 7. Similar techniques can be applied to the multi-factor models of Chapter 8, but in order to focus on the ideas and keep the notation simple we will consider only one-factor models. The approach taken in this chapter is basically to 'stretch' an equilibrium model by introducing some particular time-dependent functions in the dynamics of the state variable. A more natural approach for obtaining a model that fits the term structure is taken in Chapters 10 and 11, where the dynamics of the entire yield curve is modelled in an arbitrage-free way assuming that the initial yield curve is the one currently observed in the market.

9.2 TIME-INHOMOGENEOUS AFFINE MODELS

Replacing the constants in the time-homogeneous affine model (7.4) by deterministic functions, we get the short rate dynamics

$$dr_t = \left(\hat{\varphi}(t) - \hat{\kappa}(t)r_t\right) dt + \sqrt{\delta_1(t) + \delta_2(t)r_t}\, dz_t^{\mathbb{Q}} \tag{9.1}$$

under the risk-neutral probability measure $\mathbb{Q}$. In this extended version of the model, the distribution of the short rate $r_{t+\tau}$ prevailing τ years from now will depend both on the time horizon τ and the current calendar time t. In the time-homogeneous models the distribution of $r_{t+\tau}$ is independent of t. Despite the extension, we obtain more or less the same pricing results as for the time-homogeneous affine models. Analogously to Theorem 7.1, we have the following characterization of bond prices:

Theorem 9.1 *In the model* (9.1) *the time t price of a zero-coupon bond maturing at T is given as $B_t^T = B^T(r_t, t)$, where*

$$B^T(r,t) = e^{-a(t,T)-b(t,T)r}$$

and the functions $a(t,T)$ and $b(t,T)$ satisfy the following system of differential equations:

$$\frac{1}{2}\delta_2(t)b(t,T)^2 + \hat{\kappa}(t)b(t,T) - \frac{\partial b}{\partial t}(t,T) - 1 = 0, \tag{9.2}$$

$$\frac{\partial a}{\partial t}(t,T) + \hat{\varphi}(t)b(t,T) - \frac{1}{2}\delta_1(t)b(t,T)^2 = 0 \tag{9.3}$$

with the conditions $a(T,T) = b(T,T) = 0$.

The only difference relative to the result for time-homogeneous models is that the functions a and b (and hence the bond price) now depend separately on t and T, not just on the difference $T-t$. The proof is almost identical to the proof of Theorem 7.1 and is therefore omitted. The functions $a(t,T)$ and $b(t,T)$ can be determined from the Equations (9.2) and (9.3) by first solving (9.2) for $b(t,T)$ and then substituting that solution into (9.3), which can then be solved for $a(t,T)$.

It follows immediately from the above theorem that the zero-coupon yields and the forward rates are given by

$$y^T(r,t) = \frac{a(t,T)}{T-t} + \frac{b(t,T)}{T-t} r$$

and

$$f^T(r,t) = \frac{\partial a}{\partial T}(t,T) + \frac{\partial b}{\partial T}(t,T) r.$$

Both expressions are affine in r.

Next, let us look at the term structures of volatilities in these models, that is the volatilities on zero-coupon bond prices $B_t^{t+\tau}$, zero-coupon yields $y_t^{t+\tau}$, and forward rates $f_t^{t+\tau}$ as functions of the time to maturity τ. These volatilities involve the volatility of the short rate $\beta(r,t) = \sqrt{\delta_1(t) + \delta_2(t) r}$ and the function $b(t,T)$. The dynamics of the zero-coupon bond prices is

$$dB_t^{t+\tau} = B_t^{t+\tau} \left[(r_t - \lambda(r_t,t)\beta(r_t,t) b(t,t+\tau))\, dt - b(t,t+\tau)\beta(r_t,t)\, dz_t \right],$$

while the dynamics of zero-coupon yields and forward rates is given by

$$dy_t^{t+\tau} = \ldots\, dt + \frac{b(t,t+\tau)}{\tau} \beta(r_t,t)\, dz_t \tag{9.4}$$

and

$$df_t^{t+\tau} = \ldots\, dt + \left. \frac{\partial b(t,T)}{\partial T} \right|_{T=t+\tau} \beta(r_t,t)\, dz_t,$$

respectively. Focusing on the volatilities, we have omitted the rather complicated drift terms.

From (9.2) we see that if the functions $\delta_2(t)$ and $\hat{\kappa}(t)$ are constant, then we can write $b(t,T)$ as $b(T-t)$ where the function $b(\tau)$ solves the same differential equation as in the time-homogeneous affine models, that is the ordinary differential equation (7.6). If $\delta_1(t)$ is also constant, the short rate volatility $\beta(r_t,t) = \sqrt{\delta_1(t) + \delta_2(t) r_t}$ will be time-homogeneous. Consequently, when $\hat{\kappa}(t)$, $\delta_1(t)$, and $\delta_2(t)$—but not necessarily $\hat{\varphi}(t)$—are constants, the term structures of volatilities of the model are time-homogeneous in the sense that the volatilities of $B_t^{t+\tau}$, $y_t^{t+\tau}$, and $f_t^{t+\tau}$ depend only on τ and the current short rate, not on t. Due to the time-inhomogeneity, the future volatility structure can be very different from the current volatility structure, even for a similar yield curve. This property is inappropriate and not realistic. Furthermore, the prices of many derivative securities are highly dependent on the evolution of volatilities, see, for example Carverhill (1995) and Hull and White (1995). A model with unreasonable volatility structures will probably produce unreasonable prices and hedge strategies.

For these reasons it is typically only the parameter $\hat{\varphi}$ that is allowed to depend on time. Below we will discuss such extensions of the models of Merton, of Vasicek, and of Cox, Ingersoll, and Ross. For a particular choice of the function $\hat{\varphi}(t)$ these extended models are able to match the observed yield curve exactly, that is the models are calibrated to the market yield curve.

Note that if only $\hat{\varphi}$ depends on time, the function $b(\tau)$ is just as in the original time-homogeneous version of the model, whereas the a function will be different. Since

$$a(T,T) - a(t,T) = \int_t^T \frac{\partial a}{\partial u}(u,T)\,du$$

and $a(T,T) = 0$, Equation (9.3) implies that

$$a(t,T) = \int_t^T \hat{\varphi}(u)b(T-u)\,du - \frac{\delta_1}{2}\int_t^T b(T-u)^2\,du.$$

In particular,

$$a(0,T) = \int_0^T \hat{\varphi}(t)b(T-t)\,dt - \frac{\delta_1}{2}\int_0^T b(T-t)^2\,dt. \tag{9.5}$$

We want to pick the function $\hat{\varphi}(t)$ so that the current (time 0) model prices on zero-coupon bonds, $B^T(r_0,0)$, are identical to the observed prices, $\bar{B}(T)$, that is

$$a(0,T) = -b(T)r_0 - \ln \bar{B}(T) \tag{9.6}$$

for any time-to-maturity T. We can then determine $\hat{\varphi}(t)$ by comparing (9.5) and (9.6). In the extensions of the models of Merton and Vasicek we are able to find an explicit expression for $\hat{\varphi}(t)$, whereas numerical methods must be applied in the extension of the CIR model.

9.3 THE HO–LEE MODEL (EXTENDED MERTON)

Ho and Lee (1986) developed a recombining binomial model for the evolution of the entire yield curve taking the currently observed yield curve as given. Subsequently, Dybvig (1988) has demonstrated that the continuous time limit of their binomial model is a model with

$$dr_t = \hat{\varphi}(t)\,dt + \beta\,dz_t^{\mathbb{Q}}, \tag{9.7}$$

which extends Merton's model described in Section 7.3. The prices of zero-coupon bonds are of the form

$$B^T(r,t) = e^{-a(t,T)-b(T-t)r},$$

where $b(\tau) = \tau$ just as in Merton's model, and

$$a(t,T) = \int_t^T \hat{\varphi}(u)(T-u)\,du - \frac{1}{2}\beta^2\int_t^T (T-u)^2\,du, \tag{9.8}$$

see the discussion in the preceding section. The following theorem shows how to choose the function $\hat{\varphi}(t)$ in order to match any given initial yield curve.

Theorem 9.2 *Let $t \mapsto \bar{f}(t)$ be the current term structure of forward rates and assume that this function is differentiable. Then the term structure of interest rates in the Ho–Lee model* (9.7) *with*

$$\hat{\varphi}(t) = \bar{f}^{\,\prime}(t) + \beta^2 t \tag{9.9}$$

for all t will be identical to the current term structure. In this case we have

$$a(t, T) = -\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right) - (T - t)\bar{f}(t) + \frac{1}{2}\beta^2 t(T - t)^2,$$

where $\bar{B}(t) = \exp\{-\int_0^t \bar{f}(s)\, ds\}$ denotes the current zero-coupon bond prices.

Proof: Substituting $b(T - t) = T - t$ and $\delta_1 = \beta^2$ into (9.5), we obtain

$$a(0, T) = \int_0^T \hat{\varphi}(t)(T - t)\, dt - \frac{1}{2}\beta^2 \int_0^T (T - t)^2\, dt = \int_0^T \hat{\varphi}(t)(T - t)\, dt - \frac{1}{6}\beta^2 T^3.$$

Computing the derivative with respect to T, using Leibnitz' rule as described in Section 3.6.2, we obtain

$$\frac{\partial a}{\partial T}(0, T) = \int_0^T \hat{\varphi}(t)\, dt - \frac{1}{2}\beta^2 T^2,$$

and another differentiation gives

$$\frac{\partial^2 a}{\partial T^2}(0, T) = \hat{\varphi}(T) - \beta^2 T. \tag{9.10}$$

We wish to satisfy the relation (9.6), that is

$$a(0, T) = -Tr_0 - \ln \bar{B}(T).$$

Recall from (1.9) the following relation between the discount function $\bar{B}$ and the term structure of forward rates $\bar{f}$:

$$-\frac{\partial \ln \bar{B}(T)}{\partial T} = -\frac{\bar{B}'(T)}{\bar{B}(T)} = \bar{f}(T).$$

Hence, $a(0, T)$ must satisfy that

$$\frac{\partial a}{\partial T}(0, T) = -r_0 + \bar{f}(T)$$

and therefore

$$\frac{\partial^2 a}{\partial T^2}(0, T) = \bar{f}^{\,\prime}(T), \tag{9.11}$$

where we have assumed that the term structure of forward rates is differentiable. Comparing (9.10) and (9.11), we obtain the stated result.

Substituting (9.9) into (9.8), we get

$$a(t,T) = \int_t^T \bar{f}'(u)(T-u)\,du + \beta^2 \int_t^T u(T-u)\,du - \frac{1}{2}\beta^2 \int_t^T (T-u)^2\,du.$$

Partial integration gives

$$\int_t^T \bar{f}'(u)(T-u)du = -(T-t)\bar{f}(t) + \int_t^T \bar{f}(u)\,du = -(T-t)\bar{f}(t) - \ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right),$$

where we have used the relation between forward rates and zero-coupon bond prices to conclude that

$$\int_t^T \bar{f}(u)\,du = \int_0^T \bar{f}(u)\,du - \int_0^t \bar{f}(u)\,du = -\ln \bar{B}(T) + \ln \bar{B}(t) = -\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right). \tag{9.12}$$

Furthermore, tedious calculations yield that

$$\beta^2 \int_t^T u(T-u)\,du - \frac{1}{2}\beta^2 \int_t^T (T-u)^2\,du = \frac{1}{2}\beta^2 t(T-t)^2.$$

Now, $a(t,T)$ can be written as stated in the Theorem. □

In the Ho–Lee model the short rate follows a generalized Brownian motion (with a time-dependent drift). From the analysis in Chapter 3 we get that the future short rate is normally distributed, that is the Ho-Lee model is a Gaussian model. The pricing of European options is similar to the original Merton model. The price of a call option on a zero-coupon bond is given by (7.31) with the same expression for the variance $v(t,T,S)^2$. As usual, the price of a call option on a coupon bond follows from Jamshidian's trick.

9.4 THE HULL–WHITE MODEL (EXTENDED VASICEK)

When we replace the parameter $\hat{\theta}$ in Vasicek's model (7.35) by a time-dependent function $\hat{\theta}(t)$, we get the following short rate dynamics under the risk-neutral (spot martingale) measure:

$$dr_t = \kappa\left[\hat{\theta}(t) - r_t\right]dt + \beta\,dz_t^{\mathbb{Q}}. \tag{9.13}$$

This model was introduced by Hull and White (1990a) and is called the Hull–White model or the extended Vasicek model. As in the original Vasicek model, the process has a constant volatility $\beta > 0$ and exhibits mean reversion with a constant speed of adjustment $\kappa > 0$, but in the extended version the long-term level is time-dependent. The risk-adjusted process (9.13) may be the result of a real-world dynamics of

$$dr_t = \kappa\left[\theta(t) - r_t\right]dt + \beta\,dz_t,$$

and an assumption that the market price of risk depends at most on time, $\lambda(t)$. In that case we will have

$$\hat{\theta}(t) = \theta(t) - \frac{\beta}{\kappa}\lambda(t).$$

Despite the small extension, it follows from the discussion in Section 3.8.2 that the model remains Gaussian. To be more precise, the future short rate r_T is normally distributed with the same variance as in the original Vasicek model,

$$\mathrm{Var}_{r,t}[r_T] = \mathrm{Var}^{\mathbb{Q}}_{r,t}[r_T] = \beta^2 \int_t^T e^{-2\kappa[T-u]}\, du = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa[T-t]}\right),$$

but a different mean, namely

$$\mathrm{E}^{\mathbb{Q}}_{r,t}[r_T] = e^{-\kappa[T-t]}r + \kappa \int_t^T e^{-\kappa[T-u]}\hat{\theta}(u)\, du$$

under the risk-neutral measure and

$$\mathrm{E}_{r,t}[r_T] = e^{-\kappa[T-t]}r + \kappa \int_t^T e^{-\kappa[T-u]}\theta(u)\, du$$

under the real-world probability measure.

According to Theorem 9.1 and the subsequent discussion, the zero-coupon bond prices in the Hull–White model are given by

$$B^T(r,t) = e^{-a(t,T)-b(T-t)r},$$

where

$$\begin{aligned} b(\tau) &= \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right), \\ a(t,T) &= \kappa \int_t^T \hat{\theta}(u)b(T-u)\, du + \frac{\beta^2}{4\kappa}b(T-t)^2 + \frac{\beta^2}{2\kappa^2}\left(b(T-t) - (T-t)\right). \end{aligned} \tag{9.14}$$

This expression holds for any given function $\hat{\theta}$. Now assume that at time 0 we observe the current short rate r_0 and the entire discount function $T \mapsto \bar{B}(T)$ or, equivalently, the term structure of forward rates $T \mapsto \bar{f}(T)$. The following result shows how to choose the function $\hat{\theta}$ so that the model discount function matches the observed discount function exactly.

Theorem 9.3 *Let $t \mapsto \bar{f}(t)$ be the current (time* 0*) term structure of forward rates and assume that this function is differentiable. Then the term structure of interest rates in the Hull–White model* (9.13) *with*

$$\hat{\theta}(t) = \bar{f}(t) + \frac{1}{\kappa}\bar{f}'(t) + \frac{\beta^2}{2\kappa^2}\left(1 - e^{-2\kappa t}\right) \tag{9.15}$$

will be identical to the current term structure of interest rates. In this case we have

$$a(t,T) = -\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right) - b(T-t)\bar{f}(t) + \frac{\beta^2}{4\kappa}b(T-t)^2\left(1-e^{-2\kappa t}\right). \tag{9.16}$$

Proof: From Equation (9.14) it follows that

$$a(0,t) = \kappa\int_0^t \hat{\theta}(u)b(t-u)\,du + \frac{\beta^2}{4\kappa}b(t)^2 + \frac{\beta^2}{2\kappa^2}\left(b(t)-t\right).$$

Repeated differentiations yield

$$\frac{\partial a}{\partial t}(0,t) = \kappa\int_0^t \hat{\theta}(u)e^{-\kappa[t-u]}\,du - \frac{1}{2}\beta^2 b(t)^2$$

and

$$\begin{aligned}\frac{\partial^2 a}{\partial t^2}(0,t) &= \kappa\hat{\theta}(t) - \kappa^2\int_0^t \hat{\theta}(u)e^{-\kappa[t-u]}\,du - \beta^2 b(t)e^{-\kappa t}\\ &= \kappa\hat{\theta}(t) - \kappa\frac{\partial a}{\partial t}(0,t) - \frac{\beta^2}{2\kappa}\left(1-e^{-2\kappa t}\right),\end{aligned}$$

where we have applied Leibnitz' rule (see Section 3.6.2). Consequently,

$$\hat{\theta}(t) = \frac{\beta^2}{2\kappa^2}\left(1-e^{-2\kappa t}\right) + \frac{1}{\kappa}\frac{\partial^2 a}{\partial t^2}(0,t) + \frac{\partial a}{\partial t}(0,t). \tag{9.17}$$

Differentiation of the expression (9.6), which we want to be satisfied, yields

$$\frac{\partial a}{\partial t}(0,t) = -\frac{\bar{B}'(t)}{\bar{B}(t)} - r_0 e^{-\kappa t} = \bar{f}(t) - r_0 e^{-\kappa t}$$

and

$$\frac{\partial^2 a}{\partial t^2}(0,t) = \bar{f}'(t) + \kappa r_0 e^{-\kappa t}.$$

Substituting these expressions into (9.17), we obtain (9.15).

Substituting (9.17) into (9.14), we get

$$\begin{aligned}a(t,T) &= \int_t^T \bar{f}(u)\left(1-e^{-\kappa[T-u]}\right)du + \frac{1}{\kappa}\int_t^T \bar{f}'(u)\left(1-e^{-\kappa[T-u]}\right)du\\ &\quad + \frac{\beta^2}{2\kappa^2}\int_t^T\left(1-e^{-2\kappa u}\right)\left(1-e^{-\kappa[T-u]}\right)du + \frac{\beta^2}{4\kappa}b(T-t)^2\\ &\quad + \frac{\beta^2}{2\kappa^2}\left(b(T-t)-(T-t)\right).\end{aligned}$$

Note that

$$\int_t^T \bar{f}'(u)\,du = \bar{f}(T) - \bar{f}(t).$$

Partial integration yields

$$\frac{1}{\kappa}\int_t^T \bar{f}'(u)e^{-\kappa[T-u]}\,du = \frac{1}{\kappa}\bar{f}(T) - \frac{1}{\kappa}\bar{f}(t)e^{-\kappa[T-t]} - \int_t^T \bar{f}(u)e^{-\kappa[T-u]}\,du.$$

From (9.12) we get that

$$\begin{aligned} a(t,T) = &-\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right) - \bar{f}(t)b(T-t) + \frac{\beta^2}{2\kappa^2}\int_t^T \left(1-e^{-2\kappa u}\right)\left(1-e^{-\kappa[T-u]}\right) du \\ &+ \frac{\beta^2}{4\kappa}b(T-t)^2 + \frac{\beta^2}{2\kappa^2}\left(b(T-t)-(T-t)\right). \end{aligned}$$

After some straightforward, but tedious, manipulations we arrive at the desired relation (9.16). □

Due to the fact that the Hull–White model is Gaussian, the prices of European call options on zero-coupon bonds can be derived just as in the Vasicek model. Since the b function and the variance of the future short rate are the same in the Hull–White model as in Vasicek's model, we obtain exactly the same option pricing formula, that is

$$C^{K,T,S}(r,t) = B^S(r,t)N(d_1) - KB^T(r,t)N(d_2), \tag{9.18}$$

where

$$\begin{aligned} d_1 &= \frac{1}{v(t,T,S)}\ln\left(\frac{B^S(r,t)}{KB^T(r,t)}\right) + \frac{1}{2}v(t,T,S), \\ d_2 &= d_1 - v(t,T,S), \\ v(t,T,S) &= \frac{\beta}{\sqrt{2\kappa^3}}\left(1-e^{-\kappa[S-T]}\right)\left(1-e^{-2\kappa[T-t]}\right)^{1/2}. \end{aligned}$$

The only difference to the Vasicek case is that the Hull–White model justifies the use of *observed* bond prices in this formula. Since the zero-coupon bond price is a decreasing function of the short rate, we can apply Jamshidian's trick stated in Theorem 7.3 for the pricing of European options on coupon bonds in terms of a portfolio of European options on zero-coupon bonds.

9.5 THE EXTENDED CIR MODEL

Extending the CIR model analysed in Section 7.5 in the same way as we extended the models of Merton and Vasicek, the short rate dynamics becomes[1]

$$dr_t = \left(\kappa\theta(t) - \hat{\kappa} r_t\right) dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}}.$$

For the process to be well-defined $\theta(t)$ has to be non-negative. This will ensure a non-negative drift when the short rate is zero so that the short rate stays non-negative and the square-root term makes sense. To ensure strictly positive interest rates we must further require that $2\kappa\theta(t) \geq \beta^2$ for all t.

For an arbitrary non-negative function $\theta(t)$ the zero-coupon bond prices are

$$B^T(r,t) = e^{-a(t,T)-b(T-t)r},$$

where $b(\tau)$ is exactly as in the original CIR model, see (7.49), while the function a is now given by

$$a(t,T) = \kappa \int_t^T \theta(u) b(T-u)\, du.$$

Suppose that the current discount function is $\bar{B}(T)$ with the associated term structure of forward rates given by $\bar{f}(T) = -\bar{B}'(T)/\bar{B}(T)$. To obtain $\bar{B}(T) = B^T(r_0, 0)$ for all T, we have to choose $\theta(t)$ so that

$$a(0,T) = -\ln \bar{B}(T) - b(T)r_0 = \kappa \int_0^T \theta(u) b(T-u)\, du, \quad T > 0.$$

Differentiating with respect to T, we get

$$\bar{f}(T) = b'(T)r_0 + \kappa \int_0^T \theta(u) b'(T-u)\, du, \quad T > 0.$$

According to Heath et al. (1992, p. 96) it can be shown that this equation has a unique solution $\theta(t)$, but it cannot be written in an explicit form so a numerical procedure must be applied. We cannot be sure that the solution complies with the condition that guarantees a well-defined short rate process. Clearly, a necessary condition for $\theta(t)$ to be non-negative for all t is that

$$\bar{f}(T) \geq r_0 b'(T), \quad T > 0. \tag{9.19}$$

Not all forward rate curves satisfy this condition, see Exercise 9.1. Consequently, in contrast to the Merton and the Vasicek models, the CIR model cannot be calibrated to any given term structure.

No explicit option pricing formulas have been found in the extended CIR model. Option prices can be computed by numerically solving the partial differential equation associated with the model, for example by using the techniques outlined in Chapter 16.

[1] This extension was suggested already in the original article by Cox et al. (1985b).

9.6 CALIBRATION TO OTHER MARKET DATA

Many practitioners want a model to be consistent with basically all 'reliable' current market data. The objective may be to calibrate a model to the prices of liquid bonds and derivative securities, for example caps, floors, and swaptions, and then apply the model for the pricing of less liquid 'exotic' derivatives. In this manner the less liquid securities are priced in a way which is consistent with the indisputable observed prices. Above we discussed how an equilibrium model can be calibrated to the current yield curve (that is current bond prices) by replacing the constant in the drift term with a time-dependent function. If we replace other constant parameters by carefully chosen deterministic functions, we can calibrate the model to further market information.

Let us take the Vasicek model as an example. If we allow both $\hat{\theta}$ and κ to depend on time, the short rate dynamics becomes

$$\begin{aligned} dr_t &= \kappa(t)\left[\hat{\theta}(t) - r_t\right] dt + \beta\, dz_t^{\mathbb{Q}} \\ &= \left[\hat{\varphi}(t) - \kappa(t) r_t\right] dt + \beta\, dz_t^{\mathbb{Q}}. \end{aligned}$$

The price of a zero-coupon bond is still given by Theorem 9.1 as $B^T(r,t) = \exp\{-a(t,T) - b(t,T)r\}$. According to Equations (15) and (16) in Hull and White (1990a), the functions $\kappa(t)$ and $\hat{\varphi}(t)$ are

$$\begin{aligned} \kappa(t) &= -\frac{\partial^2 b}{\partial t^2}(0,t) \Big/ \frac{\partial b}{\partial t}(0,t), \\ \hat{\varphi}(t) &= \kappa(t)\frac{\partial a}{\partial t}(0,t) + \frac{\partial^2 a}{\partial t^2}(0,t) - \left(\frac{\partial b}{\partial t}(0,t)\right)^2 \int_0^t \beta^2 \left(\frac{\partial b}{\partial u}(0,u)\right)^{-2} du, \end{aligned}$$

and can hence be determined from the functions $t \mapsto a(0,t)$ and $t \mapsto b(0,t)$ and their derivatives. From (9.4) we get that the model volatility of the zero-coupon yield $y_t^{t+\tau} = y^{t+\tau}(r_t, t)$ is

$$\sigma_y^{t+\tau}(t) = \frac{\beta}{\tau} b(t, t+\tau).$$

In particular, the time 0 volatility is $\sigma_y^{\tau}(0) = \beta b(0,\tau)/\tau$. If the current term structure of zero-coupon yield volatilities is represented by the function $t \mapsto \bar{\sigma}_y(t)$, we can obtain a perfect match of these volatilities by choosing

$$b(0,t) = \frac{\tau}{\beta}\bar{\sigma}_y(t).$$

The function $t \mapsto a(0,t)$ can then be determined from $b(0,t)$ and the current discount function $t \mapsto \bar{B}(t)$ as described in the previous sections. Note that the term structure of volatilities can be estimated either from historical fluctuations of the yield curve or as 'implied volatilities' derived from current prices of derivative securities. Typically the latter approach is based on observed prices of caps.

Finally, we can also let the short rate volatility be a deterministic function $\beta(t)$ so that we get the 'fully extended' Vasicek model

$$dr_t = \kappa(t)[\hat{\theta}(t) - r_t]\,dt + \beta(t)\,dz_t^{\mathbb{Q}}.$$

Choosing $\beta(t)$ in a specific way, we can calibrate the model to further market data.

Despite all these extensions, the model remains Gaussian so that the option pricing formula (9.18) still applies. However, the relevant volatility is now $v(t, T, S)$, where

$$\begin{aligned} v(t,T,S)^2 &= \int_t^T \beta(u)^2\,[b(u,S) - b(u,T)]^2\,du \\ &= [b(0,S) - b(0,T)]^2 \int_t^T \beta(u)^2 \left(\frac{\partial b}{\partial u}(0,u)\right)^{-2} du, \end{aligned}$$

see Hull and White (1990a). Jamshidian's result (7.28) for European options on coupon bonds is still valid if the estimated $b(t, T)$ function is positive.

If either κ or β (or both) are time-dependent, the volatility structure in the model becomes time-inhomogeneous, that is dependent on the calendar time, see the discussion in Section 9.2. Since the volatility structure in the market seems to be pretty stable (when interest rates are stable), this dependence on calendar time is inappropriate. Broadly speaking, to let κ or β depend on time is 'stretching the model too much'. It should not come as a surprise that it is hard to find a reasonable and very simple model which is consistent with both yield curves and volatility curves.

If only the parameter θ is allowed to depend on time, the volatility structure of the model is time-homogeneous. The drift rates of the short rate, the zero-coupon yields, and the forward rates are still time-inhomogeneous, which is certainly also unrealistic. The drift rates may change over time, but only because key economic variables change, not just because of the passage of time. However, Hull and White and other authors argue that time-inhomogeneous drift rates are less critical for option prices than time-inhomogeneous volatility structures. See also the discussion in Section 9.9 below.

9.7 INITIAL AND FUTURE TERM STRUCTURES IN CALIBRATED MODELS

In the preceding section we have implicitly assumed that the current term structure of interest rates is directly observable. In practice, the term structure of interest rates is often estimated from the prices of a finite number of liquid bonds. As discussed in Chapter 2, this is typically done by expressing the discount function or the forward rate curve as some given function with relatively few parameters. The values of these parameters are chosen to match the observed prices as closely as possible.

A cubic spline estimation of the discount function will frequently produce unrealistic estimates for the forward rate curve and, in particular, for the slope of the forward rate curve. This is problematic since the calibration of the equilibrium models depends on the forward rate curve and its slope as can be seen from the earlier sections of this chapter. In contrast, the Nelson–Siegel parametrization

$$\bar{f}(t) = c_1 + c_2 e^{-kt} + c_3 t e^{-kt}, \tag{9.20}$$

see (2.13), ensures a nice and smooth forward rate curve and will presumably be more suitable in the calibration procedure.

No matter which of these parametrizations is used, it will not be possible to match all the observed bond prices perfectly. Hence, it is not strictly correct to say that the calibration procedure provides a perfect match between model prices and market prices of the bonds. See also Exercise 9.11.

Recall that the cubic spline and the Nelson–Siegel parametrizations are not based on any economic arguments, but are simply 'curve fitting' techniques. The theoretically better-founded dynamic equilibrium models of Chapters 7 and 8 also result in a parametrization of the discount function, for example (7.46) and the associated expressions for a and b in the Cox–Ingersoll–Ross model. Why not use such a parametrization instead of the cubic spline or the Nelson–Siegel parametrization? And if the parametrization generated by an equilibrium model is used, why not use that equilibrium model for the pricing of fixed income securities rather than calibrating a different model to the chosen parameterized form? In conclusion, the objective must be to use an equilibrium model that produces yield curve shapes and yield curve movements that resemble those observed in the market. If such a model is too complex, one can calibrate a simpler model to the yield curve estimate stemming from the complex model and hope that the calibrated simpler model provides prices and hedge ratios which are reasonably close to those in the complex model.

A related question is what shapes the future yield curve may have, given the chosen parametrization of the current yield curve and the model dynamics of interest rates. For example, if we use a Nelson–Siegel parametrization (9.20) of the current yield curve and let this yield curve evolve according to a dynamic model, such as the Hull–White model, will the future yield curves also be of the form (9.20)? Intuitively, it seems reasonable to use a parametrization which is consistent with the model dynamics, in the sense that the possible future yield curves can be written on the same parameterized form, although possibly with other parameter values.

Which parametrizations are consistent with a given dynamic model? This question was studied by Björk and Christensen (1999) using advanced mathematics, so let us just list some of their conclusions:

- The simple affine parametrization $\bar{f}(t) = c_1 + c_2 t$ is consistent with the Ho-Lee model (9.7), that is if the initial forward rate curve is a straight line, then the future forward rate curves in the model are also straight lines.
- The simplest parametrization of the forward rate curve, which is consistent with the Hull–White model (9.13), is

$$\bar{f}(t) = c_1 e^{-kt} + c_2 e^{-2kt}.$$

- The Nelson-Siegel parametrization (9.20) is consistent with neither the Ho–Lee model nor the Hull–White model. However, the extended Nelson–Siegel parametrization

$$\bar{f}(t) = c_1 + c_2 e^{-kt} + c_3 t e^{-kt} + c_4 e^{-2kt}$$

is consistent with the Hull–White model.

Furthermore, it can be shown that the Nelson–Siegel parametrization is not consistent with any non-trivial one-factor diffusion model, see Filipović (1999).

9.8 CALIBRATED NON-AFFINE MODELS

In Section 7.7 we looked at some non-affine one-factor models with constant parameters. These models can also be calibrated to market data by replacing the constant parameters by time-dependent functions. The Black–Karasinksi model (7.54) can thus be extended to

$$d(\ln r_t) = \kappa(t)(\theta(t) - \ln r_t)\, dt + \beta(t)\, dz_t^{\mathbb{Q}},$$

where κ, θ, and β are deterministic functions of time. Despite the generalization, the future values of the short rate remain lognormally distributed. Black and Karasinski implement their model in a binomial tree and choose the functions κ, θ, and β so that the yields and the yield volatilities computed with the tree exactly match those observed in the market. There are no explicit pricing formulas, and the construction of the calibrated binomial tree is quite complicated.

The BDT model introduced by Black et al. (1990) is the special case of the Black–Karasinski model where $\beta(t)$ is a differentiable function and

$$\kappa(t) = \frac{\beta'(t)}{\beta(t)}.$$

Still, no explicit pricing formulas have been found, and also this model is typically implemented in a binomial tree.[2] To avoid the difficulties arising from time-dependent volatilities, $\beta(t)$ has to be constant. In that case, $\kappa(t) = 0$ in the BDT model, and the model is reduced to the simple model

$$dr_t = \frac{1}{2}\beta^2 r_t\, dt + \beta r_t\, dz_t^{\mathbb{Q}},$$

which is a special case of the Rendleman–Bartter model (7.55) and cannot be calibrated to the observed yield curve.

[2] It is not clear from Black et al. (1990) how a calibrated tree can be constructed, but Jamshidian (1991) fills this gap in their presentation.

Theorem 7.6 showed that the time-homogeneous lognormal models produce completely wrong Eurodollar-futures prices. The time-inhomogeneous versions of the lognormal models exhibit the same unpleasant property.

9.9 IS A CALIBRATED ONE-FACTOR MODEL JUST AS GOOD AS A MULTI-FACTOR MODEL?

In the opening section of Chapter 8 we argued that more than one factor is needed in order to give a reasonable description of the evolution of the term structure of interest rates. However, multi-factor models are harder to estimate and apply than one-factor models. If there are no significant differences in the prices and hedge ratios obtained in a multi-factor and a one-factor model, it will be computationally convenient to use the one-factor model. But will a simple one-factor model provide the same prices and hedge ratios as a more realistic multi-factor model? We initiated the discussion of this issue in Section 8.8, where we focused on the time-homogeneous models. Intuitively, the prices of derivative securities in two different models should be closer when the two models produce identical prices to the underlying assets. Several authors compare a time-homogeneous two-factor model to a time-inhomogeneous one-factor model which has been perfectly calibrated to the yield curve generated by the two-factor model.

Hull and White (1990a) compare prices of selected derivative securities in different models that have been calibrated to the same initial yield curve. They first assume that the time-homogeneous CIR model (with certain parameter values) provides a correct description of the term structure, and they compute prices of European call options on a 5-year bullet bond and of various caps, both with the original CIR model and the extended Vasicek model calibrated to the CIR yield curve. They find that the prices in the two models are generally very close, but that the percentage deviation for out-of-the-money options and caps can be considerable. Next, they compare prices of European call options on a 5-year zero-coupon bond in the extended Vasicek model to prices computed using two different two-factor models, namely a two-factor Gaussian-model and a two-factor CIR model. In each of the comparisons the two-factor model is assumed to provide the true yield curve, and the extended Vasicek model is calibrated to the yield curve of the two-factor model. The price differences are very small. Hence, Hull and White conclude that although the true dynamics of the yield curve is consistent with a complex one-factor (CIR) or a two-factor model, one might as well use the simple extended Vasicek model calibrated to the true yield curve.

Hull and White consider only a few different derivative securities and only two relatively simple multi-factor models, and they compare only prices, not hedge strategies. Canabarro (1995) performs a more adequate comparison. First, he argues that the two-factor models used in the comparison of Hull and White are degenerate and describe the actual evolution of the yield curve very badly. For example, he shows that, in the two-factor CIR model they use, one of the factors (with the parameter values used by Hull and White) will explain more than 99% of the total variation in the yield curve and hence the second factor explains less than 1%. As discussed in Section 8.1, he finds empirically that the most important factor can explain only 85% and the second-most important factor more than 10% of the variation in the yield curve. Moreover, Hull and White's two-factor CIR model gives unrealistically high correlations between zero-coupon yields of different maturities. For example, the correlation between the 3-month and the 30-year par yields is as high as 0.96 in that model, which is far from the empirical estimate of 0.46, see Table 8.1. Therefore, a comparison to this two-factor model will provide very little information on whether it is reasonable or not to use a simple calibrated one-factor model to represent the complex real-world dynamics. In his comparisons Canabarro also uses a different two-factor model, namely the model of Brennan and Schwartz (1979), which was briefly described at the end of Section 8.7. Despite the theoretical deficiencies of the Brennan–Schwartz model, Canabarro shows that the model provides reasonable values for the correlations and the explanatory power of each of the factors.

Each of the two two-factor models is compared to two calibrated one-factor models. The first is the extended one-factor CIR model

$$dr_t = \left(\hat{\varphi}(t) - \hat{\kappa}(t) r_t\right) dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}},$$

and the second is the BDT model

$$d(\ln r_t) = \frac{\beta'(t)}{\beta(t)} \left(\theta(t) - \ln r_t\right) dt + \beta(t)\, dz_t^{\mathbb{Q}}.$$

The time-dependent functions in these models are chosen so that the models produce the same initial yield curve and the same prices of caps with a given cap-rate, but different maturities, as the two-factor model which is assumed to be the 'true' model. Note that both these calibrated one-factor models exhibit time-dependent volatility structures, which in general should be avoided.

For both of the two benchmark two-factor models, Canabarro finds that using a calibrated one-factor model instead of the correct two-factor model results in price errors that are very small for relatively simple securities such as caps and European options on bonds. For so-called *yield curve options* that have payoffs given by the difference between a short-term and a long-term zero-coupon yield, the errors are much larger and non-negligible. These findings are not surprising since the one-factor models do not allow for twists in the yield curve, that is yield curve movements where the short end and the long end move in opposite directions. It is exactly those movements that make yield curve options valuable. Regarding the efficiency of hedging strategies, the calibrated one-factor models perform very badly. This is true even for the hedging of simple securities that resemble the securities used in the calibration of the models. Both pricing errors and hedging errors are typically larger for the BDT model than for the calibrated one-factor CIR model. In general, the errors are larger when the one-factor models have been calibrated to the more realistic Brennan–Schwartz model than to the rather degenerate two-factor CIR model used in the comparison of Hull and White.

The conclusion to be drawn from these studies is that the calibrated one-factor models should be used only for pricing securities that are closely related to the securities used in the calibration of the model. For the pricing of other securities and, in particular, for the design of hedging strategies it is important to apply multi-factor models that give a good description of the actual movements of the yield curve.

9.10 FINAL REMARKS

This chapter has shown how one-factor equilibrium models can be perfectly fitted to the observed yield curve by replacing constant parameters by certain time-dependent functions. However, we have also argued that this calibration approach has inappropriate consequences and should be used only with great caution.

Similar procedures apply to multi-factor models. Since multi-factor models typically involve more parameters than one-factor models, they can give a closer fit to any given yield curve without introducing time-dependent functions. In this sense the gain from a perfect calibration is less for multi-factor models. In the following two chapters we will look at a more direct way to construct models that are consistent with the observable yield curve.

9.11 EXERCISES

Exercise 9.1 (Calibration of the CIR model) Compute $b'(\tau)$ in the CIR model by differentiation of (7.49). Find out which types of initial forward rate curves the CIR model can be calibrated to, by computing (using a spreadsheet for example) the right-hand side of (9.19) for reasonable values of the parameters and the initial short rate. Vary the parameters and the short rate and discuss the effects.

Exercise 9.2 (The Hull–White model calibrated to the Vasicek yield curve) Suppose the observable bond prices are fitted to a discount function of the form

$$(^*)\bar{B}(t) = e^{-a(t)-b(t)r_0},$$

where

$$b(t) = \frac{1}{\kappa}\left(1 - e^{-\kappa t}\right),$$
$$a(t) = y_\infty[t - b(t)] + \frac{\beta^2}{4\kappa}b(t)^2,$$

where y_∞, κ, and β are constants. This is the discount function of the Vasicek model, see (7.36)–(7.38).

(a) Express the initial forward rates $\bar{f}(t)$ and the derivatives $\bar{f}'(t)$ in terms of the functions a and b.
(b) Show by substitution into (9.15) that the function $\hat{\theta}(t)$ in the Hull–White model will be given by the constant

$$\hat{\theta}(t) = y_\infty + \frac{\beta^2}{2\kappa^2},$$

when the initial 'observable' discount function is of the form (*), that is as in the Vasicek model.

10

Heath–Jarrow–Morton Models

10.1 INTRODUCTION

In Chapter 7 and Chapter 8 we discussed various models of the term structure of interest rates which assume that the entire term structure is governed by a low-dimensional Markov vector diffusion process of state variables. Among other things, we concluded from those chapters that a time-homogeneous diffusion model generally cannot produce a term structure consistent with all observed bond prices, but as discussed in Chapter 9 a simple extension to a time-inhomogeneous model allows for a perfect fit to any (or almost any) given term structure of interest rates. A more natural way to achieve consistency with observed prices is to start from the observed term structure and then model the evolution of the entire term structure of interest rates in a manner that precludes arbitrage. This is the approach introduced by Heath et al. (1992), henceforth abbreviated HJM.[1] The HJM models are relative pricing models and focus on the pricing of derivative securities.

This chapter gives an overview of the HJM class of term structure models. We will discuss the main characteristics, advantages, and drawbacks of the general HJM framework and consider several model specifications in more detail. In particular, we shall study the relationship between HJM models and the diffusion models discussed in previous chapters. We take an applied perspective and, although the exposition is quite mathematical, we shall not go deeply into all technicalities, but refer the interested reader to the original HJM paper and the other references given below for details.

10.2 BASIC ASSUMPTIONS

As before, we let f_t^T be the (continuously compounded) instantaneous forward rate prevailing at time t for a loan agreement over an infinitesimal time interval starting at time $T \geq t$. We shall refer to f_t^T as the T-maturity forward rate at time t. Suppose that we know the term structure of interest rates at time 0 represented by the forward rate function $T \mapsto f_0^T$. Assume that, for any fixed T, the T-maturity forward rate evolves according to

[1] The binomial model of Ho and Lee (1986) can be seen as a forerunner of the more complete and thorough HJM-analysis.

$$df_t^T = \alpha\left(t, T, (f_t^s)_{s\geq t}\right)\, dt + \sum_{i=1}^{n} \beta_i\left(t, T, (f_t^s)_{s\geq t}\right)\, dz_{it}, \quad 0 \leq t \leq T, \tag{10.1}$$

where $z_1, \dots, z_n$ are n independent standard Brownian motions under the real-world probability measure. The $(f_t^s)_{s\geq t}$ terms indicate that both the forward rate drift α and the forward rate sensitivity terms β_i at time t may depend on the entire forward rate curve present at time t.

We call (10.1) an n-factor HJM model of the term structure of interest rates. Note that n is the number of random shocks (Brownian motions), and that all the forward rates are affected by the same n shocks. The diffusion models discussed in Chapter 7 and Chapter 8 are based on the evolution of a low-dimensional vector diffusion process of state variables. The general HJM model does not fit into that framework. It is not possible to use the partial differential equation approach for pricing in such general models, but we can still price by computing relevant expectations under the appropriate martingale measures. However, we can think of the general model (10.1) as an infinite-dimensional diffusion model, since the infinitely many forward rates can affect the dynamics of any forward rate.[2] In Section 10.6 we shall discuss when an HJM model can be represented by a low-dimensional diffusion model.

The basic idea of HJM is to directly model the entire term structure of interest rates. Recall from Chapter 1 that the term structure at some time t is equally well represented by the discount function $T \mapsto B_t^T$ or the yield curve $T \mapsto y_t^T$ as by the forward rate function $T \mapsto f_t^T$, due to the following relations

$$\begin{aligned}
B_t^T &= e^{-\int_t^T f_t^s\, ds} = e^{-y_t^T (T-t)}, \qquad (10.2)\\
f_t^T &= -\frac{\partial \ln B_t^T}{\partial T} = y_t^T + (T-t)\frac{\partial y_t^T}{\partial T},\\
y_t^T &= \frac{1}{T-t}\int_t^T f_t^s\, ds = -\frac{1}{T-t}\ln B_t^T.
\end{aligned}$$

We could therefore have specified the dynamics of the zero-coupon bond prices or the yield curve instead of the forward rates. However, there are (at least) three reasons for choosing the forward rates as the modelling object. Firstly, the forward rates are the most basic elements of the term structure. Both the zero-coupon bond prices and yields involve sums/integrals of forward rates. Secondly, we know from our analysis in Chapter 4 that one way of pricing derivatives is to find the expected discounted payoff under the risk-neutral probability measure (that is the spot martingale measure), where the discounting is in terms of the short-term interest rate r_t. The short rate is related to the forward rates, the yield curve, and the discount function as

[2] In fact, the results of Theorem 10.1, 10.2, and 10.4 below are valid in the more general setting, where the drift and sensitivities of the forward rates also depend the forward rate curves at previous dates. Since no models with that feature have been studied in the literature, we focus on the case where only the current forward rate curve affects the dynamics of the curve over the next infinitesimal period of time.

$$r_t = f_t^t = \lim_{T \downarrow t} y_t^T = -\lim_{T \downarrow t} \frac{\partial \ln B_t^T}{\partial T}.$$

Obviously, the relation of the forward rates to the short rate is much simpler than that of both the yield curve and the discount function, so this motivates the HJM choice of modelling basis. Thirdly, the volatility structure of zero-coupon bond prices is more complicated than that of interest rates. For example, the volatility of the bond price must approach zero as the bond approaches maturity, and, to avoid negative interest rates, the volatility of a zero-coupon bond price must approach zero as the price approaches one. Such restrictions need not be imposed on the volatilities of forward rates.

10.3 BOND PRICE DYNAMICS AND THE DRIFT RESTRICTION

In this section we will discuss how we can change the probability measure in the HJM framework to the risk-neutral measure $\mathbb{Q}$. As a first step, the following theorem gives the dynamics under the real-world probability measure of the zero-coupon bond prices B_t^T under the HJM assumption (10.1).

Theorem 10.1 *Under the assumed forward rate dynamics* (10.1), *the price* B_t^T *of a zero-coupon bond maturing at time* T *evolves as*

$$dB_t^T = B_t^T \left[\mu^T(t, (f_t^s)_{s \geq t})\, dt + \sum_{i=1}^n \sigma_i^T(t, (f_t^s)_{s \geq t})\, dz_{it} \right], \tag{10.3}$$

where

$$\mu^T(t, (f_t^s)_{s \geq t}) = r_t - \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du + \frac{1}{2} \sum_{i=1}^n \left(\int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du \right)^2,$$

$$\sigma_i^T(t, (f_t^s)_{s \geq t}) = -\int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du. \tag{10.4}$$

Proof: For simplicity we only prove the claim for the case $n = 1$, where

$$df_t^T = \alpha(t, T, (f_t^s)_{s \geq t})\, dt + \beta(t, T, (f_t^s)_{s \geq t})\, dz_t, \quad 0 \leq t \leq T,$$

for any T. Introduce the auxiliary stochastic process

$$Y_t = \int_t^T f_t^u\, du.$$

Then we have from (10.2) that the zero-coupon bond price is given by $B_t^T = e^{-Y_t}$. If we can find the dynamics of Y_t, we can therefore apply Itô's Lemma to derive the dynamics of the zero-coupon bond price B_t^T. Since Y_t is a function of infinitely many forward rates f_t^u with dynamics given by (10.1), it is however quite complicated to derive the dynamics of Y_t. Due to the fact that t appears both in the lower integration bound and in the integrand itself, we must apply Leibnitz' rule for stochastic integrals stated in Theorem 3.5, which in this case yields

$$dY_t = \left(-r_t + \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du\right) dt + \left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du\right) dz_t,$$

where we have applied that $r_t = f(t, t)$. Since $B_t^T = g(Y_t)$, where $g(Y) = e^{-Y}$ with $g'(Y) = -e^{-Y}$ and $g''(Y) = e^{-Y}$, Itô's Lemma (see Theorem 3.6) implies that the dynamics of the zero-coupon bond prices is

$$\begin{aligned} dB_t^T &= \left\{-e^{-Y_t}\left(-r_t + \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du\right)\right. \\ &\quad \left. + \frac{1}{2} e^{-Y_t}\left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du\right)^2\right\} dt \\ &\quad - e^{-Y_t}\left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du\right) dz_t \\ &= B_t^T\left[r_t - \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du + \frac{1}{2}\left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du\right)^2 dt\right. \\ &\quad \left. - \left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du\right) dz_t\right], \end{aligned}$$

which gives the one-factor version of (10.3). □

Now we turn to the behaviour under the risk-neutral probability measure $\mathbb{Q}$. The forward rate will have the same sensitivity terms $\beta_i(t, T, (f_t^s)_{s \geq t})$ as under the real-world probability measure, but a different drift. More precisely, we have from Chapter 4 that the n-dimensional process $z^{\mathbb{Q}} = (z_1^{\mathbb{Q}}, \dots, z_n^{\mathbb{Q}})^{\mathsf{T}}$ defined by

$$dz_{it}^{\mathbb{Q}} = dz_{it} + \lambda_{it}\, dt$$

is a standard Brownian motion under the risk-neutral probability measure $\mathbb{Q}$, where the λ_i processes are the market prices of risk. Substituting this into (10.1), we get

$$df_t^T = \hat{\alpha}(t, T, (f_t^s)_{s \geq t})\, dt + \sum_{i=1}^n \beta_i(t, T, (f_t^s)_{s \geq t})\, dz_{it}^{\mathbb{Q}},$$

where

$$\hat{\alpha}(t, T, (f_t^s)_{s \geq t}) = \alpha(t, T, (f_t^s)_{s \geq t}) - \sum_{i=1}^n \beta_i(t, T, (f_t^s)_{s \geq t})\lambda_{it}.$$

As in Theorem 10.1 we get that the drift rate of the zero-coupon bond price becomes

$$r_t - \int_t^T \hat{\alpha}(t, u, (f_t^s)_{s \geq t})\, du + \frac{1}{2} \sum_{i=1}^n \left(\int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du \right)^2$$

under the risk-neutral probability measure $\mathbb{Q}$. But we also know that this drift rate has to be equal to r_t. This can only be true if

$$\int_t^T \hat{\alpha}(t, u, (f_t^s)_{s \geq t})\, du = \frac{1}{2} \sum_{i=1}^n \left(\int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du \right)^2.$$

Differentiating with respect to T, we get the following key result:

Theorem 10.2 *The forward rate drift under the risk-neutral probability measure $\mathbb{Q}$ satisfies*

$$\hat{\alpha}(t, T, (f_t^s)_{s \geq t}) = \sum_{i=1}^n \beta_i(t, T, (f_t^s)_{s \geq t}) \int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du. \tag{10.5}$$

The relation (10.5) is called **the HJM drift restriction**. The drift restriction has important consequences: Firstly, the forward rate behaviour under the risk-neutral measure $\mathbb{Q}$ is fully characterized by the initial forward rate curve, the number of factors n, and the forward rate sensitivity terms $\beta_i(t, T, (f_t^s)_{s \geq t})$. The forward rate drift is not to be specified exogenously. This is in contrast to the diffusion models considered in the previous chapters, where both the drift and the sensitivity of the state variables were to be specified.

Secondly, since derivative prices depend on the evolution of the term structure under the risk-neutral measure and other relevant martingale measures, it follows that derivative prices depend only on the initial forward rate curve and the forward rate sensitivity functions $\beta_i(t, T, (f_t^s)_{s \geq t})$. In particular, derivatives prices do not depend on the market prices of risk. We do not have to make any assumptions or equilibrium derivations of the market prices of risk to price derivatives in an HJM model. In this sense, HJM models are **pure no-arbitrage models**. Again, this is in contrast with the diffusion models of Chapters 7 and 8. In the one-factor diffusion models, for example, the entire term structure is assumed to be generated by the movements of the very short end and the resulting term structure depends on the market price of short rate risk. In the HJM models we use the information contained in the current term structure and avoid separately specifying the market prices of risk.

10.4 THREE WELL-KNOWN SPECIAL CASES

Since the general HJM framework is quite abstract, in this section we will look at three specifications that result in well-known models.

10.4.1 The Ho–Lee (extended Merton) model

Let us consider the simplest possible HJM-model: a one-factor model with $\beta(t, T, (f_t^s)_{s\geq t}) = \beta > 0$, that is the forward rate volatilities are identical for all maturities (independent of T) and constant over time (independent of t). From the HJM drift restriction (10.5), the forward rate drift under the risk-neutral probability measure $\mathbb{Q}$ is

$$\hat{\alpha}(t, T, (f_t^s)_{s\geq t}) = \beta \int_t^T \beta\, du = \beta^2[T-t].$$

With this specification the future value of the T-maturity forward rate is given by

$$f_t^T = f_0^T + \int_0^t \beta^2[T-u]\, du + \int_0^t \beta\, dz_u^{\mathbb{Q}},$$

which is normally distributed with mean $f_0^T + \beta^2 t[T - t/2]$ and variance $\int_0^t \beta^2\, du = \beta^2 t$.

In particular, the future value of the short rate is

$$r_t = f_t^t = f_0^t + \frac{1}{2}\beta^2 t^2 + \int_0^t \beta\, dz_u^{\mathbb{Q}}.$$

By Itô's Lemma,

$$dr_t = \hat{\varphi}(t)\, dt + \beta\, dz_t^{\mathbb{Q}}, \tag{10.6}$$

where $\hat{\varphi}(t) = \partial f_0^t/\partial t + \beta^2 t$. From (10.6), we see that this specification of the HJM model is equivalent to the Ho–Lee extension of the Merton model, which was studied in Section 9.3. It follows that zero-coupon bond prices are given in terms of the short rate by the relation

$$B_t^T = e^{-a(t,T)-(T-t)r_t},$$

where

$$a(t, T) = \int_t^T \hat{\varphi}(u)(T-u)\, du - \frac{\beta^2}{6}(T-t)^3.$$

Furthermore, the price $C_t^{K,T,S}$ of a European call option maturing at time T with exercise price K written on the zero-coupon bond maturing at S is

$$C_t^{K,T,S} = B_t^S N(d_1) - KB_t^T N(d_2), \tag{10.7}$$

where

$$d_1 = \frac{1}{v(t,T,S)} \ln\left(\frac{B_t^S}{KB_t^T}\right) + \frac{1}{2}v(t,T,S),$$
$$d_2 = d_1 - v(t,T,S),$$
$$v(t,T,S) = \beta[S-T]\sqrt{T-t}. \tag{10.8}$$

In addition, Jamshidian's trick for the pricing of European options on coupon bonds (see Theorem 7.3) can be applied since B_T^S is a monotonic function of r_T.

10.4.2 The Hull–White (extended Vasicek) model

Next, let us consider the one-factor model with the forward rate volatility function

$$\beta(t,T,(f_t^s)_{s\geq t}) = \beta e^{-\kappa[T-t]} \tag{10.9}$$

for some positive constants β and κ. Here the forward rate volatility is an exponentially decaying function of the time to maturity. By the drift restriction, the forward rate drift under $\mathbb{Q}$ is

$$\hat{\alpha}(t,T,(f_t^s)_{s\geq t}) = \beta e^{-\kappa[T-t]} \int_t^T \beta e^{-\kappa[u-t]}\,du = \frac{\beta^2}{\kappa} e^{-\kappa[T-t]}\left(1 - e^{-\kappa[T-t]}\right)$$

so that the future value of the T-maturity forward rate is

$$f_t^T = f_0^T + \int_0^t \frac{\beta^2}{\kappa} e^{-\kappa[T-u]}\left(1 - e^{-\kappa[T-u]}\right)du + \int_0^t \beta e^{-\kappa[T-u]}\,dz_u^{\mathbb{Q}}.$$

In particular, the future short rate is

$$r_t = f_t^t = g(t) + \beta e^{-\kappa t}\int_0^t e^{\kappa u}\,dz_u^{\mathbb{Q}},$$

where the deterministic function g is defined by

$$g(t) = f_0^t + \int_0^t \frac{\beta^2}{\kappa} e^{-\kappa[t-u]}\left(1 - e^{-\kappa[t-u]}\right)du = f_0^t + \frac{\beta^2}{2\kappa^2}\left(1 - e^{-\kappa t}\right)^2.$$

Again, the future values of the forward rates and the short rate are normally distributed.

Let us find the dynamics of the short rate. Writing $R_t = \int_0^t e^{\kappa u}\,dz_u^{\mathbb{Q}}$, we have $r_t = G(t,R_t)$, where $G(t,R) = g(t) + \beta e^{-\kappa t}R$. We can now apply Itô's Lemma with $\partial G/\partial t = g'(t) - \kappa\beta e^{-\kappa t}R$, $\partial G/\partial R = \beta e^{-\kappa t}$, and $\partial^2 G/\partial R^2 = 0$. Since $dR_t = e^{\kappa t}\,dz_t^{\mathbb{Q}}$ and

$$g'(t) = \frac{\partial f_0^t}{\partial t} + \frac{\beta^2}{\kappa}e^{-\kappa t}\left(1 - e^{-\kappa t}\right),$$

we get

$$
\begin{aligned}
dr_t &= \left[g'(t) - \kappa\beta e^{-\kappa t} R_t\right] dt + \beta e^{-\kappa t} e^{\kappa t}\, dz_t^{\mathbb{Q}} \\
&= \left[\frac{\partial f_0^t}{\partial t} + \frac{\beta^2}{\kappa} e^{-\kappa t}\left(1 - e^{-\kappa t}\right) - \kappa\beta e^{-\kappa t} R_t\right] dt + \beta\, dz_t^{\mathbb{Q}}.
\end{aligned}
$$

Inserting the relation $r_t - g(t) = \beta e^{-\kappa t} R_t$, we can rewrite the above expression as

$$
\begin{aligned}
dr_t &= \left[\frac{\partial f_0^t}{\partial t} + \frac{\beta^2}{\kappa} e^{-\kappa t}\left(1 - e^{-\kappa t}\right) - \kappa[r_t - g(t)]\right] dt + \beta\, dz_t^{\mathbb{Q}} \\
&= \kappa[\hat{\theta}(t) - r_t]\, dt + \beta\, dz_t^{\mathbb{Q}},
\end{aligned}
$$

where

$$
\hat{\theta}(t) = f_0^t + \frac{1}{\kappa}\frac{\partial f_0^t}{\partial t} + \frac{\beta^2}{2\kappa^2}\left(1 - e^{-2\kappa t}\right).
$$

A comparison with Section 9.4 reveals that the HJM one-factor model with forward rate volatilities given by (10.9) is equivalent to the Hull–White (or extended-Vasicek) model. Therefore, we know that the zero-coupon bond prices are given by

$$
B_t^T = e^{-a(t,T) - b(T-t) r_t},
$$

where

$$
b(\tau) = \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right),
$$

$$
a(t,T) = \kappa \int_t^T \hat{\theta}(u) b(T-u)\, du + \frac{\beta^2}{4\kappa} b(T-t)^2 + \frac{\beta^2}{2\kappa^2}\left(b(T-t) - (T-t)\right).
$$

The price of a European call on a zero-coupon bond is again given by (10.7), but where

$$
v(t,T,S) = \frac{\beta}{\sqrt{2\kappa^3}}\left(1 - e^{-\kappa[S-T]}\right)\left(1 - e^{-2\kappa[T-t]}\right)^{1/2}. \tag{10.10}
$$

Again, Jamshidian's trick can be used for European options on coupon bonds.

10.4.3 The extended CIR model

We will now discuss the relation between the HJM models and the Cox–Ingersoll–Ross (CIR) model studied in Section 7.5 with its extension examined in Section 9.5. In the extended CIR model the short rate is assumed to follow the process

$$
dr_t = \left(\kappa\theta(t) - \hat{\kappa} r_t\right) dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}}
$$

under the risk-neutral probability measure $\mathbb{Q}$. The zero-coupon bond prices are of the form $B^T(r_t, t) = \exp\{-a(t,T) - b(T-t)r_t\}$, where

$$b(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \hat{\kappa})(e^{\gamma\tau} - 1) + 2\gamma}$$

with $\gamma = \sqrt{\hat{\kappa}^2 + 2\beta^2}$, and the function a is not important for what follows. Therefore, the volatility of the zero-coupon bond price is (the absolute value of)

$$\sigma^T(r_t, t) = -b(T-t)\beta\sqrt{r_t}.$$

On the other hand, in a one-factor HJM set-up the zero-coupon bond price volatility is given in terms of the forward rate volatility function $\beta(t, T, (f_t^s)_{s\geq t})$ by (10.4). To be consistent with the CIR model, the forward rate volatility must hence satisfy the relation

$$\int_t^T \beta(t, u, (f_t^s)_{s\geq t})\, du = b(T-t)\beta\sqrt{r_t}.$$

Differentiating with respect to T, we get

$$\beta(t, T, (f_t^s)_{s\geq t}) = b'(T-t)\beta\sqrt{r_t}.$$

A straightforward computation of $b'(\tau)$ allows this condition to be rewritten as

$$\beta(t, T, (f_t^s)_{s\geq t}) = \frac{4\gamma^2 e^{\gamma[T-t]}}{\left((\gamma + \hat{\kappa})(e^{\gamma[T-t]} - 1) + 2\gamma\right)^2}\beta\sqrt{r_t}. \tag{10.11}$$

As discussed in Section 9.5, such a model does not make sense for all types of initial forward rate curves.

10.5 GAUSSIAN HJM MODELS

In the first two models studied in the previous section, the future values of the forward rates are normally distributed. Models with this property are called Gaussian. Clearly, Gaussian models have the unpleasant and unrealistic feature of yielding negative interest rates with a strictly positive probability, see the discussion in Chapter 7. On the other hand, Gaussian models are highly tractable.

An HJM model is Gaussian if the forward rate sensitivities β_i are deterministic functions of time and maturity, that is

$$\beta_i(t, T, (f_t^s)_{s\geq t}) = \beta_i(t, T), \quad i = 1, 2, \ldots, n.$$

To see this, first note that from the drift restriction (10.5) it follows that the forward rate drift under the risk-neutral probability measure $\mathbb{Q}$ is also a deterministic function of time and maturity:

$$\hat{\alpha}(t,T) = \sum_{i=1}^{n} \beta_i(t,T) \int_t^T \beta_i(t,u)\, du.$$

It follows that, for any T, the T-maturity forward rates evolves according to

$$f_t^T = f_0^T + \int_0^t \hat{\alpha}(u,T)\, du + \sum_{i=1}^{n} \int_0^t \beta_i(u,T)\, dz_{iu}^{\mathbb{Q}}.$$

Because $\beta_i(u,T)$ at most depends on time, the stochastic integrals are normally distributed, compare Theorem 3.3. The future forward rates are therefore normally distributed under $\mathbb{Q}$. The short-term interest rate is $r_t = f_t^t$, that is

$$r_t = f_0^t + \int_0^t \hat{\alpha}(u,t)\, du + \sum_{i=1}^{n} \int_0^t \beta_i(u,t)\, dz_{iu}^{\mathbb{Q}}, \quad 0 \le t,$$

which is also normally distributed under $\mathbb{Q}$. In particular, there is a positive probability of negative interest rates.[3]

To demonstrate the high degree of tractability of the general Gaussian HJM framework, the following theorem provides a closed-form expression for the price $C_t^{K,T,S}$ of a European call on the zero-coupon bond maturing at S.

Theorem 10.3 *In the Gaussian n-factor HJM model in which the forward rate sensitivity coefficients $\beta_i(t,T,(f_t^s)_{s\ge t})$ only depend on time t and maturity T, the price of a European call option maturing at T written with exercise price K on a zero-coupon bond maturing at S is given by*

$$C_t^{K,T,S} = B_t^S N(d_1) - K B_t^T N(d_2), \tag{10.12}$$

where

$$d_1 = \frac{1}{v(t,T,S)} \ln\left(\frac{B_t^S}{K B_t^T}\right) + \frac{1}{2} v(t,T,S),$$

$$d_2 = d_1 - v(t,T,S),$$

$$v(t,T,S) = \left(\sum_{i=1}^{n} \int_t^T \left[\int_T^S \beta_i(u,y)\, dy\right]^2 du\right)^{1/2}.$$

Proof: We will apply the same procedure as we did in the diffusion models of Chapter 7, see, for example the derivation of the option price in the Vasicek model in Section 7.4.5. The option price is given by

$$C_t^{K,T,S} = B_t^T\, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max\left(B_T^S - K, 0\right)\right] = B_t^T\, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max\left(F_T^{T,S} - K, 0\right)\right],$$

[3] Of course, this does not imply that interest rates are necessarily normally distributed under the true, real-world probability measure $\mathbb{P}$, but since the probability measures $\mathbb{P}$ and $\mathbb{Q}$ are equivalent, a positive probability of negative rates under $\mathbb{Q}$ implies a positive probability of negative rates under $\mathbb{P}$.

where $\mathbb{Q}^T$ denotes the T-forward martingale measure introduced in Section 4.4.2. We will find the distribution of the underlying bond price B_T^S at expiration of the option, which is identical to the forward price of the bond with immediate delivery, $F_T^{T,S}$. The forward price for delivery at T is given at time t as $F_t^{T,S} = B_t^S / B_t^T$. We know that the forward price is a $\mathbb{Q}^T$-martingale, and by Itô's Lemma we can express the sensitivity terms of the forward price by the sensitivity terms of the bond prices, which according to (10.4) are given by $\sigma_i^S(t) = -\int_t^S \beta_i(t,y)\,dy$ and $\sigma_i^T(t) = -\int_t^T \beta_i(t,y)\,dy$. Therefore, we get that

$$dF_t^{T,S} = \sum_{i=1}^n \left(\sigma_i^S(t) - \sigma_i^T(t)\right) F_t^{T,S}\,dz_{it}^T = \underbrace{-\left(\int_T^S \beta_i(t,y)\,dy\right)}_{h_i(t)} F_t^{T,S}\,dz_{it}^T.$$

It follows (see Chapter 3) that

$$\ln F_T^{T,S} = \ln F_t^{T,S} - \frac{1}{2}\sum_{i=1}^n \int_t^T h_i(u)^2\,du + \sum_{i=1}^n \int_t^T h_i(u)\,dz_{iu}^T.$$

From Theorem 3.3 we get that $\ln B_T^S = \ln F_T^{T,S}$ is normally distributed with variance

$$v(t,T,S)^2 = \sum_{i=1}^n \int_t^T h_i(u)^2\,du = \sum_{i=1}^n \int_t^T \left(\int_T^S \beta_i(u,y)\,dy\right)^2 du.$$

The result now follows from an application of Theorem A.4 in Appendix A. □

Consider, for example, a two-factor Gaussian HJM model with forward rate sensitivities

$$\beta_1(t,T) = \beta_1 \qquad \text{and} \qquad \beta_2(t,T) = \beta_2 e^{-\kappa[T-t]},$$

where β_1, β_2, and κ are positive constants. This is a combination of two one-factor examples of Section 10.4. In this model we have

$$\begin{aligned} v(t,T,S)^2 &= \int_t^T \left[\int_T^S \beta_1\,dy\right]^2 du + \int_t^T \left[\int_T^S \beta_2 e^{-\kappa[y-u]}\,dy\right]^2 du \\ &= \beta_1^2 [S-T]^2 [T-t] + \frac{\beta_2^2}{2\kappa^3}\left(1 - e^{-\kappa[S-T]}\right)^2 \left(1 - e^{-2\kappa[T-t]}\right), \end{aligned}$$

see (10.8) and (10.10).

It is generally not possible to express the future zero-coupon bond price B_T^S as a monotonic function of r_T, not even when we restrict ourselves to a Gaussian model. Therefore, we can generally not use Jamshidian's trick to price European options on coupon bonds.

Exercise 10.1 considers some aspects of another Gaussian model initially proposed by Mercurio and Moraleda (2000).

10.6 DIFFUSION REPRESENTATIONS OF HJM MODELS

As discussed immediately below the basic assumption (10.1), the HJM models are generally not diffusion models in the sense that the relevant uncertainty is captured by a finite-dimensional diffusion process. For computational purposes there is a great advantage in applying a low-dimensional diffusion model, as we will argue below. As discussed earlier in this chapter, we can think of the entire forward rate curve as following an infinite-dimensional diffusion process. On the other hand, we have already seen some specifications of the HJM model framework which imply that the short-term interest rate follows a diffusion process. In this section, we will discuss when such a low-dimensional diffusion representation of an HJM model is possible.

10.6.1 On the use of numerical techniques for diffusion and non-diffusion models

In all dynamic term structure models, some securities can only be priced using numerical techniques. The main numerical techniques are

1. approximation of the relevant processes by a discrete-time tree-structure and then invoking a recursive valuation procedure through the tree,
2. solving the fundamental partial differential equation with the asset-specific boundary conditions,
3. applying Monte Carlo simulation to approximate the appropriately discounted and risk-adjusted payoffs with a sample average.

These numerical methods will be explained in detail in Chapter 16.

For the purpose of using numerical techniques for derivative pricing, whether or not the relevant uncertainty can be described by some low-dimensional diffusion process is crucial. A diffusion process can be approximated by a recombining tree, whereas a non-recombining tree must be used for processes for which the future evolution can depend on the path followed thus far. The number of nodes in a non-recombining tree explodes. A one-variable binomial tree with m time steps has $m+1$ endnodes if it is recombining, but 2^m endnodes if it is non-recombining. This makes it practically impossible to use trees to compute prices of long-term derivatives in non-diffusion term structure models.

In a diffusion model we can use partial differential equations (PDEs) for pricing, see the analysis in Section 4.8. Such PDEs can be efficiently solved by numerical methods for both European- and American-type derivatives as long as the dimension of the state variable vector does not exceed three or maybe four. If it is impossible to express the model in some low-dimensional vector of state variables, the PDE approach does not work.

The third frequently used numerical pricing technique is the Monte Carlo simulation approach. The Monte Carlo approach can be applied even for non-diffusion models. The basic idea is to simulate, from now and to the maturity date of the contingent claim, the underlying Brownian motions and, hence, the relevant underlying interest rates, bond prices, and so on, under an appropriately chosen martingale measure. Then the payoff from the contingent claim can be computed for this particular simulated path of the underlying variables. Doing this a large number of times, the average of the computed payoffs leads to a good approximation to the theoretical value of the claim. In its original formulation, Monte Carlo simulation can only be applied to European-style derivatives. The wish to price American-type derivatives in non-diffusion HJM models has recently induced some suggestions on the use of Monte Carlo methods for American-style assets. Generally, Monte Carlo pricing of even European-style assets in non-diffusion HJM models is computationally intensive since the entire term structure has to be simulated, not just one or two variables.

10.6.2 In which HJM models does the short rate follow a diffusion process?

We seek to find conditions under which the short-term interest rate in an HJM model follows a Markov diffusion process. First, we will find the dynamics of the short rate in the general HJM framework (10.1). For the pricing of derivatives, it is the dynamics under the risk-neutral probability measure or related martingale measures which is relevant. The following theorem gives the short rate dynamics under the risk-neutral measure $\mathbb{Q}$.

Theorem 10.4 *In the general HJM framework* (10.1) *the dynamics of the short rate r_t under the risk-neutral measure is given by*

$$\begin{aligned} dr_t = &\left\{ \frac{\partial f_0^t}{\partial t} + \sum_{i=1}^n \int_0^t \frac{\partial \beta_i(u, t, (f_u^s)_{s\geq u})}{\partial t} \left[\int_u^t \beta_i(u, x, (f_u^s)_{s\geq u})\, dx \right] du \right.\\ &\left. + \sum_{i=1}^n \int_0^t \beta_i(u, t, (f_u^s)_{s\geq u})^2\, du + \sum_{i=1}^n \int_0^t \frac{\partial \beta_i(u, t, (f_u^s)_{s\geq u})}{\partial t}\, dz_{iu}^{\mathbb{Q}} \right\} dt \\ &+ \sum_{i=1}^n \beta_i(t, t, (f_t^s)_{s\geq t})\, dz_{it}^{\mathbb{Q}}. \end{aligned} \tag{10.13}$$

Proof: For each T, the dynamics of the T-maturity forward rate under the risk-neutral measure $\mathbb{Q}$ is

$$df_t^T = \hat{\alpha}(t, T, (f_t^s)_{s\geq t})\, dt + \sum_{i=1}^n \beta_i(t, T, (f_t^s)_{s\geq t})\, dz_{it}^{\mathbb{Q}},$$

where $\hat{\alpha}$ is given by the drift restriction (10.5). This implies that

$$f_t^T = f_0^T + \int_0^t \hat{\alpha}(u, T, (f_u^s)_{s\geq u})\, du + \sum_{i=1}^n \int_0^t \beta_i(u, T, (f_u^s)_{s\geq u})\, dz_{iu}^{\mathbb{Q}}.$$

Since the short rate is simply the 'zero-maturity' forward rate, $r_t = f_t^t$, it follows that

$$\begin{aligned}
r_t &= f_0^t + \int_0^t \hat{\alpha}(u, t, (f_u^s)_{s\geq u})\, du + \sum_{i=1}^n \int_0^t \beta_i(u, t, (f_u^s)_{s\geq u})\, dz_{iu}^{\mathbb{Q}} \\
&= f_0^t + \sum_{i=1}^n \int_0^t \beta_i(u, t, (f_u^s)_{s\geq u}) \left[\int_u^t \beta_i(u, x, (f_u^s)_{s\geq u})\, dx\right] du \\
&\quad + \sum_{i=1}^n \int_0^t \beta_i(u, t, (f_u^s)_{s\geq u})\, dz_{iu}^{\mathbb{Q}}. \qquad (10.14)
\end{aligned}$$

To find the dynamics of r, we proceed as in the simple examples of Section 10.4. Let $R_{it} = \int_0^t \beta_i(u, t, (f_u^s)_{s\geq u})\, dz_{iu}^{\mathbb{Q}}$ for $i = 1, 2, \dots, n$. Then

$$dR_{it} = \beta_i(t, t, (f_t^s)_{s\geq t})\, dz_{it}^{\mathbb{Q}} + \left[\int_0^t \frac{\partial \beta_i(u, t, (f_u^s)_{s\geq u})}{\partial t}\, dz_{iu}^{\mathbb{Q}}\right] dt$$

by Leibnitz' rule for stochastic integrals (see Theorem 3.5). Define the function

$$G_i(t) = \int_0^t \beta_i(u, t, (f_u^s)_{s\geq u}) H_i(u, t)\, du,$$

where $H_i(u, t) = \int_u^t \beta_i(u, x, (f_u^s)_{s\geq u})\, dx$. By Leibnitz' rule for ordinary integrals,

$$\begin{aligned}
G_i'(t) &= \beta_i(t, t, (f_t^s)_{s\geq t}) H_i(t, t) + \int_0^t \frac{\partial}{\partial t}\left[\beta_i(u, t, (f_u^s)_{s\geq u}) H_i(u, t)\right] du \\
&= \int_0^t \left[\frac{\partial \beta_i(u, t, (f_u^s)_{s\geq u})}{\partial t} H_i(u, t) + \beta_i(u, t, (f_u^s)_{s\geq u}) \frac{\partial H_i(u, t)}{\partial t}\right] du \\
&= \int_0^t \left[\frac{\partial \beta_i(u, t, (f_u^s)_{s\geq u})}{\partial t} \int_u^t \beta_i(u, x, (f_u^s)_{s\geq u})\, dx + \beta_i(u, t, (f_u^s)_{s\geq u})^2\right] du,
\end{aligned}$$

where we have used the chain rule and the fact that $H_i(t, t) = 0$. Note that

$$r_t = f_0^t + \sum_{i=1}^n G_i(t) + \sum_{i=1}^n R_{it},$$

where the G_i's are deterministic functions and $R_i(t)$ are stochastic processes. By Itô's Lemma, we get

$$dr_t = \left[\frac{\partial f_0^t}{\partial t} + \sum_{i=1}^n G_i'(t)\right] dt + \sum_{i=1}^n dR_{it}.$$

Substituting in the expressions for $G_i'(t)$ and dR_{it}, we arrive at the expression (10.13). □

From (10.13) we see that the drift term of the short rate generally depends on past values of the forward rate curve and past values of the Brownian motion. Therefore, the short rate process is generally not a diffusion process in an HJM model. However, if we know that the initial forward rate curve belongs to a certain family, the short rate may be Markovian. If, for example, the initial forward rate curve is on the form generated by the original one-factor CIR diffusion model, then the short rate in the one-factor HJM model with forward rate sensitivity given by (10.11) will, of course, be Markovian since the two models are then indistinguishable.

Under what conditions on the forward rate sensitivity functions $\beta_i(t, T, (f_t^s)_{s\geq t})$ will the short rate follow a diffusion process for *any* initial forward rate curve? Hull and White (1993) and Carverhill (1994) answer this question. Their conclusion is summarized in the following theorem.

Theorem 10.5 *Consider an n-factor HJM model. Suppose that deterministic functions g_i and h exist such that*

$$\beta_i(t, T, (f_t^s)_{s\geq t}) = g_i(t)h(T), \quad i = 1, 2, \dots, n,$$

and h is continuously differentiable, non-zero, and never changing sign.[4] *Then the short rate has dynamics*

$$dr_t = \left[\frac{\partial f_0^t}{\partial t} + h(t)^2 \sum_{i=1}^n \int_0^t g_i(u)^2\, du + \frac{h'(t)}{h(t)}(r_t - f_0^t)\right] dt + \sum_{i=1}^n g_i(t)h(t)\, dz_{it}^{\mathbb{Q}}, \tag{10.15}$$

so that the short rate follows a diffusion process for any given initial forward rate curve.

Proof: We will only consider the case $n = 1$ and show that r_t indeed is a Markov diffusion process when

$$\beta(t, T, (f_t^s)_{s\geq t}) = g(t)h(T), \tag{10.16}$$

where g and h are deterministic functions and h is continuously differentiable, non-zero, and never changing sign. First note that (10.14) and (10.16) imply that

$$r_t = f_0^t + h(t) \int_0^t g(u)^2 \left[\int_u^t h(x)\, dx\right] du + h(t) \int_0^t g(u)\, dz_u^{\mathbb{Q}},$$

and, thus,

$$\int_0^t g(u)\, dz_u^{\mathbb{Q}} = \frac{1}{h(t)}(r_t - f_0^t) - \int_0^t g(u)^2 \left[\int_u^t h(x)\, dx\right] du. \tag{10.17}$$

[4] Carverhill claims that the h function can be different for each factor, that is $\beta_i(t, T, (f_t^s)_{s\geq t}) = g_i(t)h_i(T)$, but this is incorrect.

The dynamics of r in Equation (10.13) specializes to

$$dr_t = \left[\frac{\partial f_0^t}{\partial t} + h'(t) \int_0^t g(u)^2 \left[\int_u^t h(x)\, dx \right] du + h(t)^2 \int_0^t g(u)^2\, du \right.$$
$$\left. + h'(t) \int_0^t g(u)\, dz_u^{\mathbb{Q}} \right] dt + g(t)h(t)\, dz_t^{\mathbb{Q}},$$

which by applying (10.17) can be written as the one-factor version of (10.15). □

Note that the Ho–Lee model and the Hull–White model studied in Section 10.4 both satisfy the condition (10.16).

Obviously, the HJM models where the short rate is Markovian are members of the Gaussian class of models discussed in Section 10.5. In particular, the price of a European call on a zero-coupon bond is given by (10.12). It can be shown that with a volatility specification of the form (10.16), the future price B_t^T of a zero-coupon bond can be expressed as a monotonic function of time and the short rate r_t at time t. It follows that Jamshidian's trick introduced in Section 7.2.3 can be used for pricing European options on coupon bonds in this special setting.

The Markov property is one attractive feature of a term structure model. We also want a model to exhibit time-homogeneous volatility structures in the sense that the volatilities of, for example, forward rates, zero-coupon bond yields, and zero-coupon bond prices do not depend on calendar time in itself, see the discussion in Chapter 9. For the forward rate sensitivities in an HJM model to be time-homogeneous, $\beta_i(t, T, (f_t^s)_{s\geq t})$ must be of the form $\beta_i(T - t, (f_t^s)_{s\geq t})$. It then follows from (10.4) that the zero-coupon bond prices B_t^T will also have time-homogeneous sensitivities. Similarly for the zero-coupon yields y_t^T. Hull and White (1993) have shown that there are only two models of the HJM-class that have both a Markovian short rate and time-homogeneous sensitivities, namely the Ho–Lee model and the Hull–White model of Section 10.4.

As discussed above, the HJM models with a Markovian short rate are Gaussian models. While Gaussian models have a high degree of computational tractability, they also allow negative rates, which certainly is an unrealistic feature of a model. Furthermore, the volatility of the short rate and other interest rates seems to depend empirically on the short rate itself. Therefore, we seek to find HJM models with non-deterministic forward rate sensitivities that are still computationally tractable.

10.6.3 A two-factor diffusion representation of a one-factor HJM model

Ritchken and Sankarasubramanian (1995) show that in a one-factor HJM model with a forward rate volatility of the form

$$\beta(t, T, (f_t^s)_{s\geq t}) = \beta(t, t, (f_t^s)_{s\geq t}) e^{-\int_t^T \kappa(x)\, dx} \tag{10.18}$$

for some deterministic function κ, it is possible to capture the path dependence of the short rate by a single variable, and that this is only possible, when (10.18) holds. The evolution of the term structure will depend only on the current value of the short rate and the current value of this additional variable. The additional variable needed is

$$\varphi_t = \int_0^t \beta(u, t, (f_u^s)_{s \geq u})^2 \, du = \int_0^t \beta(u, u, (f_u^s)_{s \geq u})^2 e^{-2\int_u^t \kappa(x)\, dx} \, du,$$

which is the accumulated forward rate variance.

The future zero-coupon bond price B_t^T can be expressed as a function of r_t and φ_t in the following way:

$$B_t^T = e^{-a(t,T) - b_1(t,T) r_t - b_2(t,T)\varphi_t}, \tag{10.19}$$

where

$$a(t, T) = -\ln\left(\frac{B_0^T}{B_0^t}\right) - b_1(t, T) f_0^t,$$

$$b_1(t, T) = \int_t^T e^{-\int_t^u \kappa(x)\, dx} \, du,$$

$$b_2(t, T) = \frac{1}{2} b_1(t, T)^2.$$

The dynamics of r and φ under the risk-neutral measure $\mathbb{Q}$ is given by

$$dr_t = \left(\frac{\partial f_0^t}{\partial t} + \varphi_t - \kappa(t)[r_t - f_0^t]\right) dt + \beta(t, t, (f_t^s)_{s \geq t})\, dz_t^{\mathbb{Q}}, \tag{10.20}$$

$$d\varphi_t = \left(\beta(t, t, (f_t^s)_{s \geq t})^2 - 2\kappa(t)\varphi_t\right) dt. \tag{10.21}$$

These results are to be shown in Exercise 10.2. The two-dimensional process (r, φ) will be Markov if the short rate volatility depends on, at most, the current values of r_t and φ_t, that is if there is a function β_r such that

$$\beta(t, t, (f_t^s)_{s \geq t}) = \beta_r(r_t, \varphi_t, t).$$

In that case, we can price derivatives by two-dimensional recombining trees or by numerical solutions of two-dimensional PDEs (no closed-form solutions have been reported).[5] One admissible specification is $\beta_r(r, \varphi, t) = \beta r^\gamma$ for some non-negative constants β and γ, which, for example, includes a CIR-type volatility structure (for $\gamma = \frac{1}{2}$).

The volatilities of the forward rates are related to the short rate volatility through the deterministic function κ, which must be specified. If κ is constant, the forward rate volatility is an exponentially decaying function of the time to maturity. Empirically, the forward rate volatility seems to be a humped (first increasing, then decreasing) function of maturity. This can be achieved by letting the $\kappa(x)$ function be negative for small values of x and positive for large values of x. Also note that the volatility of some T-maturity forward rate f_t^T is not allowed to depend on the forward rate f_t^T itself, but only the short rate r_t and time.

[5] Li et al. (1995) show how to build a tree for this model, in which both European- and American-type term structure derivatives can be efficiently priced.

For further discussion of the circumstances under which an HJM model can be represented as a diffusion model, the reader is referred to Jeffrey (1995), Cheyette (1996), Bhar and Chiarella (1997), Inui and Kijima (1998), Bhar et al. (2000), and Björk and Landén (2002).

10.7 HJM MODELS WITH FORWARD-RATE DEPENDENT VOLATILITIES

In the models considered until now, the forward rate volatilities are either deterministic functions of time (the Gaussian models) or a function of time and the current short rate (the extended CIR model and the Ritchken–Sankarasubramanian model). The most natural way to introduce non-deterministic forward rate volatilities is to let them be a function of time and the current value of the forward rate itself, that is of the form

$$\beta_i(t, T, (f_t^s)_{s\geq t}) = \beta_i(t, T, f_t^T). \tag{10.22}$$

A model of this type, inspired by the Black–Scholes' stock option pricing model, is obtained by letting

$$\beta_i(t, T, f_t^T) = \gamma_i(t, T) f_t^T, \tag{10.23}$$

where $\gamma_i(t, T)$ is a positive, deterministic function of time. The forward rate drift will then be

$$\hat{\alpha}(t, T, (f_t^s)_{s\geq t}) = \sum_{i=1}^{n} \gamma_i(t, T) f_t^T \int_t^T \gamma_i(t, u) f_t^u \, du.$$

The specification (10.23) will ensure non-negative forward rates (starting with a term structure of positive forward rates) since both the drift and sensitivities are zero for a zero forward rate. Such models have a serious drawback, however. A process with the drift and sensitivities given above will *explode* with a strictly positive probability in the sense that the value of the process becomes infinite.[6] With a strictly positive probability of infinite interest rates, bond prices must equal zero, and this, obviously, implies arbitrage opportunities.

Heath et al. (1992) discuss the simple one-factor model with a capped forward rate volatility,

$$\beta(t, T, f_t^T) = \beta \min(f_t^T, \xi),$$

where β and ξ are positive constants, that is the volatility is proportional for 'small' forward rates and constant for 'large' forward rates. They showed that with this specification the forward rates do not explode, and, furthermore, they stay non-negative. The assumed forward rate volatility is rather far-fetched, however, and seems unrealistic. Miltersen (1994) provides a set of sufficient conditions for HJM-models of the type (10.22) to yield non-negative and non-exploding interest rates.

[6] This was shown by Morton (1988).

One of the conditions is that the forward rate volatility is bounded from above. This is, obviously, not satisfied for proportional volatility models, that is models where (10.23) holds.

10.8 HJM MODELS WITH UNSPANNED STOCHASTIC VOLATILITY

As explained in Section 8.7.1 empirical studies indicate unspanned stochastic volatility in bond prices and yields, that is there are stochastic factors driving yield and bond price volatilities that do not have a contemporaneous impact on the level of bond prices and yields. Moreover, this feature is only shared by some, carefully specified diffusion-type models. In contrast, it is easy to formulate HJM-type models with unspanned stochastic volatility. Following Collin-Dufresne and Goldstein (2002a), suppose that the risk-neutral forward rate dynamics is of the form

$$df_t^T = \hat{\alpha}\,(t, T, v_t)\; dt + \beta\,(t, T, v_t)\; dz_{1t}^{\mathbb{Q}}, \quad 0 \le t \le T,$$

where

$$dv_t = \mu_v(v_t)\, dt + \sigma_v(v_t) \left(\rho\, dz_{1t}^{\mathbb{Q}} + \sqrt{1-\rho^2}\, dz_{2t}^{\mathbb{Q}} \right),$$

and $z_1^{\mathbb{Q}}$ and $z_2^{\mathbb{Q}}$ are independent standard Brownian motions. $\hat{\alpha}$ is determined by the drift restriction. Then the bond price dynamics is

$$dB_t^T = B_t^T \left[r_t\, dt + \sigma^T(t, v_t)\, dz_{1t}^{\mathbb{Q}} \right],$$

where

$$\sigma^T(t, v_t) = -\int_t^T \beta(t, u, v_t)\, du,$$

as in Theorem 10.1. Unless $\rho = \pm 1$, bond prices will exhibit unspanned stochastic volatility.

Various specific HJM-models featuring unspanned stochastic volatility are analysed by Collin-Dufresne and Goldstein (2002a), Casassus et al. (2005), and Trolle and Schwartz (2009). An example is considered in Exercise 10.3.

10.9 CONCLUDING REMARKS

Empirical studies of various specifications of the HJM model framework have been performed on a variety of data sets by, for example, Amin and Morton (1994), Flesaker (1993), Heath et al. (1990), Miltersen (1998), and Pearson and Zhou (1999). However, these papers do not give a clear picture of how the forward rate volatilities should be specified.

To implement an HJM-model one must specify both the forward rate sensitivity functions $\beta_i(t, T, (f_t^s)_{s \ge t})$ *and* an initial forward rate curve $u \mapsto f_0^u$ given as a parameterized function of maturity. In the time-homogeneous Markov diffusion models studied in the Chapters 7 and 8, the forward rate curve in a given model can at all points in time be described by the same parametrization, although possibly with different parameters at different points in time due to changes in the state variable(s). For example, in the Vasicek one-factor model, we know from (7.42) that the forward rates at time t are given by

$$f_t^T = \left(1 - e^{-\kappa[T-t]}\right)\left(y_\infty + \frac{\beta^2}{2\kappa^2}e^{-\kappa[T-t]}\right) + e^{-\kappa[T-t]}r_t$$

$$= y_\infty + \left(\frac{\beta^2}{2\kappa^2} + r_t - y_\infty\right)e^{-\kappa[T-t]} - \frac{\beta^2}{2\kappa^2}e^{-2\kappa[T-t]},$$

which is always the same kind of function of time to maturity $T - t$, although the multiplier of $e^{-\kappa[T-t]}$ is non-constant over time due to changes in the short rate. As discussed in Section 9.7 time-inhomogeneous diffusion models generally do not have this nice property, and neither do the HJM-models studied in this chapter. If we use a given parametrization of the initial forward curve, then we cannot be sure that the future forward curves can be described by the same parametrization even if we allow the parameters to be different. We will not discuss this issue further but simply refer the interested reader to Björk and Christensen (1999), who study when the initial forward rate curve and the forward rate sensitivity are consistent in the sense that future forward rate curves have the same form as the initial curve.

If the initial forward rate curve is taken to be of the form given by a time-homogeneous diffusion model and the forward rate volatilities are specified in accordance with that model, then the HJM-model will be indistinguishable from that diffusion model. For example, the time 0 forward rate curve in the one-factor CIR model is of the form

$$f_0^T = r_0 + \hat{\kappa}\left[\hat{\theta} - r\right]b(T) - \frac{1}{2}\beta^2 rb(T)^2,$$

see (7.51), where the function $b(T)$ is given by (7.49). With such an initial forward rate curve, the one-factor HJM model with forward rate volatility function given by (10.11) is indistinguishable from the original time-homogeneous one-factor CIR model.

10.10 EXERCISES

Exercise 10.1 (A Gaussian HJM) Assume that the risk-neutral forward rate dynamics is given by

$$df_t^T = \hat{\alpha}\,(T-t)\;dt + \beta(T-t)\;dz_t^{\mathbb{Q}}, \quad 0 \le t \le T,$$

where $z^{\mathbb{Q}}$ is a standard Brownian motion under $\mathbb{Q}$, and

$$\beta(\tau) = (1 + \gamma\tau)\sigma e^{-\frac{\nu}{2}\tau}$$

for non-negative constants σ, γ, and ν with $2\gamma > \nu$.

(a) Show that the forward rate volatility function $\beta(\tau)$ is humped, that is a $\tau^* > 0$ exists so that β is increasing for $\tau < \tau^*$ and decreasing for $\tau > \tau^*$.
(b) Compute the risk-neutral drift $\hat{\alpha}(\tau)$.
(c) What is the price of a European call option on a zero-coupon bond under the assumptions of this model?

Exercise 10.2 (The Ritchken–Sankarasubramanian model) Show the Equations (10.19), (10.20), and (10.21).

Exercise 10.3 (HJM with USV) Let the risk-neutral forward rate dynamics be given by

$$df_t^T = \hat{\alpha}\,(t, T, v_t)\; dt + \sqrt{v_t} e^{-\kappa[T-t]}\, dz_{1t}^{\mathbb{Q}}, \quad 0 \le t \le T,$$

where

$$dv_t = \mu_v(v_t)\, dt + \sigma_v(v_t)\, dz_{2t}^{\mathbb{Q}},$$

and $z_1^{\mathbb{Q}}$ and $z_2^{\mathbb{Q}}$ are independent standard Brownian motions under $\mathbb{Q}$.

(a) Show that

$$\hat{\alpha}(t, T, v_t) = v_t \frac{1}{\kappa} \left(e^{-\kappa[T-t]} - e^{-2\kappa[T-t]} \right).$$

(b) Show that the risk-neutral dynamics of the zero-coupon bond price is

$$dB_t^T = B_t^T \left[r_t\, dt - b_1(T-t) \sqrt{v_t}\, dz_{1t}^{\mathbb{Q}} \right],$$

where $b_1(\tau) = (1 - e^{-\kappa\tau})/\kappa$.
(c) Show that the short rate can be written as

$$r_t = f_0^t + H_t + I_t,$$

where

$$H_t = \int_0^t \frac{1}{\kappa} e^{-\kappa[t-s]} v_s\, ds + \int_0^t e^{-\kappa[t-s]} \sqrt{v_s}\, dz_{1s}^{\mathbb{Q}}, \quad I_t = \int_0^t \frac{1}{\kappa} e^{-2\kappa[t-s]} v_s\, ds.$$

(d) Show that

$$dr_t = \kappa[\theta_t - r_t]\, dt + \sqrt{v_t}\, dz_{1t}^{\mathbb{Q}},$$

where $\theta_t = \frac{1}{\kappa}\frac{\partial f_0^t}{\partial t} + I_t + f_0^t$.
(e) Explain why zero-coupon bond prices will be of the form

$$B_t^T = e^{-a(t,T) - b_1(T-t) r_t - b_2(T-t)\theta_t},$$

where

$$
\begin{aligned}
b_2(\tau) &= \frac{1}{2\kappa}\left(1 - e^{-\kappa\tau}\right)^2, \\
a(t, T) &= \int_t^T \gamma(s) b_2(T - s)\, ds, \\
\gamma(t) &= \frac{1}{\kappa}\frac{\partial^2 f_0^t}{\partial t^2} + 3\frac{\partial f_0^t}{\partial t} + 2\kappa f_0^t.
\end{aligned}
$$

11

Market Models

11.1 INTRODUCTION

The term structure models studied in the previous chapters have involved assumptions about the evolution in one or more continuously compounded interest rates, either the short rate r_t or the instantaneous forward rates f_t^T. However, many securities traded in the money markets, for example caps, floors, swaps, and swaptions, depend on periodically compounded interest rates such as spot or forward LIBOR rates and spot or forward swap rates. For the pricing of these securities it seems appropriate to apply models that are based on assumptions on the LIBOR rates or the swap rates. Also note that these interest rates are directly observable in the market, whereas the short rate and the instantaneous forward rates are theoretical constructs and not directly observable.

We will use the term **market models** for models based on assumptions on periodically compounded interest rates. All the models studied in this chapter take the currently observed term structure of interest rates as given and are therefore to be classified as relative pricing models or pure no-arbitrage models. Consequently, they offer no insights into the determination of the current interest rates. We will distinguish between **LIBOR market models** that are based on assumptions on the evolution of the forward LIBOR rates and **swap market models** that are based on assumptions on the evolution of the forward swap rates. By construction, the market models are not suitable for the pricing of futures and options on government bonds and similar contracts that do not depend on the money market interest rates.

Several market models have been suggested, but most attention has been given to the so-called lognormal LIBOR market models. In such a model the volatilities of a relevant selection of the forward LIBOR rates are assumed to be proportional to the level of the forward rate so that the distribution of the future forward LIBOR rates is lognormal under an appropriate forward martingale measure. As discussed in Section 7.7, lognormally distributed *continuously compounded* interest rates have unpleasant consequences, but Sandmann and Sondermann (1997) show that models with lognormally distributed *periodically compounded* rates are not subject to the same problems. Below, we will demonstrate that a lognormal assumption on the distribution of forward LIBOR rates implies pricing formulas for caps and floors that are identical to Black's pricing formulas stated in Chapter 6. Similarly, lognormal swap market models imply European swaption prices consistent with the Black formula for swaptions. Hence, the lognormal market models provide some support for the widespread use of Black's formula for fixed income securities. However, the assumptions of the lognormal market models are not necessarily descriptive of the empirical evolution of LIBOR rates, and therefore we will also briefly discuss alternative market models.

The chapter is organized in the following way. Section 11.2 presents the general LIBOR market framework and derives and discusses some basic properties of such models. Section 11.3 specializes to the lognormal LIBOR market models, whereas alternative model specifications are briefly described in Section 11.4. The swap market modelling framework is presented in Section 11.5. Section 11.6 contains some final remarks.

11.2 GENERAL LIBOR MARKET MODELS

In this section we will introduce a general LIBOR market model, describe some of the model's basic properties, and discuss how derivative securities can be priced within the framework of the model. The presentation is inspired by Jamshidian (1997) and Musiela and Rutkowski (1997, Chapters 14 and 16).

11.2.1 Model description

As described in Section 6.4, a cap is a contract that protects a floating rate borrower against paying an interest rate higher than some given rate K, the so-called cap rate. We let $T_1, \dots, T_n$ denote the payment dates and assume that $T_i - T_{i-1} = \delta$ for all i. In addition we define $T_0 = T_1 - \delta$. At each time T_i ($i = 1, \dots, n$) the cap gives a payoff of

$$\mathcal{C}^i_{T_i} = H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right) = H\delta \max\left(L^{T_i-\delta, T_i}_{T_i-\delta} - K, 0\right).$$

Here, recall the notation l^T_t for the spot LIBOR rate prevailing at time t for the time period $[t, T]$, and the notation $L^{T,S}_t$ for the forward LIBOR rate prevailing at time t for the time period $[T, S]$ with $L^{t,T}_t = l^T_t$, of course. Moreover, H represents the face value of the cap. A cap can be considered as a portfolio of caplets, namely one caplet for each payment date.

As discussed in Section 6.4, the value of the above payoff can be found as the product of the expected payoff computed under the T_i-forward martingale measure and the current discount factor for time T_i payments:

$$\mathcal{C}^i_t = H\delta B^{T_i}_t \,\mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(L^{T_i-\delta, T_i}_{T_i-\delta} - K, 0\right)\right], \quad t < T_i - \delta. \tag{11.1}$$

The price of a cap can therefore be determined as

$$\mathcal{C}_t = H\delta \sum_{i=1}^{n} B^{T_i}_t \,\mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(L^{T_i-\delta, T_i}_{T_i-\delta} - K, 0\right)\right], \quad t < T_0. \tag{11.2}$$

For $t \geq T_0$ the first-coming payment of the cap is known so that its present value is obtained by multiplication by the risk-free discount factor, while the remaining payoffs are valued as above. For more details see Section 6.4. The price of the corresponding floor is

$$\mathcal{F}_t = H\delta \sum_{i=1}^{n} B_t^{T_i} \, \mathrm{E}_t^{\mathbb{Q}^{T_i}} \left[\max\left(K - L_{T_i-\delta}^{T_i-\delta,T_i}, 0 \right) \right], \quad t < T_0. \tag{11.3}$$

In order to compute the cap price from (11.2), we need knowledge of the distribution of $L_{T_i-\delta}^{T_i-\delta,T_i}$ under the T_i-forward martingale measure $\mathbb{Q}^{T_i}$ for each $i = 1, \dots, n$. For this purpose it is natural to model the evolution of $L_t^{T_i-\delta,T_i}$ under $\mathbb{Q}^{T_i}$. The following argument shows that, under the $\mathbb{Q}^{T_i}$ probability measure, the drift rate of $L_t^{T_i-\delta,T_i}$ is zero, that is $L_t^{T_i-\delta,T_i}$ is a $\mathbb{Q}^{T_i}$-martingale. Remember from Equation (1.6) that

$$L_t^{T_i-\delta,T_i} = \frac{1}{\delta}\left(\frac{B_t^{T_i-\delta}}{B_t^{T_i}} - 1 \right). \tag{11.4}$$

Under the T_i-forward martingale measure $\mathbb{Q}^{T_i}$ the ratio between the price of any asset and the zero-coupon bond price $B_t^{T_i}$ is a martingale. In particular, the ratio $B_t^{T_i-\delta}/B_t^{T_i}$ is a $\mathbb{Q}^{T_i}$-martingale so that the expected change of the ratio over any time interval is equal to zero under the $\mathbb{Q}^{T_i}$ measure. From the formula above it follows also that the expected change (over any time interval) in the periodically compounded forward rate $L_t^{T_i-\delta,T_i}$ is zero under $\mathbb{Q}^{T_i}$. We summarize the result in the following theorem:

Theorem 11.1 *The forward rate $L_t^{T_i-\delta,T_i}$ is a $\mathbb{Q}^{T_i}$-martingale.*

Consequently, a LIBOR market model is fully specified by the number of factors (that is the number of standard Brownian motions) that influence the forward rates and the forward rate volatility functions. For simplicity, we focus on the one-factor models

$$dL_t^{T_i-\delta,T_i} = \beta\left(t, T_i - \delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j \ge t} \right) dz_t^{T_i}, \quad t < T_i - \delta, \quad i = 1, \dots, n, \tag{11.5}$$

where z^{T_i} is a one-dimensional standard Brownian motion under the T_i-forward martingale measure $\mathbb{Q}^{T_i}$. The symbol $(L_t^{T_j,T_j+\delta})_{T_j \ge t}$ indicates (as in Chapter 10) that the time t value of the volatility function β can depend on the current values of all the modeled forward rates.[1] In the lognormal LIBOR market models we will study in Section 11.3, we have

$$\beta\left(t, T_i - \delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j \ge t} \right) = \gamma(t, T_i - \delta, T_i) L_t^{T_i-\delta,T_i}$$

[1] As for the HJM models in Chapter 10, the general results for the market models hold even when earlier values of the forward rates affect the current dynamics of the forward rates, but such a generalization seems worthless.

for some deterministic function γ. However, until then we continue to discuss the more general specification (11.5).

We see from the general cap pricing formula (11.2) that the cap price also depends on the current discount factors $B_t^{T_1}, B_t^{T_2}, \ldots, B_t^{T_n}$. From (11.4) it follows that $B_t^{T_i} = B_t^{T_i-\delta}(1 + \delta L_t^{T_i-\delta,T_i})$ so that the relevant discount factors can be determined from $B_t^{T_0}$ and the current values of the modelled forward rates, that is $L_t^{T_0,T_1}, L_t^{T_1,T_2}, \ldots, L_t^{T_{n-1},T_n}$. Similarly to the HJM models in Chapter 10, the LIBOR market models take the currently observable values of these rates as given.

11.2.2 The dynamics of all forward rates under the same probability measure

The basic assumption (11.5) of the LIBOR market model involves n different forward martingale measures. In order to better understand the model and to simplify the numerical computation of some security prices we will describe the evolution of the relevant forward rates under the same common probability measure. As discussed in the next subsection, Monte Carlo simulation is often used to compute prices of certain securities in LIBOR market models (Monte Carlo simulation is presented in detail in Chapter 16). It is much simpler to simulate the evolution of the forward rates under a common probability measure than to simulate the evolution of each forward rate under the martingale measure associated with the forward rate. One possibility is to choose one of the n different forward martingale measures used in the assumption of the model. Note that the T_i-forward martingale measure only makes sense up to time T_i. Therefore, it is appropriate to use the forward martingale measure associated with the last payment date, that is the T_n-forward martingale measure $\mathbb{Q}^{T_n}$, since this measure applies to the entire relevant time period. In this context $\mathbb{Q}^{T_n}$ is sometimes referred to as the **terminal measure**. Another obvious candidate for the common probability measure is the spot martingale measure. Let us look at these two alternatives in more detail.

11.2.2.1 The terminal measure

We wish to describe the evolution in all the modelled forward rates under the T_n-forward martingale measure. For that purpose we shall apply the following theorem which outlines how to shift between the different forward martingale measures of the LIBOR market model.

Theorem 11.2 *Assume that the evolution in the LIBOR forward rates $L_t^{T_i-\delta,T_i}$ for $i = 1, \ldots, n$, where $T_i = T_{i-1} + \delta$, is given by* (11.5). *Then the processes $z^{T_i-\delta}$ and z^{T_i} are related as follows:*

$$dz_t^{T_i} = dz_t^{T_i-\delta} + \frac{\delta\beta\left(t, T_i - \delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j \geq t}\right)}{1 + \delta L_t^{T_i-\delta,T_i}}\, dt. \tag{11.6}$$

Proof: From Section 4.4.2 we have that the T_i-forward martingale measure $\mathbb{Q}^{T_i}$ is characterized by the fact that the process z^{T_i} is a standard Brownian motion under $\mathbb{Q}^{T_i}$, where

$$dz_t^{T_i} = dz_t + \left(\lambda_t - \sigma_t^{T_i}\right)\, dt.$$

Here, $\sigma_t^{T_i}$ denotes the volatility of the zero-coupon bond maturing at time T_i, which may itself be stochastic. Similarly,

$$dz_t^{T_i-\delta} = dz_t + \left(\lambda_t - \sigma_t^{T_i-\delta}\right)\, dt.$$

A simple computation gives that

$$dz_t^{T_i} = dz_t^{T_i-\delta} + \left[\sigma_t^{T_i-\delta} - \sigma_t^{T_i}\right]\, dt. \tag{11.7}$$

As shown in Theorem 11.1, $L_t^{T_i-\delta,T_i}$ is a $\mathbb{Q}^{T_i}$-martingale and, hence, has an expected change of zero under this probability measure. According to (11.4) the forward rate $L_t^{T_i-\delta,T_i}$ is a function of the zero-coupon bond prices $B_t^{T_i-\delta}$ and $B_t^{T_i}$ so that the volatility follows from Itô's Lemma. In total, the dynamics is

$$\begin{aligned} dL_t^{T_i-\delta,T_i} &= \frac{B_t^{T_i-\delta}}{\delta B_t^{T_i}}\left(\sigma_t^{T_i-\delta} - \sigma_t^{T_i}\right)\, dz_t^{T_i} \\ &= \frac{1}{\delta}(1+\delta L_t^{T_i-\delta,T_i})\left(\sigma_t^{T_i-\delta} - \sigma_t^{T_i}\right)\, dz_t^{T_i}. \end{aligned}$$

Comparing with (11.5), we can conclude that

$$\sigma_t^{T_i-\delta} - \sigma_t^{T_i} = \frac{\delta\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)}{1+\delta L_t^{T_i-\delta,T_i}}. \tag{11.8}$$

Substituting this relation into (11.7), we obtain the stated relation between the processes z^{T_i} and $z^{T_i-\delta}$. □

Using (11.6) repeatedly, we get that

$$dz_t^{T_n} = dz_t^{T_i} + \sum_{j=i}^{n-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta f_s(t, T_j, T_{j+1})}\, dt.$$

Consequently, for each $i = 1, \ldots, n$, we can write the dynamics of $L_t^{T_i-\delta,T_i}$ under the $\mathbb{Q}^{T_n}$-measure as

$$
\begin{aligned}
dL_t^{T_i-\delta,T_i} &= \beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)\, dz_t^{T_i} \\
&= \beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right) \\
&\quad \left[dz_t^{T_n} - \sum_{j=i}^{n-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta L_t^{T_j,T_{j+1}}}\, dt \right] \\
&= -\sum_{j=i}^{n-1} \frac{\delta\beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta L_t^{T_j,T_{j+1}}}\, dt \\
&\quad + \beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)\, dz_t^{T_n}.
\end{aligned}
\tag{11.9}
$$

Note that the drift may involve some or all of the other modelled forward rates. Therefore, the vector of all the forward rates $(L_t^{T_0,T_1},\dots,L_t^{T_{n-1},T_n})$ will follow an n-dimensional diffusion process so that a LIBOR market model can be represented as an n-factor diffusion model. Security prices are hence solutions to a partial differential equation (PDE), but in typical applications the dimension n, that is the number of forward rates, is so big that neither explicit nor numerical solution of the PDE is feasible.[2] For example, to price caps, floors, and swaptions that depend on 3-month interest rates and have maturities of up to 10 years, one must model 40 forward rates so that the model is a 40-factor diffusion model!

Next, let us consider an asset with a single payoff at some point in time $T \in [T_0, T_n]$. The payoff H_T may in general depend on the value of all the modelled forward rates at and before time T. Let P_t denote the time t value of this asset (measured in monetary units, for example dollars). From the definition of the T_n-forward martingale measure $\mathbb{Q}^{T_n}$ it follows that

$$\frac{P_t}{B_t^{T_n}} = \mathrm{E}_t^{\mathbb{Q}^{T_n}}\left[\frac{H_T}{B_T^{T_n}}\right],$$

and hence

$$P_t = B_t^{T_n}\, \mathrm{E}_t^{\mathbb{Q}^{T_n}}\left[\frac{H_T}{B_T^{T_n}}\right].$$

In particular, if T is one of the time points of the tenor structure, say $T = T_k$, we get

$$P_t = B_t^{T_n}\, \mathrm{E}_t^{\mathbb{Q}^{T_n}}\left[\frac{H_{T_k}}{B_{T_k}^{T_n}}\right].$$

[2] However, Andersen and Andreasen (2000) introduce a trick that can reduce the computational complexity considerably.

From (11.4) we have that

$$\begin{aligned}
\frac{1}{B_{T_k}^{T_n}} &= \frac{B_{T_k}^{T_k}}{B_{T_k}^{T_{k+1}}} \frac{B_{T_k}^{T_{k+1}}}{B_{T_k}^{T_{k+2}}} \cdots \frac{B_{T_k}^{T_{n-1}}}{B_{T_k}^{T_n}} \\
&= \left[1 + \delta L_{T_k}^{T_k,T_{k+1}}\right]\left[1 + \delta L_{T_k}^{T_{k+1},T_{k+2}}\right] \cdots \left[1 + \delta L_{T_k}^{T_{n-1},T_n}\right] \\
&= \prod_{j=k}^{n-1} \left[1 + \delta L_{T_k}^{T_j,T_{j+1}}\right]
\end{aligned}$$

so that the price can be rewritten as

$$P_t = B_t^{T_n} \, \mathrm{E}_t^{\mathbb{Q}^{T_n}} \left[H_{T_k} \prod_{j=k}^{n-1} \left[1 + \delta L_{T_k}^{T_j,T_{j+1}}\right] \right].$$

The right-hand side may be approximated using Monte Carlo simulations in which the evolution of the forward rates under $\mathbb{Q}^{T_n}$ is used, as outlined in (11.9).

If the security matures at time T_n, the price expression is even simpler:

$$P_t = B_t^{T_n} \, \mathrm{E}_t^{\mathbb{Q}^{T_n}} \left[H_{T_n}\right].$$

In that case it suffices to simulate the evolution of the forward rates that determine the payoff of the security.

11.2.2.2 The spot LIBOR martingale measure

The risk-neutral or spot martingale measure $\mathbb{Q}$, which we defined and discussed in Chapter 4, is associated with the use of a bank account earning the continuously compounded short rate as the numeraire, see the discussion in Section 4.4. However, the LIBOR market model does not at all involve the short rate so the traditional spot martingale measure does not make sense in this context. The LIBOR market counterpart is a **roll over** strategy in the shortest zero-coupon bonds. To be more precise, the strategy is initiated at time T_0 by an investment of 1 dollar in the zero-coupon bond maturing at time T_1, which allows for the purchase of $1/B_{T_0}^{T_1}$ units of the bond. At time T_1 the payoff of $1/B_{T_0}^{T_1}$ dollars is invested in the zero-coupon bond maturing at time T_2, and so on. Let us define

$$I(t) = \min\{i \in \{1, 2, \ldots, n\} : T_i \geq t\}$$

so that $T_{I(t)}$ denotes the next payment date after time t. In particular, $I(T_i) = i$ so that $T_{I(T_i)} = T_i$. At any time $t \geq T_0$ the strategy consists of holding

$$N_t = \frac{1}{B_{T_0}^{T_1}} \frac{1}{B_{T_1}^{T_2}} \cdots \frac{1}{B_{T_{I(t)-1}}^{T_{I(t)}}}$$

units of the zero-coupon bond maturing at time $T_{I(t)}$. The value of this position is

$$A_t^* = B_t^{T_{I(t)}} N_t = B_t^{T_{I(t)}} \prod_{j=0}^{I(t)-1} \frac{1}{B_{T_j}^{T_{j+1}}} = B_t^{T_{I(t)}} \prod_{j=0}^{I(t)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right],$$

where the last equality follows from the relation (11.4). Since A_t^* is positive, it is a valid numeraire. The corresponding martingale measure is called the **spot LIBOR martingale measure** and is denoted by $\mathbb{Q}^*$.

Let us look at a security with a single payment at a time $T \in [T_0, T_n]$. The payoff H_T may depend on the values of all the modelled forward rates at and before time T. Let us by P_t denote the dollar value of this asset at time t. From the definition of the spot LIBOR martingale measure $\mathbb{Q}^*$ it follows that

$$\frac{P_t}{A_t^*} = \mathrm{E}_t^{\mathbb{Q}^*}\left[\frac{H_T}{A_T^*}\right],$$

and hence

$$P_t = \mathrm{E}_t^{\mathbb{Q}^*}\left[\frac{A_t^*}{A_T^*} H_T\right].$$

From the calculation

$$\begin{aligned}
\frac{A_t^*}{A_T^*} &= \frac{B_t^{T_{I(t)}} \prod_{j=0}^{I(t)-1}\left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]}{B_T^{T_{I(T)}} \prod_{j=0}^{I(T)-1}\left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]} \\
&= \frac{B_t^{T_{I(t)}}}{B_T^{T_{I(T)}}} \prod_{j=I(t)}^{I(T)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]^{-1},
\end{aligned}$$

we get that the price can be rewritten as

$$P_t = B_t^{T_{I(t)}} \mathrm{E}_t^{\mathbb{Q}^*}\left[\frac{H_T}{B_T^{T_{I(T)}}} \prod_{j=I(t)}^{I(T)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]^{-1}\right].$$

In particular, if T is one of the dates in the tenor structure, say $T = T_k$, we get

$$P_t = B_t^{T_{I(t)}} \mathrm{E}_t^{\mathbb{Q}^*}\left[H_{T_k} \prod_{j=I(t)}^{k-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]^{-1}\right] \tag{11.10}$$

since $I(T_k) = k$ and $B_{T_k}^{T_{I(T_k)}} = B_{T_k}^{T_k} = 1$.

In order to compute (typically by simulation) the expected value on the right-hand side, we need to know the evolution of the forward rates $L_t^{T_j,T_{j+1}}$ under the spot LIBOR martingale measure $\mathbb{Q}^*$. It can be shown (see Exercise 11.1) that the process z^* defined by

$$dz_t^* = dz_t^{T_i} - \left[\sigma_t^{T_{I(t)}} - \sigma_t^{T_i}\right] dt \tag{11.11}$$

is a standard Brownian motion under the probability measure $\mathbb{Q}^*$. As usual, σ_t^T denotes the volatility of the zero-coupon bond maturing at time T. Repeated use of (11.8) yields

$$\sigma_t^{T_{I(t)}} - \sigma_t^{T_i} = \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta L_t^{T_j,T_{j+1}}}$$

so that

$$dz_t^* = dz_t^{T_i} - \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta L_t^{T_j,T_{j+1}}}\, dt.$$

Substituting this relation into (11.5), we can rewrite the dynamics of the forward rates under the spot LIBOR martingale measure as

$$\begin{aligned}
dL_t^{T_i-\delta,T_i} &= \beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right) dz_t^{T_i}\\
&= \beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)\\
&\quad \left[dz_t^* + \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta L_t^{T_j,T_{j+1}}}\, dt\right]\\
&= \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right)}{1+\delta L_t^{T_j,T_{j+1}}}\, dt\\
&\quad +\beta\left(t, T_i-\delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k\geq t}\right) dz_t^*.
\end{aligned}$$

Note that the drift in the forward rates under the spot LIBOR martingale measure follows from the specification of the volatility function β and the current forward rates. The relation between the drift and the volatility is the market model counterpart to the drift restriction of the HJM models, see (10.5).

11.2.3 Consistent pricing

As indicated above, the model can be used for the pricing of all securities that only have payment dates in the set $\{T_1, T_2, \ldots, T_n\}$, and where the size of the payment only depends on the modelled forward rates and no other random variables. This is true for caps and floors on δ-period interest rates of different maturities where the price can be computed from (11.2) and (11.3). The model can also be used for the pricing of swaptions that expire on one of the dates $T_0, T_1, \dots, T_{n-1}$, and where the underlying swap has payment dates in the set $\{T_1, \dots, T_n\}$ and is based on the δ-period interest rate. For European swaptions the price can be written as (11.10). For Bermuda swaptions that can be exercised at a subset of the swap payment dates $\{T_1, \dots, T_n\}$, one must maximize the right-hand side of (11.10) over all feasible exercise strategies. See Andersen (2000) for details and a description of a relatively simple Monte Carlo based method for the approximation of Bermuda swaption prices.

The LIBOR market model (11.5) is built on assumptions about the forward rates over the time intervals $[T_0, T_1], [T_1, T_2], \dots, [T_{n-1}, T_n]$. However, these forward rates determine the forward rates for periods that are obtained by connecting succeeding intervals. For example, we have from Equation (1.6) that the forward rate over the period $[T_0, T_2]$ is uniquely determined by the forward rates for the periods $[T_0, T_1]$ and $[T_1, T_2]$ since

$$
\begin{aligned}
L_t^{T_0,T_2} &= \frac{1}{T_2 - T_0}\left(\frac{B_t^{T_0}}{B_t^{T_2}} - 1\right) \\
&= \frac{1}{T_2 - T_0}\left(\frac{B_t^{T_0}}{B_t^{T_1}}\frac{B_t^{T_1}}{B_t^{T_2}} - 1\right) \\
&= \frac{1}{2\delta}\left(\left[1 + \delta L_t^{T_0,T_1}\right]\left[1 + \delta L_t^{T_1,T_2}\right] - 1\right),
\end{aligned}
\tag{11.12}
$$

where $\delta = T_1 - T_0 = T_2 - T_1$ as usual. Therefore, the distributions of the forward rates $L_t^{T_0,T_1}$ and $L_t^{T_1,T_2}$ implied by the LIBOR market model (11.13), determine the distribution of the forward rate $L_t^{T_0,T_2}$. A LIBOR market model based on 3-month interest rates can hence also be used for the pricing of contracts that depend on 6-month interest rates, as long as the payment dates for these contracts are in the set $\{T_0, T_1, \dots, T_n\}$. More generally, in the construction of a model, one is only allowed to make exogenous assumptions about the evolution of forward rates for non-overlapping periods.

11.3 THE LOGNORMAL LIBOR MARKET MODEL

11.3.1 Model description

The market standard for the pricing of caplets is Black's formula, that is formula (6.24). As discussed in Section 4.8, the traditional derivation of Black's formula is based on inappropriate assumptions. The lognormal LIBOR market model provides a more reasonable framework in which the Black cap formula is valid. The model was originally developed by Miltersen, Sandmann, and Sondermann (1997), whereas Brace, Gatarek, and Musiela (1997) sort out some technical details and introduce an explicit, but approximative, expression for the prices of European swaptions in the lognormal LIBOR market model. Miltersen, Sandmann, and Sondermann (1997) derive the cap price formula using PDEs, but we will follow Brace, Gatarek, and Musiela (1997) and use the forward martingale measure technique discussed in Chapter 4 since this simplifies the analysis considerably.

Looking at the general cap pricing formula (11.2), it is clear that we can obtain a pricing formula of the same form as Black's formula by assuming that $L_{T_i-\delta}^{T_i-\delta,T_i}$ is lognormally distributed under the T_i-forward martingale measure $\mathbb{Q}^{T_i}$. This is exactly the assumption of the **lognormal LIBOR market model**:

$$dL_t^{T_i-\delta,T_i} = L_t^{T_i-\delta,T_i}\gamma(t,T_i-\delta,T_i)\,dz_t^{T_i}, \quad i=1,2,\dots,n, \tag{11.13}$$

where $\gamma(t,T_i-\delta,T_i)$ is a bounded, deterministic function. Here we assume that the relevant forward rates are only affected by one Brownian motion, but below we shall briefly consider multi-factor lognormal LIBOR market models.

A familiar application of Itô's Lemma implies that

$$d(\ln L_t^{T_i-\delta,T_i}) = -\frac{1}{2}\gamma(t,T_i-\delta,T_i)^2\,dt + \gamma(t,T_i-\delta,T_i)\,dz_t^{T_i},$$

from which we see that

$$\begin{aligned}\ln L_{T_i-\delta}^{T_i-\delta,T_i} &= \ln L_t^{T_i-\delta,T_i} - \frac{1}{2}\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du \\ &\quad + \int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)\,dz_u^{T_i}.\end{aligned}$$

Because γ is a deterministic function, it follows from Theorem 3.3 that

$$\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)\,dz_u^{T_i} \sim N\left(0,\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du\right)$$

under the T_i-forward martingale measure. Hence,

$$\begin{aligned}\ln L_{T_i-\delta}^{T_i-\delta,T_i} \sim N\Bigg(\ln L_t^{T_i-\delta,T_i} &- \frac{1}{2}\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du, \\ &\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du\Bigg)\end{aligned}$$

so that $L_{T_i-\delta}^{T_i-\delta,T_i}$ is lognormally distributed under $\mathbb{Q}^{T_i}$. The following result should now come as no surprise:

Theorem 11.3 *Under the assumption* (11.13) *the price at time $t < T_i-\delta$ of the caplet with payment date T_i is given by*

$$C_t^i = H\delta B_t^{T_i}\left[L_t^{T_i-\delta,T_i}N(d_{1i}) - KN(d_{2i})\right], \tag{11.14}$$

where

$$d_{1i} = \frac{\ln\left(L_t^{T_i-\delta,T_i}/K\right)}{v_L(t,T_i-\delta,T_i)} + \frac{1}{2}v_L(t,T_i-\delta,T_i), \tag{11.15}$$

$$d_{2i} = d_{1i} - v_L(t,T_i-\delta,T_i), \tag{11.16}$$

$$v_L(t,T_i-\delta,T_i) = \left(\int_t^{T_i-\delta} \gamma(u,T_i-\delta,T_i)^2\,du\right)^{1/2}. \tag{11.17}$$

Proof: It follows from Theorem A.4 in Appendix A that

$$\begin{aligned} \mathrm{E}_t^{\mathbb{Q}^{T_i}}\left[\max\left(L_{T_i-\delta}^{T_i-\delta,T_i} - K, 0\right)\right] &= \mathrm{E}_t^{\mathbb{Q}^{T_i}}\left[L_{T_i-\delta}^{T_i-\delta,T_i}\right]N(d_{1i}) - KN(d_{2i}) \\ &= L_t^{T_i-\delta,T_i}N(d_{1i}) - KN(d_{2i}), \end{aligned}$$

where the last equality is due to the fact that $L_t^{T_i-\delta,T_i}$ is a $\mathbb{Q}^{T_i}$-martingale. The claim now follows from (11.1). □

Note that $v_L(t,T_i-\delta,T_i)^2$ is the variance of $\ln L_{T_i-\delta}^{T_i-\delta,T_i}$ under the T_i-forward martingale measure given the information available at time t. The expression (11.14) is identical to Black's formula (6.24) if we insert $\sigma_i = v_L(t,T_i-\delta,T_i)/\sqrt{T_i-\delta-t}$. An immediate consequence of the theorem above is the following cap pricing formula in the lognormal one-factor LIBOR market model:

Theorem 11.4 *Under the assumption* (11.13) *the price of a cap at any time* $t < T_0$ *is given as*

$$\mathcal{C}_t = H\delta\sum_{i=1}^{n} B_t^{T_i}\left[L_t^{T_i-\delta,T_i}N(d_{1i}) - KN(d_{2i})\right], \tag{11.18}$$

where d_{1i} *and* d_{2i} *are as in* (11.15) *and* (11.16).

For $t \geq T_0$ the first-coming payment of the cap is known and is therefore to be discounted with the risk-free discount factor, while the remaining payments are to be valued as above. For details, see Section 6.4.

Analogously, the price of a floor under the assumption (11.13) is

$$\mathcal{F}_t = H\delta\sum_{i=1}^{n} B_t^{T_i}\left[KN(-d_{2i}) - L_t^{T_i-\delta,T_i}N(-d_{1i})\right], \quad t < T_0.$$

The deterministic function $\gamma(t,T_i-\delta,T_i)$ remains to be specified. We will discuss this matter in Section 11.6.

If the term structure is affected by d exogenous standard Brownian motions, the assumption (11.13) is replaced by

$$dL_t^{T_i-\delta,T_i} = L_t^{T_i-\delta,T_i}\sum_{j=1}^{d}\gamma_j(t,T_i-\delta,T_i)\,dz_{jt}^{T_i},$$

where all $\gamma_j(t, T_i - \delta, T_i)$ are bounded and deterministic functions. Again, the cap price is given by (11.18) with the small change that $v_L(t, T_i - \delta, T_i)$ is to be computed as

$$v_L(t, T_i - \delta, T_i) = \left(\sum_{j=1}^{d} \int_t^{T_i - \delta} \gamma_j(u, T_i - \delta, T_i)^2 \, du \right)^{1/2}. \tag{11.19}$$

11.3.2 The pricing of other securities

No exact, explicit solution for European swaptions has been found in the lognormal LIBOR market setting. In particular, Black's formula for swaptions is not correct under the assumption (11.13). The reason is that when the forward LIBOR rates $L_t^{T_i-\delta,T_i}$ have volatilities proportional to their level, the volatility of the forward swap rate $\tilde{L}_t^{T_0,\delta}$ will not be proportional to the level of the forward swap rate. As described in Section 11.2, the swaption price can be approximated by a Monte Carlo simulation, which is often quite time-consuming. Brace, Gatarek, and Musiela (1997) derive the following Black-type approximation to the price of a European payer swaption with expiration date T_0 and exercise rate K under the lognormal LIBOR market model assumptions:

$$\mathcal{P}_t = H\delta \sum_{i=1}^{n} B_t^{T_i} \left[L_t^{T_i-\delta,T_i} N(d_{1i}^*) - KN(d_{2i}^*) \right], \quad t < T_0, \tag{11.20}$$

where d_{1i}^* and d_{2i}^* are quite complicated expressions involving the variances and covariances of the time T_0 values of the forward rates involved. These variances and covariances are determined by the γ-function of the assumption (11.13). This approximation delivers the price much faster than a Monte Carlo simulation. Brace et al. provide numerical examples in which the price computed using the approximation (11.20) is very close to the correct price (computed using Monte Carlo simulations). Of course, a similar approximation applies to the European receiver swaption. The market models are not constructed for the pricing of bond options, but due to the link between caps/floors and European options on zero-coupon bonds it is possible to derive some bond option pricing formulas, see Exercise 11.2.

As argued in Section 11.2, in any LIBOR market model based on the δ-period interest rates, one can also price securities that depend on interest rates over periods of length 2δ, 3δ, and so on, as long as the payment dates of these securities are in the set $\{T_0, T_1, \dots, T_n\}$. Of course, this is also true for the lognormal LIBOR market model. For example, let us consider contracts that depend on interest rates covering periods of length 2δ. From (11.12) we have that

$$L_t^{T_0,T_2} = \frac{1}{2\delta} \left(\left[1 + \delta L_t^{T_0,T_1} \right] \left[1 + \delta L_t^{T_1,T_2} \right] - 1 \right). \tag{11.21}$$

According to the assumption (11.13) of the lognormal δ-period LIBOR market model, each of the forward rates on the right-hand side has a volatility proportional to the level of the forward rate. An application of Itô's Lemma to the above relation shows that the same proportionality does not hold for the 2δ-period forward rate $L_t^{T_0,T_2}$ (see Exercise 11.3). Consequently, Black's cap formula cannot be correct both for caps on the 3-month rate and caps on the 6-month rate. To price caps on the 6-month rate consistently with the assumptions of the lognormal LIBOR market model for the 3-month rate one must resort to numerical methods, such as Monte Carlo simulation.

It follows from the above considerations that the model cannot justify practitioners' frequent use of Black's formula for both caps and swaptions and for contracts with different frequencies δ. Of course, the differences between the prices generated by Black's formula and the correct prices according to some reasonable model may be so small that this inconsistency can be ignored, but apparently this issue has not been fully investigated in the literature.

11.4 ALTERNATIVE LIBOR MARKET MODELS

The lognormal LIBOR market model specifies the forward rate volatility in the general LIBOR market model (11.5) as

$$\beta\left(t, T_i - \delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j \geq t}\right) = L_t^{T_i-\delta,T_i} \gamma(t, T_i - \delta, T_i),$$

where γ is a deterministic function. As we have seen, this specification has the advantage that the prices of (some) caps and floors are given by Black's formula. However, observed cap prices are not completely consistent with Black's formula, which motivates the search for alternative, more realistic volatility specifications.

European stock option prices are often transformed into implicit volatilities using the Black–Scholes–Merton formula. Similarly, for each caplet we can determine an implicit volatility for the corresponding forward rate as the value of the parameter σ_i that makes the caplet price computed using Black's formula (6.24) identical to the observed market price. Suppose that several caplets are traded on the same forward rate and with the same payment date, but with different cap rates (that is exercise rates) K. Then we get a relation $\sigma_i(K)$ between the implicit volatilities and the cap rate. If the forward rate has a proportional volatility, Black's model will be correct for all these caplets. In that case all the implicit volatilities will be equal so that $\sigma_i(K)$ corresponds to a flat line. However, according to Andersen and Andreasen (2000), $\sigma_i(K)$ is typically decreasing in K, which is referred to as a **volatility skew**. Such a skew is inconsistent with the volatility assumption of the LIBOR market model (11.13).[3] See also the empirical evidence in Jarrow et al. (2007).

Andersen and Andreasen (2000) consider a so-called CEV LIBOR market model where the forward rate volatility is given as

$$\beta\left(t, T_i - \delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j \geq t}\right) = \left(L_t^{T_i-\delta,T_i}\right)^{\alpha} \gamma(t, T_i - \delta, T_i), \quad i = 1, \dots, n,$$

[3] Hull (2009, Ch. 18) has a detailed discussion of the similar phenomenon for stock and currency options.

so that each forward rate follows a CEV process[4]

$$dL_t^{T_i-\delta,T_i} = \left(L_t^{T_i-\delta,T_i}\right)^{\alpha} \gamma(t, T_i-\delta, T_i)\, dz_t^{T_i}.$$

Here α is a positive constant and γ is a bounded, deterministic function, which in general may be vector-valued, but here we have assumed that it takes values in $\mathbb{R}$. For $\alpha = 1$, the model is identical to the lognormal LIBOR market model. Andersen and Andreasen first discuss properties of CEV processes. When $0 < \alpha < 1/2$, several processes may have the dynamics given above, but a unique process is fixed by requiring that zero is an absorbing boundary for the process. Imposing this condition, the authors are able to state in closed form the distribution of future values of the process for any positive α. For $\alpha \neq 1$, this distribution is closely linked to the distribution of a non-centrally χ^2-distributed random variable.

Based on their analysis of the CEV process, Andersen and Andreasen next show that the price of a caplet will have the form

$$C_t^i = H\delta B_t^{T_i}\left[L_t^{T_i-\delta,T_i}\left(1-\chi^2(a;b,c)\right) - K\chi^2(c;b',a)\right]$$

for some auxiliary parameters a, b, b', and c that we leave unspecified here. The pricing formula is very similar to Black's formula, but the relevant probabilities are given by the distribution function for a non-central χ^2-distribution. Their numerical examples document that a CEV model with $\alpha < 1$ can generate the volatility skew observed in practice. In addition, they give an explicit approximation to the price of a European swaption in their CEV LIBOR market model. Also this pricing formula is of the same form as Black's formula, but involves the distribution function for the non-central χ^2-distribution instead of the normal distribution.

Andersen and Brotherton-Ratcliffe (2001), Hagan, Kumar, Lesniewski, and Woodward (2002), Glasserman and Kou (2003), Joshi and Rebonato (2003), and Jarrow, Li, and Zhao (2007) have suggested and studied various LIBOR market models with stochastic volatility and possible jumps.

[4] CEV is short for Constant Elasticity of Variance. This term arises from the fact that the elasticity of the volatility with respect to the forward rate level is equal to the constant α since

$$\begin{aligned}
&\frac{\partial\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)/\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)}{\partial L_t^{T_i-\delta,T_i}/L_t^{T_i-\delta,T_i}} \\
&\quad = \frac{\partial\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)}{\partial L_t^{T_i-\delta,T_i}} \frac{L_t^{T_i-\delta,T_i}}{\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)} \\
&\quad = \alpha\left(L_t^{T_i-\delta,T_i}\right)^{\alpha-1}\gamma(t, T_i-\delta, T_i)\frac{L_t^{T_i-\delta,T_i}}{\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)} = \alpha.
\end{aligned}$$

Cox and Ross (1976) study a similar variant of the Black–Scholes–Merton model for stock options.

11.5 SWAP MARKET MODELS

Jamshidian (1997) introduced the so-called swap market models that are based on assumptions about the evolution of certain forward swap rates. Under the assumption of a proportional volatility of these forward swap rates, the models will imply that Black's formula for European swaptions, Equation (6.37), is correct, at least for some swaptions.

Given time points $T_0, T_1, \dots, T_n$, where $T_i = T_{i-1} + \delta$ for all $i = 1, \dots, n$, we will refer to a payer swap with start date T_k and final payment date T_n (that is payment dates $T_{k+1}, \dots, T_n$) as a (k, n)-payer swap. Here we must have $1 \leq k < n$. Let us by $\tilde{L}_t^{T_k,\delta}$ denote the forward swap rate prevailing at time $t \leq T_k$ for a (k, n)-swap. Analogous to (6.34), we have that

$$\tilde{L}_t^{T_k,\delta} = \frac{B_t^{T_k} - B_t^{T_n}}{\delta G_t^{k,n}}, \tag{11.22}$$

where we have introduced the notation

$$G_t^{k,n} = \sum_{i=k+1}^{n} B_t^{T_i}$$

for the value of an annuity bond paying 1 dollar at each date $T_{k+1}, \dots, T_n$.

A European payer (k, n)-swaption gives the right at time T_k to enter into a (k, n)-payer swap where the fixed rate K is identical to the exercise rate of the swaption. From (6.35) we know that the value of this swaption at the expiration date T_k is given by

$$\mathcal{P}_{T_k}^{k,n} = G_{T_k}^{k,n} H\delta \max\left(\tilde{L}_{T_k}^{T_k,\delta} - K, 0\right). \tag{11.23}$$

As discussed in Section 6.5.2, it is computationally convenient to use the annuity as the numeraire. We refer to the corresponding martingale measure $\mathbb{Q}^{k,n}$ as the (k, n)-**swap martingale measure**. Since $G_t^{k,k+1} = B_t^{T_{k+1}}$, we have in particular that the $(k, k+1)$-swap martingale measure $\mathbb{Q}^{k,k+1}$ is identical to the T_{k+1}-forward martingale measure $\mathbb{Q}^{T_{k+1}}$.

By the definition of $\mathbb{Q}^{k,n}$, the time t price P_t of a security paying H_{T_k} at time T_k is given by

$$\frac{P_t}{G_t^{k,n}} = \mathrm{E}_t^{\mathbb{Q}^{k,n}}\left[\frac{H_{T_k}}{G_{T_k}^{k,n}}\right],$$

and hence

$$P_t = G_t^{k,n}\, \mathrm{E}_t^{\mathbb{Q}^{k,n}}\left[\frac{H_{T_k}}{G_{T_k}^{k,n}}\right]. \tag{11.24}$$

The pricing formula (11.24) is particularly convenient for the (k,n)-swaption. Inserting the payoff from (11.23), we obtain a price of

$$\mathcal{P}_t^{k,n} = G_t^{k,n} H\delta \, \mathrm{E}_t^{\mathbb{Q}^{k,n}} \left[\max\left(\tilde{L}_{T_k}^{T_k,\delta} - K, 0 \right) \right]. \tag{11.25}$$

To price the swaption it suffices to know the distribution of the swap rate $\tilde{L}_{T_k}^{T_k,\delta}$ under the (k,n)-swap martingale measure $\mathbb{Q}^{k,n}$. Here the following result comes in handy:

Theorem 11.5 *The forward swap rate $\tilde{L}_t^{T_k,\delta}$ is a $\mathbb{Q}^{k,n}$-martingale.*

Proof: According to (11.22), the forward swap rate is given as

$$\tilde{L}_t^{T_k,\delta} = \frac{B_t^{T_k} - B_t^{T_n}}{\delta G_t^{k,n}} = \frac{1}{\delta} \left(\frac{B_t^{T_k}}{G_t^{k,n}} - \frac{B_t^{T_n}}{G_t^{k,n}} \right).$$

By definition of the (k,n)-swap martingale measure the price of any security relative to the annuity is a martingale under this probability measure. In particular, both $B_t^{T_k}/G_t^{k,n}$ and $B_t^{T_n}/G_t^{k,n}$ are $\mathbb{Q}^{k,n}$-martingales. Therefore, the expected change in these ratios is zero under $\mathbb{Q}^{k,n}$. It follows from the above formula that the expected change in the forward swap rate $\tilde{L}_t^{T_k,\delta}$ is also zero under $\mathbb{Q}^{k,n}$ so that $\tilde{L}_t^{T_k,\delta}$ is a $\mathbb{Q}^{k,n}$-martingale. □

Consequently, the evolution in the forward swap rate $\tilde{L}_t^{T_k,\delta}$ is fully specified by (i) the number of Brownian motions affecting this and other modelled forward swap rates, and (ii) the sensitivity functions that show how the forward swap rates react to the exogenous shocks. Let us again focus on a one-factor model. A swap market model is based on the assumption

$$d\tilde{L}_t^{T_k,\delta} = \beta^{k,n}\left(t, (\tilde{L}_t^{T_j,\delta})_{T_j \geq t}\right) dz_t^{k,n},$$

where $z^{k,n}$ is a Brownian motion under the (k,n)-swap martingale measure $\mathbb{Q}^{k,n}$, and the volatility function $\beta^{k,n}$ through the term $(\tilde{L}_t^{T_j,\delta})_{T_j \geq t}$ can depend on the current values of all the modelled forward swap rates.

Under the assumption that $\beta^{k,n}$ is proportional to the level of the forward swap rate, that is

$$d\tilde{L}_t^{T_k,\delta} = \tilde{L}_t^{T_k,\delta} \gamma^{k,n}(t)\, dz_t^{k,n} \tag{11.26}$$

where $\gamma^{k,n}(t)$ is a bounded, deterministic function, we get that the future value of the forward swap rate is lognormally distributed. This model is therefore referred to as the **lognormal swap market model**. In such a model the swaption price in Equation (11.25) can be computed explicitly:

Theorem 11.6 *Under the assumption* (11.26) *the price of a European* (k,n)*-payer swaption is given by*

$$\mathcal{P}_t^{k,n} = \left(\sum_{i=k+1}^{n} B_t^{T_i} \right) H\delta \left[\tilde{L}_t^{T_k,\delta} N(d_1) - KN(d_2) \right], \quad t < T_k,$$

where

$$d_1 = \frac{\ln \left(\tilde{L}_t^{T_k,\delta} / K \right)}{v_{k,n}(t)} + \frac{1}{2} v_{k,n}(t),$$

$$d_2 = d_1 - v_{k,n}(t),$$

$$v_{k,n}(t) = \left(\int_t^{T_k} \gamma^{k,n}(u)^2 \, du \right)^{1/2}.$$

The proof of this result is analogous to the proof of Theorem 11.3 and is therefore omitted. The pricing formula is identical to Black's Equation (6.37) with σ given by $\sigma = v_{k,n}(t)/\sqrt{T_k - t}$. Hence, the lognormal swap market model provides some theoretical support for the Black swaption pricing formula.

In a previous section we concluded that in a LIBOR market model it is not justifiable to exogenously specify the processes for all forward rates, only the processes for non-overlapping periods. In a swap market model (Musiela and Rutkowski, 1997, Section 14.4) demonstrate that the processes for the forward swap rates $\tilde{L}_t^{T_1,\delta}, \tilde{L}_t^{T_2,\delta}, \dots, \tilde{L}_t^{T_{n-1},\delta}$ can be modelled independently. These are forward swap rates for swaps with the same final payment date T_n, but with different start dates $T_1, \dots, T_{n-1}$ and hence different maturities. In particular, the lognormal assumption (11.26) can hold for all these forward swap rates, which implies that all the swaption prices $\mathcal{P}_t^{1,n}, \dots, \mathcal{P}_t^{n-1,n}$ are given by Black's swaption pricing formula. However, under such an assumption neither the forward LIBOR rates $L_t^{T_{i-1},T_i}$ nor the forward swap rates for swaps with other final payment dates can have proportional volatilities. Consequently, Black's formula cannot be correct for caps and floors, nor for swaptions with other maturity dates. The correct prices of these securities must be computed using numerical methods, such as Monte Carlo simulation. Also in this case it is not clear by how much the Black pricing formulas miss the theoretically correct prices.

In the context of the LIBOR market models we have derived relations between the different forward martingale measures. For the swap market models we can derive similar relations between the different swap martingale measures and hence describe the dynamics of all the forward swap rates $\tilde{L}_t^{T_1,\delta}, \tilde{L}_t^{T_2,\delta}, \dots, \tilde{L}_t^{T_{n-1},\delta}$ under the same probability measure. Then all the relevant processes can be simulated under the same probability measure. For details the reader is referred to Jamshidian (1997) and Musiela and Rutkowski (1997, Section 14.4).

11.6 FURTHER REMARKS

De Jong, Driessen, and Pelsser (2001) investigate the extent to which different lognormal LIBOR and swap market models can explain empirical data consisting of forward LIBOR interest rates, forward swap rates, and prices of caplets and European swaptions. The observations are from the U.S. market in 1995 and 1996. For the lognormal one-factor LIBOR market model (11.13) they find that it is empirically more appropriate to use a γ-function which is exponentially decreasing in the time-to-maturity $T_i - \delta - t$ of the forward rates,

$$\gamma(t, T_i - \delta, T_i) = \gamma e^{-\kappa[T_i - \delta - t]}, \quad i = 1, \ldots, n,$$

than to use a constant, $\gamma(t, T_i - \delta, T_i) = \gamma$. This is related to the well-documented mean reversion of interest rates that makes 'long' interest rates relatively less volatile than 'short' interest rates. They also calibrate two similar model specifications perfectly to observed caplet prices, but find that in general the prices of swaptions in these models are further from the market prices than are the prices in the time-homogeneous models above. In all cases the swaption prices computed using one of these lognormal LIBOR market models exceed the market prices, that is the lognormal LIBOR market models overestimate the swaption prices. All their specifications of the lognormal one-factor LIBOR market model give a relatively inaccurate description of market data and are rejected by statistical tests. De Jong, Driessen, and Pelsser (2001) also show that two-factor lognormal LIBOR market models are not significantly better than the one-factor models and conclude that the lognormality assumption is probably inappropriate. Finally, they present similar results for lognormal swap market models and find that these models are even worse than the lognormal LIBOR market models when it comes to fitting the data. For other empirical studies involving market models, see Gupta and Subrahmanyam (2005) and Jarrow, Li, and Zhao (2007).

For additional information on market models, the reader may consult the detailed presentation of Brigo and Mercurio (2006).

11.7 EXERCISES

Exercise 11.1 (Spot LIBOR measure) Explain why the process z^* defined by (11.11) is a standard Brownian motion under the probability measure $\mathbb{Q}^*$.

Exercise 11.2 (Caplets and options on zero-coupon bonds) Assume that the lognormal LIBOR market model holds. Use the caplet formula (11.14) and the relations between caplets, floorlets, and European bond options known from Chapter 6 to show that the following pricing formulas for European options on zero-coupon bonds are valid:

$$C_t^{K,T_i-\delta,T_i} = (1-K)B_t^{T_i}N(e_{1i}) - K[B_t^{T_i-\delta} - B_t^{T_i}]N(e_{2i}),$$

$$\pi_t^{K,T_i-\delta,T_i} = K[B_t^{T_i-\delta} - B_t^{T_i}]N(-e_{2i}) - (1-K)B_t^{T_i}N(-e_{1i}),$$

where

$$e_{1i} = \frac{1}{v_L(t, T_i - \delta, T_i)} \ln\left(\frac{(1-K)B_t^{T_i}}{K[B_t^{T_i - \delta} - B_t^{T_i}]}\right) + \frac{1}{2} v_L(t, T_i - \delta, T_i),$$

$$e_{2i} = e_{1i} - v_L(t, T_i - \delta, T_i),$$

and $v_L(t, T_i - \delta, T_i)$ is given by (11.17) in the one-factor setting and by (11.19) in the multi-factor setting. Note that these pricing formulas only apply to options expiring at one of the time points $T_0, T_1, \ldots, T_{n-1}$, and where the underlying zero-coupon bond matures at the following date in this sequence. In other words, the time distance between the maturity of the option and the maturity of the underlying zero-coupon bond must be equal to δ.

Exercise 11.3 (Lognormal inconsistency) Assume that the forward LIBOR rates $L_t^{T_0,T_1}$ and $L_t^{T_1,T_2}$ have proportional volatilities as in the lognormal LIBOR market model. The forward LIBOR rate $L_t^{T_0,T_2}$ is then given by (11.21). Apply Itô's Lemma to determine the volatility of $L_t^{T_0,T_2}$.

12

The Measurement and Management of Interest Rate Risk

12.1 INTRODUCTION

The values of bonds and other fixed income securities vary over time primarily due to changes in the term structure of interest rates. Most investors want to measure and compare the sensitivities of different securities to term structure movements. The interest rate risk measures of the individual securities are needed in order to obtain an overview of the total interest rate risk of the investors' portfolio and to identify the contribution of each security to this total risk. Many institutional investors are required to produce such risk measures for regulatory authorities and for publication in their accounting reports. In addition, such risk measures constitute an important input to portfolio management decisions.

In this chapter we will discuss how to quantify the interest rate risk of bonds and how these risk measures can be used in the management of the interest rate risk of portfolios. First, Section 12.2 describes the traditional, but still widely used, duration and convexity measures and discusses their relations to the dynamics of the term structure of interest rates. Next, Section 12.3 introduces risk measures that are more directly linked to the dynamic term structure models we have analysed in the previous chapters. Here we focus on one-factor diffusion models. Section 12.4 illustrates the use of these risk measures in the construction of so-called immunization strategies. The extension of risk measures to multi-factor diffusion models is outlined in Section 12.5, whereas Section 12.6 shows how the duration measure can be useful for the pricing of European options on bonds and hence the pricing of European swaptions. Finally, Section 12.7 briefly discusses some alternative risk measures.

12.2 TRADITIONAL MEASURES OF INTEREST RATE RISK

12.2.1 Macaulay duration and convexity

The Macaulay duration of a bond was defined by Macaulay (1938) as a weighted average of the time distance to the payment dates of the bond, that is an 'effective time-to-maturity'. As shown by Hicks (1939), the Macaulay duration also measures the sensitivity of the bond value with respect to changes in its own yield. Let us consider a bond with payment dates $T_1, \dots, T_n$, where we assume that $T_1 < \cdots < T_n$. The payment at time T_i is denoted by Y_i. The time t value of the bond is denoted by B_t. We let y_t^B denote the yield of the bond at time t, computed using continuous compounding so that

$$B_t = \sum_{T_i > t} Y_i e^{-y_t^B (T_i - t)},$$

where the sum is over all the future payment dates of the bond.

The **Macaulay duration** D_t^{Mac} of the bond is defined as

$$D_t^{\text{Mac}} = -\frac{1}{B_t}\frac{dB_t}{dy_t^B} = \frac{\sum_{T_i > t}(T_i - t) Y_i e^{-y_t^B (T_i - t)}}{B_t} = \sum_{T_i > t} w^{\text{Mac}}(t, T_i)(T_i - t), \tag{12.1}$$

where $w^{\text{Mac}}(t, T_i) = Y_i e^{-y_t^B (T_i - t)}/B_t$ which is the ratio between the value of the i'th payment and the total value of the bond. Since $w^{\text{Mac}}(t, T_i) > 0$ and $\sum_{T_i > t} w^{\text{Mac}}(t, T_i) = 1$, we see from (12.1) that the Macaulay duration has the interpretation of a weighted average time-to-maturity. For a bond with only one remaining payment the Macaulay duration is equal to the time-to-maturity. A simple manipulation of the definition of the Macaulay duration yields

$$\frac{dB_t}{B_t} = -D_t^{\text{Mac}}\, dy_t^B$$

so that the relative price change of the bond due to an instantaneous, infinitesimal change in its yield is proportional to the Macaulay duration of the bond.

Frequently, the Macaulay duration is defined in terms of the bond's annually compounded yield $\hat{y}_t^B$. By definition,

$$B_t = \sum_{T_i > t} Y_i (1 + \hat{y}_t^B)^{-(T_i - t)}$$

so that

$$\frac{dB_t}{d\hat{y}_t^B} = -\sum_{T_i > t}(T_i - t) Y_i (1 + \hat{y}_t^B)^{-(T_i - t) - 1}.$$

The Macaulay duration is then often defined as

$$\begin{aligned} D_t^{\text{Mac}} &= -\frac{1 + \hat{y}_t^B}{B_t}\frac{dB_t}{d\hat{y}_t^B} = \frac{\sum_{T_i > t}(T_i - t) Y_i (1 + \hat{y}_t^B)^{-(T_i - t)}}{B_t} \\ &= \sum_{T_i > t} w^{\text{Mac}}(t, T_i)(T_i - t), \end{aligned}$$

where the weights $w^{\mathrm{Mac}}(t, T_i)$ are the same as before since $e^{y_t^B} = (1 + \hat{y}_t^B)$. Therefore the two definitions provide precisely the same value for the Macaulay duration. Because $y_t^B = \ln(1 + \hat{y}_t^B)$ and hence $dy_t^B/d\hat{y}_t^B = 1/(1 + \hat{y}_t^B)$, we have that

$$\frac{dB_t}{B_t} = -D_t^{\mathrm{Mac}} \frac{d\hat{y}_t^B}{1 + \hat{y}_t^B}.$$

For bullet bonds, annuity bonds, and serial bonds an explicit expression for the Macaulay duration can be derived.[1] In some newspapers the Macaulay duration of each bond is listed next to the price of the bond.

The Macaulay duration is defined as a measure of the price change induced by an infinitesimal change in the yield of the bond. For a non-infinitesimal change, a first-order approximation gives that

$$\Delta B_t \approx \frac{dB_t}{dy_t^B} \Delta y_t^B,$$

and hence

$$\frac{\Delta B_t}{B_t} \approx -D_t^{\mathrm{Mac}} \Delta y_t^B.$$

An obvious way to obtain a better approximation is to include a second-order term:

$$\Delta B_t \approx \frac{dB_t}{dy_t^B} \Delta y_t^B + \frac{1}{2} \frac{d^2 B_t}{d(y_t^B)^2} \left(\Delta y_t^B\right)^2.$$

Defining the **Macaulay convexity** by

$$K_t^{\mathrm{Mac}} = \frac{1}{2B_t} \frac{d^2 B_t}{d(y_t^B)^2} = \frac{1}{2} \sum_{T_i > t} w^{\mathrm{Mac}}(t, T_i)(T_i - t)^2,$$

we can write the second-order approximation as

$$\frac{\Delta B_t}{B_t} \approx -D_t^{\mathrm{Mac}} \Delta y_t^B + K_t^{\mathrm{Mac}} \left(\Delta y_t^B\right)^2.$$

Note that the approximation only describes the price change induced by an instantaneous change in the yield. In order to evaluate the price change over some time interval, the effect of the reduction in the time-to-maturity of the bond should be included, for example by adding the term $\frac{\partial B_t}{\partial t} \Delta t$ on the right-hand side.

The Macaulay measures are not directly informative of how the price of a bond is affected by a change in the zero-coupon yield curve and are therefore not a valid basis for comparing the interest rate risk of different bonds. The problem is that the Macaulay measures are defined in terms of the bond's own yield, and a given change in the zero-coupon yield curve will generally result in different changes in the yields of different bonds. It is easy to show (see, for example Ingersoll et al.

[1] The formula for the Macaulay duration of a bullet bond can be found in many textbooks, for example Fabozzi (2010).

(1978, Thm. 1)) that the changes in the yields of all bonds will be the same if and only if the zero-coupon yield curve is always flat. In particular, the yield curve is only allowed to move by parallel shifts. Such an assumption is not only unrealistic, it also conflicts with the no-arbitrage principle, as we shall demonstrate in Section 12.2.3.

12.2.2 The Fisher–Weil duration and convexity

Macaulay (1938) also defined an alternative duration measure based on the zero-coupon yield curve rather than the bond's own yield. After decades of neglect, this duration measure was revived by Fisher and Weil (1971) who demonstrated the relevance of the measure for constructing immunization strategies. We will refer to this duration measure as the **Fisher–Weil duration**. The precise definition is

$$D_t^{\mathrm{FW}} = \sum_{T_i > t} w(t, T_i)(T_i - t), \tag{12.2}$$

where $w(t, T_i) = Y_i e^{-y_t^{T_i}(T_i - t)}/B_t$. Here, $y_t^{T_i}$ is the zero-coupon yield prevailing at time t for the period up to time T_i. Relative to the Macaulay duration, the weights are different. $w(t, T_i)$ is computed using the true present value of the i'th payment since the payment is multiplied by the market discount factor for time T_i payments, $B_t^{T_i} = e^{-y_t^{T_i}(T_i - t)}$. In the weights used in the computation of the Macaulay measures the payments are discounted using the yield of the bond. However, for typical yield curves the two sets of weights and hence the two duration measures will be very close, see, for example Table 12.1 later in this chapter.

If we think of the bond price as a function of the relevant zero-coupon yields $y_t^{T_1}, \dots, y_t^{T_n}$,

$$B_t = \sum_{T_i > t} Y_i e^{-y_t^{T_i}(T_i - t)},$$

we can write the relative price change induced by an instantaneous change in the zero-coupon yields as

$$\frac{dB_t}{B_t} = \sum_{T_i > t} \frac{1}{B_t} \frac{\partial B_t}{\partial y_t^{T_i}} dy_t^{T_i} = -\sum_{T_i > t} w(t, T_i)(T_i - t) dy_t^{T_i}.$$

If the changes in all the zero-coupon yields are identical, the relative price change is proportional to the Fisher–Weil duration. Consequently, the Fisher-Weil duration represents the price sensitivity towards infinitesimal parallel shifts of the zero-coupon yield curve. Note that an infinitesimal parallel shift of the curve of continuously compounded yields corresponds to an infinitesimal proportional shift in the curve of yearly compounded yields. This follows from the relation $y_t^{T_i} = \ln(1 + \hat{y}_t^{T_i})$ between the continuously compounded zero-coupon rate $y_t^{T_i}$ and the yearly compounded zero-coupon rate $\hat{y}_t^{T_i}$, which implies that $dy_t^{T_i} = d\hat{y}_t^{T_i}/(1 + \hat{y}_t^{T_i})$ so that $dy_t^{T_i} = k$ implies $d\hat{y}_t^{T_i} = k(1 + \hat{y}_t^{T_i})$.

We can also define the **Fisher–Weil convexity** as

$$K_t^{\mathrm{FW}} = \frac{1}{2} \sum_{T_i > t} w(t, T_i)(T_i - t)^2.$$

The relative price change induced by a non-infinitesimal parallel shift of the yield curve can then be approximated by

$$\frac{\Delta B_t}{B_t} \approx -D_t^{\mathrm{FW}} \Delta y_t^* + K_t^{\mathrm{FW}} \left(\Delta y_t^*\right)^2,$$

where Δy_t^* is the common change in all the zero-coupon yields. Again the reduction in the time-to-maturity should be included to approximate the price change over a given period.

12.2.3 The no-arbitrage principle and parallel shifts of the yield curve

In this section we will investigate under which assumptions the zero-coupon yield curve can only change in the form of parallel shifts. The analysis follows Ingersoll, Skelton, and Weil (1978). If the yield curve only changes in the form of infinitesimal parallel shifts, the curve must have exactly the same shape at all points in time. Hence, we can write any zero-coupon yield $y_t^{t+\tau}$ as a sum of the current short rate and a function which only depends on the 'time-to-maturity' of the yield, that is

$$y_t^T = r_t + h(T - t),$$

where $h(0) = 0$. In particular, the evolution of the yield curve can be described by a model where the short rate is the only state variable and has risk-neutral dynamics of the type

$$dr_t = \hat{\alpha}(r_t, t)\, dt + \beta(r_t, t)\, dz_t^{\mathbb{Q}},$$

where $z^{\mathbb{Q}}$ is a standard Brownian motion under the risk-neutral probability measure $\mathbb{Q}$.

In such a model the price of any fixed income security will be given by a function solving the fundamental partial differential equation (7.2). In particular, the price function of any zero-coupon bond $B^T(r, t)$ satisfies

$$\frac{\partial B^T}{\partial t}(r,t) + \hat{\alpha}(r,t)\frac{\partial B^T}{\partial r}(r,t) + \frac{1}{2}\beta(r,t)^2\frac{\partial^2 B^T}{\partial r^2}(r,t) - rB^T(r,t) = 0,$$
$$(r,t) \in \mathcal{S} \times [0, T),$$

and the terminal condition $B^T(r, T) = 1$. However, we know that the zero-coupon bond price is of the form

$$B^T(r,t) = e^{-y_t^T (T-t)} = e^{-r[T-t] - h(T-t)[T-t]}.$$

Substituting the relevant derivatives into the partial differential equation, we get that

$$h'(T-t)(T-t) + h(T-t) = \hat{\alpha}(r,t)(T-t) - \frac{1}{2}\beta(r,t)^2(T-t)^2,$$
$$(r,t) \in \mathcal{S} \times [0,T).$$

Since this holds for all r, the right-hand side must be independent of r. This can only be the case for all t if both $\hat{\alpha}$ and β are independent of r. Consequently, we get that

$$h'(T-t)(T-t) + h(T-t) = \hat{\alpha}(t)(T-t) - \frac{1}{2}\beta(t)^2(T-t)^2, \quad t \in [0,T).$$

The left-hand side depends only on the time difference $T-t$ so this must also be the case for the right-hand side. This will only be true if neither $\hat{\alpha}$ nor β depend on t. Therefore, $\hat{\alpha}$ and β have to be constants.

It follows from the above arguments that the dynamics of the short rate is of the form

$$dr_t = \hat{\alpha}\, dt + \beta\, dz_t^{\mathbb{Q}},$$

otherwise non-parallel yield curve shifts would be possible. This short rate dynamics is the basic assumption of the Merton model studied in Section 7.3. There we found that the zero-coupon yields are given by

$$y_t^{t+\tau} = r + \frac{1}{2}\hat{\alpha}\tau - \frac{1}{6}\beta^2\tau^2,$$

which corresponds to $h(\tau) = \frac{1}{2}\hat{\alpha}\tau - \frac{1}{6}\beta^2\tau^2$. We can therefore conclude that all yield curve shifts will be infinitesimal parallel shifts if and only if the yield curve at any point in time is a parabola with downward sloping branches and the short-term interest rate follows the dynamics described in Merton's model. These assumptions are highly unrealistic. Furthermore, Ingersoll, Skelton, and Weil (1978) show that non-infinitesimal parallel shifts of the yield curve conflict with the no-arbitrage principle. The bottom line is therefore that the Fisher–Weil risk measures do not measure the bond price sensitivity towards realistic movements of the yield curve. The Macaulay risk measures are not consistent with *any* arbitrage-free dynamic term structure model.

12.3 RISK MEASURES IN ONE-FACTOR DIFFUSION MODELS

12.3.1 Definitions and relations

To obtain measures of interest rate risk that are more in line with a realistic evolution of the term structure of interest rates, it is natural to consider uncertain price movements in reasonable dynamic term structure models. In a model with one or more state variables we focus on the sensitivity of the prices with respect to a change in the state variable(s). In this section we consider the one-factor diffusion models studied in Chapters 7 and 9. Risk measures in multi-factor models are discussed in Section 12.5.

We assume that the short rate r_t is the only state variable, and that it follows a process of the form

$$dr_t = \alpha(r_t, t)\,dt + \beta(r_t, t)\,dz_t$$

under the real-world probability measure. For an asset with price $B_t = B(r_t, t)$, Itô's Lemma implies that

$$\begin{aligned} dB_t = &\left(\frac{\partial B}{\partial t}(r_t, t) + \alpha(r_t, t)\frac{\partial B}{\partial r}(r_t, t) + \frac{1}{2}\beta(r_t, t)^2 \frac{\partial^2 B}{\partial r^2}(r_t, t)\right) \\ &dt + \frac{\partial B}{\partial r}(r_t, t)\beta(r_t, t)\,dz_t, \end{aligned}$$

and hence

$$\begin{aligned} \frac{dB_t}{B_t} = &\left(\frac{1}{B(r_t, t)}\frac{\partial B}{\partial t}(r_t, t) + \alpha(r_t, t)\frac{1}{B(r_t, t)}\frac{\partial B}{\partial r}(r_t, t)\right. \\ &\left. + \frac{1}{2}\beta(r_t, t)^2 \frac{1}{B(r_t, t)}\frac{\partial^2 B}{\partial r^2}(r_t, t)\right)\,dt \\ &+ \frac{1}{B(r_t, t)}\frac{\partial B}{\partial r}(r_t, t)\beta(r_t, t)\,dz_t. \end{aligned}$$

For a bond, the derivative $\frac{\partial B}{\partial r}(r, t)$ is negative in the models we have considered, so the volatility of the bond is given by[2] $-\frac{1}{B(r_t,t)}\frac{\partial B}{\partial r}(r_t, t)\beta(r_t, t)$. It is natural to use the asset-specific part of the volatility as a risk measure. Therefore we define the **duration** of the asset as

$$D(r, t) = -\frac{1}{B(r, t)}\frac{\partial B}{\partial r}(r, t). \tag{12.3}$$

Note the similarity to the definition of the Macaulay duration. The unexpected relative return on the asset is equal to minus the product of its duration, $D(r, t)$, and the unexpected change in the short rate, $\beta(r_t, t)\,dz_t$.

Furthermore, we define the **convexity** as

$$K(r, t) = \frac{1}{2B(r, t)}\frac{\partial^2 B}{\partial r^2}(r, t) \tag{12.4}$$

and the **time value** as

$$\Theta(r, t) = \frac{1}{B(r, t)}\frac{\partial B}{\partial t}(r, t).$$

[2] Recall that the volatility of an asset is defined as the standard deviation of the return on the asset over the next instant.

Consequently, the rate of return on the asset over the next infinitesimal period of time can be written as

$$\frac{dB_t}{B_t} = \left(\Theta(r_t,t) - \alpha(r_t,t)D(r_t,t) + \beta(r_t,t)^2 K(r_t,t)\right)\, dt - D(r_t,t)\beta(r_t,t)\, dz_t. \tag{12.5}$$

The duration of a portfolio of interest rate dependent securities is given by a value-weighted average of the durations of the individual securities. For example, let us consider a portfolio of two securities, namely N_1 units of asset 1 with a unit price of $B_1(r,t)$ and N_2 units of asset 2 with a unit price of $B_2(r,t)$. The value of the portfolio is $\Pi(r,t) = N_1 B_1(r,t) + N_2 B_2(r,t)$. The duration $D_\Pi(r,t)$ of the portfolio can be computed as

$$\begin{aligned} D_\Pi(r,t) &= -\frac{1}{\Pi(r,t)} \frac{\partial \Pi}{\partial r}(r,t) \\ &= -\frac{1}{\Pi(r,t)} \left(N_1 \frac{\partial B_1}{\partial r}(r,t) + N_2 \frac{\partial B_2}{\partial r}(r,t) \right) \\ &= \frac{N_1 B_1(r,t)}{\Pi(r,t)} \left(-\frac{1}{B_1(r,t)} \frac{\partial B_1}{\partial r}(r,t) \right) + \frac{N_2 B_2(r,t)}{\Pi(r,t)} \left(-\frac{1}{B_2(r,t)} \frac{\partial B_2}{\partial r}(r,t) \right) \\ &= \eta_1(r,t) D_1(r,t) + \eta_2(r,t) D_2(r,t), \end{aligned} \tag{12.6}$$

where $\eta_i(r,t) = N_i B_i(r,t)/\Pi(r,t)$ is the portfolio weight of the i'th asset, and $D_i(r,t)$ is the duration of the i'th asset, $i = 1, 2$. Obviously, we have $\eta_1(r,t) + \eta_2(r,t) = 1$. Similarly for the convexity and the time value. In particular, the duration of a coupon bond is a value-weighted average of the durations of the zero-coupon bonds maturing at the payment dates of the coupon bond.

By definition of the market price of risk $\lambda(r_t,t)$, we know that the expected rate of return on any asset minus the product of the market price of risk and the volatility of the asset must equal the short-term interest rate. From (12.5) we therefore obtain

$$\Theta(r,t) - \alpha(r,t)D(r,t) + \beta(r,t)^2 K(r,t) - \left(-D(r,t)\beta(r,t)\right)\lambda(r,t) = r$$

or

$$\Theta(r,t) - \hat{\alpha}(r,t)D(r,t) + \beta(r,t)^2 K(r,t) = r, \tag{12.7}$$

where $\hat{\alpha}(r,t) = \alpha(r,t) - \beta(r,t)\lambda(r,t)$ is the risk-neutral drift of the short rate. We could arrive at the same relation by substituting into the partial differential equation

$$\frac{\partial B}{\partial t}(r,t) + \hat{\alpha}(r,t)\frac{\partial B}{\partial r}(r,t) + \frac{1}{2}\beta(r,t)^2 \frac{\partial^2 B}{\partial r^2}(r,t) - rB(r,t) = 0$$

that we know $B(r,t)$ solves. The relation (12.7) between the time value, the duration, and the convexity holds for all interest rate dependent securities and hence also for all portfolios of interest rate dependent securities.[3]

[3] In the Black–Scholes–Merton model the time value and the so-called Δ and Γ values are related in a similar way, see Hull (2009, Section 17.7). Apparently, Christensen and Sørensen (1994) were the

Note that the rate of return on the security over the next instant can be rewritten as

$$\frac{dB_t}{B_t} = (r_t - \lambda(r_t,t)\beta(r_t,t)D(r_t,t))\ dt - D(r_t,t)\beta(r_t,t)\, dz_t,$$

which only involves the duration, and not the convexity nor the time value. We also know that in order to replicate a given fixed income security in a one-factor model, one must form a portfolio that, at any point in time, has the same volatility and therefore the same duration as that security. This is a consequence of the proof of the fundamental partial differential equation, see Theorem 4.10 and the subsequent discussion of hedging. However, a perfect hedge requires continuous rebalancing of the portfolio. Due to transaction costs and other practical issues such a continuous rebalancing is not implementable. Differentiation implies (see Exercise 12.1) that

$$\frac{\partial D}{\partial r}(r,t) = D(r,t)^2 - 2K(r,t) \tag{12.8}$$

so that the convexity can be seen as a measure of the interest rate sensitivity of the duration. If, at each time the portfolio is rebalanced, the convexities of the portfolio and of the position to be hedged are matched, their durations are likely to stay close until the following rebalancing of the portfolio. The convexity is therefore also of practical use in the interest rate risk management.

The duration, the convexity, and the time value can also be used for speculation, that is for setting up a portfolio which will provide a high return if some specific expectations of the future term structure are realized. For example, by constructing a portfolio with a zero duration and a large positive convexity, one will obtain a high return over a period with a large change (positive or negative) in the short rate. It follows from the relation (12.7) that for such a portfolio the time value will be negative. Consequently, the portfolio will give a negative return over a period where the short rate does not change significantly.

The Macaulay duration, defined in (12.1), and the Fisher–Weil duration, defined in (12.2), are measured in time units (typically years) and can be interpreted as measures of the 'effective' time-to-maturity of a bond. The duration defined in (12.3) is not measured in time units, but it can be transformed into a time-denominated duration. Following Cox, Ingersoll, and Ross (1979), we define the **time-denominated duration** of a coupon bond as the time-to-maturity of the zero-coupon bond that has the same duration as the coupon bond. If we denote the time-denominated duration by $D^*(r,t)$, the defining relation can be stated as

$$\frac{1}{B(r,t)}\frac{\partial B}{\partial r}(r,t) = \frac{1}{B^{t+D^*(r,t)}(r,t)}\frac{\partial B^{t+D^*(r,t)}}{\partial r}(r,t).$$

For bonds with only one remaining payment the time-denominated duration is equal to the time-to-maturity, just as for the Macaulay–duration and the Fisher–Weil duration.

first to discover this relation in the context of term structure models and the importance of taking the time value into account in the construction of interest rate risk hedging strategies.

Cox, Ingersoll, and Ross (1979) used the term **stochastic duration** for the time-denominated duration $D^*(r,t)$ to indicate that this duration measure is based on the stochastic evolution of the term structure. Other authors use the term stochastic duration for the original duration $D(r,t)$. Note that both these duration concepts are defined in relation to a specific term structure model, and the duration measures therefore indicate the sensitivity of the bond price to the yield curve movements consistent with the model. The traditional Macaulay and Fisher–Weil durations can be computed without reference to a specific model, but, on the other hand, they only measure the price sensitivity to a particular type of yield curve movements that is not consistent with any reasonable interest rate dynamics. Another advantage of the risk measures introduced in this section is that they are well-defined for all types of interest rate dependent securities, whereas the Macaulay and Fisher–Weil risk measures are only meaningful for bonds.[4] For the risk management of portfolios of many different fixed income securities we need risk measures for all the individual securities, for example futures, caps/floors, and swaptions.

12.3.2 Computation of the risk measures in affine models

In the time-homogeneous affine one-factor diffusion models, such as the Vasicek model and the CIR model, the zero-coupon bond prices are of the form

$$B^{T_i}(r,t) = e^{-a(T_i-t)-b(T_i-t)r}.$$

The price of a coupon bond with payment Y_i at time T_i, $i=1,\dots,n$, is

$$B(r,t) = \sum_{T_i>t} Y_i B^{T_i}(r,t).$$

Consequently, the duration of the coupon bond is

$$\begin{aligned} D(r,t) &= -\frac{1}{B(r,t)}\frac{\partial B}{\partial r}(r,t) = \frac{1}{B(r,t)}\sum_{T_i>t} b(T_i-t) Y_i B^{T_i}(r,t) \\ &= \sum_{T_i>t} w(r,t,T_i) b(T_i-t), \end{aligned}$$

where $w(r,t,T_i) = Y_i B^{T_i}(r,t)/B(r,t)$ is the i'th payment's share of the total present value of the bond. Note that the duration of a zero-coupon bond maturing at time T is $b(T-t)$, which is different from $T-t$ (except in the unrealistic Merton model). The convexity can be computed as

$$K(r,t) = \frac{1}{2}\sum_{T_i>t} w(r,t,T_i) b(T_i-t)^2.$$

[4] The duration $D(r,t)$ is well-defined by (12.3) for any security. Since the volatilities of zero-coupon bonds are bounded from above in many models, the time-denominated duration can only be defined for securities with a volatility below that upper bound. This is always true for coupon bonds, but not for all derivative securities. See Exercise 12.2.

The time value of the coupon bond is given by

$$\Theta(r,t) = \sum_{T_i>t} w(r,t,T_i)\left(a'(T_i-t)+b'(T_i-t)r\right) = \sum_{T_i>t} w(r,t,T_i) f^{T_i}(r,t),$$

where $f^{T_i}(r,t)$ is the forward rate at time t for the maturity date T_i. The time-denominated duration $D^*(r,t)$ is the solution to the equation

$$\sum_{T_i>t} w(r,t,T_i) b(T_i-t) = b(D^*(r,t)).$$

If b is invertible, we can write the time-denominated duration of a coupon bond explicitly as

$$D^*(r,t) = b^{-1}\left(\sum_{T_i>t} w(r,t,T_i) b(T_i-t)\right). \tag{12.9}$$

Example 12.1 In the Vasicek model we know from Section 7.4 that the b-function is given by

$$b(\tau) = \frac{1}{\kappa}\left(1-e^{-\kappa\tau}\right)$$

so that the duration of a coupon bond is

$$D(r,t) = \sum_{T_i>t} w(r,t,T_i)\frac{1}{\kappa}\left(1-e^{-\kappa[T_i-t]}\right) = \frac{1}{\kappa}\left(1-\sum_{T_i>t} w(r,t,T_i)e^{-\kappa[T_i-t]}\right).$$

Since

$$\frac{1}{\kappa}\left(1-e^{-\kappa\tau}\right) = y \quad \Leftrightarrow \quad \tau = -\frac{1}{\kappa}\ln(1-\kappa y),$$

we have that

$$b^{-1}(y) = -\frac{1}{\kappa}\ln(1-\kappa y),$$

and by (12.9) the time-denominated duration of a coupon bond is

$$\begin{aligned} D^*(r,t) &= -\frac{1}{\kappa}\ln\left(1-\kappa\sum_{T_i>t} w(r,t,T_i) b(T_i-t)\right) \\ &= -\frac{1}{\kappa}\ln\left(1-\sum_{T_i>t} w(r,t,T_i)(1-e^{-\kappa[T_i-t]})\right) \end{aligned}$$

$$= -\frac{1}{\kappa} \ln \left(\sum_{T_i > t} w(r, t, T_i) e^{-\kappa [T_i - t]} \right).$$

For the extended Vasicek model (the Hull–White model) we get the same expression since the b-function in that model is the same as in the original Vasicek model.

Example 12.2 In the CIR model studied in Section 7.5 the b-function is given by

$$b(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \hat{\kappa})(e^{\gamma\tau} - 1) + 2\gamma}$$

so that the duration of a coupon bond is

$$D(r, t) = \sum_{T_i > t} w(r, t, T_i) \frac{2(e^{\gamma[T_i - t]} - 1)}{(\gamma + \hat{\kappa})(e^{\gamma[T_i - t]} - 1) + 2\gamma}.$$

In Exercise 12.3, you are asked to verify that the time-denominated duration of a coupon bond is

$$D^*(r, t) = \frac{1}{\gamma} \ln \left(1 + 2\gamma \left[\frac{2}{\sum_{T_i > t} w(r, t, T_i) b(T_i - t)} - (\hat{\kappa} + \gamma) \right]^{-1} \right). \tag{12.10}$$

12.3.3 A comparison with traditional durations

Munk (1999) shows analytically that, for any bond, the time-denominated duration in the Vasicek model is smaller than the Fisher–Weil duration. This is also true for the CIR model if the parameter $\hat{\kappa} = \kappa + \lambda$ is positive, which is consistent with typical parameter estimates. Therefore, the Fisher–Weil duration over-estimates the interest rate risk of coupon bonds. Except for extreme yield curves, the Macaulay duration and the Fisher–Weil duration will be very, very close so that the above conclusion also applies to the Macaulay duration.

Table 12.1 shows the different duration measures for bullet bonds of different maturities under the assumption that the yield curve and its dynamics are consistent with the CIR model with given, realistic parameter values. It is clear from the table that, for all bonds, the Macaulay duration and the Fisher–Weil duration are very close. For relatively short-term bonds the time-denominated duration is close to the traditional durations, but for longer-term bonds the time-denominated duration is significantly lower than the Macaulay and Fisher–Weil durations. In particular, we see that the interest rate sensitivity, and therefore also the time-denominated duration, for bullet bonds first increases and then decreases as the time-to-maturity increases.

What is the explanation for the differences in the duration measures? As discussed in Section 12.2.3, the Fisher–Weil duration is only a reasonable interest rate risk measure if the yield curve evolves as in the Merton model where both the drift and the volatility of the short rate are assumed to be constant. In the

Table 12.1: A comparison of duration measures for different bonds under the assumption that the CIR model with the parameters $\kappa = 0.36$, $\theta = 0.05$, $\beta = 0.1185$, and $\lambda = -0.1302$ provides a correct description of the yield curve and its dynamics. The current short rate is 0.04. The bonds are bullet bonds with a coupon rate of 5%, a face value of 100, one annual payment date, and exactly one year until the next payment date.

Maturity (years)	price	yield (%)	D^{Mac}	D^{FW}	D^*	D
1	100.48	4.50	1.00	1.00	1.00	0.89
2	100.31	4.84	1.95	1.95	1.95	1.56
3	99.70	5.11	2.86	2.86	2.83	2.05
4	98.81	5.34	3.72	3.72	3.63	2.41
5	97.75	5.53	4.54	4.54	4.34	2.67
6	96.60	5.68	5.32	5.31	4.95	2.86
8	94.24	5.93	6.74	6.72	5.86	3.09
10	91.96	6.10	8.01	7.97	6.40	3.21
12	89.87	6.22	9.14	9.07	6.68	3.26
15	87.15	6.35	10.57	10.45	6.83	3.28
20	83.63	6.48	12.39	12.16	6.80	3.28
25	81.13	6.56	13.65	13.30	6.71	3.26

Merton model the volatility of a zero-coupon bond is proportional to the time-to-maturity of the bond, see (7.15) and (7.29). On the other hand, in the CIR model the volatility of a zero-coupon bond with time-to-maturity τ equals $b(\tau)\beta\sqrt{r}$, where the b-function is given by (7.49). It can be shown that b is an increasing, concave function with $b'(\tau) < 1$ for all τ. Hence, the volatility of the zero-coupon bonds increases with the time-to-maturity, but less than proportionally. It can also be shown that the b-function in the CIR model is a decreasing function of the speed-of-adjustment parameter κ so the stronger the mean-reversion, the further apart the bond volatilities in the two models. Consequently, the distance between the time-denominated duration D_t^* and the Fisher–Weil duration will typically increase with the speed-of-adjustment parameter, although this probably cannot be proved analytically due to the complicated expression for D_t^* in (12.10).

12.4 IMMUNIZATION

12.4.1 Construction of immunization strategies

In many situations an individual or corporate investor will invest in the bond market either in order to ensure that some future liabilities can be met or just to obtain some desired future cash flow. For example, a pension fund will often have a relatively precise estimate of the size and timing of the future pension payments to its customers. For such an investor it is important that the value of the investment portfolio remains close to the value of the liabilities. Some financial institutions are even required by law to keep the value of the investment portfolio at any point in time above the value of the liabilities by some percentage margin.

A cash flow or portfolio is said to be immunized (against interest rate risk) if the value of the cash flow or portfolio is not negatively affected by any possible change in the term structure of interest rates. An investor who has to pay a given cash flow can obtain an immunized total position by investing in a portfolio of interest rate dependent securities that perfectly replicates that cash flow. For example, if an investor has to pay 10 million dollars in 5 years, he can make sure that this will be possible by investing in default-free 5-year zero-coupon bonds with a total face value of 10 million dollars. The present value of his total position will be completely immune to interest rate movements. An investor who has a desired cash flow consisting of several future payments can obtain perfect immunization by investing in a portfolio of zero-coupon bonds that exactly replicates the cash flow. In many cases, however, all the necessary zero-coupon bonds are neither traded on the bond market nor possible to construct by a static portfolio of traded coupon bonds. Therefore, the desired cash flow can only be matched by constructing a dynamically rebalanced portfolio of traded securities.

We know from the discussion in Section 4.8 that if the term structure follows a one-factor diffusion model, any interest rate dependent security (or portfolio) can be perfectly replicated by a particular portfolio of any two other interest rate dependent securities. The portfolio weights have to be adjusted continuously so that the volatility of the portfolio value will always be identical to the volatility of the value of the cash flow which is to be replicated. In other words, the duration of the portfolio must match the duration of the desired cash flow at any point in time. If we let $\eta(r,t)$ denote the value weight of the first security in the immunizing portfolio, the second security will have a value weight of $1-\eta(r,t)$. According to (12.6), the duration of the portfolio is

$$D_\Pi(r,t) = \eta(r,t)D_1(r,t) + (1-\eta(r,t))D_2(r,t),$$

where $D_1(r,t)$ and $D_2(r,t)$ are the durations of each of the securities in the portfolio. If $\bar{D}(r,t)$ denotes the duration of the cash flow to be matched, we want to make sure that

$$\eta(r,t)D_1(r,t) + (1-\eta(r,t))D_2(r,t) = \bar{D}(r,t)$$

for all r and t. This relation will hold if the portfolio weight $\eta(r,t)$ is chosen so that

$$\eta(r,t) = \frac{\bar{D}(r,t) - D_2(r,t)}{D_1(r,t) - D_2(r,t)}. \tag{12.11}$$

Suppose that (i) the portfolio is initially constructed with these relative weights and scaled so that the total amount invested is equal to the present value of the cash flow to be matched, and (ii) the portfolio is continuously rebalanced so that (12.11) holds at any point in time. Then the desired cash flow will be matched with certainty, that is the position is perfectly immunized against interest rate movements.

Of course, continuous rebalancing of a portfolio is not practically implementable (or desirable considering real-world transaction costs). If the portfolio is only rebalanced periodically, a perfect immunization cannot be guaranteed. The durations may be matched each time the portfolio is rebalanced, but between these dates the durations may diverge due to interest rate movements and the passage of time. With different durations the portfolio and the desired cash flow will not have the same sensitivity towards another interest rate change.

As shown in (12.8), the convexity measures the sensitivity of the duration towards changes in the term structure of interest rates. If both the durations and the convexities of the portfolio and the cash flow are matched each time the portfolio is rebalanced, the durations are likely to stay close even after several interest rate changes. Therefore, matching the convexities should improve the effectiveness of the immunization strategy. Note that when both durations and convexities are matched, it follows from (12.7) that the time values are also identical. Matching both durations and convexities requires a portfolio of three securities. Let us write the durations and convexities of the three securities in the portfolio as $D_i(r,t)$ and $K_i(r,t)$, respectively. The value weights of the three securities are denoted by $\eta_i(r,t)$, and the convexity of the desired cash flow is denoted by $\bar{K}(r,t)$. Since $\eta_3(r,t) = 1 - \eta_1(r,t) - \eta_2(r,t)$, durations and convexities will be matched if $\eta_1(r,t)$ and $\eta_2(r,t)$ are chosen such that

$$\eta_1(r,t)D_1(r,t) + \eta_2(r,t)D_2(r,t) + [1 - \eta_1(r,t) - \eta_2(r,t)]\, D_3(r,t) = \bar{D}(r,t),$$

$$\eta_1(r,t)K_1(r,t) + \eta_2(r,t)K_2(r,t) + [1 - \eta_1(r,t) - \eta_2(r,t)]\, K_3(r,t) = \bar{K}(r,t).$$

This equation system has the unique solution

$$\eta_1(r,t) = \frac{(\bar{D}(r,t) - D_3(r,t))(K_2(r,t) - K_3(r,t)) - (D_2(r,t) - D_3(r,t))(\bar{K}(r,t) - K_3(r,t))}{(D_1(r,t) - D_3(r,t))(K_2(r,t) - K_3(r,t)) - (D_2(r,t) - D_3(r,t))(K_1(r,t) - K_3(r,t))},$$

$$\eta_2(r,t) = \frac{(D_1(r,t) - D_3(r,t))(\bar{K}(r,t) - K_3(r,t)) - (\bar{D}(r,t) - D_3(r,t))(K_1(r,t) - K_3(r,t))}{(D_1(r,t) - D_3(r,t))(K_2(r,t) - K_3(r,t)) - (D_2(r,t) - D_3(r,t))(K_1(r,t) - K_3(r,t))}.$$

If only durations are matched, the convexity of the portfolio and hence the effectiveness of the immunization strategy will be highly dependent on which two securities the portfolio consists of. If the convexity of the investment portfolio is larger than the convexity of the cash flow, a big change (positive or negative) in the short rate will induce an increase in the net value of the total position. On the other hand, if the short rate stays almost constant, the net value of the position will decrease since the time value of the portfolio is then lower than the time value of the cash flow, see (12.7). The converse conclusions hold in case the convexity of the portfolio is less than the convexity of the cash flow.

Traditionally, immunization strategies have been constructed on the basis of Macaulay durations instead of the stochastic durations as above. The Macaulay duration of a portfolio is typically very close to, but not exactly equal to, the value-weighted average of the Macaulay durations of the securities in the portfolio. We will ignore the small errors induced by this approximation, just as practitioners seem to do. The immunization strategy based on Macaulay durations is then defined by Equation (12.11) where Macaulay durations are used on the right-hand side. Immunization strategies based on the Fisher–Weil duration can be constructed in a similar manner. In earlier sections of this chapter we have argued that the Macaulay and Fisher–Weil risk measures are inappropriate for realistic yield curve movements. Consequently, immunization strategies based on those measures are likely to be ineffective. Below we perform an experiment that illustrates how far off the mark the traditional immunization strategies are.

12.4.2 An experimental comparison of immunization strategies

For simplicity, let us consider an investor who seeks to match a payment of 1,000 dollars exactly 10 years from now. We assume that the CIR model

$$dr_t = \kappa[\theta - r_t]\, dt + \beta\sqrt{r_t}\, dz_t = (\kappa\theta - [\kappa + \lambda]r_t)\; dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}}$$

with the parameter values $\kappa = 0.3$, $\theta = 0.05$, $\beta = 0.1$, and $\lambda = -0.1$ provides a correct description of the evolution of the term structure of interest rates. The asymptotic long-term yield y_∞ is then 6.74%. The zero-coupon yield curve will be increasing if the current short rate is below 6.12% and decreasing if the current short rate is above 7.50%. For intermediate values of the short rate, the yield curve will have a small hump.

12.4.2.1 Duration matching immunization

In the following we will compare the effectiveness of duration matching immunization strategies based on the Macaulay duration, the Fisher–Weil duration, and the stochastic duration derived from the CIR model. In addition to the duration measure applied, the immunization strategy is characterized by the rebalancing frequency and by the securities that constitute the portfolio. We will consider strategies with 2, 12, and 52 equally spaced annual portfolio adjustments. We consider only portfolios of two bullet bonds of different maturities. The bonds have a coupon rate of 5%, one annual payment date, and exactly 1 year to the next payment date. We assume that the investor is free to pick two such bonds among the bonds that have time-to-maturities in the set $\{1, 2, \dots\}$ at the time when the strategy is initiated. We apply two criteria for the choice of maturities. One criterion is to choose the maturity of one of the bonds so that the Macaulay duration of the bond is less than, but as close as possible to, the Macaulay duration of the liability to be matched. The other bond is chosen to be the bond with Macaulay duration above, but as close as possible to, the Macaulay duration of the liability. This criterion has the nice implication that the Macaulay convexities of the portfolio and the liability will be close, which should improve the effectiveness of periodically adjusted immunization portfolios. We will refer to this criterion as the *Macaulay criterion*. The other criterion is to choose a short-term bond and a long-term bond so that the convexity of the portfolio will be significantly higher than the convexity of the liability. The short-term bond has a time-to-maturity of 1 year at the most, while the long bond matures 5 years after the liability is due. We will refer to this criterion as the *short-long criterion*. Irrespective of the criterion used, we assume that 1 year before the liability is due, the portfolio is replaced by a position in the bond with only 1 year to maturity. Consequently, the strategies are not affected by interest rate movements in the final year.

The effectiveness of the different immunization strategies is studied by performing 30,000 simulations of the evolution of the yield curve in the CIR model over the

10-year period to the due date of the liability (see Chapter 16 for details on Monte Carlo simulation). In the simulations we use 360 time steps per year. Table 12.2 illustrates the effectiveness of the different immunization strategies. The left part of the table contains results based on the Macaulay bond selection criterion, whereas the right part is based on the short-long bond second criterion. In order to explain the numbers in the table, let us take the right-most column as an example. The numbers in this column are from an immunization strategy based on matching CIR durations using a portfolio of a short-term and a long-term bond. For two annual portfolio adjustments the average of the 30,000 simulated terminal portfolio values was 1000.01, which is very close to the desired value of 1,000. The average absolute deviation was 0.124% of the desired portfolio value. In 29.1% of the

Table 12.2: Results from the immunization of a 10-year liability based on 30,000 simulations of the CIR model. The current short rate is 5%. The parameter values are $\kappa = 0.3$, $\theta = 0.05$, $\beta = 0.1$, and $\lambda = -0.1$.

	2 portfolio adjustments per year					
	Macaulay criterion			short-long criterion		
	Mac	FW	CIR	Mac	FW	CIR
Avg. terminal value	994.41	994.42	999.99	968.32	968.41	1000.01
Avg. absolute deviation (%)	1.28	1.27	0.072	5.65	5.60	0.124
Dev. $<$ 0.05%	2.2	2.2	45.2	0.4	0.4	29.1
Dev. $<$ 0.1%	4.3	4.3	76.2	0.9	0.8	53.3
Dev. $<$ 0.5%	21.5	21.5	99.7	4.5	4.5	98.6
Dev. $<$ 1.0%	42.4	42.6	100.0	8.9	8.9	99.9
Dev. $<$ 5.0%	99.6	99.6	100.0	45.9	46.2	100.0
	12 portfolio adjustments per year					
	Macaulay criterion			short-long criterion		
	Mac	FW	CIR	Mac	FW	CIR
Avg. terminal value	994.54	994.43	1000.00	968.62	968.53	1000.01
Avg. absolute deviation (%)	1.28	1.27	0.032	5.61	5.59	0.053
Dev. $<$ 0.05%	2.2	2.2	80.5	0.4	0.4	59.6
Dev. $<$ 0.1%	4.5	4.4	97.1	0.9	0.8	86.0
Dev. $<$ 0.5%	21.7	21.5	100.0	4.5	4.5	100.0
Dev. $<$ 1.0%	42.7	42.7	100.0	9.0	9.0	100.0
Dev. $<$ 5.0%	99.6	99.6	100.0	46.4	46.3	100.0
	52 portfolio adjustments per year					
	Macaulay criterion			short-long criterion		
	Mac	FW	CIR	Mac	FW	CIR
Avg. terminal value	994.50	994.47	1000.00	968.55	968.53	1000.01
Avg. absolute deviation (%)	1.27	1.26	0.015	5.61	5.58	0.026
Dev. $<$ 0.05%	2.1	2.2	97.3	0.4	0.4	87.1
Dev. $<$ 0.1%	4.3	4.4	99.9	0.9	0.9	98.6
Dev. $<$ 0.5%	21.2	21.3	100.0	4.3	4.3	100.0
Dev. $<$ 1.0%	42.7	42.8	100.0	8.7	8.6	100.0
Dev. $<$ 5.0%	99.7	99.7	100.0	46.8	47.0	100.0

30,000 simulated outcomes the absolute deviation was less than 0.05%. In 53.3% of the simulated outcomes the absolute deviation was less than 0.1%, and so on.[5]

The strategies based on the Macaulay and the Fisher–Weil durations generate results very similar to each other due to the fact that these duration measures typically are very close. The effectiveness of these strategies seems independent of the rebalancing frequency. The choice of bonds applied in the strategy is more important. The deviations from the target are generally significantly larger for a portfolio of a short and a long bond (a high convexity portfolio) than for a portfolio of bonds with very similar maturities (a low convexity portfolio). The high convexity strategy deviates by more than 5% in more than half of all cases.

The CIR strategy of matching stochastic durations is far more effective than matching Macaulay or Fisher–Weil durations. This can be seen both from the average terminal portfolio value, the average absolute deviation, and the listed fractiles from the distribution of the absolute deviations. Even with just two annual portfolio adjustments the CIR strategy will miss the target by less than 0.5% in more than 98% of all outcomes, no matter which bond selection criterion is used. The Macaulay and Fisher–Weil strategies miss the mark by more than 1% in more than 50% of all outcomes, even when the bonds are selected according to their Macaulay durations. Clearly, the effectiveness of the CIR strategy increases with the frequency of the portfolio adjustments. In particular, frequent rebalancing is advantageous if the immunization portfolio has a relatively high convexity. However, the effectiveness of the CIR strategy seems to depend less on the bonds chosen than does the effectiveness of the traditional strategies.

Simulations using other initial short rates, and therefore different initial yield curves, have shown that the average terminal value of the Macaulay strategy is highly dependent on the initial short rate. For a nearly flat initial yield curve the average terminal value is very close to the targeted value of 1,000, but the average absolute deviation is not smaller than for other initial yield curves. The CIR strategy is far more effective than the Macaulay strategy also for a nearly flat initial yield curve. The effectiveness of the strategies decreases with the current interest rate level due to the fact that the interest rate volatility is assumed to increase with the level in the CIR model. Furthermore, the accuracy of the immunization strategies will typically be decreasing in β and θ and increasing in κ.

12.4.2.2 Duration and convexity matching immunization strategies

In the following we consider the case where both the duration and the convexity of the liability are matched by the investment portfolio. In our experiment we assume that the portfolio consists of a bond with a time-to-maturity of at most 1 year, a bond maturing 2 years after the liability is due, and a bond maturing 10 years after the liability is due. Table 12.3 illustrates the gain in efficiency by matching both duration and convexity instead of just matching duration using a portfolio of the short and the long bond. For the Macaulay strategy the average deviation is reduced by a factor 10. The Fisher–Weil strategy generates almost identical results and is therefore omitted. For the CIR strategy the relative improvement is even

[5] Even with 30,000 simulations the specific fractiles are quite uncertain, but the averages are quite reliable. Experiments with other sequences of random numbers and a larger number of simulations have resulted in very similar fractiles.

Table 12.3: Results from the immunization of a 10-year liability of 1,000 dollars based on 30,000 simulations of the CIR model. The current short rate is 5%.

	12 portfolio adjustments per year					
	Identical durations			Identical durations + convexities		
	Mac	CIR	HW	Mac	CIR	HW
Avg. terminal value	968.62	1000.01	1005.39	997.11	1000.00	999.78
Avg. absolute deviation (%)	5.61	0.053	0.98	0.57	0.0002	0.33
Dev. $< 0.05\%$	0.4	59.6	2.5	4.7	100.0	10.1
Dev. $< 0.1\%$	0.9	86.0	5.0	9.5	100.0	19.7
Dev. $< 0.5\%$	4.5	100.0	26.1	47.6	100.0	77.6
Dev. $< 1.0\%$	9.0	100.0	53.4	86.9	100.0	97.5
Dev. $< 5.0\%$	46.4	100.0	100.0	100.0	100.0	100.0

more dramatic and in all of the 30,000 simulated outcomes the deviation is less than 0.05% although the portfolio is only rebalanced once a month! The numbers under the column heading HW (for Hull–White) are explained below.

12.4.2.3 Model uncertainty

The results above clearly show that if the CIR model gives a correct description of the term structure dynamics, an immunization strategy based on the CIR risk measures is far more effective than strategies based on the traditional risk measures. However, if the CIR model does not provide a good description of the evolution of the term structure, an immunization strategy based on the stochastic durations computed using the CIR model will be less successful. Since the CIR model in any case is closer to the true dynamics of the short rate than the Merton model underlying the Fisher–Weil duration, the CIR strategy is still expected to be more effective than traditional strategies.

Our analysis indicates that for immunization purposes it is important to apply risk measures that are related to the dynamics of the term structure. Therefore, it is important to identify an empirically reasonable model and then to implement immunization strategies (and hedge strategies in general) based on the relevant risk measures associated with the model.

How effective is an immunization strategy based on risk measures associated with a model which does not give a correct description of the yield curve dynamics? To investigate this issue, we assume that the CIR model is correct, but that the immunization strategy is constructed using risk measures from the Hull–White model (the extended Vasicek model). Just before each portfolio adjustment the Hull–White model is calibrated to the true yield curve, that is the yield curve of the CIR model. Table 12.3 shows the results of such an immunization strategy under the column heading HW. As expected, the strategy is far less effective than the strategy based on the true yield curve dynamics, but the Hull–White strategy is still far better than the traditional Macaulay strategy. So even though we base our immunization strategy on a model which, in some sense, is far from the true model, we still obtain a much more effective immunization than we would by using the traditional immunization strategy.

12.5 RISK MEASURES IN MULTI-FACTOR DIFFUSION MODELS

12.5.1 Factor durations, convexities, and time value

In multi-factor diffusion models it is natural to measure the sensitivity of a security price with respect to changes in the different state variables. For illustration, let us consider a two-factor diffusion model where the state variables x_1 and x_2 are assumed to develop as

$$dx_{1t} = \alpha_1(x_{1t}, x_{2t})\, dt + \beta_{11}(x_{1t}, x_{2t})\, dz_{1t} + \beta_{12}(x_{1t}, x_{2t})\, dz_{2t}, \tag{12.12}$$

$$dx_{2t} = \alpha_2(x_{1t}, x_{2t})\, dt + \beta_{21}(x_{1t}, x_{2t})\, dz_{1t} + \beta_{22}(x_{1t}, x_{2t})\, dz_{2t}. \tag{12.13}$$

For a security with the price $B_t = B(x_{1t}, x_{2t}, t)$, Itô's Lemma implies that

$$\begin{aligned} \frac{dB_t}{B_t} = \dots\, dt &- D_1(x_{1t}, x_{2t}, t)\, [\beta_{11}(x_{1t}, x_{2t})\, dz_{1t} + \beta_{12}(x_{1t}, x_{2t})\, dz_{2t}] \\ &- D_2(x_{1t}, x_{2t}, t)\, [\beta_{21}(x_{1t}, x_{2t})\, dz_{1t} + \beta_{22}(x_{1t}, x_{2t})\, dz_{2t}], \end{aligned} \tag{12.14}$$

where we have omitted the drift and introduced the notation

$$D_1(x_1, x_2, t) = -\frac{1}{B(x_1, x_2, t)} \frac{\partial B}{\partial x_1}(x_1, x_2, t),$$

$$D_2(x_1, x_2, t) = -\frac{1}{B(x_1, x_2, t)} \frac{\partial B}{\partial x_2}(x_1, x_2, t).$$

We will refer to D_1 and D_2 as the **factor durations** of the security. In such a two-factor model any interest rate dependent security can be perfectly replicated by a portfolio that always has the same factor durations as the given security. Again continuous rebalancing is needed.

In the practical implementation of hedge strategies it is relevant to include second-order derivatives just as we did in the one-factor models above. In a two-factor model we have three relevant second-order derivatives that lead to the following **factor convexities**:

$$K_1(x_1, x_2, t) = \frac{1}{2B(x_1, x_2, t)} \frac{\partial^2 B}{\partial x_1^2}(x_1, x_2, t),$$

$$K_2(x_1, x_2, t) = \frac{1}{2B(x_1, x_2, t)} \frac{\partial^2 B}{\partial x_2^2}(x_1, x_2, t),$$

$$K_{12}(x_1, x_2, t) = \frac{1}{B(x_1, x_2, t)} \frac{\partial^2 B}{\partial x_1 \partial x_2}(x_1, x_2, t).$$

Defining the time value as

$$\Theta(x_1, x_2, t) = \frac{1}{B(x_1, x_2, t)} \frac{\partial B}{\partial t}(x_1, x_2, t),$$

we get the following relation:

$$\begin{aligned}\Theta(x_1, x_2, t) - \hat{\alpha}_1(x_1, x_2) D_1(x_1, x_2, t) - \hat{\alpha}_2(x_1, x_2) D_2(x_1, x_2, t)\\ + \gamma_1(x_1, x_2)^2 K_1(x_1, x_2, t) + \gamma_2(x_1, x_2)^2 K_2(x_1, x_2, t)\\ + \gamma_{12}(x_1, x_2) K_{12}(x_1, x_2, t) = r(x_1, x_2).\end{aligned}$$

Here the terms $\gamma_1^2 = \beta_{11}^2 + \beta_{12}^2$ and $\gamma_2^2 = \beta_{21}^2 + \beta_{22}^2$ are the variance rates of changes in the first and the second state variables, and $\gamma_{12} = \beta_{11}\beta_{21} + \beta_{12}\beta_{22}$ is the covariance rate between these changes.

In a two-factor affine model the prices of zero-coupon bonds are of the form

$$B^T(x_1, x_2, t) = e^{-a(T-t) - b_1(T-t)x_1 - b_2(T-t)x_2}.$$

Therefore, the factor durations of a zero-coupon bond are given by $D_j(x_1, x_2, t) = b_j(T - t)$ for $j = 1, 2$. For a coupon bond with the price $B(x_1, x_2, t) = \sum_{T_i > t} Y_i B^{T_i}(x_1, x_2, t)$ the factor durations are

$$D_j(x_1, x_2, t) = -\frac{1}{B(x_1, x_2, t)} \frac{\partial B}{\partial x_j}(x_1, x_2, t) = \sum_{T_i > t} w(x_1, x_2, t, T_i) b_j(T_i - t),$$

where $w(x_1, x_2, t, T_i) = Y_i B^{T_i}(x_1, x_2, t)/B(x_1, x_2, t)$. The convexities and the time value are

$$K_j(x_1, x_2, t) = \sum_{T_i > t} w(x_1, x_2, t, T_i) b_j(T_i - t)^2, \quad j = 1, 2,$$

$$K_{12}(x_1, x_2, t) = \sum_{T_i > t} w(x_1, x_2, t, T_i) b_1(T_i - t) b_2(T_i - t),$$

$$\Theta(x_1, x_2, t) = \sum_{T_i > t} w(x_1, x_2, t, T_i) \left(a'(T_i - t) + b_1'(T_i - t)x_1 + b_2'(T_i - t)x_2\right).$$

The factor durations defined above can be transformed into time-denominated factor durations in the following manner. For each state variable or factor j we define the time-denominated factor duration $D_j^* = D_j^*(x_1, x_2, t)$ as the time-to-maturity of the zero-coupon bond with the same price sensitivity and hence the same factor duration relative to this state variable:

$$\frac{1}{B(x_1, x_2, t)} \frac{\partial B}{\partial x_j}(x_1, x_2, t) = \frac{1}{B^{t+D_j^*}(x_1, x_2, t)} \frac{\partial B^{t+D_j^*}}{\partial x_j}(x_1, x_2, t).$$

In an affine model this equation reduces to

$$\sum_{T_i>t} w(x_1, x_2, t, T_i) b_j(T_i - t) = b_j(D_j^*)$$

so that

$$D_j^* = D_j^*(x_1, x_2, t) = b_j^{-1}\left(\sum_{T_i>t} w(x_1, x_2, t, T_i) b_j(T_i - t)\right)$$

under the assumption that b_j is invertible.

12.5.2 One-dimensional risk measures in multi-factor models

For practical purposes it may be relevant to summarize the risks of a given security in a single (one-dimensional) risk measure. The volatility of the security is the most natural choice. By definition the volatility of a security is the standard deviation of the rate of return on the security over the next instant. In the two-factor model given by (12.12) and (12.13) the variance of the rate of return can be computed from (12.14):

$$\begin{aligned}
\mathrm{Var}_t\left[\frac{dB_t}{B_t}\right] &= \mathrm{Var}_t\left([D_1\beta_{11} + D_2\beta_{21}]\; dz_{1t} + [D_1\beta_{12} + D_2\beta_{22}]\; dz_{2t}\right) \\
&= \left([D_1\beta_{11} + D_2\beta_{21}]^2 + [D_1\beta_{12} + D_2\beta_{22}]^2\right)\, dt \\
&= \left(D_1^2\gamma_1^2 + D_2^2\gamma_2^2 + 2D_1D_2\gamma_{12}\right)\, dt,
\end{aligned}$$

where, for notational simplicity, we have omitted the arguments of the D- and β-functions. The volatility is thus given by

$$\begin{aligned}
\sigma_B(x_1, x_2, t) = \Big(&D_1(x_1, x_2, t)^2\gamma_1(x_1, x_2)^2 + D_2(x_1, x_2, t)^2\gamma_2(x_1, x_2)^2 \\
&+ 2D_1(x_1, x_2, t)D_2(x_1, x_2, t)\gamma_{12}(x_1, x_2)\Big)^{1/2}.
\end{aligned}$$

Also this risk measure can be transformed into a time-denominated risk measure, namely the time-to-maturity of the zero-coupon bond having the same volatility as the security considered. Letting $\sigma^T(x_1, x_2, t)$ denote the volatility of the zero-coupon bond maturing at time T, the time-denominated duration $D^*(x_1, x_2, t)$ is given as the solution $D^* = D^*(x_1, x_2, t)$ to the equation

$$\sigma_B(x_1, x_2, t) = \sigma^{t+D^*}(x_1, x_2, t)$$

or, equivalently,

$$\sigma_B(x_1, x_2, t)^2 = \sigma^{t+D^*}(x_1, x_2, t)^2.$$

Table 12.4: Duration measures for 5% bullet bonds with one annual payment date assuming the Longstaff–Schwartz model with the parameter values $\beta_1^2 = 0.005$, $\beta_2^2 = 0.0814$, $\kappa_1 = 0.3299$, $\hat{\kappa}_2 = 14.4277$, $\varphi_1 = 0.020112$, and $\varphi_2 = 0.26075$ provides a correct description of the yield curve dynamics. The current short rate is $r = 0.05$ with an instantaneous variance rate of $v = 0.002$.

Maturity (years)	price	yield (%)	D^{Mac}	D^{FW}	D_1^*	D_2^*	D^*
1	99.83	5.18	1.00	1.00	1.00	1.00	1.00
2	98.94	5.58	1.95	1.95	1.94	1.21	1.94
3	97.60	5.90	2.86	2.86	2.81	1.21	2.81
4	96.01	6.16	3.72	3.71	3.59	1.21	3.59
5	94.30	6.37	4.53	4.52	4.25	1.21	4.25
6	92.56	6.54	5.30	5.29	4.79	1.21	4.79
8	89.20	6.79	6.70	6.68	5.52	1.20	5.52
10	86.15	6.97	7.94	7.89	5.87	1.20	5.87
12	83.45	7.09	9.01	8.94	5.99	1.20	5.99
15	80.07	7.22	10.36	10.23	6.00	1.20	6.00
20	75.88	7.34	11.99	11.75	5.89	1.19	5.89

This equation can only be solved numerically. For an affine two-factor model the equation is of the form

$$\gamma_1^2 f_1(t)^2 + \gamma_2^2 f_2(t)^2 + 2\gamma_{12} f_1(t) f_2(t) = \\ b_1(D^*)^2 \gamma_1^2 + b_2(D^*)^2 \gamma_2^2 + 2b_1(D^*) b_2(D^*) \gamma_{12},$$

where

$$f_j(t) = \sum_{T_i > t} w(x_1, x_2, t, T_i) b_j(T_i - t), \quad j = 1, 2.$$

Some basic properties of the time-denominated duration were derived by Munk (1999). The time-denominated duration is a theoretically better founded one-dimensional risk measure than the traditional Macaulay and Fisher–Weil durations. Furthermore, the time-denominated duration is closely related to the volatility concept, which most investors are familiar with.

Table 12.4 lists different duration measures based on the two-factor model of Longstaff and Schwartz (1992a) studied in Section 8.4.2. The parameters of the models are fixed at the values estimated by Longstaff and Schwartz (1992b), which generates a reasonable distribution of the future values of the state variables r and v. For 5% bullet bonds of different maturities the table shows the price, the yield, the Macaulay duration D^{Mac}, the Fisher–Weil duration D^{FW}, the time-denominated factor durations D_1^* and D_2^*, and the one-dimensional time-denominated duration D^*. Also in this case the traditional duration measures are overestimating the risk of long-term bonds. Also note that with the parameter values applied in the computations, the first time-denominated factor duration D_1^* and the one-dimensional time-denominated duration D^* are basically identical. The reason is that the sensitivity to the second factor depends very little on the time-to-maturity. This is not necessarily the case for other parameter values.

12.6 DURATION-BASED PRICING OF OPTIONS ON BONDS

12.6.1 The general idea

In the framework of one-factor diffusion models Wei (1997) suggests that the price of a European call option on a coupon bond can be approximated by the price of a European call option on a particular zero-coupon bond, namely the zero-coupon bond having the same (stochastic) duration as the coupon bond underlying the option to be priced. According to Section 6.5.2, this approximation can also be applied to the pricing of European swaptions. As usual, we let $C_t^{K,T,S}$ be the time t price of a European call option with expiration time T and exercise price K, written on a zero-coupon bond maturing at time $S > T$. Furthermore, $C_t^{K,T,\mathrm{cpn}}$ is the time t price of a European call option with expiration time T and exercise price K, written on a given coupon bond. We denote by B_t the time t value of the payments of the coupon bond after expiration of the option, that is $B_t = \sum_{T_i > T} Y_i B_t^{T_i}$ where Y_i is the payment at time T_i. Wei's approximation is then given by the following relation:

$$C_t^{K,T,\mathrm{cpn}} \approx \tilde{C}_t^{K,T,\mathrm{cpn}} = \frac{B_t}{B_t^{t+D_t^*}} C_t^{K^*,T,t+D_t^*}, \tag{12.15}$$

where $K^* = KB_t^{t+D_t^*}/B_t$, and where D_t^* denotes the time-denominated duration of the cash flow of the underlying coupon bond after expiration of the option.

Wei does not motivate the approximation, but shows by numerical examples in the one-factor models of Vasicek (1977) and Cox, Ingersoll, and Ross (1985b) that the approximation is very accurate. The advantage of using the approximation in these two models is that the price of only one call option on a zero-coupon bond needs to be computed. To apply Jamshidian's trick (see Section 7.2.3) we have to compute a zero-coupon bond option price for each of the payment dates of the coupon bond after the expiration date of the option. In addition, one equation in one unknown has to be solved numerically to determine the critical interest rate r^*. Nevertheless, the exact price can be very quickly computed by Jamshidian's formula, but if many options on coupon bonds (or swaptions) have to priced, the slightly faster approximation may be relevant to use.

The intuition behind the accuracy of the approximation is that the underlying zero-coupon bond of the approximating option is chosen to match the volatility of the underlying coupon bond for the option we want to price. Since we know that the volatility of the underlying asset is an extremely important factor for the price of an option, this choice makes good sense.

Munk (1999) studies the approximation in more detail, gives an analytical argument for its accuracy, and illustrates the precision in multi-factor models in several numerical examples. Note that the computational advantage of using the approximation is much bigger in multi-factor models than in one-factor models, since no explicit formula for European options on coupon bonds has been found for any multi-factor model. Whereas the alternative to the approximation in the one-factor models is a slightly more complicated explicit expression, the alternative in the multi-factor models is to use a numerical technique, such as Monte Carlo simulation or numerical solution of the relevant multi-dimensional partial differential equation. Below we go through the analytical argument for the applicability of the approximation. After that we will illustrate the accuracy of the approximation in numerical examples.

It should be noted that several other techniques for approximating prices of European options on coupon bonds have been suggested in the literature. For example, in the framework of affine models Collin-Dufresne and Goldstein (2002b) and Singleton and Umantsev (2002) introduce two approximations that may dominate the duration-based approach discussed here with respect to accuracy and computational speed, but these approximations are much harder to understand.

12.6.2 A mathematical analysis of the approximation

Let us first study the error in using the approximation

$$C_t^{K,T,\mathrm{cpn}} \approx \frac{B_t}{B_t^S} C_t^{K_S,T,S}, \tag{12.16}$$

where S is any given maturity date of the underlying zero-coupon bond of the approximating option, and where $K_S = KB_t^S/B_t$. Afterwards we will argue that the error will be small when $S = t + D_t^*$, which is exactly the approximation (12.15).

Both the correct option price and the price of the approximating option can be written in terms of expected values under the S-forward martingale measure $\mathbb{Q}^S$. Under this measure the price of any asset relative to the zero-coupon bond price B_t^S is by definition a martingale. Hence, the correct price of the option can be written as

$$C_t^{K,T,\mathrm{cpn}} = B_t^S \,\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{\max(B_T - K, 0)}{B_T^S}\right],$$

while the price of the approximating option is

$$C_t^{K_S,T,S} = B_t^S \,\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{\max\left(B_T^S - \frac{KB_t^S}{B_t}, 0\right)}{B_T^S}\right] = B_t^S \,\mathrm{E}_t^{\mathbb{Q}^S}\left[\max\left(1 - \frac{KB_t^S}{B_t B_T^S}, 0\right)\right].$$

The dollar error incurred by using the approximation (12.16) is therefore equal to

$$\begin{aligned} C_t^{K,T,\mathrm{cpn}} - \frac{B_t}{B_t^S} C_t^{K_S,T,S} &= B_t^S \left(\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{\max(B_T - K, 0)}{B_T^S}\right] \right. \\ &\qquad \left. - \frac{B_t}{B_t^S} \mathrm{E}_t^{\mathbb{Q}^S}\left[\max\left(1 - \frac{KB_t^S}{B_t B_T^S}, 0\right)\right]\right) \\ &= B_t^S \,\mathrm{E}_t^{\mathbb{Q}^S}\left[\max\left(\frac{B_T}{B_T^S} - \frac{K}{B_T^S}, 0\right) - \max\left(\frac{B_t}{B_t^S} - \frac{K}{B_T^S}, 0\right)\right]. \end{aligned} \tag{12.17}$$

From the definition of the S-forward martingale measure it follows also that

$$\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{B_T}{B_T^S}\right] = \frac{B_t}{B_t^S} \tag{12.18}$$

and that

$$\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{K}{B_T^S}\right] = \frac{KB_t^T}{B_t^S}. \tag{12.19}$$

For **deep-in-the-money** call options both max-terms in (12.17) will with a high probability return the first argument, and it follows then from (12.18) that the dollar error will be close to zero. Since the option price in this case is relatively high, the percentage error will be very close to zero. For **deep-out-of-the-money** call options both max-terms will with a high probability return zero so that the dollar error again is close to zero. The option price will also be close to zero, so the percentage error may be substantial.

The error is due to the outcomes where only one and not both max-terms is different from zero. This will be the case when the realized values of B_T and B_T^S are such that the ratio K/B_T^S lies between B_T/B_T^S and B_t/B_t^S. As indicated by (12.18) and (12.19), this affects the value of **forward near-the-money** options where $B_t \approx KB_t^T$. We will therefore expect the dollar pricing errors to be largest for such options.

The considerations above are valid for any S. In order to reduce the probability of ending up in the outcomes that induce the error, we seek to choose S so that B_T/B_t and B_T^S/B_t^S are likely to end up close to each other. As a first attempt to achieve this we could try to pick S such that the variance $\mathrm{Var}_t^{\mathbb{Q}^S}\left[B_T/B_t - B_T^S/B_t^S\right]$ is minimized, but this idea is not implementable due to the typically very complicated expressions for B_T^S and, in particular, for B_T. Alternatively, we can choose S so that the relative changes in B_t and B_t^S over the next instant are close to each other. This is exactly what we achieve by using $S = t + D_t^*$.

Another promising choice is $S = T^{\mathrm{mv}}$ which is the value of S that minimizes the variance of the difference in the relative price change over the next instant, that is $\mathrm{Var}_t^{\mathbb{Q}^S}\left[\frac{dB_t}{B_t} - \frac{dB_t^S}{B_t^S}\right]$. This idea also gives rise to an alternative time-denominated duration measure, $D_t^{\mathrm{mv}} = T^{\mathrm{mv}} - t$, which we could call the variance-minimizing duration. It can be shown (see Munk (1999)) that for one-factor models the two duration measures are identical, $D_t^* = D_t^{\mathrm{mv}}$. In multi-factor models the two measures will typically be close to each other, and consequently the accuracy of the approximation will typically be the same no matter which duration measure is used to fix the maturity of the zero-coupon bond. In the extreme cases where the measures differ significantly, the approximation based on D_t^* seems to be more accurate.

Note that the analysis in this subsection applies to all term structure models. We have *not* assumed that the evolution of the term structure can be described by a one-factor diffusion model. Therefore we can expect the approximation to be accurate in all models. Below we will investigate the accuracy of the approximation in a specific term structure model, namely the two-factor model of Longstaff and Schwartz discussed in Section 8.4.2. These results are taken from Munk (1999), who also presents similar results for a two-factor Gaussian Heath–Jarrow–Morton model (see Chapter 10 for an introduction to these models). Wei (1997) studies the accuracy of the approximation in the one-factor models of Vasicek and of Cox, Ingersoll, and Ross.

12.6.3 The accuracy of the approximation in the Longstaff–Schwartz model

According to (8.28), the price of a European call option on a zero-coupon bond in the Longstaff–Schwartz model can be written as

$$C_t^{K,T,S} = B_t^S \chi_1^2 - K B_t^T \chi_2^2,$$

where χ_1^2 and χ_2^2 are two probabilities taken from the two-dimensional non-central χ^2-distribution. No explicit formula for the price of a European call option on a coupon bond has been found. Consequently, an approximation like (12.15) will be very valuable if it is sufficiently accurate.

To estimate the accuracy, we will compare the approximate price $\tilde{C}_t^{K,T,\text{cpn}}$ to a 'correct' price $C_t^{K,T,\text{cpn}}$ computed using Monte Carlo simulation (see Chapter 16).[6] Of course, in the practical use of the approximation, the approximate price will be computed using the explicit formula for the price of the option on the zero-coupon bond. But to make a fair comparison, we will compute the approximate price using the same simulated sample paths as used for computing the correct option price. In this way our evaluation of the approximation is not sensitive to a possible bias in the correct price induced by the simulation technique.

We will consider European call options with an expiration time of 2 or 6 months written on an 8% bullet bond with a single annual payment date and a time-to-maturity of 2 or 10 years. The parameters in the dynamics of the state variables, see (8.25) and (8.26), are taken to be $\beta_1^2 = 0.01$, $\beta_2^2 = 0.08$, $\varphi_1 = 0.001$, $\varphi_2 = 1.28$, $\kappa_1 = 0.33$, $\kappa_2 = 14$, and $\lambda = 0$. These values are close to the parameter values estimated by Longstaff and Schwartz in their original article. The current short rate is assumed to be $r = 0.08$ with an instantaneous variance of $v = 0.002$. The accuracy of the approximation does not seem to depend on these values in any systematic way. Table 12.5 lists results for options on the 2-year bond for various exercise prices around the forward-at-the-money value of K, that is B_t/B_t^T. The corresponding results for options on the 10-year bond are shown in Table 12.6. The absolute deviation shown in the tables is defined as the approximate price minus the correct price, whereas the relative deviation is computed as the absolute deviation divided by the correct price. The tables also show the standard deviation of the simulated difference between the correct and the approximate price.

[6] The results shown are based on simulations of 10,000 pairs of antithetic sample paths of the two state variables r and v. The time period until the expiration date of the option is divided into approximately 100 subintervals per year.

Table 12.5: Prices of 2- and 6-month European call options on a 2-year bullet 8% bond in the Longstaff–Schwartz model. The underlying bond has a current price of 89.3400, a 2-month forward price of 91.2042, a 6-month forward price of 95.7687, and a time-denominated stochastic duration of 1.9086 years.

		2-month options		
K	appr. price	abs. dev.	rel. dev. (%)	std. dev.
86	5.08407	$0.1 \cdot 10^{-5}$	0.000	$1.8 \cdot 10^{-4}$
87	4.10368	$0.2 \cdot 10^{-5}$	0.000	$1.7 \cdot 10^{-4}$
88	3.12553	$0.7 \cdot 10^{-5}$	0.000	$1.5 \cdot 10^{-4}$
89	2.16242	$2.1 \cdot 10^{-5}$	0.001	$1.2 \cdot 10^{-4}$
90	1.26608	$3.2 \cdot 10^{-5}$	0.003	$1.0 \cdot 10^{-4}$
91	0.56030	$0.7 \cdot 10^{-5}$	0.001	$0.9 \cdot 10^{-4}$
92	0.15992	$-2.7 \cdot 10^{-5}$	−0.017	$0.7 \cdot 10^{-4}$
93	0.02442	$-2.0 \cdot 10^{-5}$	−0.083	$0.7 \cdot 10^{-4}$
94	0.00163	$-0.4 \cdot 10^{-5}$	−0.253	$0.4 \cdot 10^{-4}$
95	0.00001	$-0.0 \cdot 10^{-5}$	−1.545	$0.1 \cdot 10^{-4}$
		6-month options		
K	appr. price	abs. dev.	rel. dev. (%)	std. dev.
91	4.45364	$2.0 \cdot 10^{-5}$	0.000	$4.8 \cdot 10^{-4}$
92	3.53250	$4.9 \cdot 10^{-5}$	0.001	$4.0 \cdot 10^{-4}$
93	2.63434	$8.2 \cdot 10^{-5}$	0.003	$3.4 \cdot 10^{-4}$
94	1.79287	$9.6 \cdot 10^{-5}$	0.005	$3.2 \cdot 10^{-4}$
95	1.06678	$5.5 \cdot 10^{-5}$	0.005	$3.1 \cdot 10^{-4}$
96	0.52036	$-3.3 \cdot 10^{-5}$	−0.006	$2.4 \cdot 10^{-4}$
97	0.19074	$-9.6 \cdot 10^{-5}$	−0.050	$2.0 \cdot 10^{-4}$
98	0.04576	$-7.7 \cdot 10^{-5}$	−0.168	$2.0 \cdot 10^{-4}$
99	0.00576	$-2.7 \cdot 10^{-5}$	−0.474	$1.5 \cdot 10^{-4}$
100	0.00021	$-0.2 \cdot 10^{-5}$	−1.051	$0.5 \cdot 10^{-4}$

All the approximate prices are correct to three decimals, and the percentage deviations are also very small. In all cases the absolute deviation is considerably smaller than the standard deviation of the Monte Carlo simulated differences. Based on the mathematical analysis of the approximation we expect the errors to be smaller for shorter maturities of the option and the underlying bond than for longer maturities. This expectation is confirmed by our examples. Also in line with our discussion, we see that the absolute deviation is largest for forward-near-the-money options and smallest for deep-in- and deep-out-of-the-money options.

Figure 12.1 illustrates how the precision of the approximation depends on the exercise price for different time-to-maturities of the zero-coupon bond underlying the approximating option. The figure is based on 2-month options on the 2-year bullet bond, but a similar picture can be drawn for the other options considered. For deep-out-of- and deep-in-the-money options the approximation is very accurate no matter which zero-coupon bond is used in the approximation, but for near-the-money options it is important to choose the right zero-coupon bond, namely

Table 12.6: Prices on 2- and 6-month European call options on a 10-year bullet 8% bond in the Longstaff–Schwartz model. The underlying bond has a current price of 76.9324, a 2-month forward price of 78.5377, a 6-month forward price of 82.4682, and a time-denominated stochastic duration of 4.8630 years.

		2-month options		
K	appr. price	abs. dev.	rel. dev. (%)	std. dev.
74	4.42874	$1.2 \cdot 10^{-4}$	0.003	$1.9 \cdot 10^{-3}$
75	3.46569	$2.5 \cdot 10^{-4}$	0.007	$1.6 \cdot 10^{-3}$
76	2.53643	$3.9 \cdot 10^{-4}$	0.015	$1.4 \cdot 10^{-3}$
77	1.69005	$4.0 \cdot 10^{-4}$	0.024	$1.4 \cdot 10^{-3}$
78	0.98799	$1.8 \cdot 10^{-4}$	0.018	$1.3 \cdot 10^{-3}$
79	0.48542	$-1.7 \cdot 10^{-4}$	−0.036	$1.0 \cdot 10^{-3}$
80	0.19080	$-3.9 \cdot 10^{-4}$	−0.202	$0.9 \cdot 10^{-3}$
81	0.05666	$-3.2 \cdot 10^{-4}$	−0.570	$0.9 \cdot 10^{-3}$
82	0.01267	$-1.6 \cdot 10^{-4}$	−1.263	$0.8 \cdot 10^{-3}$
83	0.00185	$-0.5 \cdot 10^{-4}$	−2.424	$0.5 \cdot 10^{-3}$
		6-month options		
K	appr. price	abs. dev.	rel. dev. (%)	std. dev.
78	4.27344	$1.1 \cdot 10^{-3}$	0.027	$4.4 \cdot 10^{-3}$
79	3.42836	$1.3 \cdot 10^{-3}$	0.037	$4.3 \cdot 10^{-3}$
80	2.64289	$1.2 \cdot 10^{-3}$	0.045	$4.3 \cdot 10^{-3}$
81	1.93654	$0.8 \cdot 10^{-3}$	0.042	$4.2 \cdot 10^{-3}$
82	1.33393	$0.2 \cdot 10^{-3}$	0.015	$3.8 \cdot 10^{-3}$
83	0.85064	$-0.5 \cdot 10^{-3}$	−0.063	$3.1 \cdot 10^{-3}$
84	0.49430	$-1.1 \cdot 10^{-3}$	−0.220	$2.8 \cdot 10^{-3}$
85	0.25641	$-1.3 \cdot 10^{-3}$	−0.508	$2.6 \cdot 10^{-3}$
86	0.11491	$-1.2 \cdot 10^{-3}$	−1.001	$2.7 \cdot 10^{-3}$
87	0.04372	$-0.8 \cdot 10^{-3}$	−1.786	$2.6 \cdot 10^{-3}$

the zero-coupon bond with a time-to-maturity equal to the time-denominated stochastic duration of the underlying coupon bond. Also these results are consistent with the analytical arguments and the discussion in the preceding subsection.

12.7 ALTERNATIVE MEASURES OF INTEREST RATE RISK

In this chapter we have focused on measures of interest rate risk in arbitrage-free dynamic diffusion models of the term structure. Similar risk measures can be defined in Heath–Jarrow–Morton (HJM) models and market models which, as discussed in Chapters 10 and 11, do not necessarily fit into the diffusion setting. In an HJM model where all the instantaneous forward rates are affected by a single Brownian motion,

$$df_t^T = \alpha(t, T, (f_t^s)_{s \geq t})\, dt + \beta(t, T, (f_t^s)_{s \geq t})\, dz_t,$$

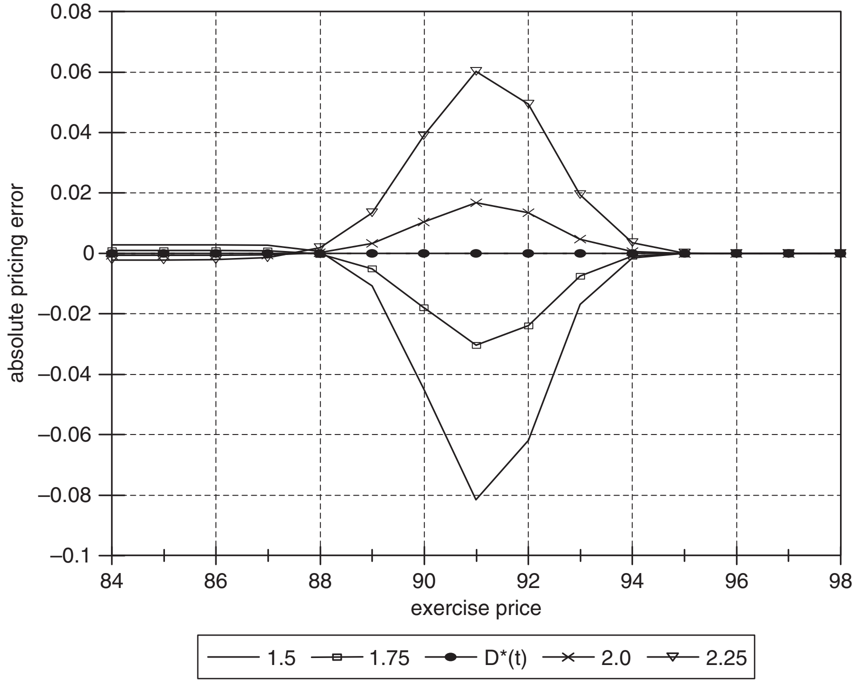

Fig. 12.1: The absolute price errors for 2-month options on 2-year bullet bonds in the Longstaff–Schwartz-model for different maturities of the zero-coupon bond underlying the approximating option. The time-denominated stochastic duration of the underlying coupon bond is $D_t^* = 1.9086$ years, and the 2-month forward price of the coupon bond is 91.2042.

the price of any fixed income security will have a dynamics of the form

$$\frac{dB_t}{B_t} = \mu_B(t, (f_t^s)_{s\geq t})\, dt + \sigma_B(t, (f_t^s)_{s\geq t})\, dz_t.$$

Here the volatility σ_B is an obvious candidate for measuring the interest rate risk of the security, and the time-denominated duration $D^* = D^*(t, (f_t^s)_{s\geq t})$ can be defined implicitly by the equation

$$\sigma_B(t, (f_t^s)_{s\geq t}) = \sigma^{t+D^*}(t, (f_t^s)_{s\geq t}),$$

where $\sigma^T(t, (f_t^s)_{s\geq t}) = -\int_t^T \beta(t, u, (f_t^s)_{s\geq t})\, du$ is the volatility of the zero-coupon bond maturing at time T, see Theorem 10.1.[7] Similar risk measures can be defined for HJM models involving more than one Brownian motion.

In the more practically oriented part of the literature several alternative measures of interest rate risk have been suggested. A seemingly popular approach is to use the so-called **key rate durations** introduced by Ho (1992). The basic idea

[7] If β is positive, the volatility of the zero-coupon bond is strictly speaking $-\sigma^T$.

is to select a number of key interest rates, that is zero-coupon yields for certain representative maturities, for example 1, 2, 5, 10, and 20 years. A change in one of these key rates is assumed to affect the yields for nearby maturities. For example, with the key rates listed above, a change in the 2-year zero-coupon yield is assumed to affect all zero-coupon yields of maturities between 1 year and 5 years. The change in those yields is assumed to be proportional to the maturity distance to the key rate. For example, a change of 0.01 (100 basis points) in the 2-year rate is assumed to cause a change of 0.005 (50 basis points) in the 1.5-year rate since 1.5 years is halfway between 2 years and the preceding key rate maturity of one year. Similarly, the change in the 2-year rate is assumed to cause a change of approximately 0.0033 (33 basis points) in the 4-year rate. A simultaneous change in several key rates will cause a piecewise linear change in the entire yield curve. With sufficiently many key rates, any yield curve change can be well approximated in this way. It is relatively simple to measure the sensitivities of zero-coupon and coupon bonds with respect to changes in the key rates. These sensitivities are called the key rate durations. When different bonds and positions in derivative securities are combined, the total key rate durations of a portfolio can be controlled so that the investor can hedge against (or speculate in) specific yield curve movements.

The key rate durations are easy to compute and relate to, but there are several practical and theoretical problems in applying these durations. The individual key rates do not move independently, and hence we have to consider which combinations of key rate changes are realistic and do not conflict with the no-arbitrage principle. Furthermore, to evaluate the interest rate risk of a security or a portfolio, we must specify the probability distribution of the possible key rate changes. Practitioners often assume that the changes in the different key rates can be described by a multi-variate normal distribution and estimate the means, variances, and covariances of the distribution from historical data. While the normal distribution is very tractable, empirical studies cannot support such a distributional assumption.

An investor who believes that the yield curve dynamics can be represented by the evolution of some selected key rates should use a theoretically better-founded model for pricing and risk management, for example an arbitrage-free dynamic model using these key rates as state variables, see the short discussion in Section 8.7.3. In such a model all yield curve movements are consistent with the no-arbitrage principle. Furthermore, for the points on the yield curve that lie in between the key rate maturities, such a model will give a more reasonable description than does the simple linear interpolation assumed in the computation of the key rate durations. Finally, the model can be specified using relatively few parameters and still provide a good description of the covariance structure of the key rates.

Other authors suggest duration measures that represent the price sensitivity towards changes in the level, the slope, and the curvature of the yield curve, see, for example, Willner (1996) and Phoa and Shearer (1997). This seems like a good idea since these factors empirically provide a good description of the shape and movements of the yield curve, see the discussion in Section 8.1. However, these duration measures should also be computed in the setting of a realistic, arbitrage-free dynamics of these characteristic variables. This can be ensured by constructing a term structure model using these factors as state variables.

12.8 EXERCISES

Exercise 12.1 (Duration-convexity link) Show Equation (12.8).

Exercise 12.2 (Time-denominated duration for non-bonds) Assume that interest rates follow the Vasicek model. Can you define the time-denominated duration for any bond futures or European bond option? If so, derive the time-denominated duration for these assets.

Exercise 12.3 (Time-denominated duration in CIR) Show Equation (12.10).

13

Defaultable Bonds and Credit Derivatives

13.1 INTRODUCTION

In the valuation models developed in previous chapters we have assumed that the issuer of the security being valued will always deliver the payments described in the contractual description of the security on a timely basis. An issuer not delivering the promised payments on a contract is said to default on the contract. The risk that this may happen is referred to as **default risk** or **credit risk**. A default does not imply that the holder of that contract walks away empty-handed. The holder will typically get a **recovery payment** either in cash or in the form of a new claim.

The assumption of no default on a bond is in most cases very reasonable if the bond is issued by the government or treasury department of a country having a limited public debt relative to the tax incomes or the GDP of the country. If a government has issued bonds and subsequently faces financial troubles, it may raise taxes, cut public spending, or—if the bonds are denominated in the domestic currency—print enough money so that it can honour its nominal debts. Of course in the latter case the purchasing power of the money received by the bondholders may be lower than expected. It should be noted however that there are plenty of historical examples of countries defaulting on part or all of their debt.[1]

For many contracts issued by private corporations it is necessary to take default risk into account. The prime example is corporate bonds, where a firm has borrowed money by issuing bonds promising a prespecified future payment stream. For various reasons the firm may end up in a situation where it cannot or will not continue paying the promised amounts to bondholders and therefore the firm defaults on its debt. Of course, when potential investors value a corporate bond they will anticipate the possibility of a default of the issuing firm before the maturity date of the bond. The main purpose of this chapter is to discuss how the

[1] Tomz and Wright (2007) report that 106 countries have defaulted a total of 250 times in the period 1820–2004. Recent examples of such defaults include the Russian government's default on the domestically issued GKO bonds in August 1998 and Argentina's default on USD 142 billion of domestic public debt in 2001 and roughly USD 1 billion debt to the World Bank in 2002. Government debt defaults are often due to a period of bad performance of the domestic economy, but Tomz and Wright (2007) point out that there are also many examples of countries defaulting in 'good times', for example, following a major change in the political regime in the country. In the worldwide financial crisis that started in 2007, some investors have been concerned about the possible default of various countries including Iceland, Greece, Ireland, Portugal, several eastern European countries, and even the United States. These fears are reflected by the CDS spreads on the bonds issued by those countries. The CDS spread will be explained later in this chapter.

default risk can be modelled and combined with our general valuation techniques from earlier chapters.

Section 13.2 introduces some basic notation and concepts needed in such models. Credit ratings and some historical statistics regarding defaults are also presented.

Two main model classes are used for pricing defaultable bonds. The structural models considered in Section 13.3 are based on assumptions about the specific issuing firm, for example an uncertain flow of earnings and a given or optimally derived capital structure. Default is then represented by the event that the value of the assets of the firm becomes 'small' compared to the outstanding debt. The reduced-form models explored in Section 13.4 try to avoid the very complicated, detailed modelling of each firm and are based on an exogenous specification of the dynamics of the default probabilities of the relevant firms. For both of the two classes we will present examples of models that can be used in practical applications. Section 13.5 presents a short introduction to hybrid models with features from both of the main model classes. It turns out to be challenging to build reasonable models capturing correlation between the defaults of different firms. Many practitioners use so-called copulas for this purpose. We describe and discuss this approach in Section 13.6.

In the recent decade or so, huge markets for credit derivatives have evolved. A credit derivative is a contract which gives a payoff that depends on whether or not certain **credit events** related to one or more pre-specified entities (typically corporations) have happened or not over a given period of time. The bankruptcy of an entity is surely a credit event, but depending on the contractual specifications events such as a major restructuring of the entity and the failure of the entity to make due payments on certain obligations may also count as credit events. Investors can use credit derivatives to reduce their exposure to the possible default of specific firms, but they can also speculate in the default of a firm by taking an appropriate position in credit derivatives. This is no different from other derivatives markets. By trading options on a stock index, investors can either reduce their exposure to changes in the stock market or speculate in specific changes. Section 13.7 provides some information and statistics about the markets for credit derivatives. The main type of credit derivative is the credit default swap. In Section 13.8 we explain how credit default swaps work and how they can be priced using reduced-form models. We discuss another important credit-related security, the collateralized debt obligation, in Section 13.9.

Section 13.10 concludes the chapter. Although the chapter is fairly long, a lot of interesting material has been left out. This includes numerous specific theoretical models and empirical investigations, various advanced credit-related securities, an in-depth coverage of practical credit risk management, as well as many interesting and complicated mathematical aspects of the models considered. For more information, we refer the reader to specialized credit risk textbooks such as Bielecki and Rutkowski (2002), Duffie and Singleton (2003), and Lando (2004) and to the articles referenced throughout the chapter.

Credit risk and credit derivatives, in particular those related to mortgages, play important roles in the financial crisis that erupted in 2007 with a devastating impact on both the financial industry and macroeconomic performance in all parts of the world. Throughout the chapter we comment on the contributions of credit derivatives, rating agencies, and valuation methods to the crisis, but we do not attempt to provide a full overview, documentation, or in-depth analysis of the crisis. Recent academic papers and books discussing the financial crisis include Bhansali, Gingrich, and Longstaff (2008), Crouhy, Jarrow, and Turnbull (2008), Demyanyk and Van Hemert (2008), Brunnermeier (2009), Gorton (2009), and Shiller (2008). See also the discussions and references in Chapter 14.

13.2 SOME BASIC CONCEPTS, RELATIONS, AND PRACTICAL ISSUES

13.2.1 Default time and default probabilities

In any credit risk model for a single defaultable security we will have to specify the time of default, which we shall denote by τ. We do not know in advance *if* default will happen and, if it does, we will typically not know *when* it happens, so τ is random. In technical terms, the default time τ is a **stopping time**. Let $\mathbf{1}_{\{\tau>t\}}$ be the associated default indicator process which is zero until default and then jumps to one. Given a probability measure $\mathbb{Q}$, the probability that default occurs in a time interval $[t, t']$, given that it has not occurred up to time t, is

$$\begin{aligned}\mathbb{Q}(t < \tau < t' | \tau > t) &= \frac{\mathbb{Q}(t < \tau < t')}{\mathbb{Q}(\tau > t)} = \frac{\mathbb{Q}(\tau > t) - \mathbb{Q}(\tau > t')}{\mathbb{Q}(\tau > t)} \\ &= 1 - \frac{\mathbb{Q}(\tau > t')}{\mathbb{Q}(\tau > t)},\end{aligned} \tag{13.1}$$

where the first equality is due to Bayes' rule. Note that $\mathbb{Q}(\tau > t')$ is the probability that the firm survives at least until time t'. Investors will be interested both in real-world default probabilities, as they define the true probability distribution of gains and losses, and in default probabilities under relevant risk-adjusted measures, as they affect the arbitrage-free prices of defaultable securities. The **default intensity** h_t at a given time t is defined via

$$h_t = \lim_{\Delta t \to 0} \frac{1}{\Delta t} \mathbb{Q}(t < \tau < t + \Delta t | \tau > t), \tag{13.2}$$

that is the default intensity is the probability of default in the very near future per time period. At least for short periods Δt, the default probability can then be approximated by

$$\mathbb{Q}(t < \tau < t + \Delta t | \tau > t) \approx h_t \Delta t.$$

The reduced-form models in Section 13.4 are based on a direct modelling of the default intensity.

13.2.2 Bond prices and credit spreads

First some notation. As before B_t^T is the price at time t of a zero-coupon bond with a risk-free unit payment at time $T \geq t$ and $y_t^T = -(\ln B_t^T)/(T-t)$ is the associated continuously compounded yield. As always, the default-free zero-coupon bond price can be characterized as $B_t^T = \mathrm{E}_t^{\mathbb{Q}}[e^{-\int_t^T r_u\,du}]$, where $\mathbb{Q}$ is a risk-neutral probability measure and $r = (r_t)$ is the default-free short-rate process. Let $\tilde{B}_t^T$ and $\tilde{y}_t^T = -(\ln \tilde{B}_t^T)/(T-t)$ denote the time t price and yield of a defaultable zero-coupon bond promising a unit payment at time $T \geq t$. Defining the **credit spread** or yield spread for maturity T as

$$\zeta_t^T = \tilde{y}_t^T - y_t^T,$$

we see that

$$\frac{\tilde{B}_t^T}{B_t^T} = e^{-\zeta_t^T (T-t)}. \tag{13.3}$$

Obviously, the prices of defaultable bonds depend on the assumptions regarding the time of default and the recovery payment in case of default as well as the relations under a risk-neutral probability measure between the time of default, the recovery, and default-free interest rates used for discounting.

As an example, suppose that the bondholder will get a payment of R at the maturity date T if the firm defaults before the bond matures.[2] The recovery payment R is in general unknown up to the payment date. Of course, if there is no default up to the maturity date, that is $\tau > T$, then the holder of the defaultable bond will get the promised unit payment at time T. The time t price of the defaultable zero-coupon bond conditional on no default up to time t is therefore

$$\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(\mathbf{1}_{\{\tau > T\}} + R\mathbf{1}_{\{\tau \leq T\}}\right)\right],$$

which is equivalent to

$$\begin{aligned}\tilde{B}_t^T &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(1 - (1-R)\mathbf{1}_{\{\tau \leq T\}}\right)\right] \\ &= B_t^T - \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}(1-R)\mathbf{1}_{\{\tau \leq T\}}\right],\end{aligned} \tag{13.4}$$

where B_t^T is the price of a default-free zero-coupon bond maturing at time T. We can think of $1 - R$ as the **loss given default**. We cannot say much about the latter expectation without further assumptions.

Suppose that the default-free short rate, the loss given default, and the default time are all independent of each other under the risk-neutral probability measure $\mathbb{Q}$. Since the (conditional) expectation of the indicator for some event is equal to the (conditional) probability of the event, Equation (13.4) yields

[2] This recovery assumption is called **recovery of treasury**. We will discuss this and two alternative recovery assumptions in Section 13.4.1. In the models considered in the following sections, the recovery payment is often assumed to be made at the time of default instead of the scheduled maturity date.

$$\begin{aligned}\tilde{B}_t^T &= B_t^T - \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right]\mathrm{E}_t^{\mathbb{Q}}[1-R]\,\mathrm{E}_t^{\mathbb{Q}}\left[\mathbf{1}_{\{\tau\le T\}}\right]\\ &= B_t^T - B_t^T\,\mathrm{E}_t^{\mathbb{Q}}[1-R]\mathbb{Q}_t(\tau\le T|\tau>t).\end{aligned}$$

Using this together with (13.3), we can express the credit spread as

$$\begin{aligned}\zeta_t^T &= -\frac{1}{T-t}\ln\left(1-\mathrm{E}_t^{\mathbb{Q}}[1-R]\mathbb{Q}_t(\tau\le T|\tau>t)\right)\\ &\approx \frac{\mathrm{E}_t^{\mathbb{Q}}[1-R]\mathbb{Q}_t(\tau\le T|\tau>t)}{T-t},\end{aligned}\tag{13.5}$$

applying the approximation $\ln(1-x)\approx -x$ for $x\approx 0$. On the other hand, we can write the risk-neutral default probability as

$$\mathbb{Q}_t(\tau\le T|\tau>t) = \frac{1-\tilde{B}_t^T/B_t^T}{\mathrm{E}_t^{\mathbb{Q}}[1-R]} = \frac{1-e^{-\zeta_t^T(T-t)}}{\mathrm{E}_t^{\mathbb{Q}}[1-R]} \approx \frac{\zeta_t^T(T-t)}{\mathrm{E}_t^{\mathbb{Q}}[1-R]},\tag{13.6}$$

using the approximation $e^x\approx 1+x$ for $x\approx 0$. Given a risk-neutral expected loss in case of default, the credit spread provides an estimate of the risk-neutral default probability.

As an alternative to independence assume, as is often done in practice, that the recovery payment is non-stochastic. Then (13.4) implies that

$$\begin{aligned}\tilde{B}_t^T &= B_t^T - (1-R)\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\mathbf{1}_{\{\tau\le T\}}\right]\\ &= B_t^T - (1-R)B_t^T\,\mathrm{E}_t^{\mathbb{Q}^T}\left[\mathbf{1}_{\{\tau\le T\}}\right]\\ &= B_t^T - (1-R)B_t^T\mathbb{Q}_t^T(\tau\le T|\tau>t),\end{aligned}$$

where $\mathbb{Q}^T$ is the T-forward martingale measure introduced in Section 4.4.2. Thus, $\mathbb{Q}_t^T(\tau\le T|\tau>t)$ is the conditional default probability under the T-forward measure, which can then be estimated from observed bond prices or yields and a given recovery rate as

$$\mathbb{Q}_t^T(\tau\le T|\tau>t) = \frac{1-\tilde{B}_t^T/B_t^T}{1-R} \approx \frac{\zeta_t^T(T-t)}{1-R},\tag{13.7}$$

which is similar, but not identical, to the relation (13.6) derived under the independence assumption. We will explore the links between credit spreads, recoveries, and risk-adjusted default probabilities more closely in Section 13.4.

It is natural to compare credit spreads across maturities. The credit spread curve for a given issuer at a given point in time t is the (graph of the) function $u\mapsto\zeta_t^{t+u}$. It is tempting, but also potentially misleading, to think of the credit spread curve as being similar to a yield curve or term structure of interest rates. Spread curves produced by models are better interpreted as *risk structures* as they reflect how the pricing of a single zero-coupon bond varies with the maturity on the bond, not how various outstanding debt contracts with different maturities are simultaneously valued. Also note that when a firm issues a new bond, it is likely to change default probabilities and recovery rates, and therefore it should not be valued using the spread curve based on the firm's existing debt.

13.2.3 Coupon bonds

Now consider a general promised stream of fixed payments Y_i at time t_i with $t_1 < t_2 < \cdots < t_n$. First allow for different recovery rates on the different promised payments so that if the issuer defaults at time $\tau \in (t_{j-1}, t_j)$, then for each $t_i \geq t_j$ the scheduled payment at time t_i is replaced by the payment $R^i Y_i$ at time t_i. The value of this claim at time $t < t_n$, given no default so far, is then

$$\begin{aligned}\tilde{B}_t &= \sum_{t_i > t} Y_i \, \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t_i} r_u \, du} \left(1 - (1 - R^i)\mathbf{1}_{\{\tau < t_i\}}\right)\right] \\ &= \sum_{t_i > t} Y_i \, \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t_i} r_u \, du} \right] - \sum_{t_i > t} Y_i \, \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t_i} r_u \, du} (1 - R^i)\mathbf{1}_{\{\tau < t_i\}} \right].\end{aligned}$$

The first expression shows that the claim can be seen as a portfolio of defaultable zero-coupon bonds with potentially different recovery rates. The second line expresses the value of the defaultable claim as the value of the corresponding default-free claim minus the present value of the loss in case of default. Combining the first expression with (13.4), we can write the price of the defaultable coupon bond as

$$\tilde{B}_t = \sum_{t_i > t} Y_i \tilde{B}_t^{t_i}, \tag{13.8}$$

which is similar to the relation between default-free coupon and zero-coupon bonds. However, it is important to notice that the defaultable zero-coupon bond price $\tilde{B}_t^{t_i}$ on the right-hand side should be computed using the recovery rate R^i appropriate for the coupon payment Y_i, which may differ from the recovery payment on a zero-coupon bond maturing at t_i issued by the same company. Moreover, Equation (13.4) and thus (13.8) presuppose that recovery payments are made at the scheduled payment dates. In reality, the default on a coupon bond will give a single recovery payment compensating for all lost remaining payments; see Jarrow (2004) for more on the validity of (13.8).

Typical corporate bonds are bullet bonds. If the face value is 1 and the coupon is q, then $Y_1 = \cdots = Y_{n-1} = q$ and $Y_n = 1 + q$. Upon default, claims to future coupon payments are often compensated with the same recovery for all coupons and this recovery is typically lower (sometimes even zero) than the recovery compensation of the claim to the face value, see Helwege and Turner (1999). Let R^{cpn} denote the proportion of the promised coupon recovered and R denote the recovery of the face value. The above valuation equation implies that

$$\tilde{B}_t = B_t - q \sum_{t_i > t} \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t_i} r_u \, du} (1 - R^{\text{cpn}})\mathbf{1}_{\{\tau < t_i\}} \right] - \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t_n} r_u \, du} (1 - R)\mathbf{1}_{\{\tau < t_n\}} \right],$$

where $B_t = q\sum_{t_i>t} B_t^{t_i} + B_t^{t_n}$ is the price of the corresponding default-free coupon bond. If the recovery rates are non-random, we can rewrite the value as

$$\tilde{B}_t = B_t - q(1-R^{\text{cpn}})\sum_{t_i>t} B_t^{t_i}\mathbb{Q}_t^{t_i}(\tau < t_i|\tau > t) - (1-R)B_t^{tn}\mathbb{Q}_t^{t_n}(\tau < t_n|\tau > t), \tag{13.9}$$

where $\mathbb{Q}_t^{t_i}(\tau < t_i|\tau > t)$ is the probability of default in the time interval (t, t_i) under the t_i-forward martingale measure. Alternatively, by assuming independence of the recovery payment, the default-free short rate, and the default time, we obtain

$$\tilde{B}_t = B_t - q\sum_{t_i>t} B_t^{t_i}\,\mathrm{E}_t^{\mathbb{Q}}\left[1 - R^{\text{cpn}}\right]\mathbb{Q}_t\left(\tau < t_i|\tau > t\right)$$
$$-B_t^{t_n}\,\mathrm{E}_t^{\mathbb{Q}}[1-R]\mathbb{Q}_t\left(\tau < t_n|\tau > t\right),$$

where the default probabilities are all computed under the risk-neutral measure $\mathbb{Q}$. The yields on the defaultable and the default-free coupon bonds are $\tilde{y}_t$ and y_t defined via

$$\tilde{B}_t = \sum_{t_i>t} Y_i e^{-\tilde{y}_t[t_i-t]}, \qquad B_t = \sum_{t_i>t} Y_i e^{-y_t[t_i-t]},$$

and a coupon bond credit spread can be defined as $\zeta_t = \tilde{y}_t - y_t$. However, it is generally not possible to derive nice and simple expressions for this spread.

As we can see from the equations derived in this section, we need to model the default time and the recovery rate in case of default, and their relations to the default-free interest rates before we can assess the default probabilities and value the defaultable bonds (and other defaultable securities). We will discuss such models in Sections 13.3–13.5.

13.2.4 Default correlation

In order to evaluate the potential losses and the value of a portfolio of defaultable loans or bonds, it is necessary to model the correlations between defaults of the individual loans. Moreover, various credit derivatives are related to possible defaults of multiple firms. If defaults and recovery rates of the different firms were completely independent, valuation of claims or loan portfolios involving multiple issuers would typically not be more complicated than valuing a claim on a single issuer. However, defaults tend to be correlated. Economy-wide shocks may affect virtually all issuers, as reflected, for example, by the market-betas in the classic CAPM or consumption-betas in the consumption-based CAPM. Companies in the same industry are affected by the same external events and it is therefore conceivable that they will face financial difficulties at the same time. The actual default of one firm may also affect the default probability of another firm. The default by one company may lead an economically related company, such as a major supplier or customer, closer to default. This is referred to as a contagion effect. Conversely, in some cases, the surviving companies in an industry may benefit from the default of a competitor leading to lower default probabilities.

The commonality in default risk across companies implies that credit risk cannot be completely diversified away by forming portfolios of corporate loans or bonds. Since the economy-wide default risk is high in bad times, investors will require a default risk premium. Consequently, risk-neutral default probabilities will normally exceed real-world default probabilities. Of course some firms are countercyclical in the sense that they tend to do well when the overall economy is in crisis, and the converse, and for such firms the risk-neutral default probability will be smaller than the real-world default probability. Moreover, diversifying away the idiosyncratic default risk in corporate bonds is more challenging than diversifying away idiosyncratic equity risk. The reason is the difference in the typical return distributions of stocks and corporate bonds. Stock return distributions tend to be smooth and fairly symmetric, not far from a normal or lognormal distribution. In contrast, return distributions of corporate bonds are much less smooth and more skewed. If, for example, you hold a corporate bond to maturity and it does not default, you receive the 'promised' return indicated by the yield of the bond. If it defaults, you get the recovery payment corresponding to a large negative return. It requires a lot of corporate bonds to smooth the return distribution and diversify away the default risk. Since it is also costly to form large portfolios of corporate bonds, investors may require some return premium—in addition to the premium for systematic default risk—to compensate them for the idiosyncratic default risk.

Consider two firms and let the random variables τ_1 and τ_2 be their default times with associated default indicators $\mathbf{1}_{\{\tau_1 \le t\}}$ and $\mathbf{1}_{\{\tau_2 \le t\}}$. The real-world default correlation over the time interval $[0, t]$ is

$$\rho_{12}^D(0,t) \equiv \operatorname{Corr}\left[\mathbf{1}_{\{\tau_1 \le t\}}, \mathbf{1}_{\{\tau_2 \le t\}}\right] = \frac{\mathrm{E}\left[\mathbf{1}_{\{\tau_1 \le t\}}\mathbf{1}_{\{\tau_2 \le t\}}\right] - \mathrm{E}\left[\mathbf{1}_{\{\tau_1 \le t\}}\right]\mathrm{E}\left[\mathbf{1}_{\{\tau_2 \le t\}}\right]}{\sqrt{\operatorname{Var}\left[\mathbf{1}_{\{\tau_1 \le t\}}\right]\operatorname{Var}\left[\mathbf{1}_{\{\tau_2 \le t\}}\right]}}.$$

Since the indicators are either zero or one, the expectations and variances can be stated as[3]

$$\begin{aligned}
\mathrm{E}\left[\mathbf{1}_{\{\tau_1 \le t\}}\mathbf{1}_{\{\tau_2 \le t\}}\right] &= \mathrm{E}\left[\mathbf{1}_{\{\tau_1 \le t, \tau_2 \le t\}}\right] = \mathbb{P}\left(\tau_1 \le t, \tau_2 \le t\right), \\
\mathrm{E}\left[\mathbf{1}_{\{\tau_i \le t\}}\right] &= \mathbb{P}\left(\tau_i \le t\right), \\
\operatorname{Var}\left[\mathbf{1}_{\{\tau_i \le t\}}\right] &= \mathbb{P}\left(\tau_i \le t\right)\left[1 - \mathbb{P}\left(\tau_i \le t\right)\right].
\end{aligned}$$

Consequently, the default correlation becomes

$$\rho_{12}^D(0,t) = \frac{\mathbb{P}\left(\tau_1 \le t, \tau_2 \le t\right) - \mathbb{P}\left(\tau_1 \le t\right)\mathbb{P}\left(\tau_2 \le t\right)}{\sqrt{\mathbb{P}\left(\tau_1 \le t\right)\left[1 - \mathbb{P}\left(\tau_1 \le t\right)\right]\mathbb{P}\left(\tau_2 \le t\right)\left[1 - \mathbb{P}\left(\tau_2 \le t\right)\right]}}. \tag{13.10}$$

The probability of both firms defaulting in $[0, t]$ is therefore

$$\begin{aligned}
\mathbb{P}\left(\tau_1 \le t, \tau_2 \le t\right) &= \mathbb{P}\left(\tau_1 \le t\right)\mathbb{P}\left(\tau_2 \le t\right) \\
&\quad + \rho_{12}^D(0,t)\sqrt{\mathbb{P}\left(\tau_1 \le t\right)\left[1 - \mathbb{P}\left(\tau_1 \le t\right)\right]\mathbb{P}\left(\tau_2 \le t\right)\left[1 - \mathbb{P}\left(\tau_2 \le t\right)\right]}.
\end{aligned}$$

[3] Note that the expressions $\tau_1 \le t, \tau_2 \le t$ in the indicator and the probability are to be understood as $\tau_1 \le t$ *and* $\tau_2 \le t$, that is both events occur.

The probability that at least one of the firms default in $[0, t]$ is

$$\mathbb{P}(\tau_1 \le t \text{ or } \tau_2 \le t) = \mathbb{P}(\tau_1 \le t) + \mathbb{P}(\tau_2 \le t) - \mathbb{P}(\tau_1 \le t, \tau_2 \le t), \tag{13.11}$$

where the last term is given by the previous equation.

Let us consider a small numerical example. Suppose that each of the firms have a 5% default probability over the given time horizon, that is $\mathbb{P}(\tau_i \le t) = 0.05$ with corresponding variance $0.05(1 - 0.05) = 0.0475$. If the default correlation is zero, then the probability that both firms default is $(0.05)^2 = 0.0025$, that is 0.25%, and the probability that at least one of the two firms defaults is $0.05 + 0.05 - 0.0025 = 0.0975$, that is 9.75%. If instead the default correlation is 0.4, the joint default probability becomes 2.15% which is much higher than before, whereas the probability of at least one default only changes to 7.85%.

Since defaults are rare and multiple defaults by the same firm (maybe after some reorganization of the company after a previous default) are even more rare, it is not possible to estimate default correlations directly from historical data. It is necessary to set up a model relating the default times of the firms in question to some variables that are easier to get reliable data on. We return to this issue later in the chapter. Note that if the recovery rates of the different loans or bonds are stochastic, any correlation between the recovery rates can also affect the valuation of a claim related to those loans and should therefore also be modelled.

13.2.5 Credit ratings

In the trading and risk management of defaultable securities, credit ratings play an important role. After evaluating the credit risk, credit rating agencies assign ratings to issuers of defaultable securities and to individual defaultable securities themselves. The main credit rating agencies are the private U.S. based companies Standard & Poor's, Moody's Investor Service, and Fitch, but there are also other such agencies both in the U.S. and in other countries. Table 13.1 lists the primary

Table 13.1: Rating categories of the main credit rating agencies. The best rating is the upper (Aaa, AAA) reflecting a very small risk of default. Lower ratings reflect increasingly higher default risks.

Moody's	S&P and Fitch
Aaa	AAA
Aa	AA
A	A
Baa	BBB
Ba	BB
B	B
Caa	CCC
Ca	CC
C	C

rating categories of the main agencies. Some of the rating categories are sometimes subdivided into finer groups either by adding a plus or a minus (S&P and Fitch) or by adding a 1, 2, or 3 (Moody's). Securities with a rating in the upper four categories are called investment grade securities, while those with a lower rating are often referred to as non-investment grade, speculative grade, or *junk* securities. Ratings for short-term bonds are often given according to slightly different categorizations. Ratings are assigned to bonds of all sorts of issuers, including national and local governments, private corporations, and non-profit organizations. Note that bonds of the same issuer may have different terms (maturity, priority, etc.) that make them more or less credit risky and they can therefore have different ratings. Ratings are also routinely assigned to collateralized debt obligations (see Section 13.9) and to mortgage-backed bonds and collateralized mortgage obligations (see Chapter 14).

If the credit rating agencies produce trustworthy ratings providing good forecasts of default frequencies, they definitely serve to increase the efficiency of financial markets. On a market basis, it is much more cost-effective to centralize the evaluation of credit risk (based on complicated and time-consuming analyses) of the thousands of different credit risk related securities, than to have all market participants performing similar credit evaluations. Without the credit rating industry, it could be impossible for small—and in particular new—companies and other entities to borrow money by issuing bonds in the financial markets. Credit ratings play an obvious role as a credit risk measure that investors can apply in their investment and risk management decisions. Some investors are restricted to securities rated above some level, for example to investment grade securities. Ratings are also used for regulatory purposes. For example, when banks calculate their capital reserves as is required by regulatory authorities, they can use credit ratings from certain approved rating agencies. The lower the rating, the higher the required capital reserves.

Table 13.2 indicates that the credit ratings on average provide a reasonable ranking of default risk. The historical default frequencies are decreasing in the

Table 13.2: The average cumulative global default rates (in per cent) based on Moody's ratings for 1970–2008.

	Maturity (years)								
	1	2	3	4	5	7	10	15	20
Aaa	0.000	0.013	0.013	0.037	0.107	0.250	0.508	0.955	1.139
Aa	0.017	0.054	0.087	0.157	0.234	0.388	0.551	1.074	2.194
A	0.025	0.118	0.272	0.432	0.612	1.025	1.752	3.111	5.102
Baa	0.164	0.472	0.877	1.356	1.824	2.770	4.397	8.009	11.303
Ba	1.113	2.971	5.194	7.523	9.639	13.263	18.276	27.220	34.845
B	4.333	9.752	15.106	19.864	24.175	32.164	41.088	52.190	56.101
Caa-C	16.015	25.981	34.154	40.515	45.800	52.702	63.275	68.873	70.922
Investment Grade	0.068	0.215	0.416	0.651	0.894	1.399	2.237	3.966	5.952
Speculative Grade	4.113	8.372	12.467	16.093	19.245	24.520	30.637	39.343	45.498
All rated	1.401	2.844	4.193	5.360	6.344	7.938	9.802	12.608	15.125

Source: Emery, Ou, Tennant, Matos, and Cantor (2009).

rating, that is higher for lower rated companies and close to zero for top-rated companies. However, the credit rating agencies have been criticized for being too slow in downgrading companies. An example, probably the most severe, is Enron which remained investment grade until 4 days before its 2001 bankruptcy, although its financial problems were apparently well-known to the agencies long before. The lag in rating changes can at least be partially explained by the fact that credit rating agencies take a 'through-the-cycle' perspective so that credit ratings are set to reflect long-term default risk rather than the short-term default probability which might be more relevant to some investors. There is some empirical evidence that the market-determined credit spread increases *before* the rating of the bond is downgraded when the credit quality of a bond-issuing company deteriorates.

Rating agencies are sometimes accused of having too close relations to the companies they are rating, which might generate conflicts of interest that call into question their objectivity. Rating agencies are paid by the issuing companies and can be tempted to provide ratings that please issuers rather than serve the investors and regulatory bodies relying on the ratings. And if an agency tells an issuer that it will be poorly rated, the issuer might take its business to a less pessimistic (or less honest) rating agency. In the current financial crisis (starting around 2007) the major credit rating agencies have been exposed to harsh critique, in particular for their high original ratings of structured products, such as collateralized debt obligations and collateralized mortgage obligations, of which many have later been substantially downgraded or even defaulted. Rating structured products became a very profitable business for the credit rating agencies who often worked together with the issuer to structure the product such that the agency was willing to give a certain high rating, see Section 13.9. Independence between the credit rating agency and the issuer was illusory. The so-called Dodd–Frank Wall Street Reform and Consumer Protection Act, signed by President Obama in July 2010, strengthens the requirements on the nationally recognized rating agencies in the U.S. Among other things, the agencies have to disclose the data and methodologies they apply when setting the ratings, disclose performance statistics, as well as any conflicts of interest with respect to sales and marketing practices.

13.2.6 Historical default statistics

Table 13.2 shows the historical default rates for different Moody's ratings over different time horizons. For example, of all securities initially rated A only 0.025% have defaulted within 1 year and 0.118% within 2 years. Of course, these default frequencies increase with the time horizon. For securities with high initial ratings the default frequency increases convexly with the horizon, whereas the relation is concave for low-rated securities. For example, for an A-rated security there is only 0.612% chance that default will happen in the first 5 years, but then there is an additional 1.752-0.612=1.140% chance that default will happen over the next 5 years. The 1.140% is the *unconditional* probability of default in years 6–10. Since the probability of survival through the first 5 years is $100 - 0.612 = 99.388\%$, the *conditional* probability of default in years 6-10 is 1.140/99.388, that is 1.147%. Over the first 5–year period the credit quality of some of the originally top-rated firms will deteriorate and their default risk will increase. In contrast, for a B-rated security

Table 13.3: Average rating migration rates over 1- and 5-year periods based on Moody's ratings for 1970–2008. 'WR' is short for without rating.

	Average 1-year rating migration rates, 1970–2008									
From/To:	Aaa	Aa	A	Baa	Ba	B	Caa	Ca-C	Default	WR
Aaa	88.494	7.618	0.650	0.026	0.028	0.002	0.002	0.000	0.000	3.179
Aa	1.047	86.817	7.077	0.288	0.042	0.016	0.008	0.001	0.016	4.688
A	0.066	2.832	87.274	4.961	0.473	0.086	0.028	0.003	0.024	4.253
Baa	0.043	0.191	4.786	84.382	4.165	0.781	0.203	0.021	0.163	5.265
Ba	0.008	0.056	0.395	5.678	76.054	7.070	0.549	0.061	1.084	9.045
B	0.011	0.037	0.133	0.346	5.034	73.939	5.090	0.620	4.165	10.624
Caa	0.000	0.026	0.037	0.222	0.484	8.928	60.781	3.589	13.122	12.810
Ca-C	0.000	0.000	0.000	0.000	0.331	2.790	9.446	39.479	30.033	17.921
	Average 5-year rating migration rates, 1970–2004									
From/To:	Aaa	Aa	A	Baa	Ba	B	Caa	Ca-C	Default	WR
Aaa	54.006	23.725	5.327	0.470	0.265	0.040	0.038	0.000	0.085	16.044
Aa	3.322	50.875	21.090	3.219	0.540	0.165	0.031	0.000	0.174	20.585
A	0.224	8.378	53.060	14.324	2.781	0.877	0.169	0.013	0.448	19.725
Baa	0.262	1.204	13.616	46.491	9.082	3.039	0.535	0.070	1.729	23.971
Ba	0.046	0.199	2.375	11.885	27.468	10.865	1.386	0.151	8.024	37.602
B	0.044	0.065	0.312	1.712	6.857	22.207	4.536	0.663	20.932	42.672
Caa	0.000	0.000	0.072	1.235	2.222	6.110	7.354	1.010	39.627	42.370
Ca-C	0.000	0.000	0.000	0.308	0.540	1.747	1.850	2.158	47.867	45.529

Source: Emery, Ou, Tennant, Matos, and Cantor (2009).

the default probability is 24.175% over the first 5 years and only an additional 41.088-24.175= 16.913% probability that default occurs in the following 5 years. Some of the originally B-rated firms that survive the first 5 years will be in better shape afterwards and have a lower default risk. This is also reflected by the credit migration statistics of Table 13.3. The table shows that, in the period 1970–2008, 88.494% of all firms rated Aaa at the beginning of a year were also rated Aaa at the end of the year, 7.618% were downgraded to Aa, 0.650% to A, and so on. Clearly, ratings are typically very stable from year to year, but big changes in ratings do occur. Of course, the probability that a firm is moved to another rating category is higher over a 5-year period than over a 1-year period, so the 5-year migration rate matrix is less concentrated on the diagonal.

When a firm defaults on some of the claims it has issued, it means that the claimants do not receive the contractually promised payments on a timely basis, but it does not necessarily mean that the claimants walk away empty-handed or that the issuing firm seizes to exist. If the financial problems of the issuing firm are likely to be temporary, the firm and the claimants will often try to reach an agreement so that the firm can continue its operations. This agreement will sometimes include a reorganization of the firm. The original claim is replaced by a new claim. For example, if the firm has defaulted on bonds it had issued, the bondholders will normally receive a new, longer-term bond with no or smaller payments in the near future and/or receive stocks issued by the firm. If the financial distress of the defaulted firm is more severe or no agreement between the firm and its claimants can be reached, the firm is liquidated so that its assets are sold and the net proceeds are then distributed to the claimants. The value of the new claim or the liquidation payment received by the owner of a claim on a defaulted firm constitutes the recovery payment. Recovery payments are often normalized by the face value of the claim and are then referred to as a recovery rate. Note that some claims may have a higher priority than other claims and will therefore receive a higher recovery payment. Stockholders have lowest priority. According to the absolute priority rule, the net proceeds from a liquidation should first be used to satisfy the claim with highest priority such as senior secured debt. If there is still something left, the claim with second-highest priority will be considered, and so forth. Typically nothing should be left for stockholders. However, the absolute priority rule is sometimes violated, see, for example, Eberhart and Senbet (1993), Eberhart and Weiss (1998), and Bebchuk (2002).

Table 13.4 lists the average recovery rates in Europe and North America on corporate loans and bonds over the period 1985–2008. The recovery rates exhibit the expected ranking in terms of priority of the debt. North American recovery rates tend to be higher than European ones. There is considerable variation in recovery rates across rating classes, within rating classes, and over time. Recovery rates tend to be negatively correlated with default rates so that recovery rates tend to be low at times where many firms default, see, for example the discussion and references in Altman, Brady, Resti, and Sironi (2005).

Market prices of defaultable securities reflect default probabilities (and recovery risk) under appropriately risk-adjusted probabilities. In particular, under some

Table 13.4: Average recovery rates for European and North American Defaulters, 1985–2008. The recovery rate is estimated using 30-day post-default bid prices on defaulted debt.

Class	Europe	North America
Senior secured loan	58.7	68.9
Senior unsecured loan	45.9	56.2
Senior secured bond	42.4	53.5
Senior unsecured bond	28.6	37.2
Senior subordinated bond	32.6	31.6
Subordinated bond	23.8	30.1

Source: Parwani, Emery, and Cantor (2009, Exhibit 15).

Table 13.5: 7-year average historical and implied default probabilities.

Rating	Historical default intensity	Implied default intensity from bonds	Ratio	Difference
Aaa	0.04	0.60	16.7	0.56
Aa	0.05	0.74	14.6	0.68
A	0.11	1.16	10.5	1.04
Baa	0.43	2.13	5.0	1.71
Ba	2.16	4.67	2.2	2.54
B	6.10	7.97	1.3	1.98
Caa-C	13.07	18.16	1.4	5.50

Source: Hull (2009, Table 22.4)

fairly restrictive assumptions, the risk-neutral default probabilities implied by market prices can be computed from Equation (13.6). Table 13.5 lists historical rates and implied probabilities of corporate bond defaults over a 7-year horizon for the different rating classes. The observed credit spreads imply default probabilities that are markedly higher than the historical default rates. The ratio of the implied default probability to the historical default rate declines as the credit quality declines, whereas the difference increases as credit quality declines. The historical default frequencies and recovery rates would only justify a much smaller credit spread than observed. In other words, corporate bonds seem to be priced too low and offer too high average returns. But, as explained in Section 13.2.4, credit risk has a systematic component with higher default rates in bad times, where investors would appreciate payments more than in good times. The risk-neutral default probabilities are therefore higher than the real-world default probabilities. Still, the basic pricing models for defaultable bonds are incapable of explaining credit spreads that are as high as observed. This phenomenon is known as the **credit spread puzzle**. We will discuss risk premia again in Section 13.4.6 and refer to the credit spread puzzle throughout the modelling sections.

13.3 STRUCTURAL MODELS

In order to quantify the default risk of an issuer of some contract, it is natural to model the assets and liabilities of the issuer since defaults happen when the assets are small relative to the liabilities. This is the approach taken by the class of so-called structural models, sometimes also referred to as firm-value models. As we shall see below, structural models apply standard option pricing techniques for valuing the equity and debt of the company, and therefore this methodology is sometimes called the option-based approach to the valuation of corporate claims.

The advantage of structural models is that they provide insights regarding when and in what situations firms default on their debt and they can potentially be used for predicting defaults of individual firms. At least in principle, they can be used for analysing how a change in the capital structure of the firm or in the asset dynamics affects default probabilities, credit spreads, and so on. The main problem is that both the assets and the liabilities of firms are typically complex and difficult to represent in a tractable mathematical model and some of the inputs to such a model are hard to observe or estimate. Moreover, many investors hold bonds and other financial contracts issued by many different companies and the assets and liabilities of some of these companies might very well be interrelated, which calls for simultaneous modelling of many companies leading to huge and very complicated models. While such large-scale models may be useful or even necessary for rating agencies and very big investors, even simple structural models offer some fundamental understanding of relevance for all types of investors and issuers.

We will first present the path-breaking structural model of Merton (1974) and then discuss some of the many extensions. Along the way we will comment on practical applications and the empirical performance of the models. In most of this section we focus on a single issuer, but at the end of the section we briefly discuss how to handle multiple issuers.

13.3.1 The Merton model

The seminal Merton (1974) model is basically a clever application of the Black–Scholes–Merton stock option pricing model. It therefore makes many of the same assumptions, that is continuous and frictionless trading, infinite divisibility of all assets, and constant interest rates. Merton assumes the simplest possible debt structure consisting of a single zero-coupon loan or bond. Let T denote the maturity date and F the face value (and thus the promised payment to the creditors at maturity) of the bond issued by the company. Let V_t denote the time t value of the assets of the company. Merton assumes that the company can only default on this loan at the maturity date and will do so if and only if the value of the assets is lower than the face value of the loan. By assuming no bankruptcy costs, that the so-called absolute priority rule is followed, and that the legal process induced by a default is instantaneous, the bondholders will receive V_T at time T in case of default and, of course, F if the company does not default. The final payment to the bondholders is therefore

$$\tilde{B}_T = \min(F, V_T) = V_T - \max(V_T - F, 0) = F - \max(F - V_T, 0),$$

and the value of the equity at the maturity date of the debt is

$$S_T = V_T - \tilde{B}_T = \max(V_T - F, 0).$$

Now the option analogy is obvious: the equity value at time T is equal to the payoff of a call option on the assets of the firm with an exercise price equal to the face value of the debt. The equity owners have the right to keep control of the firm, but they need to honour the claim of the debtholders. The bondholders' claim can be seen as a sure claim to F minus a put option on the value of the assets.

The option analogy immediately leads to the so-called asset substitution problem originally pointed out by Jensen and Meckling (1976): the shareholders have an incentive to increase the volatility of the assets of the firm since that will make their call option on the assets more valuable. Note however that this is only true after the debt was issued. In pricing the debt at issuance, potential bondholders will take into account that incentive and price the debt accordingly so that the proceeds to the stockholders are reduced. The asset substitution incentive motivates the introduction of bond covenants limiting the possibility of the stockholders to increase asset volatility.[4]

[4] Bond covenants can also limit the dividend payments to stockholders. If the firm is in financial problems and default is imminent, stockholders have an incentive to sell the assets of the firm and pay out the proceeds as dividends to stockholders leaving little or nothing to bondholders. Moreover, if a firm issues new debt with the same (or even higher) priority than existing debt, it will reduce (dilute) the value of the existing debt, unless the debt can be considered free of default risk. It would therefore seem natural to issue corporate bonds with a covenant prohibiting the issue of further debt contracts, unless these have lower priority. However, as discussed by, for example Collin-Dufresne and Goldstein (2001), bond covenants have been rare in the recent couple of decades.

The recovery payment and the default time in Merton's model are

$$R = V_T, \qquad \tau = \begin{cases} T, & \text{if } V_T < F, \\ \infty, & \text{otherwise.} \end{cases}$$

Assuming the existence of a risk-neutral probability measure $\mathbb{Q}$ and a constant risk-free rate r, the value of the bond at time $t < T$ will be

$$\tilde{B}_t = e^{-r(T-t)} \, \mathrm{E}_t^{\mathbb{Q}} \left[\min(F, V_T)\right].$$

The time $t < T$ value of the equity is (assuming no dividends before time T)

$$S_t = e^{-r(T-t)} \, \mathrm{E}_t^{\mathbb{Q}} \left[\max(V_T - F, 0)\right].$$

At any point in time the value of the equity and the debt add up to the value of the assets, $S_t + \tilde{B}_t = V_t$.

Merton assumes that the value of the assets of the firm follows a geometric Brownian motion,

$$dV_t = V_t \left[\mu \, dt + \sigma \, dz_t\right],$$

where σ is the asset volatility and μ is the expected rate of value changes. For simplicity we assume that the firm does not pay dividends before maturity of the debt. By assuming that the assets of the firm represent a traded security, it is clear that the risk-neutral dynamics of the asset value will be

$$dV_t = V_t \left[r \, dt + \sigma \, dz_t^{\mathbb{Q}}\right]. \tag{13.12}$$

It follows that the value of the equity at time $t < T$ is similar to the value of a stock option in the Black–Scholes–Merton framework,

$$S_t = V_t N\left(d(V_t, T-t)\right) - F e^{-r(T-t)} N\left(d(V_t, T-t) - \sigma\sqrt{T-t}\right), \tag{13.13}$$

where

$$d(V, u) = \frac{\ln\left(V/F\right) + (r + \sigma^2/2)u}{\sigma\sqrt{u}}.$$

Recall from Chapter 6 that $N\left(d(V_t, T-t) - \sigma\sqrt{T-t}\right)$ is the probability under the T-forward probability measure of the option ending in the money. Since default-free interest rates are assumed constant for now, this coincides with the risk-neutral probability of the same event, that is $\mathbb{Q}_t(V_T \geq F)$. The time t value of the bond is

$$\begin{aligned} \tilde{B}_t &= V_t - S_t \\ &= V_t N\left(-d(V_t, T-t)\right) + F e^{-r(T-t)} N\left(d(V_t, T-t) - \sigma\sqrt{T-t}\right). \end{aligned} \tag{13.14}$$

The risk-neutral probability that the company will default on the debt is given by

$$\mathbb{Q}_t(V_T < F) = 1 - \mathbb{Q}_t(V_T \geq F) = 1 - N\left(d(V_t, T-t) - \sigma\sqrt{T-t}\right)$$
$$= N\left(-\left[d(V_t, T-t) - \sigma\sqrt{T-t}\right]\right).$$

In order to calculate this probability, we need to know the current value V_t of the assets and the volatility σ of the assets, but neither of these are directly observable. If the company is publicly traded, we can observe the current market value of the equity, S_t, and we can estimate the equity volatility, σ_{St}.[5] Note that since equity is an option on a lognormal process, the equity value itself will not be a lognormal process and hence σ_{St} is not constant. An application of Itô's Lemma on (13.13) (this is Exercise 13.1) shows that the risk-neutral dynamics of the equity value is

$$dS_t = S_t\left[r\,dt + \sigma_{St}\,dz_t^{\mathbb{Q}}\right], \qquad \sigma_{St} = \sigma N\left(d(V_t, T-t)\right)\frac{V_t}{S_t}. \tag{13.15}$$

From the expression for S_t and σ_{St}, one can back out the asset value V_t and asset volatility σ.

The real-world default probability is by analogy given by

$$\mathbb{P}_t(V_T < F) = N\left(-\mathcal{D}_t\right),$$

where

$$\mathcal{D}_t = \frac{\ln(V_t/F) + (\mu + \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}} - \sigma\sqrt{T-t} = \frac{\ln(V_t/F) + (\mu - \sigma^2/2)(T-t)}{\sigma\sqrt{T-t}}$$

is sometimes called the **distance-to-default**. The higher the distance-to-default, the lower the default probability. The distance-to-default is clearly increasing in the drift rate of asset value and in the current asset–debt ratio, but decreasing in asset volatility. Note that in order to compute $\mathcal{D}_t$ and the real-world default probability, we also need to know the drift μ of the asset value. The distance-to-default measure is widely used by practitioners and is a key element in various commercial credit risk management systems such as Moody's KMV, see Crouhy, Galai, and Mark (2000).[6] Duffie, Saita, and Wang (2007) and Bharath and Shumway (2008) show in empirical studies that the Merton distance-to-default measure has some predictive power for default. It cannot match the observed default frequencies, but

[5] The equity volatility is not a constant in this model, but a stochastic process. The volatility estimation is therefore not as simple as when estimating the constant volatility of a stock price in the standard option pricing application of the Black–Scholes–Merton model, see Lando (2004, Sec. 2.11).

[6] KMV is an acronym for Kealhofer, McQuown, and Vasicek who founded the KMV corporation in 1989 which in 2002 was acquired by Moody's Corporation for more than USD 200 million. Moody's KMV Corporation uses available market prices and information in financial statements of individual firms together with a large database on defaults and losses to produce estimates of default probabilities over various time horizons, distance-to-default measures, credit spread curves, and loss given default for the firms covered. Apparently, the approach of Moody's KMV allows for more general debt structures involving different priorities and maturities, and it transforms the distance-to-default measure into a default probability by using historical default frequencies.

the distance-to-default measure produces a reasonable ranking in the sense that firms with a high distance-to-default measure tend to default less frequently than firms with a low distance-to-default.

Let $\tilde{B}_t^T$ denote the time t price of a defaultable zero-coupon bond maturing at time T and let $\tilde{y}_t^T = -\frac{1}{T-t} \ln \tilde{B}_t^T$ denote the corresponding continuously compounded yield. The value of the debt in the Merton model is then clearly $\tilde{B}_t = F\tilde{B}_t^T$ and (13.14) implies that

$$\tilde{y}_t^T = -\frac{1}{T-t} \ln \left(\frac{V_t}{F} N(-d) + e^{-r(T-t)} N(d - \sigma\sqrt{T-t}) \right).$$

Subtracting the default-free yield $y_t^T = r = -\frac{1}{T-t} \ln(e^{-r(T-t)})$, the credit spread is

$$\zeta_t^T = \tilde{y}_t^T - y_t^T = -\frac{1}{T-t} \ln \left(\frac{V_t}{Fe^{-r(T-t)}} N(-d) + N(d - \sigma\sqrt{T-t}) \right). \quad (13.16)$$

Of course, the credit spread is always non-negative and decreasing in the asset–debt ratio V_t/F. The credit spread risk structure $u \mapsto \zeta_t^{t+u}$ can be increasing, decreasing, or hump-shaped depending on the leverage and various parameter values, see Exercise 13.3. For highly leveraged firms in financial distress, the credit spread is decreasing. Due to the current distress, the short-term credit spreads are high, but if the firm survives, it is likely to be in better shape in the future so long-term credit spreads are smaller. However, Fons (1994) and Sarig and Warga (1989) report that flat or downward-sloping risk structures are not uncommon even for non-distressed firms. On the other hand, Helwege and Turner (1999) find empirical examples of upward-sloping risk structures for low-grade debt. According to Jones, Mason, and Rosenfeld (1984), Huang and Huang (2003), and Eom, Helwege, and Huang (2004), the Merton model produces too low credit spreads.

The limiting behaviour of the credit spread ζ_t^T for $t \to T$ is interesting. It can be shown that the credit spread will approach zero when $V_T > F$ and approach ∞ when $V_T < F$, see Lando (2004, Sec. 2.2.2). If the asset value is above the outstanding debt near maturity, the credit spread will be very close to zero. This is due to the fact that the asset value process in the model has continuous sample paths so that the asset value shortly before debt maturity will lead to a very good prediction of whether or not the firm defaults. In technical terms, the default time τ is a **predictable** stopping time. Empirically, however, even highly rated corporate bonds are traded with non-zero short-term credit spreads, see Duffee (1999) and Collin-Dufresne and Goldstein (2001).

13.3.2 Premature defaults when hitting a barrier

In the basic Merton model outlined above default can only happen at the maturity date of the zero-coupon debt, but this is a restrictive and unrealistic feature. Black and Cox (1976) extended Merton's setting by assuming that default can happen at any time and will do so the first time that the firm value is smaller than or equal to some default barrier. In case the barrier is hit before maturity, the bondholders immediately take over the firm. For any t, let $\underline{V}_t$ denote the default barrier. The time of default is then

$$\tau = \inf\{t : V_t \le \underline{V}_t\}$$

and the recovery payment to creditors in case of premature default is $R_\tau = \underline{V}_\tau$. As long as we stick to a model where both the firm value and the default barrier have continuous sample paths, the time of default is

$$\tau = \inf\{t : V_t = \underline{V}_t\},$$

and is therefore often called the first hitting time of the barrier. If $\tau \le T$, the bondholders receive $\underline{V}_\tau$ at time τ. If $\tau > T$, the bondholders receive $\min(F, V_T)$ at time T as before. Given no default so far, the time t value of the bond is therefore

$$\begin{aligned}\tilde{B}_t &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-r(\tau-t)}\underline{V}_\tau \mathbf{1}_{\{\tau\le T\}} + e^{-r(T-t)}\min(F, V_T)\mathbf{1}_{\{\tau>T\}}\right] \\ &= \underbrace{\mathrm{E}_t^{\mathbb{Q}}\left[e^{-r(\tau-t)}\underline{V}_\tau \mathbf{1}_{\{\tau\le T\}}\right]}_{\tilde{B}_t^{\mathrm{bar}}} + \underbrace{\mathrm{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}\min(F, V_T)\mathbf{1}_{\{\tau>T\}}\right]}_{\tilde{B}_t^{\mathrm{mat}}},\end{aligned}$$

where the first part refers to the payment when the barrier is crossed and the second part refers to the payment at maturity. The value of the potential payment at the maturity date can be further decomposed into

$$\tilde{B}_t^{\mathrm{mat}} = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}V_T\mathbf{1}_{\{\tau>T\}}\right] - \mathrm{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}\max(V_T - F, 0)\mathbf{1}_{\{\tau>T\}}\right].$$

Here the last term is the value of a so-called down-and-out call option on the firm value, that is a call option which is eliminated if the underlying firm value drops to the barrier before maturity, and this is exactly the value of the equity under the stated assumptions. Down-and-out options are in fact traded over-the-counter, for example, on foreign exchange rates. When the underlying follows a geometric Brownian motion and the barrier is constant, the price of such an option is known in closed form (Rubinstein and Reiner (1991); Björk (2009), Ch. 18).

The term $\mathrm{E}_t^{\mathbb{Q}}\left[e^{-r(T-t)}V_T\mathbf{1}_{\{\tau>T\}}\right]$ represents the present value of the firm conditional on not hitting the barrier before maturity of the debt and can also be computed in closed form under those assumptions. With the constant barrier $\underline{V}_s = \underline{v}$, where we assume $V_t > \underline{v}$, we get

$$\begin{aligned}\tilde{B}_t^{\mathrm{mat}} &= \tilde{B}^{\mathrm{mat}}(V_t, T-t, F, \underline{v}) \equiv \underline{v}H(V_t, T-t, \underline{v}) - \underline{v}\left(\frac{\underline{v}}{V_t}\right)^{\alpha} H\left(\frac{\underline{v}^2}{V_t}, T-t, \underline{v}\right) \\ &\quad + C^{\mathrm{BS}}(V_t, \underline{v}, T-t) - \left(\frac{\underline{v}}{V_t}\right)^{\alpha} C^{\mathrm{BS}}\left(\frac{\underline{v}^2}{V_t}, \underline{v}, T-t\right) - C^{\mathrm{BS}}(V_t, F, T-t) \\ &\quad + \left(\frac{\underline{v}}{V_t}\right)^{\alpha} C^{\mathrm{BS}}\left(\frac{\underline{v}^2}{V_t}, F, T-t\right),\end{aligned}$$

where $\alpha = (2r/\sigma^2) - 1$, $C^{BS}(v, K, T - t)$ is the Black–Scholes call price for exercise price K, time-to-maturity $T - t$, and a current value of the underlying equal to v, and

$$H(v, u, \underline{v}) = e^{-ru} N \left(\frac{\ln(v/\underline{v}) + (r - \frac{1}{2}\sigma^2)u}{\sigma\sqrt{u}} \right)$$

reflects the value of getting 1 at maturity if the asset value ends up above $\underline{v}$ after u years given that it currently equals v.

These expressions can be extended to an exponential default barrier, $\underline{V}_s = \underline{v}e^{-\gamma(T-s)}$, so that $\underline{v}$ is the default barrier at the maturity date of the debt, while the barrier is lower before maturity when $\gamma > 0$. The relevant value is then

$$\tilde{B}_t^{\text{mat}} = \tilde{B}^{\text{mat}} \left(e^{-\gamma(T-t)} V_t, T - t, e^{-\gamma(T-t)} F, e^{-\gamma(T-t)} \underline{v} \right),$$

where the risk-free rate r furthermore has to be replaced by $r - \gamma$.

Next consider the payment at the barrier. The risk-neutral firm value dynamics (13.12) implies that $V_s = V_t \exp\{(r - \sigma^2/2)(s - t) + \sigma(z_s^{\mathbb{Q}} - z_t^{\mathbb{Q}})\}$. The exponential barrier is then hit the first time $z_s^{\mathbb{Q}} - z_t^{\mathbb{Q}} = (1/\sigma)[\ln(\underline{v}/V_t) - \gamma(T - s) - (r - \gamma - \sigma^2/2)(s - t)]$, that is the first time a standard Brownian motion hits a boundary which is deterministic and affine in time. This is mathematically tractable and leads to the expression

$$\tilde{B}_t^{\text{bar}} = \underline{v}\, e^{-\gamma(T-t)+\beta[m-\tilde{m}]} \left(N \left(\frac{\beta - \tilde{m}(T - t)}{\sqrt{T - t}} \right) + e^{2\beta\tilde{m}} N \left(\frac{\beta + \tilde{m}(T - t)}{\sqrt{T - t}} \right) \right),$$

where

$$\beta = \frac{\ln(\underline{v}/V_t) - \gamma(T - t)}{\sigma}, \quad m = \frac{r - \gamma - \sigma^2/2}{\sigma}, \quad \tilde{m} = \sqrt{2(r - \gamma) + m^2}.$$

For details, see, for example Lando (2004, Appendix B).

Adding up $\tilde{B}_t^{\text{mat}}$ and $\tilde{B}_t^{\text{bar}}$, we find the value of the defaultable bond with the exogenously given exponential default barrier. The shareholders effectively have a down-and-out call option on the assets of the firm and this is clearly less valuable than a standard call option with the same maturity and strike, which is what the shareholders have in the Merton model. As the shareholders are worse off, the bondholders are necessarily better off in the model with potential premature defaults than in the Merton model. Therefore, the credit spreads are lower than in the Merton model, in particular for long-term debt. Obviously, the opportunity of premature defaults increases the default probabilities compared to Merton's model. However, when the default barrier is hit, the bondholders take over the firm and this more than compensates them for the increased likelihood of default. Also note that since the asset value dynamics is modelled by a diffusion, bondholders will never get less than the value at the barrier.

13.3.3 Stochastic interest rates

Default-free interest rates are assumed to be constant in the basic Merton model, but real-life default-free rates fluctuate stochastically with considerable effects on the valuation of bonds and other debt contracts as well as many other financial assets. As default probabilities and credit spreads on corporate debt may also be significantly affected by variations in default-free rates and the covariation between default-free rates and the value of the assets of the firm, the extension of Merton's model to stochastic interest rates is crucial. Interest rate uncertainty does not change the fact that the equity of the firm is equivalent to a European call option on the assets of the firm when default can only happen at maturity and the value of the corporate debt equals the value of the assets minus the value of that call option. The key problem is therefore to value the call option under stochastic interest rates. This is analogous to an extension of the Black–Scholes–Merton stock option pricing model to stochastic interest rates, an extension already considered by Merton (1973) for the case of Gaussian models of the default-free short rate. Shimko et al. (1993) provided the first detailed study of the impact of stochastic default-free rates on corporate bond valuation and credit spreads under the assumption that the default-free rates are described by the one-factor Vasicek model, which was studied in detail in Section 7.4. We will first consider the valuation problem for general dynamics of default-free rates and later see where the specialization to Gaussian dynamics is fruitful.

Equity gives right to a time T payment of $\max(V_T - F, 0)$ and thus has a value at time $t < T$ equal to

$$S_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\max(V_T - F, 0)\right] = B_t^T\,\mathrm{E}_t^{\mathbb{Q}^T}\left[\max(V_T - F, 0)\right], \tag{13.17}$$

where $\mathbb{Q}^T$ is the T-forward measure corresponding to the use of the default-free zero-coupon bond maturing at time T as the numeraire. The value of the corporate debt is $\tilde{B}_t = V_t - S_t$. To compute the last of the expectations in (13.17), we need to know the distribution of V_T under the measure $\mathbb{Q}^T$. Assuming no dividend payout, the risk-neutral dynamics of the asset value is

$$dV_t = V_t\left[r_t\,dt + \sigma\,dz_t^{\mathbb{Q}}\right].$$

In order to find the $\mathbb{Q}^T$-dynamics of V_t, we need to adjust the drift using the sensitivity of the default-free zero-coupon bond price as explained in Section 4.4.2. Write the risk-neutral price dynamics of the default-free zero-coupon bond maturing at time T as

$$dB_t^T = B_t^T\left[r_t\,dt + \rho\sigma_t^T\,dz_t^{\mathbb{Q}} + \sqrt{1-\rho^2}\sigma_t^T\,d\hat{z}_t^{\mathbb{Q}}\right],$$

where $\hat{z}^{\mathbb{Q}}$ is a standard Brownian motion independent of $z^{\mathbb{Q}}$, σ_t^T is the—potentially stochastic—instantaneous volatility of the bond, and ρ is the instantaneous correlation between the asset value and the bond price. The above structure of bond price dynamics is very general and it also includes multi-factor models.

It is useful to focus on the forward price of the assets for delivery at time T, which is $F_t^T = V_t/B_t^T$ as deduced in Section 6.2. At time T, the forward price equals the spot price, that is $F_T^T = V_T$, which we want to know the $\mathbb{Q}^T$-distribution of. We know that the forward price F_t^T is a martingale under the $\mathbb{Q}^T$-measure, that is it has zero drift. The sensitivity of the forward price to the shocks will be the same under the $\mathbb{Q}^T$-measure as under the risk-neutral measure $\mathbb{Q}$. Given the above risk-neutral processes for V_t and B_t^T, the sensitivities of the forward price $F_t^T = V_t/B_t^T$ follow from Itô's Lemma. In sum, the $\mathbb{Q}^T$-dynamics of the forward price is

$$dF_t^T = F_t^T \left[(\sigma - \rho\sigma_t^T)\, dz_t^T - \sqrt{1-\rho^2}\sigma_t^T\, d\hat{z}_t^T \right], \tag{13.18}$$

where z^T and $\hat{z}^T$ are independent standard Brownian motions under $\mathbb{Q}^T$; you are asked to show this in Exercise 13.2. A standard application of Itô's Lemma shows that

$$\begin{aligned} d(\ln F_t^T) &= -\frac{1}{2}\left((\sigma - \rho\sigma_t^T)^2 + (1-\rho^2)\sigma_t^T \right) dt + (\sigma - \rho\sigma_t^T)\, dz_t^T - \sqrt{1-\rho^2}\sigma_t^T\, d\hat{z}_t^T \\ &= -\frac{1}{2}\left(\sigma^2 + (\sigma_t^T)^2 - 2\rho\sigma\sigma_t^T \right) dt + (\sigma - \rho\sigma_t^T)\, dz_t^T - \sqrt{1-\rho^2}\sigma_t^T\, d\hat{z}_t^T \end{aligned}$$

and, consequently, the terminal value is

$$\begin{aligned} F_T^T = F_t^T \exp\Bigg\{ &-\frac{1}{2}\int_t^T \left(\sigma^2 + (\sigma_u^T)^2 - 2\rho\sigma\sigma_u^T \right) du \\ &+ \int_t^T (\sigma - \rho\sigma_u^T)\, dz_u^T - \int_t^T \sqrt{1-\rho^2}\sigma_u^T\, d\hat{z}_u^T \Bigg\}. \end{aligned} \tag{13.19}$$

For a general bond price volatility process $\sigma^T = (\sigma_t^T)$ the $\mathbb{Q}^T$-distribution of the right-hand side is unknown, but the formula is still useful as the basis of a Monte Carlo simulation of V_T to approximate the expectation $\mathrm{E}_t^{\mathbb{Q}^T}[\max(V_T - F, 0)]$ and thus the values of equity and debt.[7]

With a deterministic bond price volatility $\sigma_t^T = \sigma^T(t)$, the stochastic integrals in (13.19) are normally distributed so that $V_T = F_T^T$ is lognormally distributed with mean $\mathrm{E}_t^{\mathbb{Q}^T}[V_T] = \mathrm{E}_t^{\mathbb{Q}^T}[F_T^T] = F_t^T$ as the forward price is a $\mathbb{Q}^T$-martingale. Theorem A.4 in Appendix A then implies that the equity value in (13.17) can be written in closed form as

$$S_t = B_t^T \left(F_t^T N(d_1) - FN(d_2) \right) = V_t N(d_1) - FB_t^T N(d_2),$$

where

$$d_1 = \frac{1}{v_F(t,T)} \ln\left(\frac{V_t}{FB_t^T} \right) + \frac{1}{2} v_F(t,T), \quad d_2 = d_1 - v_F(t,T),$$

[7] In fact, a similar procedure applies when the asset volatility σ is allowed to vary stochastically according to some Itô process.

and

$$\begin{aligned}
v_F(t,T)^2 &\equiv \mathrm{Var}_t^{\mathbb{Q}^T}[\ln F_T^T] = \mathrm{Var}_t^{\mathbb{Q}^T}\left[\int_t^T (\sigma - \rho\sigma_u^T)\, dz_u^T - \int_t^T \sqrt{1-\rho^2}\sigma_u^T\, d\hat{z}_u^T\right] \\
&= \int_t^T (\sigma - \rho\sigma^T(u))^2\, du + \int_t^T (1-\rho^2)\sigma^T(u)^2\, du \\
&= \sigma^2(T-t) + \int_t^T (\sigma^T(u))^2\, du - 2\rho\sigma \int_t^T \sigma^T(u)\, du,
\end{aligned}$$

where we have exploited the independence of the Brownian motions z^T and $\hat{z}^T$ as well as Theorem 3.3. The first term in the final expression for the variance is due to the uncertainty about the future value of the underlying assets, the second term is due to the uncertainty about the default-free discount factor, and the third term is due to the covariance of the asset value and the default-free discount factor.

The bond volatility is deterministic in Gaussian models. For example, in the one-factor Vasicek model the bond price volatility is $\sigma^T(u) = \sigma_r b(T-u)$, where σ_r is the (absolute) volatility of the default-free short rate and $b(s) = \left(1 - e^{-\kappa s}\right)/s$, where κ is the mean reversion speed.

As long as the correlation between the default-free interest rate and the firm value is close to zero, allowing stochastic interest rates has very limited effects on the credit spreads. However, when the correlation is far from zero, the credit spreads become markedly different. With a positive correlation, the credit spreads are higher than for a zero correlation (or constant interest rates). The intuition is that the default-free rates will tend to fall if the asset value falls. This lowers the risk-neutral drift of the asset value so that default becomes more likely. Conversely, negative correlation induces lower credit spreads.

Stochastic interest rates have been combined with barriers for premature defaults. Longstaff and Schwartz (1995a) consider a model with Vasicek-dynamics of the default-free interest rates and assume that premature defaults happen at a constant default barrier. In case of default the bondholders are assumed to get a recovery payment at the scheduled maturity date equal to a fraction of the face value, namely $(1-w)F$. The time t value of the bond—assuming the barrier has not been hit yet—is therefore

$$\begin{aligned}
\tilde{B}_t^T &= F\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_s\, ds}\left(1 - w\mathbf{1}_{\{\tau \le T\}}\right)\right] \\
&= FB_t^T\,\mathrm{E}_t^{\mathbb{Q}^T}\left[1 - w\mathbf{1}_{\{\tau \le T\}}\right] = FB_t^T\left(1 - w\mathbb{Q}_t^T(\tau \le T)\right).
\end{aligned}$$

The $\mathbb{Q}^T$-dynamics of the firm value is

$$dV_t = V_t\left[(r_t - \sigma_r b(T-t))\, dt + \sigma\, dz_t^T\right].$$

Due to the presence of the stochastic default-free short rate in the drift, it seems impossible to derive closed-form expressions for when the process hits a constant barrier, but it can be computed by numerical methods.[8] Also assuming Vasicek-type default-free interest rates, but a stochastic default barrier proportional to the price of similar default-free debt, that is $\underline{V}_t = \alpha F B_t^T$ for some constant $\alpha \leq 1$, Briys and de Varenne (1997) are indeed able to derive a closed-form expression for the value of the defaultable debt.

Other models for defaultable bond valuation with stochastic interest rates have been studied by Kim et al. (1993) and Cathcart and El-Jahel (1998). Both papers assume CIR-type dynamics of default-free rates and a constant default barrier, but differ on some modelling details. Nielsen et al. (1993) and Saa-Requejo and Santa-Clara (1999) consider models involving both stochastic interest rates and a stochastic default barrier. Except for very special cases, debt and equity must be valued numerically in these models.

13.3.4 More general debt contracts

The basic Merton model assumes that the debt of the firm consists of a single loan with a single payment date, that is a zero-coupon loan. Obviously, real-life debt structures are more complicated and involve loans with several payment dates (coupon loans) and may also involve loans of different priorities.

First, consider the case where the firm has issued a coupon loan promising payments of Y_i at time t_i with $t_1 < t_2 < \cdots < t_n$. At time $t \in (t_{n-1}, t_n)$, we are effectively in the original Merton setting with a single remaining payment on the debt, so the value of the equity S_t is given by the Black–Scholes option pricing formula as in (13.13) with $F = Y_n$ and $T = t_n$ and the value of the debt is $\tilde{B}_t = V_t - S_t$ as in (13.14). Before time t_{n-1}, the situation is more complicated. The equity owners need to make the scheduled debt payment of Y_{n-1} at time t_{n-1} in order to keep the firm running and to get the call option on the assets of the firm up to the final debt payment date. In other words, the equity owners have a compound option: an option to purchase another option. If the last-period option has a value at time t_{n-1} below Y_{n-1}, the equity owners will choose to default. The default barrier at time t_{n-1} is the asset value $\underline{V}_{n-1}$ at which

$$Y_{n-1} = \underline{V}_{n-1} N\left(d(\underline{V}_{n-1}, \Delta t_n)\right) - Y_n e^{-r\Delta t_n} N\left(d(\underline{V}_{n-1}, \Delta t_n) - \sigma\sqrt{\Delta t_n}\right),$$

where $\Delta t_n = t_n - t_{n-1}$. Going back in time, you keep adding layers of options. At time $t \in (t_{n-3}, t_{n-2})$, the equity owners have the option to pay Y_{n-2} at time t_{n-2} to get the option to pay Y_{n-1} at time t_{n-1} to get the option on the assets expiring at time t_n with an exercise price of Y_n.

The exact valuation depends on where the money for the intermediate debt payments comes from. Suppose the equity owners pay them out of their own pockets so that the value of the assets of the firm is unaffected and still follows a geometric Brownian motion. While there is a reasonably simple closed-form solution for 'a call on a call' in this case, see Geske (1979), extending it to more

[8] In the original article, Longstaff and Schwartz (1995a) did in fact include a closed-form expression for the probability $\mathbb{Q}_t^T(\tau \leq T)$, but this expression is incorrect, see the discussion in Collin-Dufresne and Goldstein (2001).

layers of options is intractable so a numerical solution approach needs to be taken. A binomial tree approximating the asset dynamics is appropriate and the equity and the debt can be valued by backward recursion through the tree. Along the way, you will also determine when and where (that is for what asset values) the equity owners will not exercise the next option and therefore lead the company into default.

Alternatively, suppose the intermediate debt payments are financed by selling part of the assets of the firm so that firm value drops by Y_i at time t_i. Then at time t_{n-1}, the firm will survive as long as the asset value exceeds the required debt payment, since the equity owners will then have a call option on the remaining assets starting out at the positive value $V_{t_{n-1}} - Y_{n-1}$. The valuation of equity and debt at earlier dates again has to rely on numerical computations, which are more cumbersome, as the underlying value of the assets no longer follows a diffusion process but jumps at debt payment dates. The approximating binomial tree is therefore not going to be recombining, leading to increased computational complexity. The empirical analysis of Eom et al. (2004) indicates that the modelling of the compound option feature in this way leads to higher credit spreads than the original Merton model, but they conclude that credit spreads are still lower than in reality.

Next, we consider the case where a firm has issued two loans where one loan, the 'senior' loan, has a higher priority than the other loan, the 'junior' loan, in the sense that the senior loan has to be fully serviced before the junior loan can be serviced. Suppose for simplicity that both loans are zero-coupon loans maturing at the same date T. Let F_S be the face value of the senior loan and F_J the face value of the junior loan. Table 13.6 provides an overview of the state-contingent payoffs to the different claimants. If the asset value ends up above the total debt $F_S + F_J$, all creditors receive their respective face values and equityholders stay in charge of the firm. If asset value ends up below $F_S + F_J$, the firm defaults, and—assuming the absolute priority rule is followed—equity owners get nothing. If $F_S < V_T \leq F_S + F_J$, senior creditors get the promised F_S, while junior creditors receive $V_T - F_S \leq F_J$. If $V_T \leq F_S$, senior creditors get V_T, while junior creditors walk away empty-handed. The payment to equityholders is now analogous to a call option on the assets of the firm with an exercise price equal to total debt $F_S + F_J$. The payoff to the senior creditors is analogous to having all assets and writing a call option with an exercise price of F_S. Finally, the payoff to the junior creditors is equal to the payoff of a call option with exercise price F_S minus a call option with exercise price $F_S + F_J$.

Table 13.6: State-contingent payoffs and current values to senior creditors, junior creditors, and equityholders. Senior and junior debt are both zero-coupon loans maturing at time T with face values F_S and F_J, respectively. $C_t(K)$ denotes the time t price of a call option maturing at time T, having an exercise price of K, when the underlying currently has a value of V_t.

Claim	Payoffs			Value at time t
	$V_T \leq F_S$	$F_S < V_T \leq F_S{+}F_J$	$F_S{+}F_J < V_T$	
Senior debt	V_T	F_S	F_S	$V_t - C_t(F_S)$
Junior debt	0	$V_T - F_S$	F_J	$C_t(F_S) - C_t(F_S{+}F_J)$
Equity	0	0	$V_T - [F_S{+}F_J]$	$C_t(F_S{+}F_J)$

13.3.5 Stationary leverage

In the models described above, the firm value is modelled by a geometric Brownian motion so the expected firm value will increase exponentially over time. With the level of outstanding debt being constant or even decreasing (as part of the debt is repaid), the leverage ratio will tend to decrease over time. But in reality, firms issue new debt, and leverage ratios appear to be stationary with fairly little variation over time. Based on this observation, Collin-Dufresne and Goldstein (2001) modify the model so that the firm in the future will issue new debt with the same priority as the existing debt in a way that keeps leverage ratios stationary. More precisely, they assume that the risk-neutral dynamics of firm value is

$$dV_t = V_t \left[r_t\, dt + \sigma\, dz^{\mathbb{Q}}_{1t} \right],$$

that the default-free short rate r_t follows the Vasicek model

$$dr_t = \kappa(\hat{\theta} - r_t)\, dt + \beta \left[\rho_{rV}\, dz^{\mathbb{Q}}_{1t} + \sqrt{1 - \rho_{rV}^2}\, dz^{\mathbb{Q}}_{2t} \right],$$

where $z^{\mathbb{Q}}_1$ and $z^{\mathbb{Q}}_2$ are independent standard Brownian motions, and that the default barrier $\underline{V}_t$ varies according to

$$d(\ln \underline{V}_t) = k_0 \left(\ln V_t - k_1 - k_2[r_t - \hat{\theta}] - \ln \underline{V}_t \right) dt,$$

where k_0, k_1, and k_2 are positive constants. This specification has the realistic feature that the firm will tend to issue additional debt when the current firm value is sufficiently above the current default barrier so that the default barrier subsequently increases. Moreover, the drift of the (log) default barrier is a decreasing function of the default-free short rate consistent with the observation that debt issuance is typically lower in periods of high interest rates. The log-leverage ratio $\mathrm{LEV}_t = \ln(\underline{V}_t / V_t)$ will then follow the process (see Exercise 13.4)

$$d\mathrm{LEV}_t = k_0 \left[\overline{\mathrm{LEV}}(r_t) - \mathrm{LEV}_t \right] dt - \sigma\, dz^{\mathbb{Q}}_{1t}, \tag{13.20}$$

which varies around a 'target log-leverage level' $\overline{\mathrm{LEV}}(r_t)$ that is linearly decreasing in the default-free short rate. As before, default happens the first time firm value hits the default barrier, which is identical to the first time that LEV_t hits zero (from below).

For the pricing of corporate coupon bonds Collin-Dufresne and Goldstein assume no recovery of coupons and a fixed, non-random recovery of the face value. Following (13.9), the price of the defaultable coupon bond is then

$$\begin{aligned}
\tilde{B}_t &= B_t - q \sum_{t_i > t} \mathrm{E}^{\mathbb{Q}}_t \left[e^{-\int_t^{t_i} r_u\, du} \mathbf{1}_{\{\tau < t_i\}} \right] - (1 - R)\, \mathrm{E}^{\mathbb{Q}}_t \left[e^{-\int_t^{t_n} r_u\, du} \mathbf{1}_{\{\tau < t_n\}} \right] \\
&= B_t - q \sum_{t_i > t} B^{t_i}_t \mathbb{Q}^{t_i}_t \left(\tau < t_i | \tau > t \right) - (1 - R) B^{t_n}_t \mathbb{Q}^{t_n}_t \left(\tau < t_n | \tau > t \right),
\end{aligned}$$

and it is clear from the model that the expectations and default probabilities will depend only on r_t and LEV_t. In this two-dimensional setting, the density of the first hitting time of the barrier is not known in closed form. Collin-Dufresne and Goldstein implement an apparently efficient numerical technique for computing the relevant default probabilities and thus the bond prices. For a presumably reasonable choice of parameter values they show that their model generates considerably higher [lower] credit spreads for investment-grade [non-investment-grade] corporate bonds than the standard model with a constant default barrier. Intuitively, investment-grade firms are more inclined to have low current leverage relative to the long-term leverage level, so that leverage tends to increase leading to higher long-maturity credit spreads than with a constant default barrier. Conversely for the currently highly leveraged non-investment-grade firms. Moreover, the model can generate upward-sloping term structures of credit spreads for non-investment-grade bonds and credit spreads that are relatively insensitive to firm value changes. All these model properties are in line with empirical observations and are not achievable with the standard model. However, according to the empirical investigation of Eom et al. (2004), the model tends to overestimate credit spreads.

Of course, it is fair to criticize some of the model assumptions. First, the model basically assumes that new debt is issued continuously according to some exogenously fixed rule and apparently it also allows for negative debt issuance, which can be interpreted as a prepayment of part of the existing debt. Debt issuance and retirement are endogenous and strategic decisions of the management of the firm. Second, the new debt issued is assumed to have the same priority as the existing debt, which dilutes the value of the existing debt claims making it natural to think that the original debt contracts are issued with covenants restricting future debt decisions. However, in practice most corporate debt contracts do not rule out future issuance of equal priority debt. While the model therefore is best seen as a 'very reduced-form' representation of leverage dynamics, it is quite tractable and does demonstrate that the possibility of future changes to the capital structure of the firm can have a significant impact on current credit spreads.

13.3.6 Other extensions

Regarding the model of the value of the underlying assets of the firm, Zhou (2001b) adds the possibility of a lognormally distributed jump to the geometric Brownian motion used in Merton's original model. If debt has the form of a zero-coupon loan and default can only happen at maturity, the value of equity can be found in semi-closed form as an infinite series of Black–Scholes-type option prices, exactly as in the Merton (1976) jump-extension of the original Black–Scholes formula. With the possibility of premature default, equity and debt have to be valued numerically. Due to the jump element in asset value, default can occur unexpectedly, which implies positive credit spreads for even very short-term debt in line with empirical findings. Compared to the original Merton model, the jump-diffusion model generally leads to higher credit spreads and more flexibility in the shape of risk structures.

In most of the above models the capital structure of the firm is assumed to be constant except for the reduction in debt as it matures. The extension by Collin-Dufresne and Goldstein (2001) is based on an exogenous rule for changing the debt level over time. Major adjustments of the capital structure of a firm are strategic choices of top management acting on behalf of equity owners (ignoring any agency problems). A number of papers have studied the valuation of equity and debt, when equity owners can optimize the capital structure of the firm dynamically. To talk meaningfully about the optimal level of debt, we need to go beyond the idealized framework of Modiagliani and Miller (1958, 1963). Most papers assume a trade-off between a tax advantage of debt (financed by tax authorities) and costs of financial distress (including direct bankruptcy costs to lawyers and so on as well as indirect costs due, for example to the loss of customers and employees when default risk is high). Default barriers are then determined endogenously by managers maximizing equity value. Some of these models allow debt to be callable so that the firm can eliminate the existing debt contracts by paying the remaining debt plus some preset premium. As a large part of corporate bonds are callable, it is potentially important to include this feature. The interested reader is referred to the survey of Lando (2004, Ch. 3) or to original papers such as Brennan and Schwartz (1984), Fischer, Heinkel, and Zechner (1989), Leland (1994), Leland and Toft (1996), Mella-Barral and Perraudin (1997), Goldstein, Ju, and Leland (2001), and Christensen, Flor, Lando, and Miltersen (2002).

13.3.7 Default correlation in structural models

In structural models the default of each individual firm is governed by the process of the value of its assets. The correlation between the assets of two firms will thus determine the default correlation between the two firms. Assume that the asset values of the two firms evolve as

$$\begin{aligned} dV_{1t} &= V_{1t}\left[\mu_1\, dt + \sigma_1\, dz_{1t}\right], \\ dV_{2t} &= V_{2t}\left[\mu_2\, dt + \sigma_2\left(\rho_{12}\, dz_{1t} + \sqrt{1-\rho_{12}^2}\, dz_{2t}\right)\right], \end{aligned}$$

where ρ_{12} is the instantaneous correlation between the asset values. In addition, assume constant default barriers $\underline{V}_{it} = \underline{v}_i$, $i = 1, 2$. Then $X_i \equiv \left(\ln[V_{it}/V_{i0}] - [\mu_i - \sigma_i^2/2]t\right)/(\sigma_i\sqrt{t})$ is a standard normal random variable so

$$\mathbb{P}(V_{it} < \underline{v}_i) = N(-\mathcal{D}_{it}) \equiv \int_{-\infty}^{-\mathcal{D}_{it}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2}\, dx, \tag{13.21}$$

where

$$\mathcal{D}_{it} = \frac{\ln(V_{i0}/\underline{v}_i) + (\mu_i - \sigma_i^2/2)t}{\sigma_i\sqrt{t}}$$

is the distance-to-default of firm i. Moreover, (X_1, X_2) is a standard bivariate normal distribution with correlation coefficient ρ_{12}, so

$$\mathbb{P}\left(V_{1t} < \underline{v}_1, V_{2t} < \underline{v}_2\right) = N_2\left(-\mathcal{D}_{1t}, -\mathcal{D}_{2t}; \rho_{12}\right) \tag{13.22}$$

$$\equiv \int_{-\infty}^{-\mathcal{D}_{1t}} \int_{-\infty}^{-\mathcal{D}_{2t}} \frac{1}{\sqrt{2\pi(1-\rho_{12}^2)}}$$

$$\times \exp\left\{-\frac{x_1^2 + x_2^2 - 2\rho_{12}x_1x_2}{2(1-\rho_{12}^2)}\right\} dx_2\, dx_1.$$

It is now tempting to combine the above expressions with (13.10) and compute the default correlation over the time interval $[0, t]$ as

$$\rho_{12}^D(0,t) = \frac{N_2\left(-\mathcal{D}_{1t}, -\mathcal{D}_{2t}; \rho_{12}\right) - N\left(-\mathcal{D}_{1t}\right) N\left(-\mathcal{D}_{2t}\right)}{\sqrt{N\left(-\mathcal{D}_{1t}\right)\left[1 - N\left(-\mathcal{D}_{1t}\right)\right] N\left(-\mathcal{D}_{2t}\right)\left[1 - N\left(-\mathcal{D}_{2t}\right)\right]}},$$

and—according to Crouhy et al. (2000)—this is done in some commercial credit risk management systems. However, if firm i defaults the first time V_i hits $\underline{v}_i$, the asset value process can increase above $\underline{v}_i$ again later, so the firm may well be in default at time t even though $V_{it} \geq \underline{v}_i$. Hence, (13.21) is not really the probability that firm i is in default at time t and (13.22) is not the probability that both firms are in default at time t. Consequently, the above expression for the default correlation is also flawed.

Zhou (2001a) derives the correct default correlation for two firms with asset dynamics as given above under the assumption that the default barrier of firm i is given by $\underline{V}_i = \underline{v}_i \exp\{\gamma_i t\}$. The individual default probabilities are then

$$\mathbb{P}\left(\tau_i \leq t\right) = N\left(-\frac{\ln\left(\frac{V_{i0}}{\underline{v}_i}\right) + \alpha_i t}{\sigma_i\sqrt{t}}\right) + \left(\frac{V_{i0}}{\underline{v}_i}\right)^{2\alpha_i/\sigma_i^2} N\left(-\frac{\ln\left(\frac{V_{i0}}{\underline{v}_i}\right) - \alpha_i t}{\sigma_i\sqrt{t}}\right),$$

where $\alpha_i = \mu_i - \gamma_i - \frac{\sigma_i^2}{2}$. Zhou finds a quite complicated expression for the probability of at least one default, $\mathbb{P}\left(\tau_1 \leq t \text{ or } \tau_2 \leq t\right)$, involving a double integral of Bessel functions that has to be evaluated by numerical methods.[9] Given these probabilities, the default correlation $\rho_{12}^D(0,t)$ over the time interval $[0, t]$ can then be computed using (13.11) and subsequently (13.10). Zhou provides numerical results showing, among other things, that

1. the default correlation $\rho_{12}^D(0,t)$ and the asset correlation ρ_{12} are of the same sign;
2. the higher the asset correlation, the higher the default correlation;
3. default correlations tend to be smaller in magnitude (that is closer to zero) than asset correlations;
4. high credit quality implies a low default correlation over relevant horizons;
5. the default correlations vary over time as the credit quality of the firms vary.

[9] The solution is much simpler under the parameter condition $\gamma_i = \mu_i - \sigma_i^2/2$ and this solution seems to produce roughly the same default correlations as the exact complicated solution even though the parameters do not satisfy the condition.

An alternative way of capturing default dependence is by using so-called copulas. We describe this approach in Section 13.6.

13.4 REDUCED-FORM MODELS

The firm-specific inputs to the pricing of defaultable bonds and credit derivatives are the risk-neutral default probabilities and the recovery rates in case of default. Structural models attempt to explain these quantities by the fundamental assets and liabilities of each firm. The reduced-form approach pioneered by Jarrow and Turnbull (1995) is to model default probabilities and recovery rates directly. Reduced-form models are also called intensity-based models because the default probabilities over any time period are modelled via the default intensity, that is the instantaneous default probabilities per time period. A key motivation behind reduced-form models is to avoid very detailed modelling of the asset dynamics and the capital structure of each issuer, but firm-related variables like the stock price and its volatility are sometimes included as factors determining the default intensity. In reduced-form models the stopping time τ representing the time of default is assumed to be a *totally inaccessible* stopping time, which basically means that default will always come as a surprise, in contrast to the diffusion-type structural models where default is predictable.

Given a probability measure $\mathbb{Q}$, the cumulative hazard associated with default is the process $H = (H_t)$ defined by

$$H_t = -\ln \mathbb{Q}_t(\tau > t)$$

so that $\mathbb{Q}_t(\tau > t) = e^{-H_t}$. Typically the cumulative hazard is assumed to have an integral form

$$H_t = \int_0^t h_u \, du,$$

where h_t is referred to as the default intensity (or hazard rate). Suppose for now that the default intensity and, consequently, the cumulative hazard are deterministic. Then the probability of default between time t and $t + \Delta t$ conditional on no default up to time t is

$$\begin{aligned}\mathbb{Q}(t < \tau < t + \Delta t | \tau > t) &= 1 - \frac{\mathbb{Q}(\tau > t + \Delta t)}{\mathbb{Q}(\tau > t)} = 1 - e^{-(H_{t+\Delta t} - H_t)} \\ &= 1 - e^{-\int_t^{t+\Delta t} h_u \, du} \approx h_t \Delta t,\end{aligned}$$

where the first equality is due to (13.1) and the approximation is due to $e^x \approx 1 + x$ for $x \approx 0$ and $\int_t^{t+\Delta t} h_u \, du \approx h_t \Delta t$. The default intensity at time t thus captures the conditional default probability in a short (instantaneous, strictly speaking) period following time t. With a stochastic cumulative hazard and default intensity, the probability of default in an interval $[t, t']$ conditional on no default up to time t is

$$\mathbb{Q}(t < \tau < T | \tau > t) = 1 - \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T h_u \, du}\right], \tag{13.23}$$

which for a short period again is approximately proportional to the default intensity at time t,

$$\mathbb{Q}(t < \tau < t + \Delta t | \tau > t) = 1 - \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t+\Delta t} h_u\, du}\right] \approx h_t \Delta t.$$

The default intensity at time t can therefore be interpreted as the probability of default in the very near future per time period. Note that the default intensity is linked to the probability measure so, in general, a change of probability measure will also change the default intensity.

According to (13.23), the expectation $\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T h_u\, du}\right]$ is one minus the default probability, which is simply the probability that the firm survives the time interval $[t, T]$. This survival probability is similar to the general expression for the price of a default-free zero-coupon bond, $B_t^T = \mathrm{E}_t^{\mathbb{Q}}[e^{-\int_t^T r_u\, du}]$, where $r = (r_t)$ is the default-free short-term interest rate process. If the intensity is a function of a Markov diffusion process $x = (x_t)$, then it is clear that the survival probability $\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T h_u\, du}\right]$ is given by some function $f(x_t, t)$ where f satisfies a parabolic partial differential equation. If, furthermore, the intensity is an affine function of x and the dynamics of x is in the affine class, the survival probability will be an exponential-affine function of the current value x_t and the default probabilities follow immediately.

For the pricing of defaultable bonds with intensity-based models we need to make assumptions about the recovery payment in case of default. We consider the three frequently applied recovery specifications below together with the bond pricing equations they imply. An important feature of reduced-form models is the similarity of these pricing equations to the pricing equations for default-free bonds. Hence, we can apply the machinery developed in earlier chapters. Subsequently, we will discuss concrete reduced-form models for defaultable bonds.

13.4.1 Recovery assumptions and bond pricing

For simplicity consider a zero-coupon bond promising a unit payment at time T. We assume that the bondholder will get a recovery with a value of R_τ at the time of default if that occurs before T. Therefore, the value of the defaultable bond at time $t < T$ is (assuming here and in the following that default has not occurred up to time t):

$$\begin{aligned}\tilde{B}_t^T &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\, du}\mathbf{1}_{\{\tau > T\}} + e^{-\int_t^\tau r_u\, du}\mathbf{1}_{\{\tau \leq T\}} R_\tau\right] \\ &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\, du}\mathbf{1}_{\{\tau > T\}}\right] + \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^\tau r_u\, du}\mathbf{1}_{\{\tau \leq T\}} R_\tau\right].\end{aligned}$$

As shown by Lando (1998), the first term can be rewritten as

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\, du}\mathbf{1}_{\{\tau > T\}}\right] = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (r_u + h_u)\, du}\right], \tag{13.24}$$

which is again similar to the general expression for the price of a default-free zero-coupon bond. The second term, that is the value of the recovery payment, can be rewritten as

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^\tau r_u\,du}\mathbf{1}_{\{\tau\le T\}}R_\tau\right] = \int_t^T \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds}h_u R_u\right]du \tag{13.25}$$

so that

$$\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (r_u+h_u)\,du}\right] + \int_t^T \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds}h_u R_u\right]du. \tag{13.26}$$

To proceed it is necessary to make an assumption about the recovery payment. In the following we look at three alternatives often used in the credit risk literature.

Recovery of face value. The recovery payment is a fraction w_τ of the face value (or par value) of the bond and is paid immediately at the time of default. For a zero-coupon bond, this just means $R_\tau = w_\tau$ in the expressions above. With a constant recovery of face value, $w_\tau = w$, the bond price becomes

$$\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (r_u+h_u)\,du}\right] + w\int_t^T \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds}h_u\right]du. \tag{13.27}$$

If, furthermore, the default intensity and the default-free rates are assumed independent, we get

$$\begin{aligned}\tilde{B}_t^T &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right]\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T h_u\,du}\right] + w\int_t^T \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u r_s\,ds}\right]\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u h_s\,ds}h_u\right]du \\ &= B_t^T\mathbb{Q}(\tau > T|\tau > t) + w\int_t^T B_t^u\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u h_s\,ds}h_u\right]du. \end{aligned}\tag{13.28}$$

As will be discussed below, the last expectation in (13.27) or (13.28) can be computed in closed form under some assumptions on the default intensity, but still the integration needs to be carried out. Of course, with a constant default intensity, $h_u = h$, the expectation is redundant and we get

$$\tilde{B}_t^T = e^{-h(T-t)}B_t^T + wh\int_t^T e^{-h(u-t)}B_t^u\,du.$$

Recovery of treasury. Upon default the owner of the corporate bond receives a treasury bond with the same maturity but a lower face value. For a zero-coupon bond, this means $R_\tau = w_\tau B_\tau^T$, that is the recovery equals a fraction w_τ of the present value of the face value. In principle, we could proceed by substituting this into (13.26), but it is easier to use the observation that getting $w_\tau B_\tau^T$ at the default time τ is equivalent to getting w_τ at the original maturity date T. Therefore, the defaultable bond price can be expressed as

$$\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(\mathbf{1}_{\{\tau>T\}} + w_\tau\mathbf{1}_{\{\tau\le T\}}\right)\right] = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(w_\tau + (1-w_\tau)\mathbf{1}_{\{\tau>T\}}\right)\right].$$

With a constant recovery, $w_\tau = w$, this yields

$$\begin{aligned}\tilde{B}_t^T &= w\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right] + (1-w)\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\mathbf{1}_{\{\tau>T\}}\right] \\ &= wB_t^T + (1-w)\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (r_u+h_u)\,du}\right],\end{aligned} \tag{13.29}$$

where the last equality is due to (13.24). Note that the last expectation in (13.29) is similar to the first expectation in (13.27), whereas the last expectation in (13.27) is more complicated as long as the default intensity is not constant. Therefore, an assumption of a constant recovery of treasury will generally lead to a simpler pricing equation than an assumption of a constant recovery of face value.

In Section 13.2.2 we have already derived equations linking defaultable bond prices, credit spreads, and risk-neutral default probabilities. In particular, when the default intensity, the recovery payment, and the default-free rates are assumed to be independent, we can express the credit spread as

$$\begin{aligned}\zeta_t^T &= -\frac{1}{T-t}\ln\left(1-\left(1-\mathrm{E}_t^{\mathbb{Q}}[w]\right)\mathbb{Q}_t(\tau \le T|\tau > t)\right) \\ &\approx \frac{\left(1-\mathrm{E}_t^{\mathbb{Q}}[w]\right)\mathbb{Q}_t(\tau \le T|\tau > t)}{T-t},\end{aligned}$$

and we can write the risk-neutral default probability as

$$\mathbb{Q}_t(\tau \le T|\tau > t) = \frac{1-e^{-\zeta_t^T(T-t)}}{1-\mathrm{E}_t^{\mathbb{Q}}[w]} \approx \frac{\zeta_t^T(T-t)}{1-\mathrm{E}_t^{\mathbb{Q}}[w]},$$

see (13.5) and (13.6). The default probabilities in these relations can be expressed in terms of the default intensity using (13.23). Unfortunately, the independence assumption is generally not realistic. Alternatively, if we assume a non-random recovery payment and allow the default intensity and the default-free rates to be dependent under the risk-neutral measure, we have the approximation

$$\mathbb{Q}_t^T(\tau \le T|\tau > t) \approx \frac{\zeta_t^T(T-t)}{1-w},$$

see (13.7). This expression links T-forward default probabilities and credit spreads without assuming independence. The T-forward default probability is related to the T-forward default intensity similarly to (13.23), but the T-forward default intensity is generally different from the risk-neutral default intensity h.

Recovery of market value. The recovery payment is assumed to be a fraction w_τ of the market value of the bond immediately before default and is paid at default. For a zero-coupon bond this means that $R_\tau = w_\tau \tilde{B}_{\tau-}^T$, where $\tau-$ indicates 'an instant before time τ'. In other words, default implies that a fraction $\ell_\tau = 1 - w_\tau$ of the market value of the bond is lost. If we substitute the recovery payment into (13.26), we obtain a recursive valuation expression, which appears very complicated. However, Duffie and Singleton (1999) show that the solution to the recursive relation is given by

$$\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (r_u + h_u \ell_u)\,du}\right], \tag{13.30}$$

that is the price of a defaultable zero-coupon bond is linked to the 'default risk adjusted' short rate $r + h\ell$ exactly as the default-free zero-coupon bond price is linked to the default-free short rate r. You can think of $h_u \ell_u$ as a measure of the expected loss due to default. Recall that default-free zero-coupon yields approach r_t as maturity is decreased to zero. Analogously, the yield on a defaultable zero-coupon bond will approach $r_t + h_t \ell_t$, so the very short-term credit spread will be $h_t \ell_t$. If you assume that the default risk adjusted rate $r + h\ell$ has risk-neutral dynamics of the form assumed for the short rate in any affine or quadratic model, you would get an equation for the defaultable bond price directly. However, if you want to consistently price corporate bonds with different default intensities or different losses in case of default, you should apply separate dynamic models for the default-free short rate and for the expected loss $h\ell$ and maybe even employ separate models for the intensity h and the loss given default ℓ.

Due to the technical nature of the proof of (13.30), we will only provide a discrete-time argument supporting the conclusion for the special case of a constant recovery rate w and, hence, a constant loss rate $\ell = 1 - w$. Divide the time period $[t, T]$ into intervals of length Δt so that $T = t + N\Delta t$. Assume that the issuer has not defaulted up to time t. Then there is a probability of default over the next Δt interval (approximately) equal to $1 - e^{-h_t \Delta t}$ in which case the bondholder will get $w\tilde{B}_t^T$. Otherwise, with a probability of $e^{-h_t \Delta t}$, the bondholder just keeps his bond which has a value of $\tilde{B}_{t+\Delta t}^T$ at the end of this short interval. Discounting and taking expectations, we get

$$\tilde{B}_t^T = \left(1 - e^{-h_t \Delta t}\right) e^{-r_t \Delta t} w \tilde{B}_t^T + e^{-h_t \Delta t} e^{-r_t \Delta t}\, \mathrm{E}_t^{\mathbb{Q}}[\tilde{B}_{t+\Delta t}^T],$$

which implies

$$\left[e^{(r_t + h_t)\Delta t} - \left(e^{h_t \Delta t} - 1\right) w\right] \tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}[\tilde{B}_{t+\Delta t}^T].$$

If we define R_t via

$$e^{R_t \Delta t} = e^{(r_t + h_t)\Delta t} - \left(e^{h_t \Delta t} - 1\right) w,$$

we have $e^{R_t \Delta t}\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}[\tilde{B}_{t+\Delta t}^T]$ and thus $\tilde{B}_t^T = e^{-R_t \Delta t}\, \mathrm{E}_t^{\mathbb{Q}}[\tilde{B}_{t+\Delta t}^T]$. Moreover, using the approximation $e^x \approx 1 + x$ for x close to zero, we get

$$1 + R_t \Delta t \approx 1 + (r_t + h_t)\Delta t - w h_t \Delta t = 1 + (r_t + (1 - w)h_t)\Delta t,$$

and, consequently,

$$R_t \approx r_t + (1 - w)h_t \equiv r_t + \ell h_t$$

as wanted. The above link between the bond price at two consecutive dates works for all Δt intervals. For example, $\tilde{B}^T_{t+\Delta t} = e^{-R_{t+\Delta t}\Delta t}\, \mathrm{E}_{t+\Delta t}[\tilde{B}^T_{t+2\Delta t}]$. Thus we have

$$\begin{aligned}\tilde{B}^T_t &= e^{-R_t\Delta t}\, \mathrm{E}^{\mathbb{Q}}_t[\tilde{B}^T_{t+\Delta t}] = e^{-R_t\Delta t}\, \mathrm{E}^{\mathbb{Q}}_t\left[e^{-R_{t+\Delta t}\Delta t}\, \mathrm{E}_{t+\Delta t}[\tilde{B}^T_{t+2\Delta t}]\right] \\ &= \mathrm{E}^{\mathbb{Q}}_t\left[e^{-(R_t+R_{t+\Delta t})\Delta t}\tilde{B}^T_{t+2\Delta t}\right].\end{aligned}$$

If we continue the iterations and use that $\tilde{B}^T_T = 1$, we end up with

$$\tilde{B}^T_t = \mathrm{E}^{\mathbb{Q}}_t\left[\exp\left\{-\sum_{n=0}^{N-1} R_{t+n\Delta t}\Delta t\right\}\right] \approx \mathrm{E}^{\mathbb{Q}}_t\left[\exp\left\{-\sum_{n=0}^{N-1} (r_{t+n\Delta t} + \ell h_{t+n\Delta t})\,\Delta t\right\}\right],$$

which is a discrete-time version of (13.30).

In full generality, the three recovery assumptions are equivalent. However, as already indicated above, simple pricing equations typically require that the recovery fraction w_τ is a constant or a certain function of some specific variables, in which case the equivalence breaks down. For example, a recovery payment which is a constant fraction of market value, $R_\tau = w\tilde{B}^T_{\tau-}$, corresponds to a stochastic fraction $w_\tau = w\tilde{B}^T_{\tau-}$ of the face value.

No matter which recovery assumption is used, it is clear that we can exploit our modelling techniques and results developed in earlier chapters. The main model class for default-free bonds and derivatives is the class of (one- or multi-dimensional) affine models. For example, if r and h are affine functions of some, possibly multi-dimensional, diffusion x with affine drift and instantaneous variance-covariance matrix, we have

$$\mathrm{E}^{\mathbb{Q}}_t\left[e^{-\int_t^T (r_u+h_u)\,du}\right] = e^{A(T-t)+B(T-t)^{\top}x_t} \tag{13.31}$$

for functions A and B which solve certain ordinary differential equations. This takes care of the expectation in (13.29) and the first expectation in (13.27). Moreover, it follows from Duffie et al. (2000) that the second expectation in (13.27) gives

$$\mathrm{E}^{\mathbb{Q}}_t\left[e^{-\int_t^u (r_s+h_s)\,ds}h_u\right] = (\hat{A}(u-t) + \hat{B}(u-t)^{\top}x_t)e^{A(u-t)+B(u-t)^{\top}x_t}, \tag{13.32}$$

where $\hat{A}$ and $\hat{B}$ are other deterministic functions that can be found by solving some ordinary differential equations (see also Exercise 13.7). The expectation in (13.30) will also be of the exponential-affine form as long as r and $h\ell$ are affine functions of variables with affine dynamics.

When constructing such models, we should of course be aware of available empirical findings on the fluctuations in default probabilities or intensities and recovery rates as well as the link between these default-related variables and the default-free interest rates. In the following we will first take a detailed look at a very simple Vasicek-type model and then consider some more elaborate models.

13.4.2 A simple affine reduced-form model

In this section we consider a very simple affine model that serves as an illustration of the general approach. This model example was studied by Kraft and Munk (2007). Assume that the default-free short rate follows the Vasicek model of Section 7.4 with risk-neutral dynamics

$$dr_t = \kappa(\hat{\theta} - r_t)\,dt + \beta\,dz_t^{\mathbb{Q}}$$

so that the default-free zero-coupon bond price is

$$B_t^T = e^{-a(T-t)-b(T-t)r_t}$$

with

$$b(s) = \frac{1}{\kappa}\left(1 - e^{-\kappa s}\right), \qquad a(s) = \frac{1}{\kappa}\left(\hat{\theta} - \frac{\beta^2}{2\kappa}\right)(s - b(s)) + \frac{\beta^2}{4\kappa}b(s)^2.$$

Further assume that the Duffie–Singleton recovery of market value rule applies and that the default risk adjusted short rate is affine in the default-free short rate,

$$\tilde{r}_t \equiv r_t + h_t\ell_t = k_0 + k_1 r_t.$$

Such a relation is satisfied in two cases:

1. Constant loss rate, $\ell_t = L \geq 0$, and affine default intensity, $h_t = H_0 + H_1 r_t$. Then $k_0 = H_0 L$ and $k_1 = 1 + H_1 L$.
2. Constant default intensity, $h_t = H > 0$, and affine loss rate, $\ell_t = L_0 + L_1 r_t$. Then $k_0 = HL_0$ and $k_1 = 1 + HL_1$.

Under these assumptions the price of the defaultable zero-coupon bond becomes

$$\tilde{B}_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (k_0 + k_1 r_u)\,du}\right] = e^{-k_0(T-t)}\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T k_1 r_u\,du}\right].$$

Since

$$d(k_1 r_t) = k_1\,dr_t = \kappa[k_1\hat{\theta} - (k_1 r_t)]\,dt + k_1\beta\,dz_t^{\mathbb{Q}},$$

we see that $k_1 r_t$ is a Vasicek-type process with the same mean reversion speed parameter κ, but with long-run level $k_1\hat{\theta}$ instead of $\hat{\theta}$ and volatility $k_1\beta$ instead of β. It follows by analogy to the computation of the default-free bond price that

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T k_1 r_u\,du}\right] = e^{-k_1\tilde{a}(T-t)-b(T-t)k_1 r_t},$$

where

$$\tilde{a}(s) = \frac{1}{\kappa}\left(\hat{\theta} - \frac{k_1\beta^2}{2\kappa}\right)(s - b(s)) + \frac{k_1\beta^2}{4\kappa}b(s)^2.$$

The price of the defaultable zero-coupon bond is therefore

$$\tilde{B}_t^T = e^{-k_0[T-t]-k_1\tilde{a}(T-t)-b(T-t)k_1 r_t} = e^{f(T-t)}\left(B_t^T\right)^{k_1}, \tag{13.33}$$

where

$$f(s) = -k_0 s + \frac{\beta^2 k_1(k_1-1)}{2\kappa^2}\left(s - b(s) - \frac{\kappa}{2}b(s)^2\right).$$

The yields of the default-free and the defaultable zero-coupon bonds are

$$y_t^{t+s} = \frac{a(s)}{s} + \frac{b(s)}{s}r_t, \qquad \tilde{y}_t^{t+s} = k_1 y_t^{t+s} - \frac{f(s)}{s},$$

respectively, so that the yield spread becomes

$$\zeta_t^{t+s} = \tilde{y}_t^{t+s} - y_t^{t+s} = (k_1-1)y_t^{t+s} - \frac{f(s)}{s} = (k_1-1)\frac{a(s)}{s} - \frac{f(s)}{s} + (k_1-1)\frac{b(s)}{s}r_t.$$

The empirical evidence on the key parameters is mixed. Using data on U.S. Treasury and corporate bond prices over the period 1989–1998, Bakshi et al. (2006) estimate several specifications of the default risk adjusted short rate $\tilde{r}_t$ in the Duffie–Singleton framework. For a model with $\tilde{r}_t = k_0 + k_1 r_t$, consistent with our assumption above, their estimate of k_1 is 1.018 for BBB-rated bonds and 0.985 for A-rated bonds. Adding a firm-specific distress variable x_t so that $\tilde{r}_t = k_0 + k_1 r_t + k_2 x_t$, they estimate k_1 to be below one, namely in the range $[0.767, 0.910]$ for BBB bonds and in the range $[0.902, 0.966]$ for A bonds with the estimates depending on the proxy used for financial distress. Jarrow and Yildirim (2002) assume a generalized Vasicek model, a constant loss rate, and a default intensity affine in the default-free short rate. They estimate the parameters using corporate default swap quotes for 22 individual companies. Their company-specific estimates of H_1 are positive, ranging from 1.3 to 26.9 basis points. Empirical studies of Longstaff and Schwartz (1995b), Duffee (1998), and Papageorgiou and Skinner (2006) conclude that yield spreads are generally decreasing in default-free yields. Within the current setting, this will be the case if $k_1 < 1$.

For concreteness assume that version 1 of the model applies, that is that the loss rate is given by a constant L and the default intensity is $h_t = H_0 + H_1 r_t$. We assume that the current short rate is 4%, $\kappa = 0.15$, $\beta = 0.01$, and $\hat{\theta} = 0.0522$, which imply an upward-sloping default-free yield curve with an asymptotic long-term zero-coupon yield of 5%. The loss in case of default is assumed to be 40%. The fixed part of the default intensity is $H_0 = 0.025$ so that $k_0 = 0.01$. We vary H_1 between -0.5 and 0.5 corresponding to variations in default intensities (at the current short rate) between 0.005 and 0.045 and variations in k_1 between 0.8 and 1.2. Figure 13.1 shows yield spread curves for different values of H_1. All the spread curves depicted are quite flat. Clearly the default intensity is an important determinant of the level of the spreads. Note that the short-term spreads in the graph are all greater than

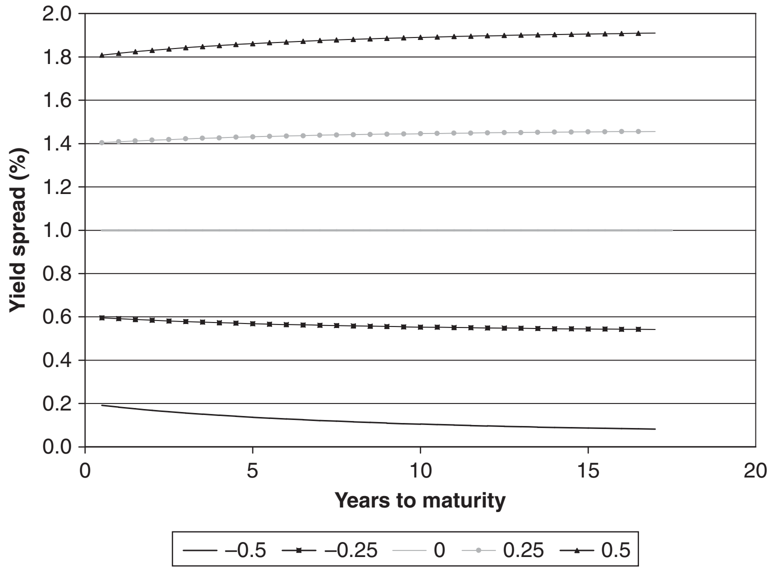

Fig. 13.1: Zero-coupon yield spreads. The figure shows the difference between the yield of a corporate zero-coupon bond and a default-free zero-coupon bond as a function of time to maturity. The different curves are for different values of the parameter H_1 ranging from -0.5 (lower curve) to 0.5 (upper curve). The current default-free short rate is 4%. The parameters of the short rate process are $\kappa = 0.15$, $\sigma_r = 0.01$, and $\hat{\theta} = 0.0522$ so that the asymptotic zero-coupon yield is 5%. The loss in case of default is $L = 0.4$ and the fixed part of the default intensity is $H_0 = 0.025$.

zero. In fact, it can be shown that $f(s)/s \to k_0$ for $s \to 0$ and, consequently, the instantaneous spread becomes

$$\lim_{s\to 0} \zeta_t^{t+s} = \lim_{s\to 0} \tilde{y}_t^{t+s} - \lim_{s\to 0} y_t^{t+s} = k_1 r_t + k_0 - r_t = (k_1 - 1)r_t + k_0,$$

which will only be zero when $\tilde{r}_t = r_t$, that is in the absence of default risk.

The limitations of the Vasicek model of default-free interest rates have been discussed extensively in earlier chapters. Therefore we focus on the assumptions regarding the default risk elements of the model. The default-adjusted short rate is assumed to depend on the default-free short rate. As indicated above, Bakshi et al. (2006) show in an empirical study that the default-free short rate is an important determinant of corporate bond yields and credit spreads, but for low-rated bonds the empirical performance is significantly improved when the default-adjusted short rate is also allowed to depend on the leverage of the issuer (book value of debt divided by total firm value) or another firm-specific distress variable. Macro-variables beyond the default-free interest rates do not seem to systematically affect the default-adjusted short rates.[10] Of course, modelling the default intensity, as

[10] Bakshi et al. (2006) assume a two-factor essentially affine Vasicek model for the default-free interest rates in contrast to the one-factor completely affine Vasicek model assumed in the simple model here.

in version 1 of the model, as an affine function of the Vasicek-driven default-free short rate implies that the intensity may take on negative values which are not only unrealistic but also nonsensical. Version 2 of the model is no better, since it allows loss rates both below zero and above one. Just cutting off the intensity at zero or the loss rate at zero and one will ruin the nice analytical expressions for the defaultable bond prices and the credit spreads.

Another problem with the precise model discussed above is that the default intensity is assumed to depend only on the default-free short rate and no other random sources, which leads to a perfect correlation not supported by the data. While—as mentioned above—no other macro-factors seem to have much influence on default intensities, there is still some stochastic variation in intensities that cannot be explained by default-free rates. Another source of stochasticity has to be introduced to improve the model in this respect.

13.4.3 Other affine models

A major problem of Gaussian models is the chance of negative interest rates and intensities. An obvious alternative is to use a square-root model such as

$$dr_t = \kappa_r[\theta - r_t]\,dt + \beta\sqrt{r_t}\,dz_{1t}^{\mathbb{Q}},$$
$$dh_t = \kappa_h[\bar{h} - h_t]\,dt + \sigma_h\sqrt{h_t}\left(\rho\,dz_{1t}^{\mathbb{Q}} + \sqrt{1-\rho^2}\,dz_{2t}^{\mathbb{Q}}\right)$$

where ρ is the instantaneous correlation between the default-free short rate and the default intensity under the risk-neutral probability measure. The square-root specification ensures positive intensities as well as positive default-free interest rates. However, the model is generally not an affine model, because the instantaneous covariance $\text{Cov}_t[dr_t, dh_t] = \rho\beta\sigma_h\sqrt{r_t}\sqrt{h_t}$ is not affine, and a true affine model requires all elements of the drift and the variance-covariance matrix to be affine functions of the state variables, that is r_t and h_t. The square-root model is only affine if the default intensity is assumed to be independent of the default-free short rate, that is when $\rho = 0$, but this is not supported by the empirical evidence.

An indirect way of generating dependence between default-free rates and default intensities is to assume that $h_t = H_0 + H_1 r_t + H_2 y_t$, where r and y are independent square-root processes. If the coefficients H_i are all positive, that specification ensures non-negative intensities and is in the affine class, and the instantaneous correlation $\text{Corr}_t[dr_t, dh_t]$ is of course positive. However, the empirical evidence on the sign of H_1 is not clear-cut as discussed in the previous subsection. Even if r and h are affine functions of multiple square-root processes, it is impossible to obtain both a non-negative intensity and a negative instantaneous correlation between the intensity and the default-free short rate, see the discussion in Duffie and Singleton (1999) and Dai and Singleton (2000).

Duffie and Singleton (1999) provide an example of an affine model that is capable of generating negative (or positive) correlation while preserving non-negative intensities. The model assumes

$$r_t = \delta_0 + \delta_1 Y_{1t} + Y_{2t} + Y_{3t}, \quad h_t = H_0 + H_1 Y_{1t} + H_2 Y_{2t},$$

and the risk-neutral dynamics of $Y_t = (Y_{1t}, Y_{2t}, Y_{3t})^{\top}$ is

$$dY_t = \mathcal{K}\left(\Theta - Y_t\right)\, dt + \Sigma\sqrt{S(Y_t)}\, dz_t^{\mathbb{Q}}$$

where $z^{\mathbb{Q}}$ is a three-dimensional standard Brownian motion, Θ is a vector of constant 'long-run averages',

$$\mathcal{K} = \begin{pmatrix} \kappa_{11} & \kappa_{12} & 0 \\ \kappa_{21} & \kappa_{22} & 0 \\ 0 & 0 & \kappa_{33} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ \sigma_{31} & \sigma_{32} & 0 \end{pmatrix},$$

and $S(Y_t)$ is a 3×3 diagonal matrix with diagonal elements

$$S_{11}(Y_t) = Y_{1t}, \quad S_{22}(Y_t) = \beta_{22}Y_{2t}, \quad S_{33}(Y_t) = \alpha_3 + \beta_{31}Y_{1t} + \beta_{32}Y_{2t}.$$

The coefficients δ_i, H_i, and β_{ij} are all strictly positive. κ_{12} and κ_{21} are non-positive to ensure that Y_1 and Y_2—and thus the intensity—stay non-negative. The signs of σ_{31} and σ_{32} are not restricted, which implies that Y_3 can be negatively correlated with Y_1 and/or Y_2. Consequently, the intensity can be negatively correlated with the default-free short rate. Note that Y_3 and thus r can take on negative values in this model.

Finally, we note that Duffie and Liu (2001) give an example of a model of the quadratic class in which it is possible to ensure that both default-free rates and default intensities are non-negative and at the same time allow for a negative correlation between the default-free short rate and the default intensity.

13.4.4 Market risk of corporate bonds

Default risk is not the only risk that corporate bonds are exposed to. As other bonds, corporate bonds are sensitive to general market risks, in particular interest rate risk. Chapter 12 discussed interest rate risk measures for non-defaultable bonds. A key measure is the duration of the bond, which in its modern version measures the percentage price change of the bond per unit change in the default-free short rate (or any other relevant state variable). More generally, we can define the duration of any asset as the negative of the percentage price sensitivity with respect to the short rate, that is if V_t denotes the time t value of the asset, the duration is defined as

$$D_t^V = -\frac{\partial V_t}{\partial r}\frac{1}{V_t}.$$

In particular, we can measure the interest rate risk of a corporate bond in this way. An interesting question is under which circumstances corporate bonds have higher or smaller durations than their default-free equivalents. This question is studied by Kraft and Munk (2007).

For concreteness consider the simple Vasicek-type model of Section 13.4.2. The price of a defaultable zero-coupon bond is given by (13.33) so the duration of this bond is

$$\tilde{D}_t^T \equiv D_t^{\tilde{B}^T} = k_1 b(T-t),$$

whereas the duration of the default-free zero-coupon bond is $b(T-t)$. In case 1 of the model the loss rate is $\ell_t = L > 0$ and the default intensity is $h_t = H_0 + H_1 r_t$ and, consequently, $k_1 = 1 + H_1 L$. The duration of the defaultable zero-coupon bond is therefore smaller [larger] than the duration of the default-free zero-coupon bond of the same maturity if $H_1 < 0$ [if $H_1 > 0$]. The duration of the corporate bond will be negative if $H_1 < -1/L$, which is theoretically possible. Also note that the sign of the duration of the corporate bond does not depend on the *level* of the default probability (as the discussion in Longstaff and Schwartz (1995b) suggests) but on the *interest rate sensitivity* of the default probability. It follows from the discussion in Section 13.4.2 that the empirical estimates for case 1 are mixed with some studies pointing to $H_1 < 0$ and others to $H_1 > 0$. In case 2 of the model the default intensity is $h_t = H > 0$ and the loss rate is $\ell_t = L_0 + L_1 r_t$) and, consequently, $k_1 = 1 + HL_1$. Then the duration of the defaultable bond is smaller [larger] than the duration of the default-free bond if $L_1 < 0$ [if $L_1 > 0$].

Considering coupon bonds, Kraft and Munk (2007) show that the following holds under some reasonable assumptions: if the duration of any defaultable zero-coupon bond is smaller than or equal to the duration of the equivalent default-free zero-coupon bond, then the duration of a corporate coupon bond is smaller than the duration of the equivalent default-free coupon bond. Even when the defaultable zero-coupon bonds have slightly higher duration than the default-free zero-coupon bonds, the defaultable coupon bond will have a lower duration than the default-free coupon bond.

In the Vasicek-type model with the parameters used for generating Fig. 13.1, the duration of a 10-year Treasury coupon bond with a 3% semi-annual coupon is 4.3099. Figure 13.2 shows how the duration of the 10-year corporate coupon bond varies with the parameter H_1. For $H_1 < 0$, the duration of a corporate zero-coupon bond is smaller than the duration of the similar default-free zero-coupon bond for any maturity, and we see from the figure that the duration of the corporate coupon bond is then also smaller than the duration of the Treasury coupon bond. For $H_1 = 0$, the durations of a corporate and a default-free zero-coupon bond of the same maturity are identical. Nevertheless, the duration of the 10-year corporate coupon bond (4.2663) is smaller than the duration of the similar default-free bond. This is also true for slightly positive values of H_1 (up to approximately 0.027). For values of H_1 higher than 0.027, the durations of both corporate zero-coupon and corporate coupon bonds are higher than those of their default-free counterparts. Given the empirical parameter estimates reported above, this may very well be the case for a number of corporate bonds.

13.4.5 Models for multiple issuers

As discussed already in Section 13.2.4, proper pricing and risk management involving portfolios of defaultable loans or bonds require a reasonable model for how defaults and recoveries across firms are related.

The empirical studies of Collin-Dufresne, Goldstein, and Martin (2001) and Elton, Gruber, Agrawal, and Mann (2001) indicate systematic variation in credit spread changes across firms. This can be captured by letting default intensities of different firms depend on one or more common factors. Collin-Dufresne,

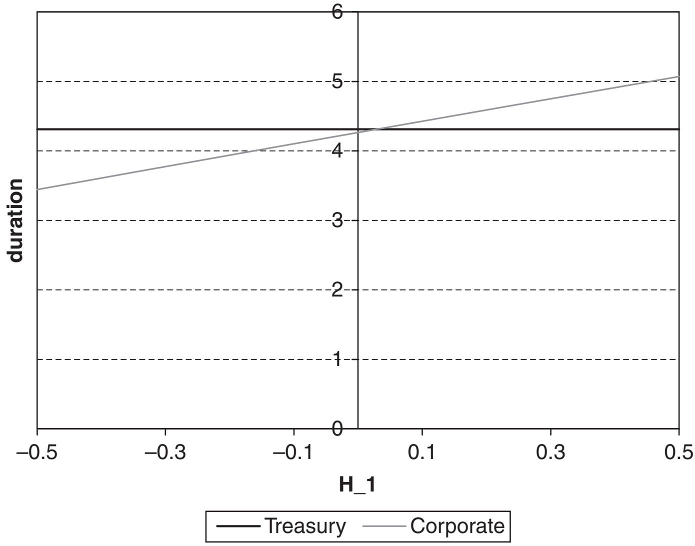

Fig. 13.2: The durations of 10-year bonds. The figure shows the duration of a 10-year corporate coupon bond as a function of the interest rate sensitivity of the default intensity rate (the parameter H_1). The horizontal line shows the duration of the similar Treasury bond. The coupon is 6% per year with semi-annual payments. The current default-free short rate is 4%. The parameters of the short rate process are $\kappa = 0.15$, $\beta = 0.01$, and $\theta = 0.0522$ so that the asymptotic zero-coupon yield is 5%. The loss in case of default is $L = 0.4$ and the fixed part of the default intensity is $H_0 = 0.025$.

Goldstein, and Martin (2001) further show that the common variation is only partially explained by an extensive set of financial and economic variables, which suggests that at least one of the common factors in the default intensities should be a *latent* factor, that is a factor that is not directly observable but can only be inferred from other observable variables. The presence of a common factor in the intensities induces correlations in defaults. Such models have been studied by, for example Duffie and Gârleanu (2001), Driessen (2005), Mortensen (2006), and Eckner (2009).

Let us look at a simple example. Assume that the risk-neutral default intensity of any given issuer, say number i, can be decomposed as

$$h_{it} = x_{it} + \beta_i y_t, \tag{13.34}$$

where $y = (y_t)$ then captures a common, systematic component in default intensities and $x_i = (x_{it})$ is an idiosyncratic, issuer-specific component. The parameter $\beta_i > 0$ is the sensitivity of issuer i with respect to the common intensity component. Both the common and all the idiosyncratic intensity components are assumed to follow independent square-root processes under the risk-neutral probability measure:[11]

[11] The papers just cited allow for 'affine' jumps in x_{it} and y_t but we disregard jumps for simplicity of the presentation.

$$dx_{it} = \kappa_i(\theta_i - x_{it})\,dt + \sigma_i\sqrt{x_{it}}\,dz^{\mathbb{Q}}_{it},$$

$$dy_t = \kappa_y(\theta_y - y_t)\,dt + \sigma_y\sqrt{y_t}\,dz^{\mathbb{Q}}_{yt}.$$

Note that the default intensities are non-negative by construction. The instantaneous variance of the default intensity is

$$\mathrm{Var}[dh_{it}] = \mathrm{Var}[dx_{it}] + \beta_i^2\,\mathrm{Var}[dy_t] = \left(\sigma_i^2 x_{it} + \beta_i^2\sigma_y^2 y_t\right)\,dt,$$

and the instantaneous covariance rate between the intensities of two different issuers is

$$\mathrm{Cov}[dh_{it}, dh_{jt}] = \beta_i\beta_j\,\mathrm{Var}[dy_t] = \beta_i\beta_j\sigma_y^2 y_t\,dt,$$

so the instantaneous correlation is

$$\mathrm{Corr}[dh_{it}, dh_{jt}] = \frac{\beta_i\beta_j\sigma_y^2 y_t}{\sqrt{\sigma_i^2 x_{it} + \beta_i^2\sigma_y^2 y_t}\sqrt{\sigma_j^2 x_{jt} + \beta_j^2\sigma_y^2 y_t}},$$

which is positive and time-varying. Eckner (2009) imposes the constraints $\kappa_i = \kappa_y$ and $\sigma_i = \sqrt{\beta_i}\sigma_y$ for all i, which reduce the number of parameters considerably and to some extent are supported by the empirical study of Feldhütter (2008b). The default correlation then reduces to

$$\mathrm{Corr}[dh_{it}, dh_{jt}] = \frac{\sqrt{\beta_i\beta_j}y_t}{\sqrt{x_{it} + \beta_i y_t}\sqrt{x_{jt} + \beta_j y_t}}.$$

In any case, the affine nature of the model produces exponential-affine expressions of default probabilities via (13.23):

$$\begin{aligned}
\mathbb{Q}(t < \tau < T | \tau > t) &= 1 - \mathrm{E}^{\mathbb{Q}}_t\left[e^{-\int_t^T h_u\,du}\right] \\
&= 1 - \mathrm{E}^{\mathbb{Q}}_t\left[e^{-\int_t^T x_{iu}\,du}\right]\mathrm{E}^{\mathbb{Q}}_t\left[e^{-\beta_i\int_t^T y_u\,du}\right] \\
&= 1 - e^{-a_i(T-t) - b_i(T-t)x_{it} - \hat{a}_i(T-t) - \hat{b}_i(T-t)y_t},
\end{aligned}$$

for suitable deterministic functions $a_i, b_i, \hat{a}_i, \hat{b}_i$. Mortensen (2006) and Eckner (2009) explain how to compute loss distributions on a portfolio of loans and price credit derivatives in this modelling framework.

In settings like the one above there are no feedback effects from the default of a firm on other firms since, conditional on the common factor y in (13.34), defaults of different firms are independent events. However, as explained in Section 13.2.4, the default intensity of a given firm may very well be affected positively or negatively by the defaults of competitors, customers, suppliers, or maybe the counterpart in a large financial transaction. Jarrow and Yu (2001) incorporate that feature by directly allowing the default intensity of one firm to depend on whether or not another firm has yet defaulted. A simple example is to specify the default intensity of firm j as

$$h_{jt} = H_0 + H_1 r_t + H_2 \mathbf{1}_{\{t \geq \tau_i\}}$$

so that the default intensity of firm j will jump by H_2 at the time of default of firm i. Jarrow and Yu provide some simple examples with closed-form expressions for defaultable bonds.

Kraft and Steffensen (2007) suggest the use of a Markov chain to capture the default status of all relevant firms at any point in time. For example, with only two firms i and j of which neither is defaulted initially, we need a Markov chain with four states, say $\{0, 1, 2, 3\}$. Let state 0 represent the state where neither firm is in default, state 1 is the state where i is in default but j is not, state 2 is the state where j is in default but i is not, and state 3 is the state where both firms are in default. The possible transitions of the Markov chain are from state 0 to state 1 or 2, from state 1 to state 3, and from state 2 to state 3.[12] The transitions from 0 to 1 and from 2 to 3 are both due to the default of firm i, but the transitions differ with respect to the status of firm j. By letting the transition probabilities for those two transitions differ, we obtain a model where the default of firm j impacts the default probability of firm i. Similarly, we can allow the default of firm i to influence the default probability of firm j by letting the transition probability from state 0 to 2 to be different from the transition probability from state 1 to 3. Default related claims can then be priced by solving a system of partial differential equations, which under suitable assumptions leads to fairly explicit pricing equations for defaultable bonds. Obviously the number of states of the Markov chain increases rapidly when more firms are included.

13.4.6 Change of measure in reduced-form models

Risk premia are needed to link the dynamics of relevant variables under the risk-neutral pricing measure $\mathbb{Q}$ and under the real-world measure $\mathbb{P}$. We have seen in earlier chapters, for example Chapter 5, that the change of measure corresponds to a drift adjustment of diffusion processes. However, models involving possible defaults are not pure diffusion models. The price of a corporate bond will jump downwards in the event of the default of the issuing firm. The default event itself constitutes a risk that may carry a risk premium. Moreover, this event risk premium enters differently than the risk premia for diffusion risks. To be more precise, let $h_t^{\mathbb{P}}$ and $h_t^{\mathbb{Q}}$ denote the default intensity under the real-world and the risk-neutral probability measure, respectively, that is

$$h_t^{\mathbb{P}} = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{P}(t < \tau < t + \Delta t | \tau > t), \quad h_t^{\mathbb{Q}} = \lim_{\Delta t \downarrow 0} \frac{1}{\Delta t} \mathbb{Q}(t < \tau < t + \Delta t | \tau > t)$$

as in (13.2). The two intensities are generally related via a multiplicative risk premium, which we will denote by λ_t^{def}, so that

[12] Simultaneous default of the two firms is not allowed.

$$h_t^{\mathbb{Q}} = (1 + \lambda_t^{\text{def}}) h_t^{\mathbb{P}}.$$

The event risk premium λ_t^{def} has to be strictly greater than -1 and can very well vary stochastically over time.

Jarrow, Lando, and Yu (2005) argue that the event risk premium is zero when default risk is 'conditionally diversifiable'. They assume that the default intensities of all firms depend on a finite-dimensional state variable X. The default risk is said to be conditionally diversifiable if there are infinitely many firms in the economy and, conditionally on the value of the state variable, the default events are independent. In that case any default risk not driven by the common state variable X can be diversified away by forming large portfolios of corporate bonds from different issuers. With a finite number of corporate bonds and/or some (conditional) dependence in defaults, the event risk premium can be non-zero.

Following Yu (2002), let us look at a small model of the Duffie–Singleton type. The default-free rates are described by a one-factor CIR model

$$dr_t = \kappa[\theta - r_t]\,dt + \beta\sqrt{r_t}\,dz_t, \quad \lambda_{zt} = \frac{\lambda\sqrt{r_t}}{\beta}.$$

The risk-neutral dynamics is then

$$dr_t = \hat{\kappa}\left[\hat{\theta} - r_t\right]dt + \beta\sqrt{r_t}\,dz_t^{\mathbb{Q}},$$

where $\hat{\kappa} = \kappa + \lambda$ and $\hat{\theta} = \theta\kappa/\hat{\kappa}$. The price of the default-free zero-coupon bond maturing at time T is given by

$$B_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right] = e^{-a(T-t)-b(T-t)r_t}, \tag{13.35}$$

where

$$b(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \hat{\kappa})(e^{\gamma\tau} - 1) + 2\gamma},$$

$$a(\tau) = -\frac{2\hat{\kappa}\hat{\theta}}{\beta^2}\left(\ln(2\gamma) + \frac{1}{2}(\hat{\kappa} + \gamma)\tau - \ln\left[(\gamma + \hat{\kappa})e^{\gamma\tau} - 1) + 2\gamma\right]\right),$$

and $\gamma = \sqrt{\hat{\kappa}^2 + 2\beta^2}$. See Section 7.5 for details. Suppose that the real-world dynamics of the risk-neutral expected loss rate $s_t^{\mathbb{Q}} = h_t^{\mathbb{Q}}\ell_t$ is

$$ds_t^{\mathbb{Q}} = \kappa^*[\theta^* - s_t^{\mathbb{Q}}]\,dt + \beta^*\sqrt{s_t^{\mathbb{Q}}}\,dz_t^*,$$

where z^* is a standard Brownian motion under $\mathbb{P}$ and independent from z. Further suppose that the market price of risk associated with z^* is $\lambda_{z^*t} = \lambda^*\sqrt{s_t^{\mathbb{Q}}}/\beta^*$. Then the risk-neutral dynamics of $s^{\mathbb{Q}}$ is

$$ds_t^{\mathbb{Q}} = \hat{\kappa}^*[\hat{\theta}^* - s_t^{\mathbb{Q}}]\,dt + \beta^*\sqrt{s_t^{\mathbb{Q}}}\,dz_t^{*\mathbb{Q}},$$

where $\hat{\kappa}^* = \kappa^* + \lambda^*$ and $\hat{\theta}^* = \theta^*\kappa^*/\hat{\kappa}^*$. By analogy to (13.35), we have

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T s_u^{\mathbb{Q}}\,du}\right] = e^{-a^*(T-t)-b^*(T-t)s_t^{\mathbb{Q}}},$$

where a^* and b^* are defined in the same manner as a and b, just with stars on all relevant parameters. From the Duffie–Singleton pricing rule (13.30) and the independence of r and $s^{\mathbb{Q}}$, we conclude that the price of the defaultable zero-coupon bond is

$$\begin{aligned}\tilde{B}_t^T &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T (r_u+s_u^{\mathbb{Q}})\,du}\right] = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right]\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T s_u^{\mathbb{Q}}\,du}\right]\\ &= e^{-a(T-t)-a^*(T-t)-b(T-t)r_t-b^*(T-t)s_t^{\mathbb{Q}}}\end{aligned}$$

if there is no default up to time t. It follows (see Exercise 13.8) that the dynamics of the bond price in the instant following time t, *conditional on no default in that period*, is

$$\begin{aligned}\frac{d\tilde{B}_t^T}{\tilde{B}_t^T} &= \left(r_t + s_t^{\mathbb{Q}} - \lambda b(T-t)r_t - \lambda^* b^*(T-t)s_t^{\mathbb{Q}}\right)\,dt\\ &\quad - b(T-t)\beta\sqrt{r_t}\,dz_t - b^*(T-t)\beta^*\sqrt{s_t^{\mathbb{Q}}}\,dz_t^*. \qquad (13.36)\end{aligned}$$

However, with a real-world probability of $h_t^{\mathbb{P}}\,dt$ the firm will default in the short period $[t, t+dt]$ in which case the bond value drops by a fraction ℓ_t. The correct expected rate of return on the bond is therefore

$$\begin{aligned}&\left(1 - h_t^{\mathbb{P}}\,dt\right)\left(r_t + s_t^{\mathbb{Q}} - \lambda b(T-t)r_t - \lambda^* b^*(T-t)s_t^{\mathbb{Q}}\right)dt + \left(h_t^{\mathbb{P}}\,dt\right)(-\ell_t)\\ &\qquad = \left(r_t + s_t^{\mathbb{Q}} - \lambda b(T-t)r_t - \lambda^* b^*(T-t)s_t^{\mathbb{Q}} - h_t^{\mathbb{P}}\ell_t\right)\,dt\\ &\qquad = \left(r_t - \lambda b(T-t)r_t - \lambda^* b^*(T-t)s_t^{\mathbb{Q}} + \lambda_t^{\mathrm{def}} h_t^{\mathbb{P}}\ell_t\right)\,dt,\end{aligned}$$

where the first equality follows by disregarding $(dt)^2$-terms, and the second equality follows by substituting in $s_t^{\mathbb{Q}} = h_t^{\mathbb{Q}}\ell_t = (1+\lambda_t^{\mathrm{def}})h_t^{\mathbb{P}}\ell_t$. The corporate bond carries the well-known risk premium due to its exposure to default-free interest rate risk. It also carries a risk premium due to the variations in the risk-neutral expected loss given default as long as λ^* is non-zero. Note that this risk premium is increasing in the maturity of the bond. If λ_t^{def} is non-zero, the defaultable bond will carry an additional risk premium associated with the default event itself. This premium is independent of the maturity of the bond and can therefore induce a positive increase in the spread across all maturities. We can thus conclude that a premium due to non-diversifiable default event risk can explain at least part of the credit spread puzzle. Driessen (2005) provides empirical support for a positive premium on default event risk, but cannot estimate the premium with high precision.

13.4.7 Rating-based models

Rating-based models attempt to derive prices for corporate bonds in the different rating categories. In the basic rating-based models all bonds with the same rating are treated symmetrically and will therefore have the same price and credit spread. The main idea is to use a Markov chain approach for modelling the transitions between different rating categories, including the default category. Risk-neutral rating transition probabilities over all short time periods will imply risk-neutral default intensities and probabilities over any time horizon, which is the key determinant of the bond price. An important issue is, of course, how real-world transition probabilities like those presented in Table 13.3 are transformed into their risk-neutral equivalents. Clearly, the rating of a bond is a crude measure of its default risk, and firms with identical ratings are not necessarily equally risky. However, a rating-based model calibrated to market data might provide a useful benchmark against which the credit spreads of individual corporate bonds can be evaluated. Furthermore, some financial contracts have build-in rating triggers so that certain changes in ratings may imply preset changes in the scheduled cash flow. A rating-based model is natural for the pricing of such contracts.

We will not dig deeper into the rating-based models, but the interested reader should consult the pioneering papers of Lando (1994) and Jarrow, Lando, and Turnbull (1997), the extensions and applications by Das and Tufano (1996), Lando (1998), Acharya, Das, and Sundaram (2002), and Lando and Mortensen (2005) as well as the textbook survey of Lando (2004, Ch. 6).

13.5 HYBRID MODELS

The intensity-based reduced-form approach is computationally attractive, but the standard reduced-form models do not link the default intensity and recovery rate associated with each firm to the relevant characteristics of the firm, such as the asset value and volatility, the magnitude and maturity distribution of the firm's debt, and so on which are essential in the structural models. It is difficult to use reduced-form models to assess the effects of changes in a firm's capital structure or operational risks on credit spreads and default probabilities. Some authors have tried to incorporate more firm-specific information in the reduced-form models by allowing the default intensity to depend on firm-specific state variables, for example the asset value, the equity value, or the equity volatility. The interested reader is referred to Madan and Unal (1998, 2000) and Bakshi, Madan, and Zhang (2006).

Duffie and Lando (2001) investigate the more fundamental links between the structural modelling approach and the reduced-form modelling approach. They set up a structural model in which market participants cannot directly observe the value of the firm's assets, but receive only periodic accounting reports of imperfect precision regarding the true asset value. Between observations, a default could occur completely unexpectedly. The default time is then a totally inaccessible stopping time with an associated intensity process which is linked to the available market information about the value of the assets. Intuitively, the uncertainty about the underlying asset value leads to higher credit spreads and therefore helps in explaining the credit spread puzzle.

Jarrow and Protter (2004) provide an interesting comparison of structural and reduced-form models focusing on the differences in their assumptions concerning the available information. See also Cetin, Jarrow, Protter, and Yildirim (2004).

13.6 COPULAS

For the management of large loan portfolios or pricing contracts depending on such portfolios, the full multivariate distribution of the default times of all loans in the portfolio has to be considered. An approach that has gained popularity among some practitioners is to use a so-called copula model of default times. We will first give a short general introduction to copulas and then discuss applications to credit risk.

A copula is a function that transforms a number of univariate (marginal) probability distribution functions into a multivariate (joint) probability distribution function. More formally, an N-dimensional copula is a function $\mathtt{C}$ from the N-dimensional unit cube $[0,1]^N$ to the interval $[0,1]$. Suppose we are given N univariate random variables X_n with marginal distribution functions F_n, that is $F_n(x_n) = \mathbb{P}(X_n \leq x_n)$. A multivariate distribution function for the N-dimensional random variable $(X_1, \ldots, X_N)$, that is $F(x_1, \ldots, x_n) = \mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n)$, can then be generated by

$$F(x_1, \ldots, x_N) = \mathtt{C}\left(F_1(x_1), \ldots, F_N(x_N)\right).$$

As distribution functions are increasing, the copula function must be increasing in each component. Since we need $F(\infty, \ldots, \infty, x_n, \infty, \ldots, \infty) = F_n(x_n)$, a copula function must satisfy the property

$$\mathtt{C}(1, \ldots, 1, y_n, 1, \ldots, 1) = y_n, \quad n = 1, \ldots, N, \quad 0 \leq y_n \leq 1.$$

According to Sklar's Theorem, any multivariate distribution function that has marginal distribution functions $F_1, \ldots, F_N$ can be generated by some copula function. Note that we can think of the copula itself as being a multivariate probability distribution function of an N-dimensional random variable valued on $[0,1]^N$ and the corresponding probability density function—if it exists—is then the derivative

$$\mathtt{c}(y_1, \ldots, y_N) = \frac{\partial^N}{\partial y_1 \ldots \partial y_N} \mathtt{C}(y_1, \ldots, y_N).$$

How is the dependence of random variables reflected by their copula? First note that if the random variables $X_1, \ldots, X_N$ are independent, the multivariate distribution function is the product of the marginals,

$$F(x_1, \ldots, x_N) = F_1(x_1) \cdots F_N(x_N),$$

so the copula corresponding to independence is $\mathtt{C}(y_1,\ldots,y_N) = y_1 \cdots y_N$ with corresponding density function $\mathtt{c}(y_1,\ldots,y_N) = 1$. The dependence of random variables is therefore captured by how the copula density deviates from 1. If the copula for two variables exceeds 1 at, say, the point $(0.9, 0.6)$, it is more likely that the first variable is at its 90% quantile and the second variable is at its 60% quantile than would be the case if the two variables were independent.

Maximum dependence between two random variables is obtained when they are always identical. Suppose X_1 and X_2 are always identical. Then

$$\begin{aligned}F(x_1,x_2) &= \mathbb{P}\left(X_1 \le x_1, X_2 \le x_2\right) = \mathbb{P}\left(X_1 \le x_1, X_1 \le x_2\right)\\ &= \mathbb{P}\left(X_1 \le \min\{x_1,x_2\}\right) = F_1(\min\{x_1,x_2\}) = \min\{F_1(x_1), F_2(x_2)\},\end{aligned}$$

where the last equality is due to the fact that the probability distribution functions of X_2 and X_1 are identical, $F_2 \equiv F_1$, and increasing. The corresponding copula is therefore

$$\mathtt{C}(y_1,y_2) = \min\{y_1,y_2\},$$

which has no density function as all the probability mass is located at the diagonal $y_1 = y_2$. The same copula works whenever X_2 is a monotonic function of X_1, which makes sense, because the two variables are also maximally dependent in that case. This also shows that copulas are in some respects better measures of dependence than correlation coefficients which, by definition, are linked to *linear* dependence. To follow an example of Brigo and Mercurio (2006, Sec. 21.1.9), it can be shown that the correlation between a standard normal random variable X and X^3 equals $\sqrt{3/5}$, which is of course less than 1, although X and X^3 are clearly maximally dependent in any reasonable sense. The copula of X and X^3 is the same as the copula of X with itself, namely the maximum dependence copula. The maximum dependence copula can easily be extended to higher dimensions.

If $X_2 = 1 - X_1$, it is clear that the correlation between X_1 and X_2 is -1, so they are in a sense minimally dependent (or maximally negatively dependent). The probability distribution function for X_2 is then

$$F_2(x_2) = \mathbb{P}(X_2 \le x_2) = \mathbb{P}(X_1 \ge 1 - x_2) = 1 - \mathbb{P}(X_1 < 1 - x_2) = 1 - F_1(1 - x_2),$$

assuming X_1 has a continuous distribution, and the joint distribution function is

$$\begin{aligned}F(x_1,x_2) &= \mathbb{P}\left(X_1 \le x_1, X_2 \le x_2\right) = \mathbb{P}\left(1 - x_2 \le X_1 \le x_1\right)\\ &= F_1(x_1) - F_1(1 - x_2) = F_1(x_1) + F_2(x_2) - 1,\end{aligned}$$

whenever $x_1 > 1 - x_2$ and $F(x_1,x_2) = 0$ otherwise. The corresponding copula is thus

$$\mathtt{C}(y_1,y_2) = \max\{y_1 + y_2 - 1, 0\}.$$

An obvious extension to higher dimensions would be $\mathtt{C}(y_1,\ldots,y_N) = \max\{\sum_n y_n - (N-1), 0\}$, and although this is harder to relate to correlations, it is still relevant due to the following result. It can be shown that any copula is bounded by the extreme copulas we have just considered:

$$\max\left\{\sum_{n=1}^{N} y_n - (N-1), 0\right\} \leq \mathtt{C}(y_1, \dots, y_N) \leq \min\{y_1, \dots, y_N\}.$$

The bounds are referred to as the Fréchet–Hoeffding bounds on copulas.

Copulas have a long history in probability and statistics, but were only recently introduced in a credit risk context by Li (2000) for modelling dependence between the default times τ_n of different issuers, $n = 1, \dots, N$. Let F_n be the risk-neutral marginal probability distribution function for τ_n, $F_n(t_n) = \mathbb{Q}(\tau_n \leq t_n)$, and let F be the risk-neutral joint distribution function for the default times $\tau_1, \dots, \tau_N$, that is

$$F(t_1, \dots, t_N) = \mathbb{Q}\left(\tau_1 \leq t_1, \dots, \tau_N \leq t_N\right).$$

Then a copula $\mathtt{C}$ links the marginal distributions to the joint distribution for all default times via

$$F(t_1, \dots, t_N) = \mathtt{C}\left(F_1(t_1), \dots, F_N(t_N)\right) = \mathtt{C}(y_1, \dots, y_N),$$

where $y_n = F_n(t_n)$, that is $t_n = F_n^{-1}(y_n)$ assuming the inverse functions F_n^{-1} exist. The choice of copula determines the dependence between the default times. Below we look at some examples. For simplicity we consider only two issuers and two default times, τ_1 and τ_2.

Many practitioners seem to use a **Gaussian copula**. The two-dimensional Gaussian copula with correlation parameter ρ is

$$\mathtt{C}(y_1, y_2) = N_2\left(N^{-1}(y_1), N^{-1}(y_2); \rho\right)$$

so that the joint distribution of default times is

$$F(t_1, t_2) = \mathtt{C}\left(F_1(t_1), F_2(t_2)\right) = N_2\left(N^{-1}(F_1(t_1)), N^{-1}(F_2(t_2)); \rho\right).$$

Here $N_2(\cdot, \cdot; \rho)$ is the probability distribution function of a bivariate standard normal random variable (X_1, X_2) with a correlation of ρ between X_1 and X_2. $N^{-1}(\cdot)$ is the inverse of the probability distribution function for a univariate standard normal random variable. Each default time τ_n is first transformed into $x_n = N^{-1}(F_n(\tau_n))$, which is a standard normal variable since

$$\begin{aligned}
\mathbb{Q}(x_n \leq k) &= \mathbb{Q}\left(N^{-1}(F_n(\tau_n)) \leq k\right) = \mathbb{Q}\left(F_n(\tau_n) \leq N(k)\right) \\
&= \mathbb{Q}\left(\tau_n \leq F_n^{-1}(N(k))\right) = F_n\left(F_n^{-1}(N(k))\right) = N(k).
\end{aligned}$$

Then the two 'normalized default times' x_1 and x_2 are linked assuming a correlation of ρ. The joint default behaviour is fully determined via the marginal default time distributions, that is F_1 and F_2, and the copula correlation parameter ρ. The density of the Gaussian copula is (see Exercise 13.9)

$$\mathtt{c}(y_1, y_2) = \frac{n_2\left(N^{-1}(y_1), N^{-1}(y_2); \rho\right)}{n\left(N^{-1}(y_1)\right) n\left(N^{-1}(y_2)\right)}, \tag{13.37}$$

where

$$n(a) = \frac{1}{\sqrt{2\pi}} \exp\left\{-\frac{a^2}{2}\right\} \quad \text{and}$$

$$n_2(a_1, a_2; \rho) = \frac{1}{\sqrt{2\pi(1-\rho^2)}} \exp\left\{-\frac{a_1^2 + a_2^2 - 2\rho a_1 a_2}{2(1-\rho^2)}\right\}$$

are the probability density functions for a one-dimensional and a two-dimensional standard normal distribution. For $\rho > 0$, the density of the copula will be highest when y_1 and y_2 are either both close to one or both close to zero, and the density is lowest when y_1 is close to zero and y_2 close to one or vice versa. This reflects a tendency for the firms to default at nearby dates. The converse is the case if the copula correlation is negative. Figure 13.3 shows the density of the Gaussian copula for $\rho = 0.2$, where the peaks near $(0, 0)$ and $(1, 1)$ are clearly visible. In practice, the value of the copula correlation is often assumed to be equal to the correlation of equity returns of the two issuing companies or the correlation of their asset values in a Merton-type structural model. Alternatively, one can choose the copula correlation so that the model matches the default correlation discussed in Sections 13.2.4 and 13.3.7 over some time interval. If reliable prices on securities depending on the joint default behaviour of the two companies can be found, the copula correlation can be chosen to get the best match with those prices, but such prices are not always available. The fortuitousness in the determination of the

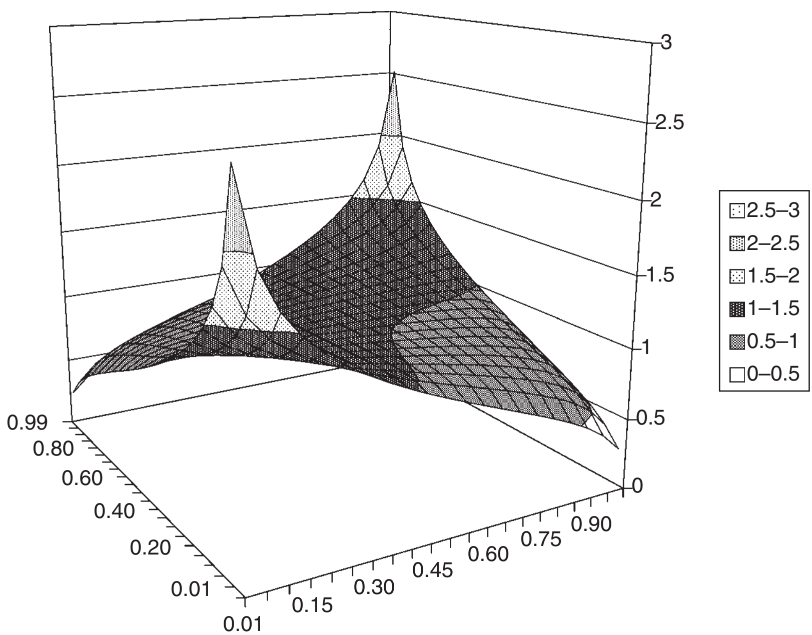

Fig. 13.3: The density of the two-dimensional Gaussian copula with copula correlation parameter $\rho = 0.2$.

copula correlation—or, more generally, the parameters defining the copula used in the modelling of default dependence—is a weakness of the whole copula approach.

If many defaults are to be modelled simultaneously, a factor structure is often imposed for parsimony. In fact, the most widely used model for multi-name credit problems among practitioners is the so-called one-factor Gaussian copula model. The default probability of each relevant entity is implicitly modelled via

$$x_n = \alpha_n M + \sqrt{1-\alpha_n^2}\varepsilon_n, \quad n = 1, \dots, N,$$

where M is a common factor and ε_n is an entity-specific factor. M and all the ε_n's are independent random variables each with a standard normal distribution. The constants α_n are between -1 and $+1$ and determine cross-section correlations: $\mathrm{Corr}[x_n, x_k] = \alpha_n \alpha_k$. Entity n defaults before time t_n, whenever $x_n \leq N^{-1}(F_n(t_n))$, which happens whenever

$$\varepsilon_n \leq \frac{N^{-1}(F_n(t_n)) - \alpha_n M}{\sqrt{1-\alpha_n^2}}.$$

Conditional on the common factor M, the probability that entity n defaults before time t_n is then

$$\mathbb{Q}(\tau_n \leq t_n | M) = N\left(\frac{N^{-1}(F_n(t_n)) - \alpha_n M}{\sqrt{1-\alpha_n^2}}\right).$$

As the conditional distributions are independent, the joint conditional distribution is just the product of the marginals,

$$\mathbb{Q}(\tau_1 \leq t_1, \dots, \tau_N \leq t_n | M) = \prod_{n=1}^{N} \mathbb{Q}(\tau_n \leq t_n | M) = \prod_{n=1}^{N} N\left(\frac{N^{-1}(F_n(t_n)) - \alpha_n M}{\sqrt{1-\alpha_n^2}}\right),$$

and the unconditional joint distribution of default times then follows from an integration over M,

$$\mathbb{Q}(\tau_1 \leq t_1, \dots, \tau_N \leq t_n) = \int_{-\infty}^{\infty} \left\{ \prod_{n=1}^{N} N\left(\frac{N^{-1}(F_n(t_n)) - \alpha_n M}{\sqrt{1-\alpha_n^2}}\right)\right\} \frac{1}{\sqrt{2\pi}} e^{-M^2/2}\, dM.$$

The integration has to be done by numerical methods. In implementations the α_n is often assumed the same for all entities, $\alpha_n = \alpha$. Letting $\rho = \mathrm{Corr}[x_n, x_k] = \alpha^2$, we then get

$$\mathbb{Q}(\tau_n \leq t_n | M) = N\left(\frac{N^{-1}(F_n(t_n)) - \sqrt{\rho} M}{\sqrt{1-\rho}}\right).$$

Also, the marginal distributions F_n are often assumed to be identical.

In order to value an asset depending on the defaults of all N entities, first the payoff structure of the asset has to be specified. Combining that with the joint distribution of the default times, the asset can—at least, in principle—be evaluated.

If the payoff is random, it is usually assumed to be independent of the default times and of the default-free interest rates under the risk-neutral pricing measure.

Other copulas can generate more pronounced dependence patterns than the fairly smooth Gaussian copula. An example is the two-dimensional t-copula defined by

$$\mathtt{C}(y_1, y_2) = t_{\nu,2}\left(t_\nu^{-1}(y_1), t_\nu^{-1}(y_2); \rho\right),$$

where t_ν is the one-dimensional t-distribution with ν degrees of freedom and $t_{\nu,2}(\cdot, \cdot; \rho)$ is the two-dimensional counterpart with correlation ρ. The density of this copula can have much higher peaks in the extremes (the corners of the square $[0,1] \times [0,1]$ on which the copula is defined) than the Gaussian copula and can thus reflect a higher dependence in the tails. Copulas do not have to be defined in terms of probability distributions. For example, the so-called generalized Clayton copula is defined by

$$\mathtt{C}(y_1, y_2) = \left\{\left[\left(y_1^{-\theta} - 1\right)^\delta + \left(y_2^{-\theta} - 1\right)^\delta\right]^{1/\delta} + 1\right\}^{-1/\theta},$$

where $\delta \geq 1$ and $\theta > 0$. Unlike the Gaussian copula and the t-copula, this copula is explicit and therefore, for some purposes, easier to handle. It can also generate very high densities near the points $(0, 0)$ and $(1, 1)$. Note that all three copulas considered can easily be extended to more issuers. The Gaussian copula and the t-copula will generally involve all pairwise correlations, but often the correlation structure is restricted to reduce the number of parameters.

To sum up, a copula gives a mathematically elegant representation of the dependence between random variables such as the default times of different entities. The default time distributions for the individual firms can come either from a structural Merton-type model or from a reduced-form model. However, it is often difficult or impossible to give an economic interpretation of why a given copula should capture default dependence, and it is difficult to determine appropriate values for the parameters that enter the copula. It can therefore easily become a black box that magically produces a joint default distribution which can then be used to price certain securities or assess the value distribution for a portfolio of corporate loans or bonds with little understanding of the procedure and the results. Moreover, copulas are static models of dependence, while default dependence often changes over time and with economic conditions. Financial correlations tend to vary over time and with economic conditions, so any correlation model calibrated to a given data set might be very misleading for another time period. And calibrating a correlation model involving a black box is even more dangerous. For more about copulas, the interested reader is referred to the books by Nelsen (1999) and Cherubini, Luciano, and Vecchiato (2004).

13.7 MARKETS FOR CREDIT DERIVATIVES

A credit derivative is a financial contract that facilitates the transfer of credit risk from one market participant to another. The cash flow and value are mainly determined by the credit performance of one or more corporations, sovereign entities, or loans. Credit derivatives allow investors to manage their exposure to credit risk. They can be used both to reduce specific credit risks (hedging) or overall credit exposure (diversification) or to increase specific credit risks or the overall credit exposure (speculation). The main actors in the credit derivative markets are large banks, insurance companies, and hedge funds, see, for example Mengle (2007). Credit derivatives are mostly traded over-the-counter, but are nevertheless relatively standardized with most contracts following the standards prescribed by the ISDA, the International Swaps and Derivatives Association. Recently (from 2007) some credit derivatives have been introduced on organized exchanges. In exchange-traded contracts, the exchange clearing house is the counterpart in any trade, which eliminates (or, at least, strongly reduces) the counterparty risk which has shown to be a very real risk following the September 2008 bankruptcy filing of Lehman Brothers, one of the most active credit market participants. The failure of Lehman Brothers caused severe damage to other financial institutions that were heavily trading credit derivatives and thus played an important role in the propagation of the financial crisis. Politicians and regulators have since been trying to reform trading practices so that more (if not all) trades in credit derivatives should go through central clearing houses.[13]

Formalized OTC markets for credit derivatives started in the mid-1990s and grew tremendously until 2007. Various types of credit derivatives are traded, but the market continues to be dominated by credit default swaps (CDSs). A CDS can either depend on the default of a single entity or several entities and can thus be divided into single-name and multi-name CDSs. According to Mengle (2007), credit default swaps of various types constituted around 76% of the total credit derivative markets in 2006, followed by a 17% market share of collateralized default obligations (CDOs). Consequently, only little trading takes place in other types of credit derivatives, such as the so-called total return swaps, asset swaps, and various option products. Figures 13.4 and 13.5 show the evolution in the notional amount outstanding and the gross market value, respectively, of credit default swaps from the second half of 2004 to the second half of 2008 according to Bank for International Settlements (2009). The notional amount outstanding peaked in 2007 at around 57 trillion US dollars, but has since declined somewhat. On the other hand, the gross market value has continued to grow into the financial crisis, as the credit spreads and CDS prices have increased due to higher perceived default risk and possibly also liquidity concerns. Note that the market for multi-name CDSs has grown faster than the market for single-name CDSs over the period, but single-name CDSs remain more popular. While the figures only document the trends over the recent years, the market growth rates were also very big in the preceding years. Based on data from the British Bankers' Association, Mengle (2007) reports that the notional amount outstanding of credit derivatives was 'only' 180 billion US dollars in 1997, compared to 20 trillion US dollars in 2006.

We will explain what CDSs and CDOs are and how they can be used and priced in the following two sections.

[13] See Duffie and Zhu (2009) for a critical evaluation of this idea.

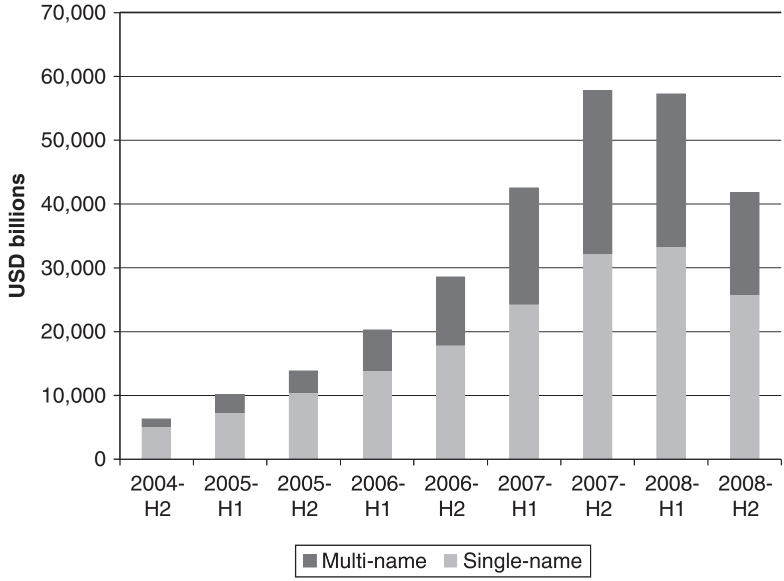

Fig. 13.4: The notional amount outstanding on credit default swaps, 2004–09.
Source: Bank for International Settlements (2009).

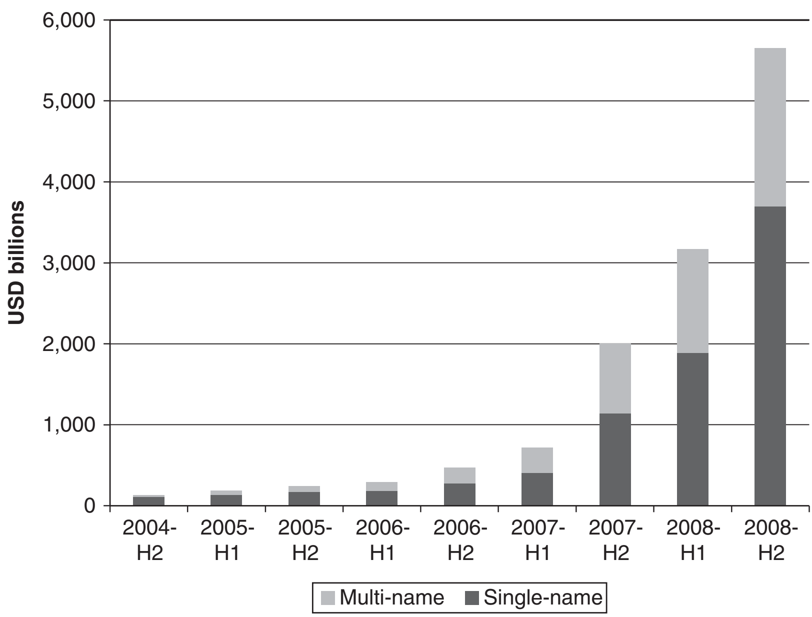

Fig. 13.5: The gross market value of credit default swaps, 2004–09.
Source: Bank for International Settlements (2009).

13.8 CREDIT DEFAULT SWAPS (CDSs)

13.8.1 Single-name CDSs

The basic credit default swap (CDS) is a contract of a certain maturity between two parties—the protection buyer and the protection seller—with a cash flow depending on whether or not a third party—the reference entity—defaults or not before the contract matures. The CDS is designed to protect the protection buyer against the default of the reference entity. The protection buyer and seller can engage in the CDS without the consent of the reference entity. As long as the reference entity has not defaulted and the CDS has not matured, the protection buyer must pay a periodic premium to the protection seller. Should the reference entity default before the CDS matures, the protection seller has to pay a default payment to the protection buyer. The default payment equals the notional amount of the CDS multiplied by the loss rate on bonds, relative to their face value, issued by the reference entity upon default.

The premium payments are typically paid in arrears on a quarterly basis on 20 March, June, September, and December. If the reference entity defaults between two premium payment dates, the protection buyer will have to pay an accrual payment to the seller for the protection in the period from the preceding premium payment to default. The CDS spread is defined as the total premium payment per year as a percentage of the notional amount. The notional amount is typically around 10–20 million US dollars for North American investment-grade credits and around 10 million Euros for European investment-grade credits with smaller notational amounts for non-investment grade reference entities, see Mengle (2007). Contracts are often initiated with 5 years to maturity, but can also be shorter or longer. Usually, a CDS does not involve any other payments than the periodic premia, the accrual premium at default, and the default payment. The CDS spread is set so that no money changes hand when the contract is made, analogous to forward contracts and interest rate swaps. CDS market makers announce bid and ask CDS spreads on various reference entities for one or more maturities and, for the most popular reference entities, the CDS markets are fairly liquid. Reference entities include major corporations and some sovereign nations. If virtually all market participants agree that there is no risk that a given corporation or nation will default over a given period, there will be no interest in trading CDSs linked to that entity, so the reference entities used typically face a real risk of default.

As an example, suppose that company A buys protection from company B on reference company C for a 5-year period starting 20 March 2010. The notional amount is 10 million dollars, and the CDS spread is 160 basis points, that is 1.6% of the notional amount. Over the 5-year period company A has to pay 0.4% of 10 million dollars, that is 40,000 dollars, to company B every quarter as long as company C does not default. The first premium payment is due 20 June 2010. Suppose now that company C defaults on 5 November 2012, and that the bonds issued by company C immediately after default have an estimated value of 40 dollars per 100 dollars of face value. The loss is then 60%, so company B has to pay company A 60% of the notional amount of 10 million dollars, that is 6 million dollars. On the other hand, at the time of default, company A has to pay half of the periodic premium—20,000 dollars—to company B as an accrual premium to cover the protection from 20 September 2012 to the default date, which is roughly half a quarter of a year. If company A possessed bonds issued by company C with a total face value of 10 million dollars, the CDS would exactly compensate for the loss on the bond position. Alternatively, company A may find that the CDS spread is underestimating the probability that company C defaults. Company A will therefore benefit (on a discounted risk-adjusted basis) from holding the CDS either until default/maturity or until the CDS spread increases and company A can sell an off-setting default protection at a higher premium than it pays on the original CDS.

A CDS can either be settled physically or by a cash payment. With a physical settlement the protection buyer has the right to sell bonds issued by the reference entity at the nominal value of the bonds to the protection seller. In this case there is often a delivery option so that the protection buyer can choose between different bonds issued by the reference entity. In a cash settlement a market auction of the reference bond takes place after the default has occurred, the benefits of which go to the protection buyer. The seller of protection then makes a cash payment to the buyer for the possible difference between the face value and the auction-determined recovery value of the reference bond. In the U.S. physical settlement has been most common, but is gradually being replaced by cash settlement, which is already the most frequent settlement method in Europe.

Naturally, it is important for the parties in a CDS to agree on what 'default' of the reference entity is supposed to mean. Other events than the actual bankruptcy may also induce a default payment in the CDS. The events that trigger a CDS default payment are called credit events. In particular, the parties have to agree on whether reorganizations of the reference entity, violation of corporate bond covenants, and so on are credit events.

The basic CDS just described is by far the most frequent credit derivative, but some slightly modified CDSs are also seen. In an **upfront CDS**, the protection buyer pays (the present value of) all premia upfront when the contract is initiated. This is sometimes used for short maturity CDSs or CDSs where the default risk of the reference entity is very high. In a **forward CDS**, the protection buyer and seller agree on a CDS that becomes effective over a future period with a CDS spread determined on the day of the agreement. In a **binary CDS**, the payment at default is preset and thus independent of the actual recovery rate of the defaulted reference bond. Finally, in a **constant maturity CDS** the protection premium is reset at regular intervals according to a prespecified benchmark CDS spread of a fixed maturity, for example the 5-year CDS spread of a particular reference entity. In the following we will focus on the valuation of the basic CDS, but the modified versions can be valued in a similar manner. Options on CDSs are also traded.[14]

13.8.2 The valuation of single-name CDSs

Let us show how a single-name CDS can be valued in the reduced-form framework. Suppose the scheduled premium payment dates are $T_1, \dots, T_n$, where, for simplicity, we assume $T_{i+1} - T_i = \delta$. As introduced in (1.12), we define $i(t) = \min\{i \in \{1, 2, \dots, n\} : T_i > t\}$, so that $T_{i(t)}$ is the nearest following scheduled pay-

[14] See Hull and White (2003) and Brigo and Mercurio (2006, Ch. 23) for an introduction to the pricing of CDS options.

ment date after time t. Let F denote the notional amount of the CDS. Suppose the CDS involves an annualized premium of k. Given no default at time t, the value of the future premia (excluding the accrual payment upon default) is

$$\begin{aligned}
V_t^{\text{prem}} &= \mathrm{E}_t^{\mathbb{Q}}\left[\sum_{i=i(t)}^{n} \delta k F e^{-\int_t^{T_i} r_u\,du} \mathbf{1}_{\{\tau > T_i\}}\right] \\
&= \delta k F \sum_{i=i(t)}^{n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{T_i} r_u\,du} \mathbf{1}_{\{\tau > T_i\}}\right] \\
&= \delta k F \sum_{i=i(t)}^{n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{T_i} (r_u + h_u)\,du}\right],
\end{aligned} \tag{13.38}$$

where the last equality is due to (13.24). The last expectation is the time t price of a defaultable zero-coupon bond with zero recovery. Let $\ell_\tau = 1 - R_\tau$ denote the loss rate on the reference entity in case of default. The CDS default payment is then $F\ell_\tau$, which has a present value of

$$V_t^{\text{prot}} = \mathrm{E}_t^{\mathbb{Q}}\left[F\ell_\tau e^{-\int_t^{\tau} r_u\,du} \mathbf{1}_{\{\tau \le T_n\}}\right] = F \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{u} (r_s + h_s)\,ds} h_u \ell_u\right] du,$$

where we have used (13.25). Often the loss is assumed to be known in advance and constant over time, in which case we get

$$V_t^{\text{prot}} = F\ell \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{u} (r_s + h_s)\,ds} h_u\right] du.$$

With default at time τ, the previous premium payment took place at time $T_{i(\tau)-1}$. The accrual payment at default is therefore $[\tau - T_{i(\tau)-1}]kF$, which has a present value of

$$\begin{aligned}
V_t^{\text{accr}} &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau} r_u\,du} [\tau - T_{i(\tau)-1}] k F \mathbf{1}_{\{\tau \le T_n\}}\right] \\
&= kF\, \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau} r_u\,du} [\tau - T_{i(\tau)-1}] \mathbf{1}_{\{\tau \le T_n\}}\right] \\
&= kF \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{u} (r_s + h_s)\,ds} h_u\right] [u - T_{i(u)-1}]\,du,
\end{aligned}$$

again using (13.25). Adding up the terms, the total present value of the CDS with spread k to the protection buyer is

$$\begin{aligned}
V_t^{\text{CDS}} &= V_t^{\text{prot}} - V_t^{\text{prem}} - V_t^{\text{accr}} \\
&= F \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{u} (r_s + h_s)\,ds} h_u \ell_u\right] du - \delta k F \sum_{i=i(t)}^{n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{T_i} (r_u + h_u)\,du}\right] \\
&\quad - kF \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{u} (r_s + h_s)\,ds} h_u\right] [u - T_{i(u)-1}]\,du.
\end{aligned}$$

The fair CDS spread at time t, for the given maturity and reference entity, is the value of k that makes $V_t^{\mathrm{CDS}} = 0$. We denote this by ζ_t^{CDS} and find that

$$\zeta_t^{\mathrm{CDS}} = \frac{\int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds} h_u \ell_u\right] du}{\delta \sum_{i=i(t)}^{n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{T_i}(r_u+h_u)\,du}\right] + \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds} h_u\right][u - T_{i(u)-1}]\,du},$$

which with a constant loss rate simplifies to

$$\zeta_t^{\mathrm{CDS}} = \frac{\ell \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds} h_u\right] du}{\delta \sum_{i=i(t)}^{n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{T_i}(r_u+h_u)\,du}\right] + \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^u (r_s+h_s)\,ds} h_u\right][u - T_{i(u)-1}]\,du}. \tag{13.39}$$

When the loss rate is assumed to be constant, and r and h are affine functions of some, possibly multi-dimensional, diffusion x with affine drift and instantaneous variance-covariance matrix, the expectations in (13.39) are given in closed form by (13.31) and (13.32). The fair CDS spread then follows from two fairly simple numerical integrations. Alternatively, assume—together with a constant loss rate—a deterministic default intensity $h(u)$ in which case the survival probability is $\mathbb{Q}(\tau > T|\tau > t) = \exp\{-\int_t^T h(u)\,du\}$. Then the fair CDS spread becomes

$$\zeta_t^{\mathrm{CDS}} = \frac{\ell \int_t^{T_n} h(u)\mathbb{Q}(\tau > u|\tau > t)B_t^u\,du}{\delta \sum_{i=i(t)}^{n} \mathbb{Q}(\tau > T_i|\tau > t)B_t^{T_i} + \int_t^{T_n} h(u)\mathbb{Q}(\tau > u|\tau > t)B_t^u[u - T_{i(u)-1}]\,du}.$$

The present value of the accrual payment is typically small relative to the other terms. Ignoring the accrual term, we obtain the simple approximation

$$\zeta_t^{\mathrm{CDS}} \approx \frac{\ell \int_t^{T_n} h(u)\mathbb{Q}(\tau > u|\tau > t)B_t^u\,du}{\delta \sum_{i=i(t)}^{n} \mathbb{Q}(\tau > T_i|\tau > t)B_t^{T_i}},$$

which gives a simple relationship between the CDS spreads and the survival (and, hence, the default) probabilities. It can be shown that

$$\zeta_t^{\mathrm{CDS}} \approx \frac{\ell \sum_{i=i(t)}^{n} \mathbb{Q}(T_{i-1} < \tau \le T_i|\tau > t)B_t^{T_i}}{\delta \sum_{i=i(t)}^{n} \mathbb{Q}(\tau > T_i|\tau > t)B_t^{T_i}} \tag{13.40}$$

if we assume that the protection payment following a default takes place at the following scheduled premium payment date. Note that Equation (13.40) looks very much like the fair swap rate in a plain vanilla interest rate swap, see (6.31). Given observed CDS spreads for different maturities, it is possible to back out implied risk-neutral default probabilities over different time horizons from these expressions.

13.8.3 Multi-name CDSs

Credit index trading was introduced in the markets in 2004 and has grown very rapidly since then. In fact, in 2006 index CDSs had roughly the same market share as single-name CDSs, see Mengle (2007). An **index CDS** contract provides protection against the defaults of a portfolio of reference entities. In the life of the contract, the protection buyer receives a default payment whenever one of the entities in the portfolio defaults. The default payment is calculated exactly as for a single-name CDS. Upon a default, the notional amount and the premium are reduced to reflect the decreased number of entities in the portfolio. As an example, consider a 5-year index CDS on 100 corporations. Suppose that the notional amount is 1 million dollars per company, that is 100 million dollars in total, and that the index CDS spread is 80 basis points per year. As long as none of the 100 reference companies default, the protection buyer has to pay a premium, say on a quarterly basis, of 0.8%/4=0.2% of 100 million, that is 200,000 U.S. dollars. When the first company defaults, the protection buyer receives a default premium equal to the loss rate on that company multiplied by the 1 million notional amount for each company. The total notional amount goes down to 99 million dollars and the quarterly premium to 0.2% of 99 million, that is 198,000 dollars, until the next default. Accrual premium payments are computed for index contracts in the same way as for single-name CDSs.

The fair index CDS spread should roughly be an average of the fair CDS spreads on the reference entities in the index portfolio. Companies with high CDS spreads should be weighted less than companies with low spreads though. This is because the high spread is not expected to be paid for as long as the low spread due to the differences in the default probabilities reflected by the spreads. Also note that there might be differences in what counts as a credit event in the index CDS and in the individual CDSs. The default correlations between the companies in the portfolio are also important when determining the index CDS spread. In principle, reduced-form models with correlated intensities across companies can be used, but the computations quickly become very messy. Most market participants rely on the Gaussian copula when handling multi-name instruments.

The two main index CDS contracts are the CDX index and the iTraxx index. The CDX index refers to a portfolio of 125 investment grade companies in North America, whereas the iTraxx index refers to a portfolio of 125 European companies, most of them being investment grade. The portfolios are updated on 20 March and September each year. There are liquid markets for CDSs on these indices for maturities of 3, 5, 7, and 10 years, typically maturing on 20 June or December. It is also possible to buy or sell protection on tranches of an index, where protection is only given for total portfolio losses in some interval, say $[F_1, F_2]$. As long as the total loss due to defaults of the companies in the portfolio are below F_1, the protection buyer receives nothing from the protection seller. When losses exceed F_1, protection is triggered, but the payment from the seller to the buyer is capped at $F_2 - F_1$.

Another type of multi-name CDSs is the basket default swap that provides some protection against defaults in a basket of (typically 5–10) specified reference entities. The most common is the **first-to-default basket CDS** in which the protection seller compensates the protection buyer for losses incurred when the first entity in the basket defaults, after which the CDS terminates. The first-to-default basket CDS is relatively tractable with a reduced-form approach as long as the defaults of the entities are conditionally independent, that is given all other uncertainty the default events are independent across entities. Let $\tau_1, \dots, \tau_K$ denote the default times of the K entities, and let $h_1, \dots, h_K$ be the associated risk-neutral default intensities. The first default time is obviously

$$\tau^* = \min\{\tau_1, \dots, \tau_K\},$$

and under the conditional independence assumption τ^* has an associated risk-neutral default intensity given by

$$h_t^* = h_{1t} + \dots + h_{Kt}.$$

Intuitively, the probability that one of the entities defaults over the next short period is the sum of the default probabilities of the individual entities. Since the payment of the protection premium stops after the first default, the premium payments can be valued as in (13.38) but using the default intensity h^*. The probability that the first default takes place in a short interval $[t, t + dt]$ and that it is entity k that defaults is $h_{kt} \exp\{-\int_0^t h_s^*\, ds\}\, dt$. If $F\ell_{kt}$ denotes the protection payment in this case, it can be shown—see, for example Lando (2004, Ch. 8)—that the value of the protection payment becomes

$$V_t^{\text{prot}} = F \sum_{k=1}^{K} \int_t^{T_n} \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^u (r_s + h_s^*)\, ds} h_{ku} \ell_{ku} \right] du.$$

The fair premium for the first-to-default basket CDS can now be computed by aligning the value of the premium payments and the value of the protection payment. However, note that the assumption of independent defaults seems unrealistic, as we have discussed earlier in this chapter. Basket CDS valuation can be very sensitive to the correlations assumed.

More advanced basket CDSs also exist. For example, the ***k*th-to-default basket CDS** provides a protection payment only when the kth default among the entities in the basket occurs and the contract is then terminated. The valuation of such contracts is very complicated, see, for example Bielecki and Rutkowski (2002).

Exercises 13.5 and 13.6 contain more information on the valuation of basket CDSs.

13.9 COLLATERALIZED DEBT OBLIGATIONS (CDOs)

A collateralized debt obligation (CDO) is a financial structure that takes the payments from a portfolio of loans and redistributes them in a pre-specified way to different securities. The loans can be ordinary bank loans to consumers and corporations, credit card loans, mortgage loans, and so on. CDOs backed by other assets than loans, such as various bonds, also exist.[15] Usually the structure is designed as follows. A bank or other financial institution has provided loans to various customers. The bank sets up a separate legal entity called a special purpose vehicle (SPV) or a conduit and sells off the loan portfolio to this entity. Usually, the bank continues to service the loans, that is it collects the payments from the borrowers, and earns a fee for doing this. The SPV then issues and sells securities with cash flows determined by the payments received on the loans in the portfolio, net of the service fees to the bank. If the originating bank does not buy any of these securities, it will be completely unexposed to the credit risk of the borrowers. The credit risk has been securitized and sold off to other investors. However, the bank will normally retain some of the issued securities as explained below.

The SPV will typically issue a low number of so-called **tranches** that have very different risk and return characteristics. Let us consider a simple example, where the underlying loan portfolio has a total face value of 100 million dollars and all loans have a zero coupon and mature at the same date T. Suppose the SPV has issued three tranches backed by this loan portfolio so that tranche 1 (called the equity tranche or junior tranche) has a face value of 5 million dollars, tranche 2 (the mezzanine tranche) has a face value of 15 million dollars, and tranche 3 (the senior tranche) has a face value of 80 million. Each tranche can then be divided into smaller identical pieces that can be traded among investors. Let R be the actual payments received from the pool of loans at time T. If all of the borrowers pay their debt in full, $R = 100$ million and all three tranches will receive their face value. With only few defaults on the loans so that R is between 95 and 100 million, both tranche 2 and tranche 3 investors get their face value, whereas tranche 1 investors only receive $R - 95 \leq 5$ million dollars. If a relatively large number of defaults occur so that R is between 80 and 95 million, tranche 1 investors receive nothing, tranche 2 investors receive only $R - 80$ million, whereas tranche 3 investors still receive the full 80 million. Finally, with many defaults so that R is less than 80 million, neither tranche 1 nor tranche 2 investors get anything, whereas tranche 3 investors get to share R, less than promised. The equity tranche 1 is hit first by defaults and thus contains most credit risk, while the senior tranche 3 is protected the most against defaults and only hit with massive defaults in the loan portfolio. Obviously, investors will demand compensation for the credit risk in terms of higher required returns.

Typically, the CDO has been constructed, often in close cooperation between the bank and a credit rating agency, so that the senior tranche can be top-rated and thus sold at a high price to investors trusting that a top-rating means virtually risk-free securities. However, even some of these tranches have later suffered from an unexpectedly high number of defaults on the underlying loans. From the first quarter of 2005 to the first quarter of 2007, 66% of CDOs of asset backed securities were downgraded and 44% were even downgraded from investment grade to non-investment grade, see Crouhy, Jarrow, and Turnbull (2008). The credit rating agencies made large profits from the banks for these services and they have been criticized for issuing much too positive ratings on many CDO tranches. A related critique is that the bank originally lending money to customers did not have to

[15] Sometimes a name reflecting the underlying type of loan or asset is used, for example collateralized loan obligation (CLO), collateralized mortgage obligation (CMO), or collateralized bond obligation (CBO).

do a proper credit evaluation of the borrowers, since the bank would afterwards form a CDO and unload the credit risk to other investors. Of course, the CDO tranche investors should then care about the default risk on the loans and price the tranches accordingly, but most investors seem to have trusted the possibly inflated ratings. The most risky tranche, the equity tranche, is not rated and is sometimes retained by the bank creating the CDO, which at least should give them incentives for undertaking some credit analysis of individual borrowers.

The medium-risky mezzanine tranches typically receive a BBB rating and have been difficult to sell off. Creative financial engineers also had an answer to this problem. They pooled together mezzanine tranches from different CDOs and constructed a new CDO based on those payments with a senior tranche, a mezzanine trance, and an equity tranche. This is referred to as a CDO-squared structure. The senior tranche in this second-dimension CDO was then also given a top-rating and could be sold off at a high price. Such a high rating might be reasonable if it is very unlikely that a substantial number of defaults occur simultaneously in all the underlying loan portfolios, but again that depends on the default correlations within each of the loan portfolios and across the different loan portfolios.

To understand the importance of default correlations for the payoffs and, consequently, the risks and prices of CDOs, let us consider the two extreme cases of independence and perfect correlation. Suppose that there is a (risk-neutral) probability of 3% over the next 5 years for each of the loans in a portfolio of 100 loans. If the defaults are independent of each other, the probability of no defaults is $(0.97)^{100} = 0.04755$, that is just below 5%. The probability of five or more defaults—which would wipe out the equity tranche in our above CDO example—is 18.2%, whereas the probability of 20 or more defaults—which would wipe out the mezzanine tranche—is $1.85 \cdot 10^{-6}$ per cent. The senior tranche in the example would then be virtually risk-free.

With perfectly correlated defaults, either none or all of the loans default. No defaults occur with a probability of 97% and all loans default with a probability of 3%. In this extreme case, the tranches of a CDO are equally risky. Compared to the case of independent defaults, the equity tranche is now less risky, while the mezzanine and senior tranches are more risky. These extremes illustrate a general point. The higher the default correlation, the less risky is the equity tranche (often retained by the CDO issuer) and the more risky is the senior tranche (sold off as virtually risk-free). It is now generally believed that default correlations were underestimated in the valuation of many CDOs and that the credit rating agencies and the CDO issuer had incentives to do so.

Our above example assumes that the underlying loans were zero-coupon loans with identical maturities. This makes it easy to define how the payments received from the loans flow to different tranches. In reality the underlying loans are typically coupon loans—maybe with floating coupon rates—which can have different maturities and amortization schedules. All or some of the CDO tranches will also receive coupon payments that can be fixed or floating. Then it is far more complicated to define and understand how the payments on the loans are redistributed to tranche holders. If no or few defaults on the loans occur early in the life of the CDO, the payments received on the loans may be sufficient to pay coupons to all tranches. But then if there are many defaults near the CDO maturity, it may be impossible to fully pay off the senior trancheholders. In that situation equity trancheholders have received some payments even though senior trancheholders do not receive their full face value, which is conflicting with the original idea of CDOs. To avoid this, some funds are retained by the SPV during the life of the CDO to ensure that nothing flows to sub-senior tranches unless the senior tranche is fully repaid. The schedule for distributing cash flows from loans to tranches is called the **waterfall** of the CDO and is often very complex. Also note that CDOs are sometimes over-collateralized in the sense that the total face value of the underlying loans exceeds the total face value on the CDO tranches, which reduces the credit risk of all tranches.

In the CDO structures described above, the SPV issuing the tranches owns the underlying loans or assets, and the structure is then called a **cash CDO**. Another structure is the **synthetic CDO** in which the tranche issuer does not have legal ownership of the loans or assets, but has a credit risk exposure to various entities via short positions in CDSs, that is the tranche issuer has sold default protection on these entities. The premia and protection payments on all these CDSs are pooled together and are divided into tranches that can be sold to other investors, just as for a cash CDO.

There are various approaches to the valuation of CDOs. The most popular among practitioners seem to be a factor copula approach, in particular the one-factor Gaussian copula described in Section 13.6 is widely used, but other copulas have also been implemented. Furthermore, the copula technique has been extended to allow for a dynamic correlation structure. For more details on the copula approach to CDO pricing, see Andersen and Sidenius (2005), Burtschell, Gregory, and Laurent (2009), Gregory and Laurent (2005), and Hull and White (2004, 2006). Duffie and Gârleanu (2001) assume that the default intensity for each loan in the portfolio is the sum of a common variable and a loan-specific variable, that is $h_{it} = X_t^c + X_t^i$, where all the variables $X^c, X^1, \dots, X^K$ are independent affine stochastic processes. The default intensities are thus correlated via the common variable X^c. The fractional loss rates given default are assumed to be uniformly distributed on the $(0, 1)$ interval. Default-free interest rates, loss rates, and default intensities are assumed to be independent (under the risk-neutral probability measure), and CDOs are then priced by Monte Carlo simulation. While simulation can also be implemented for non-affine processes, the procedure is computationally more efficient with an affine model structure. Their examples further illustrate the effects of the cross-sectional correlation of default intensities on CDO prices. Mortensen (2006) extends the approach and provides supportive empirical results. Finally, Hull, Predescu, and White (2006) suggest a structural model in which the assets of the entities in a multi-name credit structure are correlated. For modelling of CDO-squareds, see, for example Dorn (2007).

13.10 CONCLUDING REMARKS

This chapter has provided an introduction to credit risk modelling and the pricing of defaultable bonds and credit derivatives. The main modelling approaches build on techniques well-known from standard option pricing models and dynamic term structure models. A key challenge is to build models that are consistent with observed prices on defaultable bonds and credit derivatives as well as the level, shape, and dynamics of observed credit spread curves. Another key challenge is the modelling of default correlations between various firms or loans. The market practice of using Gaussian copulas seems to have misspecified the true correlation structure. However, more proper modelling easily becomes very computationally intensive with respect to both estimation of the model and pricing using the model.

Some issues not considered in the main part of the chapter deserve mentioning. The first issue is *liquidity*. Corporate bonds are typically traded with higher transaction costs, and they are much less liquid than comparable government bonds. Investors want compensation—a 'liquidity premium'—for the level of illiquidity or transaction costs and maybe also for the liquidity risk, that is the uncertainty about future liquidity, in the form of a higher expected return so that the expected return after taking trading costs and liquidity issues into account is the same as that of a perfectly liquid asset with the same exposures to market risks. Amihud, Mendelson, and Pedersen (2005) survey the theoretical and empirical literature on the general impact of liquidity on asset prices. A number of recent papers investigate empirically how liquidity is priced in the markets for corporate bonds and credit derivatives, see Driessen (2005), Longstaff, Mithal, and Neis (2005), Chen, Lesmond, and Wei (2007), Bongaerts, de Jong, and Driessen (2009), and Dick-Nielsen, Feldhütter, and Lando (2009). The liquidity premium is found to be quite high for low-rated bonds and increasing as the credit quality improves and typically also increasing in the maturity of the bond. For example, Dick-Nielsen, Feldhütter, and Lando (2009) compute liquidity premia for U.S. corporate bonds across different rating categories and maturities, both before (first quarter 2005 to first quarter 2007) and after (second quarter 2007 to end of 2008) the onset of the so-called sub-prime crisis. Among other things they find (see their Table 6) that the liquidity premium for AAA- and AA-rated bonds was 1–2 basis points before and 13–45 basis points after the sub-prime crisis, whereas for B-rated bonds the liquidity premium was 69–116 basis points before and 162–309 basis points after the crisis began. This line of research suggests that liquidity risk in corporate bonds is significant and can explain at least part of the credit spread puzzle. In the recovery of market value framework, Duffie and Singleton (1999) suggest introducing liquidity directly in the adjusted discount rate by adding a liquidity variable to the original default-risk adjusted rate, that is replace $r_u + h_u \ell_u$ in (13.30) by $r_u + h_u \ell_u + \text{liq}_u$ for some liquidity-related process liq_u. However, for applications of a model like that, it is a challenge to quantify such a liquidity variable in a simple and reasonable way.

The second issue regards the fact that many corporate bonds are *callable*, which means that the issuer has the right to pay back the debt prematurely at a prespecified price (the face value plus a call premium) during the full maturity of the bond or at least part of the life of the bond. The decision to call the bond is taken by the management of the issuing company, so it is natural to study call decisions and value callable corporate bonds in a model of the structural type, see Goldstein, Ju, and Leland (2001). The call feature can also be included in a reduced-form model by introducing a call intensity to model the probability of call decisions, see Jarrow, Li, Liu, and Wu (2010).

A third issue which we have more or less ignored is *counterparty risk*. Credit risk is present in many contractual relations besides a corporate bond or loan, for example, in over-the-counter derivative transactions. Suppose Company A has bought an option from Company B. If the option finishes in the money but Company B has defaulted, Company A will not receive the promised option payoff. Derivatives dealers typically include various clauses in their contracts to mitigate credit risk, for example netting (if a company defaults on one contract it has with a counterparty, then it must default on all outstanding contracts with the counterparty) and collateralization (contracts are marked to market periodically using a pre-agreed formula), but counterparty risk is not completely eliminated. Models of counterparty risk have been studied by, for example Duffie and Huang (1996), Huge and Lando (1999), and Hübner (2001).

13.11 EXERCISES

Exercise 13.1 Use Itô's Lemma to show (13.15).

Exercise 13.2 Show (13.18).

Exercise 13.3 Consider the basic Merton model of Section 13.3.1. Define $\mathcal{L}_t = Fe^{-r[T-t]}/V_t$, which is some normalized measure of the firm's leverage.

(a) Show that the credit spread in (13.16) can be rewritten as

$$\zeta_t^T = -\frac{1}{T-t}\ln\left(\mathcal{L}_t^{-1}N(\theta_1) + N(\theta_2)\right),$$

where

$$\theta_1 = -\frac{\frac{1}{2}\sigma^2(T-t) - \ln\mathcal{L}}{\sigma\sqrt{T-t}} \quad \text{and} \quad \theta_2 = -\frac{\frac{1}{2}\sigma^2(T-t) + \ln\mathcal{L}}{\sigma\sqrt{T-t}}.$$

(b) Draw the spread curve $u \mapsto \zeta_t^{t+u}$ (for maturities from 0 up to $u = 30$ years) for all combinations of $\sigma \in \{0.1, 0.3, 0.5\}$ and $\mathcal{L} \in \{0.25, 1, 4\}$. What can you conclude about the influence of σ and $\mathcal{L}$ on the shape of the spread curve?

Exercise 13.4 Use Itô's Lemma to show (13.20) and discuss the properties of the process LEV.

Exercise 13.5 Which contract is cheaper: a first-to-default basket CDS or a second-to-default basket CDS (on the same portfolio, with same maturity, etc.)? How does the price difference depend on the default correlations in the portfolio?

Exercise 13.6 Consider a basket of corporate bonds. Credit default swaps are traded on all individual bonds in the basket and a first-to-default basket CDS is also traded. Explain why the fair premium on the first-to-default basket CDS has to be smaller than the sum of the fair premia on the individual CDSs. Is it true that the fair first-to-default premium has to be smaller than the biggest of the fair CDS premia on the individual bonds?

Exercise 13.7 Suppose $x = (x_t)$ is an affine diffusion with dynamics

$$dx_t = (\varphi - \kappa x_t)\, dt + \sqrt{\delta_1 + \delta_2 x_t}\, dz_t.$$

Then Section 7.2 has shown that

$$\mathrm{E}_t\left[e^{-\int_t^T x_u\, du}\right] = e^{-a(T-t)-b(T-t)x_t}$$

for functions a and b that solve a system of ordinary differential equations. Show that

$$\mathrm{E}_t\left[e^{-\int_t^T x_u\, du}(\nu_1 + \nu_2 x_T)\right] = (\hat{a}(T-t) + \hat{b}(T-t)x_t)e^{-a(T-t)-b(T-t)x_t},$$

where a and b are the same as above, and where $\hat{a}$ and $\hat{b}$ are also deterministic functions. Provide a system of ordinary differential equations (with appropriate boundary conditions) which $\hat{a}$ and $\hat{b}$ must solve. *Hint:* First note that $\mathrm{E}_t\left[\exp\{-\int_t^T x_u\, du\}(\nu_1 + \nu_2 x_T)\right]$ can be written as $f(x_t, t)$ for some function f. Use Theorem 4.10 to write down a partial differential equation for f and verify that $f(x,t) = (\hat{a}(T-t) + \hat{b}(T-t)x)\exp\{-a(T-t) - b(T-t)x\}$ is a solution when the deterministic functions solve appropriate ordinary differential equations.

Exercise 13.8 Show (13.36).

Exercise 13.9 Show (13.37).

14

Mortgages and Mortgage-backed Securities

14.1 INTRODUCTION

A mortgage is a loan offered by a financial institution to the owner of a given real estate property, which is then used as collateral for the loan. In many countries, a mortgage is the standard way to finance (a large part of) the purchase of residential property, for example in the United States, the United Kingdom, Germany, Spain, France, the Netherlands, and the Scandinavian countries. In some countries mortgages are typically financed by the issuance of bonds. A large number of similar mortgages are pooled either by the original lending institution or by some other financial institution. The pooling institution issues bonds with payments that are closely linked to the payments on the underlying mortgages. Afterwards, the bonds are traded publicly. The primary purpose of this chapter is to discuss the valuation of these mortgage-backed bonds.

Mortgage-backed bonds are different from government bonds in several respects. Most importantly, the cash flow to the bond owners depends on the payments that borrowers make on the underlying mortgages. While a mortgage specifies an amortization schedule, most mortgages allow the borrow to pay back outstanding debt earlier than scheduled. This will, for example, be relevant after a drop in market interest rates, where some borrowers will prepay their existing high-rate mortgage and refinance at a lower interest rate. As we will discuss later in this chapter, mortgages are also prepaid for other reasons. The biggest challenge in the valuation of mortgage-backed bonds is to model the prepayment behaviour of the borrowers, whose mortgages are backing the bonds. Once the state-dependent cash flow of the bond has been specified, the standard valuation tools can be applied.

Section 14.2 provides an overview of some typical mortgages. Section 14.3 describes the standard class of mortgage-backed bonds, the so-called pass-throughs. Section 14.4 focuses on the prepayment option embedded in most mortgages and lists various factors that are likely to affect the prepayment activity. There are two distinct approaches to the modelling of prepayment behaviour and the effects on the valuation of mortgage-backed bonds. The option-based approach discussed in Section 14.5 focuses on determining the rational prepayment behaviour of a borrower. Since the prepayment option can be interpreted as an American call option on an interest rate dependent security, the optimal prepayment strategy can be determined exactly as one determines an optimal exercise strategy for an American option. However, this will only capture prepayments due to a drop in interest rates, while a borrower may choose to prepay for other reasons. Hence, various modifications of the basic option-based approach are also considered. The empirical approach outlined in Section 14.6 is based on historical records of actual prepayment behaviour and tries to derive a relation between the prepayment activity and various explanatory variables. Measures of the risk of investments in mortgage-backed bonds are discussed in Section 14.7. Section 14.8 offers a short introduction to other mortgage-backed securities than the standard pass-throughs, whereas Section 14.9 gives a very brief overview of the subprime crisis that hit the U.S. mortgage and housing markets in 2006. Finally, Section 14.10 summarizes the chapter.

In a single chapter it is impossible to cover the many aspects, conventions, and institutional details of the fairly advanced and innovative markets for mortgages and mortgage-backed securities. For more details on especially the U.S. mortgage industry, see Fabozzi (2010, Chs. 10–14). Further information about European markets can be found in Batten, Fetherston, and Szilagyi (2004).

14.2 MORTGAGES

A mortgage is a loan for which a specified real estate property serves as collateral. The lender is a financial institution, the borrower is the owner of the property. The borrower commits to pay back the debt according to a specified payment schedule. If the borrower cannot pay the lender according to the agreed schedule, the lender has the right to foreclose the loan and seize the property. Typically, the mortgage is initiated when the property is traded and the new owner needs to finance the purchase, but sometimes the existing owner of a property may also want to take out a (new) mortgage. Before offering a mortgage, the financial institution will typically assess the market value of the property and the creditworthiness of the potential borrower. The two key figures are the loan-to-value ratio and the payment-to-income ratio. The loan-to-value ratio is simply the ratio of the amount borrowed to the market value of the property (in case the property has not been traded recently, a market value must be estimated). The payment-to-income ratio is the ratio of the periodic payment on the mortgage to the income of the borrower. Obviously, higher values of the payment-to-income ratio and the loan-to-value ratio imply a higher likelihood of default on the loan. Legislation may impose restrictions on the mortgages that can be offered, for example by capping the original loan-to-value ratio at, for example, 80%. Additional borrowing is then possible via the standard banking system. The difference between the market value of the property and the total amount borrowed using the property as collateral is called the equity in the property.

The lender is also referred to as the mortgage originator. In the U.S., the mortgage originator will often sell off the mortgage to another financial institution, which in many cases will buy a lot of similar mortgages and use such a pool of mortgages as collateral for the issuance of securities. These securities are then referred to as mortgage-backed securities, and the mortgages are said to be securitized. Sometimes the mortgage originator themselves will issue securities backed by a pool of mortgages they have originated.

In the U.S., the business of buying and pooling mortgages and selling liquid mortgage-backed securities has long been dominated by **Fannie Mae** (more formally: the Federal National Mortgage Association) and **Freddie Mac** (the Federal Home Loan Mortgage Corporation). Both have been private corporations with stocks listed on the New York Stock Exchange, but have also been government-sponsored entities whose activities have been controlled by the federal government. The purpose of the entities is to support the mortgage market leading to lower financing rates for borrowers and thus to more widespread home ownership. When Fannie Mae and Freddie Mac issue mortgage-backed securities, they guarantee that the scheduled principal and interest on the underlying loan will be paid even if the borrower defaults. Although these guarantees have not been officially backed by the U.S. federal government, market participants have long been confident that the government would provide the entities with any funds necessary to honour their guarantees. However, the entities have purchased a lot more mortgages than they have been packing and selling off in the form of mortgage-backed securities, so they have kept a large share of the mortgages in their own accounts and have thus effectively acted as lenders to home owners rather than just efficient mortgage repackaging institutions. As a result of the substantial depreciation in house prices, Fannie Mae and Freddie Mac suffered huge losses in 2007 and 2008, and on 7 September 2008 both entities were taken over by the federal government.

Another government-related player is **Ginnie Mae** (the Governmental National Mortgage Association) which is a federal institution and thus explicitly backed by the U.S. government. Ginnie Mae does not originate or purchase mortgages, and it does not issue any securities. Ginnie Mae guarantees the timely payment of principal and interest to investors in mortgage-backed securities backed by loans insured by some other federal institutions such as the Federal Housing Administration (more on mortgage insurance and guarantees below). The Ginnie Mae guaranteed mortgage-backed bonds are issued by certain financial companies approved by Ginnie Mae. These companies can convert a large number of illiquid individual loans into highly liquid and virtually default-free securities by acquiring the Ginnie Mae guarantee for a fee. In the end, the securitization process and the guarantees involved in that process should lead to lower interest rates on the mortgages and thus make home ownership more affordable. However, if house buyers and sellers focus on the (initial) payments on the loan rather than the price of the house, reducing the financing rates will be (at least partly) capitalized in the form of higher house prices so that, in the end, existing home owners benefit more than prospective home owners.

Before we discuss the different characteristics of mortgages, let us introduce some notation. Let time 0 denote the date where the mortgage was originally issued. The scheduled payment dates of the mortgage are denoted by $t_1, t_2, \dots, t_N$, and we assume that the payment dates are equally spaced so that a $\delta > 0$ exists with $t_{i+1} - t_i = \delta$ for all i. Consequently, $t_i = i\delta$. For example, $\delta = 1/12$ reflects monthly payments and $\delta = 1/4$ reflects quarterly payments. We let $D(t)$ denote the outstanding debt at time t (immediately after any payments at time t). In particular, $D(0)$ is the original face value of the mortgage. The scheduled payment of the borrower on the mortgage at any time t_n can be split into three parts:

- an interest payment $I(t_n)$,
- a partial repayment of principal $P(t_n)$,
- a fee $F(t_n)$.

The total scheduled payment at time t_n is thus $Y(t_n) = I(t_n) + P(t_n) + F(t_n)$.

The interest payment is determined as the product of the nominal rate (also known as the mortgage rate or the contract rate) of the loan and the outstanding debt after the previous payment date. The repayment of the original face value of the loan is typically split over several dates into partial repayments as defined by the amortization type of the loan, see below. Clearly, the sum of all the partial repayments must equal the original face value, $D(0) = \sum_{n=1}^{N} P(t_n)$, and we have $D(t_{n+1}) = D(t_n) - P(t_{n+1})$, and $D(t_N) = 0$ since the loan has to be paid off in full.

The fee is intended to cover the costs due to servicing the loan, for example the actual collection of payments from borrowers, preparing information on the fiscal implications of the mortgage, and so on. The servicing of the mortgage may be done directly by the original lending institution or some other institution. Usually the fee to be paid at a given date is some percentage of the outstanding debt. It may be incorporated by increasing the nominal rate of the mortgage, in which case there is really no separate fee payment.

Mortgages are typically long-term (for example 30-year) loans with a prespecified schedule of regular (for example monthly or quarterly) interest rate payments and repayments of the principal. Mortgage loans come in a large variety of forms. Mortgages can be classified based on the following characteristics (partly following Fabozzi (2010)):

Lien status. The seniority of the loan relative to other loans collateralized by the same property. A **first lien** loan has the highest priority if the property is liquidated following a default by the borrower. **Second lien** or **junior lien** loans have lower priority.

Credit classification. If the borrower has a high credit quality (based on the loan-to-value ratio, the payment-to-income ratio, and the credit history of the borrower), the mortgage is a said to be a **prime loan**. If the borrower has a low credit quality or the loan is not a first lien loan, the mortgage is a **subprime loan**. In between the two categories are the so-called **alt-A loans**. These are prime loans according to the standard categorization but, due to lack of documentation or other uncertainty about the creditworthiness of the borrower, the loan is considered riskier than other prime loans.

Interest rate type. The nominal rate of a mortgage is either fixed for the entire maturity of the loan (a **fixed-rate mortgage**) or adjusted according to some clearly specified conditions (an **adjustable-rate mortgage**). The contract rate of an adjustable-rate mortgage is reset at prespecified dates and prespecified terms. The reset is typically done at regular intervals, for example, once a year or once very five years. The contract rate is reset to reflect current market rates so that the new contract rate is linked to some observable interest rates, for example, the yield on a relatively short-term government bond or a money market rate. Some adjustable-rate mortgages come with a cap, that is a maximum on the contract rate, either for the entire term of the mortgage or for some fixed period in the beginning of the term. The adjustable rate can also be fixed for a number of years in the beginning of the mortgage after which it is reset periodically. So-called **teaser loans** have also gained popularity, at least in the years up to the onset of the financial crisis. A teaser loan is an adjustable-rate mortgage loan in which the borrower pays a very low initial interest rate in the first few years after which the interest rate increases dramatically. If the value of the property increases sufficiently, the borrower may refinance before the high interest rate applies.

Amortization type. The amortization principle of the loan determines how the repayment of the face value is distributed over the payment dates of the mortgage. A relatively simple and popular mortgage is a loan where the sum of the interest payment and the principal repayment is the same for all payment dates. The interest payment at any given payment date is the product of a fixed periodic contract rate (the nominal rate on the loan) and the current outstanding debt. This is an **annuity** loan. In the U.S. these mortgages are called **level-payment fixed-rate mortgages**.

Let R denote the fixed periodic contract rate of the mortgage. Usually an annualized nominal rate is specified in the contract and the periodic rate is then given by the annualized rate divided by the number of payment dates per year. Let A denote the constant periodic payment comprising the interest payment and the principal repayment. Using R as the discount rate, the present value of a sequence of N payments equal to A is given by

$$A(1+R)^{-1} + A(1+R)^{-2} + \cdots + A(1+R)^{-N} = A\frac{1-(1+R)^{-N}}{R}.$$

To obtain a present value equal to $D(0)$, the periodic payment must thus be

$$A = D(0)\frac{R}{1-(1+R)^{-N}}.$$

Immediately after the n'th payment date, the remaining cash flow is an annuity with $N-n$ payments, so that the outstanding debt must be

$$D(t_n) = A\frac{1-(1+R)^{-(N-n)}}{R}. \tag{14.1}$$

The part of the payment that is due to interest is

$$I(t_{n+1}) = RD(t_n) = A\left(1-(1+R)^{-(N-n)}\right) = RD(0)\frac{1-(1+R)^{-(N-n)}}{1-(1+R)^{-N}}$$

so that the repayment must be

$$P(t_{n+1}) = A - I(t_{n+1}) = A - A\left(1 - (1+R)^{-(N-n)}\right)$$
$$= A(1+R)^{-(N-n)} = RD(0)\frac{(1+R)^{-(N-n)}}{1-(1+R)^{-N}}.$$

In particular, $P(t_{n+1}) = (1+R)P(t_n)$ so that the periodic repayment increases geometrically over the term of the mortgage.

When the nominal rate is reset on an adjustable-rate annuity mortgage, the remaining payment schedule is recalculated based on the debt and number of payment dates still outstanding and the new nominal rate under the assumption that the nominal rate will remain unchanged until the loan is fully repaid.

Note that the above equations give the *scheduled* cash flow and outstanding debt over the life of the mortgage, but the actual evolution of cash flow and outstanding debt can be different due to delayed or missing payments or to unscheduled prepayments.

In the recent 10–20 years, a substantial product innovation has been seen in the mortgage markets. A variety of mortgages with non-standard amortization schemes and rules for resetting the nominal rate of adjustable-rate mortgages have been introduced. A popular example is the **interest-only mortgage** in which the borrower is not obliged to make any repayment of debt in a certain 'lockout' period starting at the origination of the mortgage and often lasting for 5–10 years. Obviously, compared to a standard annuity loan with the same face value and nominal interest rate, the required payment on the interest-only loan is lower in the lockout period, but higher afterwards as the repayment of debt is distributed over fewer payment dates. An interest-only mortgage can allow young families to buy a bigger and more expensive house right away rather than first buying a smaller and cheaper house and then 'trading up' to a bigger house in some years with the associated transaction costs and inconvenience. The lockout period may allow the borrower to reduce other debts with higher interest rates before reducing the mortgage that typically has a relatively low interest rate. Other mortgages have **negative amortization** in a certain initial period. Not only is there no repayment of debt in this period, the total payment is smaller than what the interest payment should have been, given the nominal rate and the face value. The shortfall in the interest payment is added to the loan balance so that future payments are computed from a larger face value. The lower initial payments imply higher payments later.

Prepayments. Most mortgages come with a prepayment option. At basically any point in time the borrower may choose to make a repayment which is larger than scheduled. The outstanding debt is then decreased correspondingly, so that future interest payments become smaller and the loan is paid off earlier. In particular, the borrower may terminate the mortgage by repaying the total outstanding debt. This happens if the borrower wants to refinance his debt, for example if the current mortgage rates are lower than the interest rate on the existing mortgage. Alternatively, if the current loan-to-value ratio is rather low, the borrower may want to take a new and bigger loan, and thus transform part of the equity in the property into cash and increased consumption. In some countries, including the U.S., mortgages must be prepaid whenever the underlying property is being sold. In other countries, the new owner can take over the existing loan, but will often choose to pay off the existing loan and take out a new loan.

In addition to paying the outstanding debt, a prepaying borrower has to cover some prepayment costs. Typically, the smaller part of these costs can be attributed to the actual repayment of the existing mortgage, while the larger part is really linked to the new mortgage that normally follows a full prepayment, for example application fees, origination fees, credit evaluation charges, and so on. Some of the costs are fixed, while other costs are proportional to the loan amount. The effort required to determine whether or not to prepay and to fill out forms and so on should also be taken into account.

We will discuss motives for prepayments and how to model the prepayment option in the following sections.

Credit guarantees. As discussed above, a mortgage by itself comes with a credit risk, since the borrower may default and in that case the collateral may have a lower market value than the outstanding mortgage debt. In some countries, including the U.S., mortgage loans with a high initial loan-to-value ratio are typically insured or guaranteed by a third party so that if the borrower defaults on a mortgage, the lender or mortgage investor (the current holder of the mortgage, if it has been sold off) is compensated by the insurer. The borrower has to pay for the insurance in form of an initial lump-sum payment, periodic premium payments, or a combination of both. The guarantee is typically cancelled as soon as the loan-to-value ratio falls below a certain threshold, due to repayments of the initial debt and/or due to an increase in the value of the property. In the U.S. there are several private mortgage insurance companies. In addition, two governmental agencies provide guarantees on mortgages to certain groups of borrowers in order to make home ownership more affordable for these groups. The Federal Housing Administration (FHA) guarantees loans to low-income borrowers who can only afford a low down payment, whereas the Veterans Administration (VA) provides guarantees on loans to military veterans and reservists. Of course, if the insurer should go bankrupt, the guarantee becomes worthless, so the lender or mortgage investor should care about the default risk of the insurer. As the FHA and VA are ultimately backed by the U.S. government, their default risk is considered to be negligible. The mortgages guaranteed by FHA and VA are referred to as **government loans**, whereas other mortgages are referred to as **conventional loans**.

Conforming or non-conforming loans. The mortgages that Fannie Mae and Freddie Mac include in the pools backing the mortgage-backed securities issues must be either government loans or so-called **conforming** conventional loans, that is conventional loans meeting certain standards. Mortgages that can be included in these pools normally have lower interest rates than other mortgages, so it is attractive to qualify for a conforming loan.

Points. In the U.S. (and apparently in no other countries), some lenders offer additional flexibility in mortgage choice. For a given loan type of a given maturity, the borrower may choose between different loans characterized by the contract rate and so-called **points**. A mortgage with 2 points mean that the borrower has to pay 2% of the mortgage amount up front. The compensation is that the mortgage rate is lowered. Some lending institutions offer a menu of loans with different combinations of mortgage rates and points. Of course, the higher the points, the lower the mortgage rate. It is even possible to take a loan with negative points, but then the mortgage rate will be higher than the advertised rate associated with zero points.

When choosing between different combinations, the borrower has to consider whether he can afford to make the upfront payment and also the length of the period that he is expected to keep the mortgage, since he will benefit more from the lowered interest rate over long periods. For this reason one can expect a link between the prepayment probability of a mortgage and the number of points paid. LeRoy (1996) constructs a model in which the points serve to separate borrowers with high prepayment probabilities (low or no points and relatively high mortgage coupon rate) from borrowers with low prepayment probabilities (pay points and lower mortgage coupon rate). Stanton and Wallace (1998) provide a related analysis.

14.3 MORTGAGE-BACKED BONDS

In some countries mortgages are often pooled either by the lending institution or other financial institutions, who then issue mortgage-backed securities that have an ownership interest in a specific pool of mortgage loans. A mortgage-backed security is thus a claim to a specified fraction of the cash flows coming from a certain pool of mortgages. Usually the mortgages that are pooled together are very similar, at least in terms of maturity and contract rate. The issuer of the mortgage-backed security will provide some summary statistics about the mortgages in the underlying pool with respect to loan size, geographic location, loan-to-value ratio, exact remaining maturity, coupon rate, and so on.

Mortgage-backed bonds are by far the largest class of securities backed by mortgage payments. Basically, the payments of the borrowers in the pool of mortgages are passed through to the owners of the bonds. Therefore, standard mortgage-backed bonds are also referred to as **pass-through** bonds. Only the interest and principal payments on the mortgages are passed on to the bondholders, not the servicing fees. In particular, if the servicing fee of the borrower is included in the contract rate, this part is filtered out before the interest is passed through to bondholders. Moreover, the costs of issuance of the bonds and so on must be covered. Hence, the coupon rate of the bond will be lower (usually by half a percentage point) than the contract rate on the mortgage. The total nominal amount of the bond issued equals the total principal of all the mortgages in the pool. If the mortgages in the pool are level-payment fixed-rate mortgages with the same term and the same contract rate, then the scheduled payments to the bondholders will correspond to an annuity. There can be a slight timing mismatch of payments, in the sense that the payments that the bond issuer receives from the borrowers at a given due date are paid out to bondholders with a delay of some weeks. Note that the prepayments on the mortgages in the pool are also passed on to bond investors, so when valuing the bond they have to estimate the amount and timing of future prepayments.

Apparently the idea of issuing bonds to finance the construction or purchase of real estate dates back to 1797, where a large part of the Danish capital Copenhagen was destroyed due to a fire creating a sudden need for substantial financing of reconstruction. Currently, well-developed markets for mortgage-backed securities exist in various countries including the U.S., Germany, Denmark, and Sweden. The U.S. market initiated in the 1970s is by now far the largest of these markets. The mortgage-backed bond market is even bigger than the market for U.S. Treasury bonds. Most of the mortgage-backed securities in the U.S. are so-called agency mortgage-backed securities which means that the mortgages included in the pool are conforming loans, that is loans that satisfy the underwriting standards of the mortgage agencies Fannie Mae, Freddie Mac, and Ginnie Mae. Fannie Mae and Freddie Mac themselves issue mortgage-backed securities on a large scale, and as explained above the payments to be made on those securities can be considered virtually free of default risk. Some commercial banks and other financial institutions also issue mortgage-backed bonds. Some of these so-called non-agency mortgage-backed securities are backed by prime mortgages, others by subprime mortgages. The credit quality of these bond issues are rated by the institutions that rate other bond issues such as corporate bonds, see Chapter 13. The largest European market for mortgage-backed bonds is the German market for so-called Pfandbriefe, but relative to GDP the mortgage-backed bond markets in Denmark and Sweden are even larger since in those countries a larger fraction of the mortgages are funded by the issuance of mortgage-backed bonds.

There are some differences in the construction of pass-through bonds across countries. For example, in the U.S., the pass-through bonds are issued at par. In contrast, in Denmark, the annualized coupon rate of pass-through bonds is required to be an integer so that the bond is slightly below par when issued. The purpose of this practice is to form relatively large and liquid bond series instead of many smaller bond series.

14.4 THE PREPAYMENT OPTION

It is, of course, no problem to value the scheduled cash flow of a mortgage or a pass-through mortgage-backed bond using your favorite term structure model. If the payments to the pass-through bond are guaranteed by a government-backed agency, the bond investor does not have to worry about default risk. However, whenever the mortgages have an embedded prepayment option, bond investors have to take the prepayment option into account when valuing the pass-through bonds as the prepayments are passed on from borrowers to bondholders. This is the main challenge in the valuation of standard pass-through mortgage-backed bonds.

In case the borrower decides to prepay the mortgage in the interval $(t_{n-1}, t_n]$ we assume that he has to pay the scheduled payment $Y(t_n)$ for the current period, the outstanding debt $D(t_n)$ after the scheduled mortgage repayment at time t_n, and the associated prepayment costs. Recall that $Y(t_n) = I(t_n) + P(t_n) + F(t_n)$ and $D(t_n) = D(t_{n-1}) - P(t_n)$. Hence, the time t_n payment following a prepayment decision at time $t \in (t_{n-1}, t_n]$ can be written as $Y(t_n) + D(t_n) = D(t_{n-1}) + I(t_n) + F(t_n)$, again with the addition of prepayment costs.

Suppose that Π_{t_n} is the probability that a mortgage is prepaid in the time period $(t_{n-1}, t_n]$ given that it was not prepaid at or before time t_{n-1}. Then the expected repayment at time t_n is

$$\Pi_{t_n} D(t_{n-1}) + (1 - \Pi_{t_n}) P(t_n) = P(t_n) + \Pi_{t_n} D(t_n)$$

and the total expected payment at time t_n is

$$I(t_n) + P(t_n) + \Pi_{t_n} D(t_n) + F(t_n) = Y(t_n) + \Pi_{t_n} D(t_n)$$

plus the expected prepayment costs. If all mortgages in a pool are prepaid with the same probability, but the actual prepayment decisions of individuals are independent of each other, we can also think of Π_{t_n} as the fraction of the pool which (1) was not prepaid at or before t_{n-1} and (2) is prepaid in the time period $(t_{n-1}, t_n]$. This is known as the (periodic) **conditional prepayment rate** of the pool. Some models specify an **instantaneous** conditional prepayment rate also known as a **hazard rate**. Given a hazard rate π_t for each $t \in [0, t_N]$, the periodic conditional prepayment rates can be computed from

$$\Pi_{t_n} = 1 - e^{-\int_{t_{n-1}}^{t_n} \pi_t \, dt} \approx \int_{t_{n-1}}^{t_n} \pi_t \, dt \approx (t_n - t_{n-1}) \pi_{t_n} = \delta \pi_{t_n}. \tag{14.2}$$

Note the close relation to the default probabilities and intensities discussed in Chapter 13.

Since the prepayments of mortgages will affect the cash flow of pass-through bonds, it is important for bond investors to identify the factors determining the prepayment behaviour of borrowers. Below, we list a number of factors that can be assumed to influence the prepayment of individual mortgages and hence the prepayments from a entire pool of mortgages backing a pass-through bond.

Current refinancing rate. When current mortgage rates are below the contract rate of a borrower's mortgage, the borrower may consider prepaying the existing mortgage in full and take a new mortgage at the lower borrowing rate. In the absence of prepayment costs it is optimal to refinance if the current refinancing rate is below the contract rate. Here the relevant refinancing rate is for a mortgage identical to the existing mortgage except for the coupon rate, for example it should have the same time to maturity. This refinancing rate takes into account possible future prepayments.

We can think of the prepayment option as the option to buy a cash flow identical to the remaining scheduled cash flow of the mortgage. This corresponds to the cash flow of a hypothetical non-callable bond, that is an annuity bond in the case of a level-payment fixed-rate mortgage. So the prepayment option is like an American call option on a bond with an exercise price equal to the face value of the bond. It is well-known from option pricing theory that an American option should not be exercised as soon as it moves into the money, but only when it is sufficiently deep in the money. Hence, the present value of the scheduled future payments (the hypothetical non-callable bond) should be sufficiently higher than the outstanding debt (the face value of the hypothetical non-callable bond) before exercise is optimal. Intuitively, this will be the case when current interest rates are sufficiently low. Option pricing models can help quantify the term 'sufficiently low' and hence help explain and predict this type of prepayments. We discuss this in detail in Section 14.5.

Previous refinancing rates. Not only the current refinancing rate, but also the entire history of refinancing rates since origination of the mortgage will affect the prepayment activity in a given pool of mortgages. The current refinancing rate may well be very low relative to the contract rate, but if the refinancing rate was as low or even lower previously, a large part of the mortgages originally in the pool may have been prepaid already. The remaining mortgages are presumably given to borrowers that for some reasons are less likely to prepay. This phenomenon is referred to as **burnout**. On the other hand, if the current refinancing rate is historically low, a lot of prepayments can be expected.

If we want to include the burnout feature in a model, we have to quantify it somehow. One measure of the burnout of a pool at time t is the ratio between the currently outstanding debt in the pool, D_t, and what the outstanding debt would have been in the absence of any prepayments, D_t^*. The latter can be found from an equation like (14.1).

Slope of the yield curve. The borrower should not only consider refinancing the original mortgage with a new, but similar mortgage. He should also consider shifting to alternative mortgages. For example, when the yield curve is steeply upward-sloping a borrower with a long-term fixed-rate mortgage may find it optimal to prepay the existing mortgage and refinance with an adjustable-rate mortgage with a contract rate that is linked to short-term interest rates.

House sales. In some countries, including the U.S., mortgages must be prepaid whenever the underlying property is being sold. In other countries, the new owner can take over the existing loan, but will often choose to pay off the existing loan and take out a new loan. Hence, more prepayments are expected in periods with a high turnover in the housing market. There are seasonal variations in the number of transactions of residential property with more activity in the spring and summer months than in the fall and winter. This is also reflected in the number of prepayments.

Development in house prices. The prepayment activity is likely to increase with the level of house prices. When the market value of the property increases significantly, the owner may want to prepay the existing mortgage and take a new mortgage with a higher principal to replace other debt, to finance other investments, or simply to increase consumption. Conversely, if the market value of the property decreases significantly, the borrower may be more or less trapped. Since the mortgages offered are restricted by the market value of the property, it may not be possible to obtain a new mortgage that is large enough for the proceeds to cover the prepayment of the existing mortgage.

General economic situation of the borrower. A borrower that experiences a significant growth in income may want to sell his current house and buy a larger or better house, or he may just want to use his improved personal finances to eliminate debt. Conversely, a borrower experiencing decreasing income may want to move to a cheaper house, or he may want to refinance his existing house, for example to cut down mortgage payments by extending the term of the mortgage. Also, financially distressed borrowers may be tempted to prepay a loan when the prepayment option is only somewhat in-the-money, although not deep enough according to the optimal exercise strategy. Note, however, that the borrower needs to qualify for a new loan. If he is in financial distress, he may only be able to obtain a new mortgage at a premium rate. This may, at least in part, explain why some mortgages are not prepaid even when the current mortgage rate (for quality borrowers) is way below the contract rate. If the prepayments due to these reasons can be captured by some observable business-cycle related macroeconomic variables, it may be possible to include these in the models for the valuation of mortgage-backed bonds.

Bad advice or lack of knowledge. Most borrowers will not be aware of the finer details of American option models. Hence, they tend to consult professionals. The lending institutions will often issue general recommendations about the choice of mortgage and when borrowers should prepay existing debt. However, since these institutions benefit financially from every prepayment, their recommendations are not necessarily unbiased.

Pool characteristics. The precise composition of mortgages in a pool may be important for the prepayment activity. Other things equal, you can expect more prepayment activity in a pool based on large individual loans than in a pool with many small loans since the fixed part of the prepayment costs are less important for large loans. Also, some pools may have a larger fraction of non-residential (commercial) mortgages than other pools. Non-residential mortgages are often larger and the commercial borrowers may be more active in monitoring the profitability of a mortgage prepayment. In the U.S. there are also regional differences so that some pools are based on mortgages in a specific area or state. To the extent that there are different migration patterns or economic prospects of different regions, potential bond investors should take this into account, if possible.

14.5 RATIONAL PREPAYMENT MODELS

14.5.1 The pure option-based approach

The prepayment option essentially gives the borrower the option to buy the remaining part of the scheduled mortgage payments by paying the outstanding debt plus prepayment costs. This can be interpreted as an American call option on a bond. For a level-payment fixed-rate mortgage, the underlying bond is an annuity bond. A rather obvious strategy for modelling the prepayment behaviour of the borrowers is therefore to specify a dynamic term structure model and find the optimal exercise strategy of an American call according to this model. For a diffusion model of the term structure, the optimal exercise strategy and the present value of the mortgage can be found by solving the associated partial differential equation numerically or by constructing an approximating tree. This approach was first pursued by Dunn and McConnell 1981a, 1981b and Brennan and Schwartz (1985). Note that partial prepayments are not allowed (or not optimal) in this setting.

The prepayment costs affect the effective exercise price of the option. As discussed earlier, a prepayment may involve some fixed costs and some costs proportional to the outstanding debt. As before, we let $D(t)$ denote the outstanding debt at time t. Denote by $X(t) = X(D(t))$ the costs of prepaying at time t. Then the effective exercise price is $D(t) + X(t)$.

The borrower will maximize the value of his prepayment option. This corresponds to minimizing the present value of his mortgage. Let M_t denote the time t value of the mortgage, that is the present value of future mortgage payments using the optimal prepayment strategy. Let us assume a one-factor diffusion model with the short-term interest rate r_t as the state variable. Then $M_t = M(r_t, t)$. Note that r is not the refinancing rate, that is the contract rate for a new mortgage, but clearly lower short rates mean lower refinancing rates.

Suppose the short rate process under the risk-neutral probability measure is

$$dr_t = \hat{\alpha}(r_t)\,dt + \beta(r_t)\,dz_t^{\mathbb{Q}}.$$

Then we know from Section 4.8 that in time intervals without both prepayments and schedule mortgage payments, the mortgage value function $M(r, t)$ must satisfy the partial differential equation (PDE)

$$\frac{\partial M}{\partial t}(r,t) + \hat{\alpha}(r)\frac{\partial M}{\partial r}(r,t) + \frac{1}{2}\beta(r)^2\frac{\partial^2 M}{\partial r^2}(r,t) - rM(r,t) = 0. \tag{14.3}$$

Immediately after the last mortgage payment at time t_N, we have $M(r, t_N) = 0$, which serves as a terminal condition. At any payment date t_n there will be a discrete jump in the mortgage value,

$$M(r, t_n-) = M(r, t_n) + Y(t_n).$$

The standard approach to solving a PDE like (14.3) numerically is the finite difference approach. Below, we provide a short introduction focusing on the role of prepayments. More details and discussions about the method is given in Chapter 16. The finite difference approach is based on a discretization of time and state. For example, the valuation and possible exercise is only considered at time points $t \in \bar{\mathcal{T}} \equiv \{0, \Delta t, 2\Delta t, \dots, \bar{N}\Delta t\}$, where $\bar{N}\Delta t = t_N$. The value space of the short rate is approximated by the finite space $\bar{\mathcal{S}} \equiv \{r_{\min}, r_{\min} + \Delta r, r_{\min} + 2\Delta r, \dots, r_{\max}\}$. Hence we restrict ourselves to combinations of time points and short rates in the grid $\bar{\mathcal{S}} \times \bar{\mathcal{T}}$. For the mortgage considered here, it is helpful to have $t_n \in \bar{\mathcal{T}}$ for all payment dates t_n, which is satisfied whenever the time distance between payment dates, δ, is some multiple of the grid size, Δt. For simplicity, let us assume that these distances are identical so that we only consider prepayment and value the mortgage at the payment dates. As before, we assume that if the borrower at time t_n decides to prepay the mortgage (in full), he still has to pay the scheduled payment

$Y(t_n)$ for the period that has just passed, in addition to the outstanding debt $D(t_n)$ immediately after t_n, and the prepayment costs $X(t_n)$.

The first step in the finite difference approach is to impose

$$M(r, t_N) = 0, \quad r \in \bar{\mathcal{S}},$$

and therefore

$$M(r, t_N-) = Y(t_N), \quad r \in \bar{\mathcal{S}}.$$

Using the finite difference approximation to the PDE, we can move backwards in time, period by period. In each time step we check whether prepayment is optimal for any interest rate level. Suppose we have computed the possible values of the mortgage immediately before time t_{n+1}, that is we know $M(r, t_{n+1}-)$ for all $r \in \bar{\mathcal{S}}$. In order to compute the mortgage values at time t_n, we first use the finite difference approximation to compute the values $M^c(r, t_n)$ if we choose not to prepay at time t_n and make optimal prepayment decisions later. (Superscript 'c' for 'continue'.) Then we check for prepayment. For a given interest rate level $r \in \bar{\mathcal{S}}$, it is optimal to prepay at time t_n, if that leads to a lower mortgage value, that is

$$M^c(r, t_n) > D(t_n) + X(t_n).$$

The corresponding conditional prepayment probability $\Pi_{t_n} \equiv \Pi(r_{t_n}, t_n)$ is

$$\Pi(r, t_n) = \begin{cases} 1 & \text{if } M^c(r, t_n) > D(t_n) + X(t_n), \\ 0 & \text{if } M^c(r, t_n) \leq D(t_n) + X(t_n). \end{cases}$$

The mortgage value at time t_n is

$$\begin{aligned} M(r, t_n) &= \min\left\{M^c(r, t_n), D(t_n) + X(t_n)\right\} \\ &= (1 - \Pi(r, t_n))M^c(r, t_n) + \Pi(r, t_n)(D(t_n) + X(t_n)), \quad r \in \bar{\mathcal{S}}. \end{aligned} \tag{14.4}$$

The value just before time t_n is

$$M(r, t_n-) = M(r, t_n) + Y(t_n), \quad r \in \bar{\mathcal{S}}.$$

Since the mortgage value will be decreasing in the interest rate level, there will be a critical interest rate $r^*(t_n)$ defined by the equality $M^c(r^*(t_n), t_n) = D(t_n) + X(t_n)$ so that prepayment is optimal at time t_n if and only if the interest rate is below the critical level, $r_{t_n} < r^*(t_n)$. Note that $r^*(t_n)$ will depend on the magnitude of the prepayment costs. The higher the costs, the lower the critical rate.

The mortgage-backed bond can be valued at the same time as the mortgage itself. We have to keep in mind that the prepayment decision is made by the borrower and that the bondholders do not receive the prepayment costs. We assume that the entire scheduled payments are passed through to the bondholders, although in practice part of the mortgage payment may be retained by the original lender or the bond issuer. The analysis can easily be adapted to allow for differences in the scheduled payments of the two parties. Let $B(r, t)$ denote the value of the bond at time t when the short rate is r. If the underlying mortgage has not been prepaid, the bond value immediately before the last scheduled payment date is given by

$$B(r, t_N-) = Y(t_N), \quad r \in \bar{\mathcal{S}}.$$

At any previous scheduled payment date t_n, we first compute the continuation values of the bond, that is $B^c(r, t_n)$, $r \in \bar{\mathcal{S}}$, by the finite difference approximation. Then the bond value excluding the payment at t_n is

$$\begin{aligned} B(r, t_n) &= (1 - \Pi(r, t_n))B^c(r, t_n) + \Pi(r, t_n)D(t_n) \\ &= \begin{cases} B^c(r, t_n) & \text{if } M^c(r, t_n) \leq D(t_n) + X(t_n), \\ D(t_n) & \text{if } M^c(r, t_n) > D(t_n) + X(t_n), \end{cases} \end{aligned} \tag{14.5}$$

for any $r \in \bar{\mathcal{S}}$. Then the scheduled payment can be added:

$$B(r, t_n-) = B(r, t_n) + Y(t_n), \quad r \in \bar{\mathcal{S}}.$$

While the discussion above was based on a finite difference approach, readers familiar with tree-approximations to diffusion models will realize that a similar backward iterative valuation technique applies in an interest rate tree approximating the assumed interest rate process. For an introduction to the construction of interest rate trees, see Chapter 16.

In the Danish mortgage financing system, mortgages come with an additional option feature if they are part of a pool upon which pass-through bonds have been issued. Not only can the borrower prepay the outstanding debt (plus costs) in cash, he can also prepay by buying pass-through bonds (based on that particular pool) with a total face value equal to the outstanding debt of his mortgage. These bonds have to be delivered to the mortgage lender, which in the Danish system is also the bond issuer. This additional option will be valuable if the borrower wants to prepay (for 'suboptimal' reasons) in a situation where the market price of the bond is below the outstanding debt, which is the case for sufficiently high market interest rates. If we assume that a prepayment by bond purchase generates costs of $\hat{X}(t_n)$, such a prepayment is preferred to a cash prepayment whenever $B(r, t_n) + \hat{X}(t_n) < D(t_n) + X(t_n)$. This should be accounted for in the backwards iterative procedure.

In practice, the borrower may have to notify the lender some time before the prepayment will be effective. If the borrower wants the time t_n mortgage to be the last payment date on the mortgage, the notification may be due at time $t_n - h$ for some fixed time period $h > 0$. Then the above equations have to be modified slightly. Let us assume that $t_n - h > t_{n-1}$. It will be optimal to decide to prepay at the notification date $t_n - h$, if

$$M^c(r, t_n - h) > (D(t_n) + X(t_n) + Y(t_n))\, B^{t_n}(r, t_n - h).$$

Here the right-hand side is the value at time $t_n - h$ of the total payment at time t_n if the borrower decides to prepay. The left-hand side is the value of the mortgage at time $t_n - h$ if the borrower decides not to prepay now and makes optimal prepayment decisions in the future. This value includes the present value of the upcoming scheduled payment $Y(t_n)$. The bond values can be modified similarly.

According to the above analysis, all borrowers with identical mortgages and identical prepayment costs should prepay in the same states and same points in time, essentially when the refinancing rate is sufficiently low. The conditional prepayment rate will be one if $r_{t_n} < r^*(t_n)$ and zero otherwise. In practice, simultaneous prepayments of all mortgages in a given pool is never observed. One potential explanation is that the mortgages in a given pool are not completely identical and hence may not have the same critical interest rate. Another explanation is that borrowers prepay for other reasons than just low refinancing rates, as discussed in Section 14.4. In the following subsections we discuss how these features can be incorporated into the option-based approach.

14.5.2 Heterogeneity

The mortgages in a pool are never completely identical. In particular the prepayment costs may be different for different mortgages. To study the implications of different costs, we assume that all the mortgages have the same contract rate and the same term so that the stream of scheduled payments (relative to the outstanding debt of the mortgage) is the same for all mortgages. Suppose that the pool can be divided into M sub-pools so that at any point in time all the mortgages in a given sub-pool have identical prepayment behaviour. If the prepayment costs on any individual mortgage can be assumed to be some fixed fraction of the outstanding debt, then we must divide the pool according to the value of this fraction. All mortgages for which the costs are a fraction x_m of the outstanding debt is put into sub-pool number m. In this case the mortgages in a given sub-pool may have different face values. On the other hand, if there is also a fixed cost element of the prepayment costs, we have to divide the mortgages according to the face value of the loans. In any case let us assume that the prepayment costs for mortgages in sub-pool m are given by $X_m(t_n)$.

Note that it may be difficult to obtain the information that is necessary to implement such a categorization of the individual mortgages in a pool, but in some countries at least some useful summary statistics are published for each mortgage pool. In order to estimate the cost parameters, we need data on observed prepayments for each sub-pool. If the mortgages are newly issued, it will be necessary to use prepayment data on similar, but more mature mortgages. In order to avoid the separate estimation of cost parameters for each sub-pool, one can assume that the variations in prepayment costs across the mortgages in the pool can be described by a distribution involving only one or two parameters that may then be estimated from actual prepayment behaviour. For example, Stanton (1995) assumes that prepayment costs on each mortgage is some constant proportion of the outstanding debt, but the magnitude of this constant varies across mortgages according to a so-called beta distribution on the interval $[0, 1]$. The beta distribution is completely determined by two parameters. For implementation purposes, the distribution is approximated by a discrete distribution with M possible values $x_1, \dots, x_M$ given by certain quantiles of the full distribution.

Using the approach of the previous subsection, we can derive a critical interest rate boundary $r^*_m(t_n)$ for each sub-pool. Note that if $X_m(t_n)$ is sufficiently high, it may be suboptimal to prepay the mortgage at time t_n no matter what the interest rate will be. In that case $r^*_m(t_n)$ must be set below the minimum possible interest rate.

How do we value a pass-through bond backed by such a heterogenous pool of mortgages? We can think of a pass-through bond backed by the entire pool as a portfolio of hypothetical sub-pool specific pass-through bonds. The hypothetical bond for any sub-pool m can be valued exactly as discussed in the previous subsection, acknowledging the optimal prepayments of mortgages in that sub-pool. Let $B_m(r,t)$ be the time t price of that bond (normalized to a given face value, say 100) for a short rate of r. Suppose that w_{mt} denotes the fraction of the pool that belongs to sub-pool m at time t. By definition, $\sum_{m=1}^{M} w_{mt} = 1$. The value of the bond backed by the entire pool is then a weighted average of the values of the hypothetical sub-pool bonds:

$$B(r,t) = \sum_{i=1}^{M} w_{mt} B_m(r,t).$$

It is important to realize that the sub-pool weights w_{mt} vary over time, depending on the evolution of interest rates. The sub-pool weights at a given point in time depend on the entire history of interest rates since the original issuance of the bonds.

While it is certainly an improvement of the model to allow for heterogeneity in the prepayment costs, it is still not consistent with empirically observed prepayment behaviour. According to the model, all mortgages in the same sub-pool should be prepaid simultaneously. The first time the interest rate drops to the critical level for a given prepayment cost, all the mortgages in the sub-pool will be prepaid immediately—and all mortgages with lower prepayment costs have already been prepaid. If the interest rate then rises and drops to the same critical level, no prepayments will take place. The simultaneous exercise of a large number of mortgages will generate large, sudden moves in the bond price, when the interest rate hits a critical level for some sub-pool. This is not observed in practice.

14.5.3 Allowing for seemingly irrational prepayments

The pure option-based approach described above can by construction only generate rational prepayments, which simply means prepayments caused by the fact that the prepayment option is deep enough in the money to warrant early exercise. As discussed extensively in Section 14.4, borrowers may prepay for other reasons. Several authors have suggested minor modifications of the option-based approach in order to incorporate the seemingly irrational prepayments in a simple manner.

Dunn and McConnell (1981a, 1981b) assume that for each mortgage a suboptimal prepayment can be described by a hazard rate λ_t. Given that the mortgage has not been prepaid at time t_{n-1}, there is a probability of

$$\Pi^e_{t_n} \equiv 1 - e^{-\int_{t_{n-1}}^{t_n} \lambda_t \, dt} \approx \int_{t_{n-1}}^{t_n} \lambda_t \, dt \approx (t_n - t_{n-1})\lambda_{t_n} = \delta \lambda_{t_n} \tag{14.6}$$

that the borrower will prepay the mortgage in the interval $(t_{n-1}, t_n]$ for 'exogenous' reasons, that is whether or not it is optimal from an interest rate perspective. This can easily be included in the option-based approach as long as the hazard rate λ_t at most depends on time and the current interest rate, that is $\lambda_t = \lambda(r_t, t)$. If λ_t is the same for all mortgages in a reasonably large pool (or sub-pool), this implies that a fraction of $\lambda_t \Delta t$ of the mortgages can be expected to be prepaid over a Δt period in any case. This introduces a minimum level of prepayment activity.

Stanton (1995) adds a second source of suboptimality. He assumes that borrowers will not constantly evaluate whether a prepayment is advantageous or not. If prepayment is considered according to a hazard rate η_t, then the probability that the borrower will check for optimal prepayment over $(t_{n-1}, t_n]$ is

$$1 - e^{-\int_{t_{n-1}}^{t_n} \eta_t \, dt} \approx \int_{t_{n-1}}^{t_n} \eta_t \, dt \approx (t_n - t_{n-1})\eta_{t_n} = \delta \eta_{t_n}.$$

Continuous prepayment evaluation corresponds to $\eta_t = \infty$. The non-continuous decision-making may reflect the costs and difficulties of considering whether prepayment is optimal or not. Again, for tractability, the hazard rate η_t is assumed to depend at most on time and the interest rate level, $\eta_t = \eta(r_t, t)$. We can interpret η_t as the (expected) fraction of the pool (or sub-pool) for which the optimal prepayment rule applies. With this modification, not all mortgages will be prepaid even though prepayment is optimal from an interest rate perspective.

Combining these two modifications, the probability that a mortgage is not prepaid in a time interval $[t, t + \Delta t]$ even though it is optimal must be

$$\begin{aligned} \text{Prob(no prepayment)} &= \text{Prob((no optimal prepayment) AND} \\ &\quad \text{(no suboptimal prepayment))} \\ &= \text{Prob}\left(\text{no optimal prepayment}\right) \\ &\quad \times \text{Prob}\left(\text{no suboptimal prepayment}\right) \\ &= e^{-\eta_t \Delta t} e^{-\lambda_t \Delta t} \\ &= e^{-(\eta_t + \lambda_t)\Delta t}. \end{aligned}$$

Hence, the probability that a prepayment takes place in $(t_{n-1}, t_n]$, given that it is optimal,

$$\Pi^r_{t_n} = 1 - e^{-\int_{t_{n-1}}^{t_n} (\eta_t + \lambda_t) \, dt} \approx \int_{t_{n-1}}^{t_n} (\eta_t + \lambda_t) \, dt \approx \delta(\eta_{t_n} + \lambda_{t_n}). \tag{14.7}$$

Note that $\Pi^r_{t_n} \geq \Pi^e_{t_n}$ and also note that $\Pi^r_{t_n} \to 1$ for $\eta_t \to \infty$.

If we assume that the hazard rates η_t and λ_t are functions of at most time and the short rate, and we use the right-most approximations in Equations (14.6) and (14.7)—which would be natural approximations in an implementation—the periodic conditional prepayment rate Π_{t_n} over the period $(t_{n-1}, t_n]$ will be a function of t_n and r_{t_n}. It will be Π^e when prepayment is suboptimal and Π^r when prepayment is optimal, that is

$$\Pi(r, t_n) = \begin{cases} \Pi^e(r, t_n) & \text{if } M^c(r, t_n) \leq D(t_n) + X(t_n), \\ \Pi^r(r, t_n) & \text{if } M^c(r, t_n) > D(t_n) + X(t_n). \end{cases}$$

In the backwards valuation iterative procedure, we replace Equation (14.4) by

$$
\begin{aligned}
M(r,t_n) &= (1-\Pi(r,t_n))M^c(r,t_n) + \Pi(r,t_n)(D(t_n)+X(t_n)) \\
&= \begin{cases} (1-\Pi^e(r,t_n))M^c(r,t_n) + \Pi^e(r,t_n)(D(t_n)+X(t_n)), \\ \qquad\qquad\qquad\qquad \text{if } M^c(r,t_n) \le D(t_n)+X(t_n), \\ (1-\Pi^r(r,t_n))M^c(r,t_n) + \Pi^r(r,t_n)(D(t_n)+X(t_n)), \\ \qquad\qquad\qquad\qquad \text{if } M^c(r,t_n) > D(t_n)+X(t_n). \end{cases}
\end{aligned}
$$

There will still be a critical interest rate level $r^*(t_n)$ that gives the maximum interest rate for which it is optimal to prepay at time t_n, but it will be different to that in the pure option-based approach since the continuation value takes into account the possibility for making suboptimal prepayments or missing optimal prepayments in the future. Similarly, Equation (14.5) in the valuation of the bond has to be replaced by

$$
\begin{aligned}
B(r,t_n) &= (1-\Pi(r,t_n))B^c(r,t_n) + \Pi(r,t_n)D(t_n) \\
&= \begin{cases} (1-\Pi^e(r,t_n))B^c(r,t_n) + \Pi^e(r,t_n)D(t_n), \\ \qquad\qquad\qquad \text{if } M^c(r,t_n) < D(t_n)+X(t_n), \\ (1-\Pi^r(r,t_n))B^c(r,t_n) + \Pi^r(r,t_n)D(t_n), \\ \qquad\qquad\qquad \text{if } M^c(r,t_n) \ge D(t_n)+X(t_n). \end{cases}
\end{aligned}
$$

The other steps in the iterative valuation procedure are unaltered and it is still possible to divide mortgages into sub-pools.

The model of Stanton (1995) incorporates both heterogeneity, suboptimal prepayments, and non-continuous decision-making. He implements the model assuming that the hazard rates λ_t and η_t are constant. He estimates the values of the various parameters so that the prepayment rates predicted by the model come as close as possible to observed prepayment rates in a given sample. The estimated prepayment cost distribution has an average cost equal to 41% of the outstanding debt, which is very high, even if we take into account the implicit costs that may be associated with prepayments (time spent, and so on). The estimate for the hazard rate η implies that the average time between two successive checks for optimal prepayment (which is $1/\eta$) is 8 months, which seems to be an unreasonably long period. The estimate for λ implies that approximately 3.4% of mortgages are prepaid in a given year for exogenous reasons. Using the estimated model, Stanton predicts future prepayment rates of a given pool and compares the predictions with the realized prepayment rates. The predictions of the model are reasonably accurate and slightly better than the predictions from a purely empirical prepayment model suggested by Schwartz and Torous (1989), which we will study below.

14.5.4 An example

Consider a 30-year mortgage with quarterly payments. The mortgage is an annuity loan in the sense that the sum of the scheduled interest payment and scheduled repayment is the same for all 120 payment dates. The mortgage is issued by a

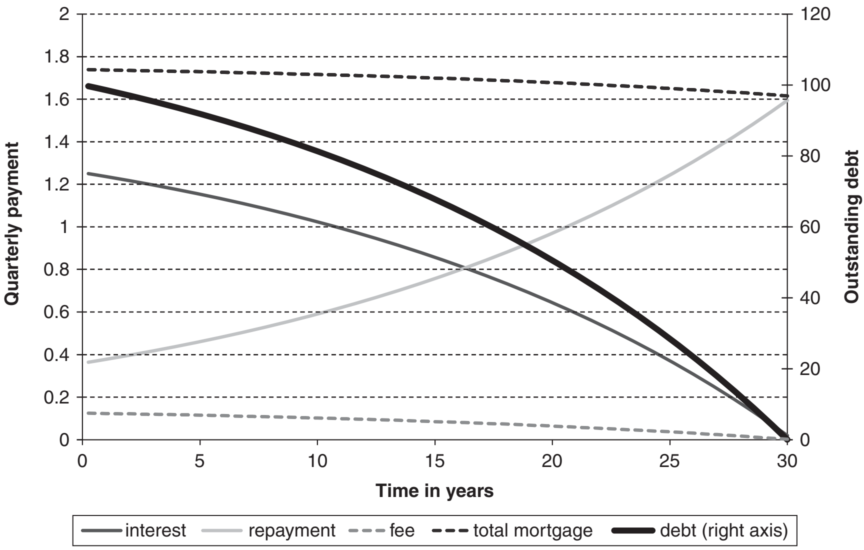

Fig. 14.1: The scheduled payments and the outstanding debt over the life of the mortgage described in the example in Section 14.5.4.

financial institution and financed by the issuance of a 'pass-through' mortgage-backed bond. For illustration purposes, we assume that the bond is backed by a single mortgage (or, equivalently, a number of identical mortgages). The annualized coupon rate is 5% and the face value is set at 100. At each of the payment dates, the borrower must also pay a fee to the financial institution. The fee equals 0.125% of the outstanding debt after the previous payment date (corresponding to an extra 0.5% on an annual basis). This fee is not passed through to bond investors. Figure 14.1 depicts the various scheduled payments and the remaining debt over the life of the mortgage (you are asked to compute these in Exercise 14.1).

The mortgage borrower has the right to prepay the loan at any payment date. In order to value the mortgage and the bond and to find the optimal prepayment strategy, we have to make an assumption about the interest rate dynamics. Suppose the CIR one-factor model is true, that is

$$dr_t = \kappa[\theta - r_t]\,dt + \beta\sqrt{r_t}\,dz_t, \qquad \lambda(r,t) = \lambda\sqrt{r}/\beta,$$

and assume that $\kappa = 0.3$, $\theta = 0.045$, $\beta = 0.15$, and $\lambda = -0.0717$. In particular, the asymptotic long-term yield y_∞ then equals 5%. The mortgage and the bond can now be valued by solving the relevant partial differential equation numerically using the implicit finite difference technique explained in Section 16.2. Assume that the prepayment costs are a fixed percentage of the outstanding debt, that is $X(t_n) = kD(t_n)$ where $D(t_n)$ is the outstanding debt immediately after t_n.

Figure 14.2 shows the current value of the mortgage (that is the present value of future payments) and the current value of the bond as a function of the current short-term interest rate for three different values of the cost parameter k. The value of the loans is increasing in the cost parameter since higher costs imply higher

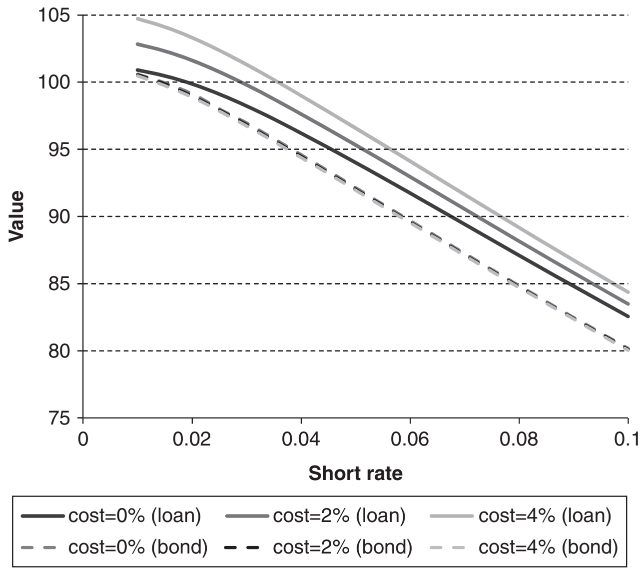

Fig. 14.2: The value of the mortgage and the mortgage-backed bond as a function of the short-term interest rate in the example in Section 14.5.4. Prepayment costs are assumed to be a percentage of the outstanding debt. Only optimal prepayments are included. The values have been computed with the implicit finite difference method.

payments in case of a prepayment. The largest effect of the costs is for low values of the current short rate where prepayment is likely to happen soon, but even for a very high current short rate the cost parameter is important for the loan value because there is still a high chance that prepayment will happen at some point during the 30-year life of the mortgage. The bond value is much less sensitive to the cost parameter, since the bondholders do not receive the prepayment costs. The effect of prepayment costs on the bond price comes only through the effect on the prepayment strategy. Figure 14.3 illustrates the critical short rate of the life of the mortgage for the three different levels of the cost parameter. The higher the cost parameter, the lower the critical short rate. When it is costly to exercise the prepayment option, the payoff from exercising has to be higher, that is the refinancing rate has to be lower.

Next, assume that there is an exogenously given prepayment probability Π^e, but prepayments can be implemented at zero cost. Figure 14.4 shows the value of the loan and the bond for exogenous prepayment probabilities of 0%, 5%, and 10%. The difference between the loan value and the bond value stems from the mortgage fees that are paid by the borrower but not received by the bondholder. The exogenous prepayment probability has a significant impact on the values, in particular for high interest rates since an exogenous prepayment implies that the relatively low present value of the future scheduled payments is replaced by the higher outstanding debt. When the borrower realizes that there is a chance that he will prepay suboptimally in the future, he will choose to exercise the prepayment option earlier, that is at higher values of the short rate. This is illustrated in Fig. 14.5.

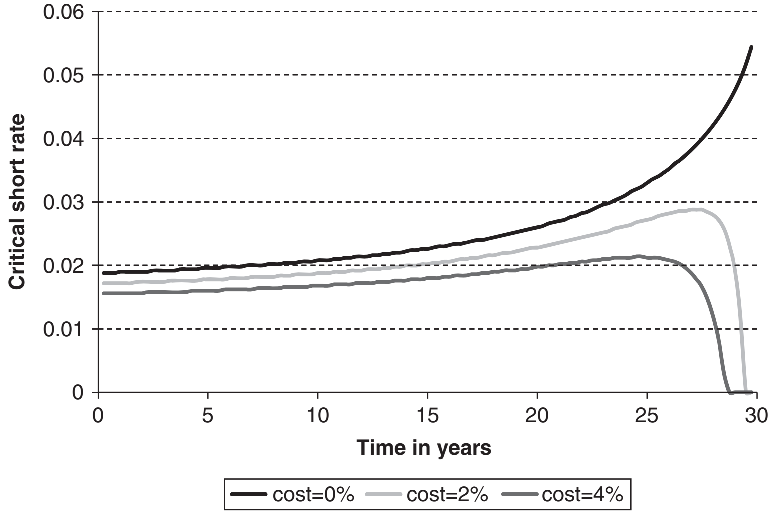

Fig. 14.3: The critical short rate of the mortgage as a function of time in the example in Section 14.5.4. Prepayment costs are assumed to be a percentage of the outstanding debt. Only optimal prepayments are included. The mortgage is optimally prepaid the first time the short rate crosses the curve from above. The values have been computed with the implicit finite difference method.

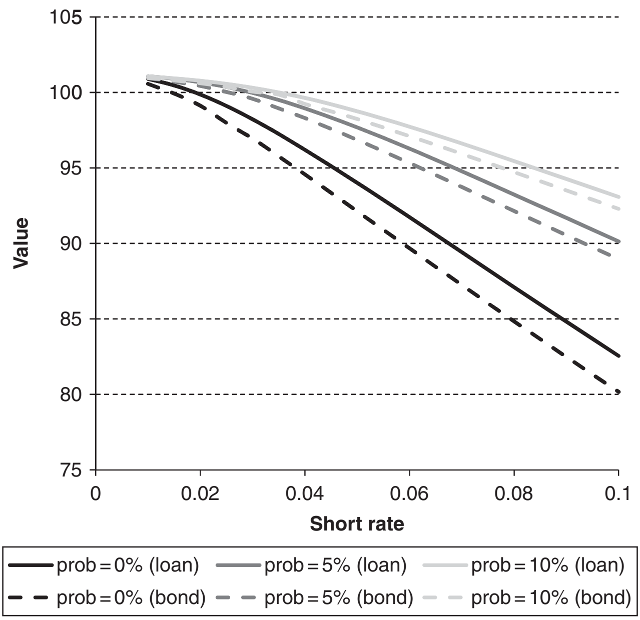

Fig. 14.4: The value of the mortgage and the mortgage-backed bond as a function of the short-term interest rate in the example in Section 14.5.4. For every payment date there is a certain probability that the mortgage is prepaid even though prepayment is suboptimal from an option-theoretic perspective. The mortgage is prepaid for sure whenever it is theoretically optimal do so. Prepayments can be implemented at zero costs. The values have been computed with the implicit finite difference method.

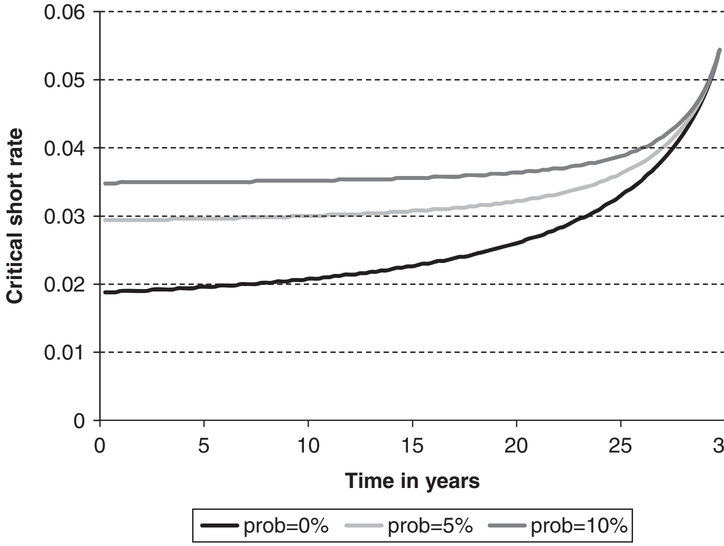

Fig. 14.5: The critical short rate of the mortgage as a function of time in the example in Section 14.5.4. For every payment date there is a certain probability that the mortgage is prepaid even though prepayment is suboptimal from an option-theoretic perspective. The mortgage is prepaid for sure whenever it is theoretically optimal do so. Prepayments can be implemented at zero costs. The mortgage is optimally prepaid the first time the short rate crosses the curve from above. The values have been computed with the implicit finite difference method.

14.5.5 The option to default

The borrower may also have an option to default on the mortgage and simply walk away. It may be optimal for the borrower to default if the market value of the mortgage exceeds the market value of the house. The downside is that the default will significantly impair the credit score of the borrower making it harder to obtain credit in the future. In order to include optimal defaults into the model, it is necessary to include some measure of house prices as another state variable. Both Kau et al. (1992) and Deng et al. (2000) find that it is important to consider the prepayment option and the default option simultaneously. According to Deng et al. (2000), the inclusion of the default option is helpful in explaining empirical behaviour. The investors in mortgage-backed securities only have to worry about mortgage defaults if the payments to security holders are not insured by a financially solid guarantor.

14.5.6 Other rational models

The option-pricing approach described above is focused on minimizing the value of the current mortgage. In most cases the prepayment of a mortgage is immediately followed by a new mortgage. In the presence of prepayment costs, it is really not reasonable to look separately at one mortgage. The timing of the prepayment decision for the current mortgage influences the contract rate of the next mortgage and, hence, the potential profitability of prepayments of that mortgage. Borrowers should have a life-time perspective and minimize the life-time mortgage costs, for example they probably want to make relatively few prepayments over their life in order to reduce the total prepayment costs. The life-cycle perspective is advocated by, for example, Stanton and Wallace (1998) and Longstaff (2005).

The mortgage prepayment decision is only one of many financial decisions in the life of any borrower. The borrower has to decide on investments in financial assets, transactions of property and other durable consumption goods, and so on. For a rational individual all these decisions are taken in order to maximize the expected life-time utility from consumption of various goods, both perishable goods (food, entertainment, and so on) and durable goods (house, car, and so on). The prepayment decisions of a borrower are not taken independently of other financial decisions. Hence, it could be useful to build a model incorporating the prepayment decisions into an optimal consumption-portfolio framework. This might give a better picture of the rational prepayments of an individual mortgage, for example it has the potential to point out in which economic scenarios the borrower will choose to default on the mortgage or will prepay for liquidity reasons, and so on. Such models could also address the choice of mortgage, for example who should prefer fixed-rate mortgages and who should prefer adjustable-rate mortgages. The models of Campbell and Cocco (2003) and Van Hemert (2009) are good examples. Any such model is bound to be quite complex, though.

For further applications and discussions of the option-theoretic approach to prepayments, see, for example, Kau et al. (1995), Azevedo-Pereira et al. (2003), Longstaff (2005), and Sharp et al. (2009).

14.6 EMPIRICAL PREPAYMENT MODELS

Historical records of prepayments clearly show that actual mortgage prepayments cannot be fully explained by the basic American option pricing models. This observation lead Green and Shoven (1986) and Schwartz and Torous 1989, 1992 to suggest a purely empirical model for prepayment behaviour. The conditional prepayment rate for a mortgage pool is assumed to be a function of some explanatory state variables that have to be specified, that is $\Pi_t = \Pi(t, \boldsymbol{v}_t)$ where $\boldsymbol{v}$ is a vector of explanatory variables. Section 14.4 offers a list of relevant candidates for explanatory variables. Given the history of the explanatory variables, the parameters of the function are determined so that it comes as close as possible to the historic prepayment rates for the pool (or a similar pool). Then the function with the estimated parameters is used to predict future prepayment rates contingent on the future values of the explanatory variables. This will determine the state-dependent cash flow of the mortgage-backed bonds. This cash flow can then be valued by standard valuation methods.

At least some of the explanatory variables will evolve stochastically over the life of the mortgage. In order to do the valuation we have to make some assumptions about the stochastic dynamics of these variables. For any reasonable empirical model the actual valuation has to be done using one of our standard numerical techniques, that is by constructing a tree or finite-difference based lattice or by performing Monte Carlo simulations, see Chapter 16. Below we will discuss how the valuation technique to be used may depend on the explanatory variables included. Regardless of which of these techniques we want to implement, we will have to discretize the time span so that we only look at time points $t \in \bar{\mathcal{T}} \equiv \{0, \Delta t, 2\Delta t, \dots, \bar{N}\Delta t\}$, where $\bar{N}\Delta t = t_N$ is the final payment date.

The original suggestion of Schwartz and Torous (1989) was to model the prepayment hazard rate as

$$\pi(t, \boldsymbol{v}_t) = \pi_0(t) e^{\boldsymbol{\theta}^{\top} \boldsymbol{v}_t},$$

where $\boldsymbol{\theta}$ is a vector of parameters and $\pi_0(t)$ is the deterministic function

$$\pi_0(t) = \frac{\gamma p (\gamma t)^{p-1}}{1 + (\gamma t)^p}$$

giving a 'base-line' prepayment rate. This function has $\pi_0(0) = 0$ and $\lim_{t\to\infty} \pi_0(t) = 0$, and it is increasing from $t = 0$ to $t = (p-1)^{1/p}/\gamma$ after which it decreases. This is consistent with the empirical observation that conditional prepayment rates tend to be very low for new and for old mortgages and higher for intermediate mortgage ages. From the prepayment hazard rate, the conditional prepayment rate over any period can be derived as in Equation (14.2). The explanatory variables chosen by Schwartz and Torous are the following:

1. the difference between the coupon rate and the current long-term interest rate (slightly lagged), reflecting the gain from refinancing,
2. the same difference raised to the power 3, reflecting a non-linearity in the relation between the potential interest rate savings and the prepayment rate,
3. the degree of burn-out measured by (the log of) the ratio between the currently outstanding debt in the pool and what the outstanding debt would have been in the absence of any prepayments,
4. a seasonal dummy, reflecting that more real estate transactions—with associated prepayments—take place in spring and summer than in winter and fall.

In the sample considered by Schwartz and Torous this prepayment function comes reasonably close to the observed prepayment rates. Note however that this is an in-sample comparison in the sense that the parameters of the function have been estimated on the basis of the same data sample. The real test of the prepayment function is to what extent it can predict prepayment behaviour *after* the estimation period.

Many recent empirical prepayment models are based on a periodic conditional prepayment rate of the form

$$\Pi(t, \boldsymbol{v}_t) = N(f(\boldsymbol{v}_t; \boldsymbol{\theta})),$$

where $N(\cdot)$ is the cumulative standard normal distribution function and f is some function to be specified. A very simple example is to let

$$\Pi(t, g_t) = N(\theta_0 + \theta_1 g_t),$$

where θ_0 and θ_1 are constants and g_t is a measure for the present value gain from a prepayment at time t.

If a heterogeneous pool of mortgages can be divided into M sub-pools of homogeneous mortgages, it may be worthwhile specifying sub-pool specific prepayment functions, $\Pi_m(t, \boldsymbol{v}_t)$. Then the conditional prepayment rate for the entire pool is a weighted average of the sub-pool prepayment functions,

$$\Pi(t, \boldsymbol{v}_t) = \sum_{m=1}^{M} w_{mt} \Pi_m(t, \boldsymbol{v}_t),$$

where w_{mt} is the relative weight of sub-pool m at time t.

It is important to realize that some of the potential explanatory variables are forward-looking and others are backward-looking and it will be difficult to include variables of both types in the same valuation model. For example, any reasonable measure of the monetary gain from a prepayment is forward-looking since it includes the present value of future payments. Such variables cannot be handled easily in a Monte Carlo based valuation technique, but is better suited for the backwards iterative procedure in lattices and trees. On the other hand the burnout factor is inherently a backward-looking variable since it depends on the previous prepayment activity, for example the path taken by interest rates. This is also true for the relative weighting of different sub-pools. Such variables are difficult to handle in a backward iterative scheme, but better suited for Monte Carlo simulation. However, as mentioned earlier, the burnout factor can be approximated by the ratio of current outstanding debt to the total outstanding debt if no prepayments have taken place so far. Since the denominator in this ratio is deterministic, the burnout factor may be captured by including the currently outstanding debt as a state variable in a tree- or lattice-based valuation.

Clearly, the main limitation of the purely empirical prepayment models is that the prepayment behaviour in the future might be very different from the prepayment behaviour in the estimation period. If the underlying economic environment changes and this is not captured by the explanatory variables included, the empirical prepayment models might generate very poor predictions of future prepayment rates.

Various empirical prepayment models have been developed by sellers and investors in mortgage-backed securities and by independent financial research companies. A widely used model is the PSA standard prepayment model introduced by the Public Securities Association which is now known as the Securities Industry and Financial Market Association (SIFMA).

In the recent literature, some papers have directly exploited the clear link between CMOs (see below) and credit derivatives and suggested reduced-form models for the valuation of CMOs closely related to the reduced-form credit models discussed in Section 13.4. A prepayment intensity process is modelled either as an independent process or as a function of other observable processes, for example, the short-term interest rate on government bonds. For some specifications of the relevant processes, the valuation of the mortgages and associated pass-through bonds becomes analytically tractable, see, for example Gorovoy and Linetsky (2007) and Liao et al. (2008). See also Exercise 14.3 at the end of the chapter.

14.7 RISK MEASURES FOR MORTGAGE-BACKED BONDS

As in Chapter 12 we can define risk measures for mortgage-backed bonds in terms of the derivatives of the value of the bond with respect to the relevant state variables, that is the factors driving the yield curve of government bonds and any mortgage-specific factors driving the prepayment activity and, possibly, the credit risk. As no closed-form expressions for the value of a mortgage-backed bond are available, risk measures cannot be stated in closed form either.

It is possible, however, to reach some conclusions regarding the impact of the prepayment option on the sensitivity of the bond price with respect to the interest rate level. The prepayment activity is much higher when the current market interest rates are low than when they are high (compared to the contract rate on the mortgage). As the interest rate level falls, prices of bonds with no prepayment option will increase in a convex way. The bond price can easily go (far) above its face value. In contrast, for mortgage-backed bonds the upside potential is limited. As interest rates fall, more and more borrowers will find it advantageous to prepay their mortgage to refinance at a lower rate. The prepayment option has to be sufficiently in-the-money before exercise is optimal, as is always the case for American-style options and also because of the prepayment costs. Hence, the market price of the bond cannot increase far above the face value. When market interest rates are lower than some threshold (which has to be computed numerically along with the bond valuation), the bond value will be a concave function of the interest rate level rather than a convex function, that is the mortgage-backed bond has negative convexity for low interest rates. For high interest rates a mortgage-backed bond behaves much like a government bond as the prepayment activity is then typically very low and the prepayment option is less valuable. The local concavity of the price of the mortgage-backed bond can be seen in Figs 14.2 and 14.4 under the assumptions of our example in Section 14.5.4 (for medium and high interest rates the bond price may appear to be a linear function, but it is in fact slightly convex).

14.8 OTHER MORTGAGE-BACKED SECURITIES

The pass-through mortgage-backed bonds represent the simplest way to distribute the payments of mortgage borrowers to a group of investors in the sense that all investors receive identical payments. With the so-called Collateralized Mortgage Obligations (CMOs) different classes or tranches of bonds are issued with reference to the same pool of mortgages. Bonds within each tranche receive identical payments, but bonds from different tranches do not. Bonds in different tranches have different risk–return characteristics and can thus attract different types of investors. In the U.S., CMOs are issued both by the government agencies and by other players in the mortgage industry.[1] For the agency CMOs the default risk is negligible, and the CMO construction mainly has to define how the interest payments and principal repayments (including prepayments) are distributed to different tranches. For uninsured non-agency CMOs the default risk is a key factor, and the CMO construction also has to specify how the different tranches are affected by mortgage defaults. This is especially important for the subprime CMOs for which the underlying mortgages are subprime loans. Note the similarity between the CMOs and the CDOs described in Section 13.9.

An example of a CMO structure is a sequential-pay CMO. The sum of the principals of the tranches equals the total principal of the mortgage pool. Each of the tranches have a coupon rate. The tranche coupon rates can differ from the interest rate on the underlying mortgages and may vary across tranches. Given the total interest payments and debt repayments on the mortgage pool in a given period, each of the tranches is allocated the promised interest payment computed from the tranche coupon rate and outstanding debt. The remaining mortgage payments are then used for principal payments on the tranches. Here the sequential-pay construction means that initially only the first tranche is allocated principal payments, whereas the other tranches receive no principal payments, until the full tranche 1 principal has been repaid and the tranche 1 bonds expires. After that, only tranche 2 bonds receive principal payments until they are fully repaid. Then tranche 3 takes over and so on. Because of prepayments on the underlying mortgages, there is uncertainty about the precise schedule of the principal repayments on the different tranches (and thus also on the future outstanding debt and the interest rate payments). Prepayment forecasting is therefore very important in the valuation of the CMO tranches.

A CMO can be structured so that a tranche is only entitled to interest payments or to principal payments. A closely related product is the so-called agency *stripped* mortgage-backed securities. Here, two bond classes are issued with reference to the same mortgage pool, the IO (interest only) class and the PO (principal only) class. All the interest payments on the mortgage pool are distributed to the IO bonds, and all the principal payments (scheduled payments as well as prepayments) are distributed to the PO bonds. Obviously, the cash flow and return to PO bond investors will depend heavily on the prepayment activity in the mortgage pool. Also, the IO bonds are highly sensitive to prepayments since the IO bondholders only receive interest on the outstanding debt, which is reduced following prepayments.

When non-agency CMOs are constructed, a key issue is the credit risk of the different tranches. Exactly as for the CDOs described in Section 13.9, there will be senior tranches with a fairly low credit risk and junior or equity tranches with a very high credit risk. The different tranches are rated by rating agencies that usually assist the issuer in the CMO construction to obtain some desired ratings, for example to ensure that the senior tranche receives the best possible rating. To reduce credit risk, subprime CMOs are commonly overcollateralized in the sense that the total face value of the underlying subprime loans exceeds the total principal

[1] The agencies issuing CMOs refer to them as REMICs (Real Estate Mortgage Investment Conduits).

on the CMO bonds. The issuers can also obtain credit enhancement by acquiring financial guarantees from so-called monoline insurance companies.

For examples on how to value various mortgage-backed securities, see McConnell and Singh (1994) and Childs et al. (1996).

14.9 THE SUBPRIME CRISIS

From 1997 to 2006 house prices in the United States increased rapidly and this phenomenon is often referred to as a housing bubble. On average, house prices more than doubled in this decade, as can be seen from Fig. 14.6. The dark grey curve shows the evolution in the S&P/Case-Shiller U.S. national home price index over the period 1987–2010. The steep increase in house prices was partly driven by a drop in interest rates, as reflected by the light grey curve in Fig. 14.6, which shows the 3-month Treasury yield. Furthermore, in the same period, the U.S. Congress pressured mortgage institutions to increase their lending to low-income groups and make their loans more affordable for these groups. The home ownership rate did in fact increase from (roughly) 64% in 1994 to 69% in 2004. A substantial proportion of the new home owners financed their investment by subprime mortgages. The subprime share of the mortgage originations grew from less than 2% in

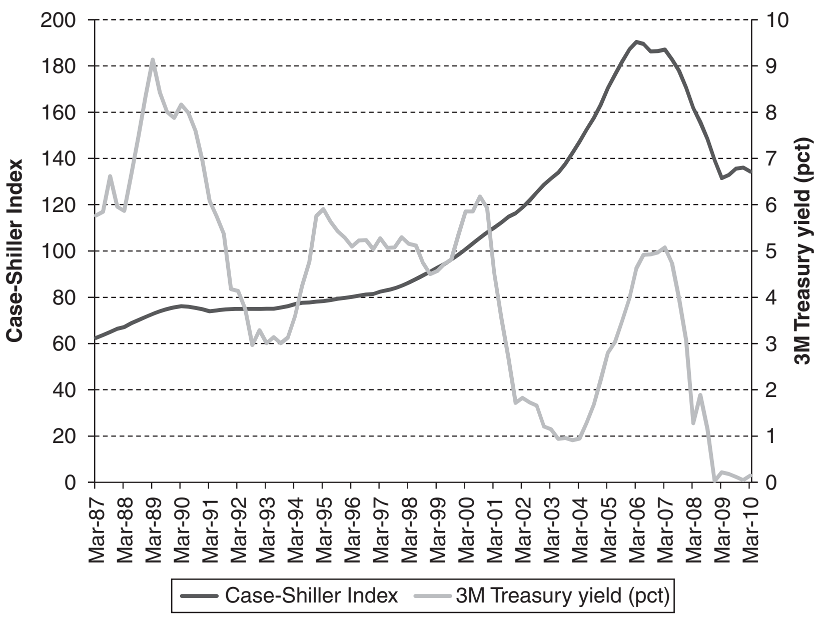

Fig. 14.6: U.S. house prices and interest rates 1987–2010. The dark curve (left axis) shows the seasonally adjusted S&P/Case-Shiller U.S. national home price index value. The light curve shows the 3-month market yield on U.S. Treasury securities.

Source: `www.standardpoors.com` and `www.federalreserve.gov`.

1994 to 8% in 2003 and on to around 20% in 2005, see the report by the Joint Center for Housing Studies of Harvard University (2008). There was also a significant increase in adjustable-rate mortgages, interest-only mortgages, teaser loans, and other mortgages with low initial payments compared to the traditional fixed-rate mortgage. Requirements on down payments were relaxed and property buyers could easily obtain 100% financing (or even more).

In light of low yields on government bonds, many investors—including large foreign investors—were eager to invest in mortgage-backed securities offering somewhat higher returns. Apparently the high demand for mortgage-backed securities contributed to lax lending standards in the mortgage industry. Mortgage originators reduced the documentation of income and assets needed to obtain a loan and were willing to lower their credit standards towards borrowers, since they would typically sell off the mortgages anyway to buyers who seemed little concerned about the default risk on the individual mortgages. Deliberately or not, the issuers and the credit rating companies may have underestimated the default risk of various mortgage-backed securities so that high-rated securities with higher (promised) returns than government bonds could be offered to potential investors. When valuing securities backed by a pool of uninsured mortgages, investors and analysts frequently applied the Gaussian copula introduced in Section 13.6 to model the default correlation between loans in the pool, and this practice was apparently severely underestimating the default correlations. As long as all the players in the mortgage industry made huge profits, few questioned the sustainability of the business model and the valuation methods applied. The spread between the interest rate on subprime mortgages and prime mortgages dropped and home ownership rates increased. The developments in the mortgage and housing markets could be seen as a perfect example of the benefits of modern financial innovation and securitization for ordinary people and society as a whole.

As long as house prices continued to increase, teaser loan borrowers could easily refinance at the end of the initial low-rate period, and interest-only borrowers could refinance before they had to start paying principal. In addition, it was easy to obtain credit via bank loans or credit cards. Saving rates went down, spending up. The apparent endless increase in house prices attracted speculators purchasing residential real estate for investment purposes. A growing group of property speculators with plenty of credit opportunities in commercial banks were making large and highly leveraged investments in commercial real estate and condominiums.

The house price rally led to a high housing turnover and a building boom. Eventually the number of unsold homes went up significantly. Media reports about the concerns of some economists and market analysts that the housing bubble would soon burst appeared more and more frequently, and these concerns spread to some market participants. Interest rates were increasing dramatically from 2004 to 2006 making financing and refinancing more expensive. The U.S. housing prices peaked in the first half of 2006. For about a year prices seemed to stabilize only slightly below the maximum, but from early 2007 prices began to drop significantly and by March 2010 the national average house price is down by 29.5% relative to the March 2006 maximum (according to the S&P/Case-Shiller seasonally adjusted national home price index). As soon as interest rates went up and the increase in house prices stopped, some borrowers were facing severe financial problems, in particular subprime borrowers with adjustable-rate mortgages and teaser loans. As house prices dropped significantly, problems spread to the prime mortgage market. The number of foreclosures and defaults rocketed. Investors in subprime mortgage-backed securities and mortgage insurers started to report huge losses. Several major financial institutions failed, others were bailed-out by governments or merged with more well-funded institutions. The complicated structure of many mortgage-related securities made it difficult to gauge the exposure of investors to house prices and to mortgage defaults, contributing to a freezing credit market among financial institutions, which again led commercial banks and mortgage originators to significantly reduce their credit to individual and corporate customers. The crisis that started in the subprime mortgage and housing markets spread and developed into a major financial and economic crisis.

While we focused on the U.S. above, many other countries have experienced similar developments in mortgage and housing markets. Moreover, because of the securitization of the mortgage products and the highly globalized financial markets, investors around the world have been exposed to the U.S. housing market and suffered substantial losses following the subprime collapse. Hence, the subsequent financial and economic crisis is a worldwide crisis.

The subprime crisis and the associated financial and economic crises are complex and we have only touched upon them here. For more details and discussions the interested reader is referred to Crouhy et al. (2008), Demyanyk and Van Hemert (2008), Sanders (2008), Shiller (2008), Gorton (2009), Swan (2009), and Berndt et al. (2010).

14.10 CONCLUDING REMARKS

The mortgage markets in the U.S. and many other developed countries have grown into a huge, complex, and innovative industry. This chapter has provided a short introduction to the main products and concepts as well as the valuation and modelling challenges. A key issue is the prepayment option embedded in many mortgages, which is notoriously difficult to include in models in a way that is consistent with both theoretical predictions on early exercise and empirically observed prepayment behaviour. Any model for the valuation of standard mortgages and mortgage-backed securities must be based on some assumptions about interest rate dynamics and prepayment activity over a 30-year horizon. Note that due to the difficulties in predicting prepayment behaviour, bond investors may incorporate a 'safety margin' in their valuation of mortgage-backed bonds, which will lead to lower bond prices and higher borrowing rates.

Another key issue in mortgage modelling is the potential default of the borrowers. In some cases the default risk is passed on to the investors in securities backed by the mortgages, whereas in other cases the default risk exposure is assumed by the original lenders or some insurers. In any case, a solid understanding and assessment of mortgage default risks is important for a sound mortgage market.

14.11 EXERCISES

Exercise 14.1 Consider the mortgage described in the example in Section 14.5.4.

(a) Compute the scheduled payments over the life of the mortgage and compare with Fig. 14.1.

Suppose a pass-through bond is backed by a pool of many, many identical mortgages with the terms as described in the example in Section 14.5.4. Let the face value of each bond be 100.

(b) Determine the cash flow to the bond if none of the mortgages are ever prepaid.

(c) Determine the cash flow to the bond if, at every payment date, 5% of the remaining mortgages are fully prepaid.

(d) Determine the cash flow to the bond if, at every payment date, 10% of the remaining mortgages are fully prepaid. Compare with the answers to the two preceding questions and comment on the impact of prepayments and prepayment risk on the bonds.

Exercise 14.2 Consider a 30-year annuity-type mortgage with a face value of USD 100,000 and monthly payments. Assume for simplicity that no fee is paid on the mortgage. Initially the nominal rate is 6% per year, but the nominal rate is adjusted every 5 years.

(a) What is the initial monthly payment?

(b) Compute the scheduled payments (split into an interest payment and a repayment of principal) over the life of the mortgage assuming the nominal rate never changes.

(c) Now suppose that after 5 years, the nominal rate is reset to 3% per year. Determine the new monthly payment and the scheduled payments over the remaining life of the mortgage.

(d) Now suppose that after 10 years, the nominal rate is reset to 6% per year. Determine the new monthly payment and the scheduled payments over the remaining life of the mortgage.

Exercise 14.3 As mentioned towards the end of Section 14.6, mortgage-backed bonds can be priced using the reduced-form modelling approach for defaultable securities discussed in Section 13.4. In particular, a prepayment intensity must then be modelled. Suppose that the government bond yield curve can be described by a one-factor diffusion model with the short rate as the state variable. How should the prepayment intensity depend on short rate? Can/should you allow the prepayment intensity to depend on other variables? Can you think of a simple and reasonable model in which such a reduced-form approach would lead to closed-form expressions for the mortgage-backed bonds?

15

Stock and Currency Derivatives When Interest Rates are Stochastic

15.1 INTRODUCTION

In the development of term structure models in Chapters 7–11 we focused on securities with payments and values that only depend on the term structure of interest rates, not on any other random variables. However, the shape and the dynamics of the yield curve will also affect the prices of securities with payments that depend on other random variables, for example stock prices and currency rates. The reason is that the present value of a security involves the discounting of the future payments, and the appropriate discount factors depend on the interest rate uncertainty as well as the correlations between interest rates and the random variables that determine the payments of the security.

Section 15.2 considers the pricing of stock options when we allow for the uncertain evolution of interest rates in contrast to the classic Black–Scholes–Merton model. We show that for Gaussian term structure models the price of a European stock option is given by a simple generalization of the Black–Scholes–Merton formula. This generalized formula corresponds to the way in which practitioners often implement the Black–Scholes–Merton formula. Section 15.3 discusses the pricing of European options on forwards and futures on stocks and provides explicit formulas in the Gaussian interest rate setting. Finally, Section 15.4 studies securities with payments related to a foreign exchange rate. With a lognormal foreign exchange rate and Gaussian interest rates we obtain simple expressions for currency futures prices and European currency option prices. Throughout the chapter we focus on European call options. The prices of the corresponding European put options follow from the relevant version of the put-call parity. As always, to price American options we generally have to resort to numerical methods.

15.2 STOCK OPTIONS

15.2.1 General analysis

Let us look at a European call option that expires at time $T > t$, is written on a stock with price process (S_t), and has an exercise price on K. We know from Section 4.4 that the time t price of this option is given by

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max \left(S_T - K, 0 \right) \right], \tag{15.1}$$

where $\mathbb{Q}^T$ is the T-forward martingale measure. For simplicity we assume that the underlying asset does not provide any payments in the life of the option. The forward price of the underlying asset for delivery at date T is given by $F_t^T = S_t / B_t^T$. In particular, $F_T^T = S_T$ so that the option price can be rewritten as

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max \left(F_T^T - K, 0 \right) \right].$$

Recall that, by definition of the T-forward martingale measure, we have that $\mathrm{E}_t^{\mathbb{Q}^T} [F_T^T] = F_t^T = S_t / B_t^T$. To compute the expected value either in closed form or by simulation, we have to know the distribution of $S_T = F_T^T$ under the T-forward martingale measure. This distribution will follow from the dynamics of the forward price F_t^T. But first we will set up a model for the price of the underlying stock and for the relevant discount factors, that is the zero-coupon bond prices.

As usual, we will stick to models where the basic uncertainty is represented by one or several standard Brownian motions. In a model with a single Brownian motion, all stochastic processes will be instantaneously perfectly correlated, compare the discussion in the introduction to Chapter 8. To price stock options in a setting with stochastic interest rates, we have to model both the stock price and the appropriate discount factor. Since these two variables are not perfectly correlated, we have to include more than one Brownian motion in our model.

Under the risk-neutral or spot martingale measure $\mathbb{Q}$ the drift of the price of any traded asset (in time intervals with no dividend payments) is equal to the short-term interest rate, r_t. The dynamics of the price of the underlying asset is assumed to be of the form

$$dS_t = S_t \left[r_t \, dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} \right)^{\mathsf{T}} \, d\boldsymbol{z}_t^{\mathbb{Q}} \right], \tag{15.2}$$

where $\boldsymbol{z}^{\mathbb{Q}}$ is a multi-dimensional standard Brownian motion under the risk-neutral measure $\mathbb{Q}$, and where $\boldsymbol{\sigma}_t^{\mathrm{st}}$ is a vector representing the sensitivity of the stock price with respect to the exogenous shocks. We will refer to $\boldsymbol{\sigma}^{\mathrm{st}}$ as the sensitivity vector of the stock price. In general, $\boldsymbol{\sigma}_t^{\mathrm{st}}$ may itself be stochastic, for example depend on the level of the stock price, but we will only derive explicit option prices in the case where $\boldsymbol{\sigma}_t^{\mathrm{st}}$ is a deterministic function of time and then we will use the notation $\boldsymbol{\sigma}^{\mathrm{st}}(t)$. It is hard to imagine that the volatility of a stock will depend directly on calendar time, so the most relevant example of a deterministic volatility is a constant sensitivity vector. We can also write (15.2) as

$$dS_t = S_t \left[r_t \, dt + \sum_{j=1}^{n} \sigma_{jt}^{\mathrm{st}} \, dz_{jt}^{\mathbb{Q}} \right],$$

where n is the number of independent one-dimensional Brownian motions in the model, and $\sigma_1^{\mathrm{st}}, \dots, \sigma_n^{\mathrm{st}}$ are the components of the sensitivity vector.

Similarly, we will assume that the price of the zero-coupon bond maturing at time T will evolve according to

$$dB_t^T = B_t^T \left[r_t\, dt + \left(\boldsymbol{\sigma}_t^T\right)^{\top} d\boldsymbol{z}_t^{\mathbb{Q}} \right], \tag{15.3}$$

where the sensitivity vector $\boldsymbol{\sigma}_t^T$ of the bond may depend on the current term structure of interest rates (and in theory also on previous term structures). Equivalently, we can write the bond price dynamics as

$$dB_t^T = B_t^T \left[r_t\, dt + \sum_{j=1}^{n} \sigma_{jt}^T\, dz_{jt}^{\mathbb{Q}} \right].$$

In the model given by (15.2) and (15.3) the variance of the instantaneous rate of return on the stock is given by

$$\mathrm{Var}_t^{\mathbb{Q}}\left[\frac{dS_t}{S_t}\right] = \mathrm{Var}_t^{\mathbb{Q}}\left[\sum_{j=1}^{n} \sigma_{jt}^{\mathrm{st}}\, dz_{jt}^{\mathbb{Q}}\right] = \sum_{j=1}^{n} \left(\sigma_{jt}^{\mathrm{st}}\right)^2 dt$$

so that the volatility of the stock is equal to the length of the vector $\boldsymbol{\sigma}_t^{\mathrm{st}}$, that is $\|\boldsymbol{\sigma}_t^{\mathrm{st}}\| = \sqrt{\sum_{j=1}^{n}\left(\sigma_{jt}^{\mathrm{st}}\right)^2}$. Similarly, the volatility of the zero-coupon bond is given by $\|\boldsymbol{\sigma}_t^T\|$. The covariance between the rate of return on the stock and the rate of return on the zero-coupon bond is $\left(\boldsymbol{\sigma}_t^{\mathrm{st}}\right)^{\top} \boldsymbol{\sigma}_t^T = \sum_{j=1}^{n} \sigma_{jt}^{\mathrm{st}} \sigma_{jt}^T$. Consequently, the instantaneous correlation is $\left(\boldsymbol{\sigma}_t^{\mathrm{st}}\right)^{\top} \boldsymbol{\sigma}_t^T / \left[\|\boldsymbol{\sigma}_t^{\mathrm{st}}\| \cdot \|\boldsymbol{\sigma}_t^T\|\right]$.

Note that if we just want to model the prices of this particular stock and this particular bond, a model with $n = 2$ is sufficient to capture the imperfect correlation. For example, if we specify the dynamics of prices as

$$dS_t = S_t \left[r_t\, dt + v_t^{\mathrm{st}}\, dz_{1t}^{\mathbb{Q}} \right],$$
$$dB_t^T = B_t^T \left[r_t\, dt + \rho v_t^T\, dz_{1t}^{\mathbb{Q}} + \sqrt{1-\rho^2}\, v_t^T\, dz_{2t}^{\mathbb{Q}} \right],$$

the volatilities of the stock and the bond are given by v_t^{st} and v_t^T, respectively, while $\rho \in [-1, 1]$ is the instantaneous correlation. However, we will stick to the more general notation introduced earlier.

Given the dynamics of the stock price and the bond price in Equations (15.2) and (15.3), we obtain the dynamics of the forward price $F_t^T = S_t/B_t^T$ under the $\mathbb{Q}^T$ probability measure by an application of Itô's Lemma for functions of two stochastic processes, see Theorem 3.7. Knowing that F_t^T is a $\mathbb{Q}^T$-martingale so that its drift is zero, we do not have to compute the drift term from Itô's Lemma. Therefore, we just have to find the sensitivity vector, which we know is the same under all the relevant probability measures. Writing $F_t^T = g(S_t, B_t^T)$, where $g(S, B) = S/B$, the relevant derivatives are $\partial g/\partial S = 1/B$ and $\partial g/\partial B = -S/B^2$ so that we obtain the following forward price dynamics:

$$
\begin{aligned}
dF_t^T &= \frac{\partial g}{\partial S}(S_t, B_t^T)\, S_t \left(\boldsymbol{\sigma}_t^{\text{st}}\right)^\top dz_t^T + \frac{\partial g}{\partial B}(S_t, B_t^T)\, B_t^T \left(\boldsymbol{\sigma}_t^T\right)^\top dz_t^T \\
&= F_t^T \left(\boldsymbol{\sigma}_t^{\text{st}} - \boldsymbol{\sigma}_t^T\right)^\top dz_t^T.
\end{aligned}
$$

A standard calculation yields

$$
d(\ln F_t^T) = -\frac{1}{2}\|\boldsymbol{\sigma}_t^{\text{st}} - \boldsymbol{\sigma}_t^T\|^2\, dt + \left(\boldsymbol{\sigma}_t^{\text{st}} - \boldsymbol{\sigma}_t^T\right)^\top dz_t^T,
$$

and hence

$$
\ln S_T = \ln F_T^T = \ln F_t^T - \frac{1}{2}\int_t^T \|\boldsymbol{\sigma}_u^{\text{st}} - \boldsymbol{\sigma}_u^T\|^2\, du + \int_t^T \left(\boldsymbol{\sigma}_u^{\text{st}} - \boldsymbol{\sigma}_u^T\right)^\top dz_u^T. \tag{15.4}
$$

In general, $\boldsymbol{\sigma}^{\text{st}}$ and $\boldsymbol{\sigma}^T$ will be stochastic, in which case we cannot identify the distribution of $\ln S_T$ and hence S_T, but Equation (15.4) provides the basis for Monte Carlo simulations of S_T and thus an approximation of the option price. Below, we discuss the case where $\boldsymbol{\sigma}^{\text{st}}$ and $\boldsymbol{\sigma}^T$ are deterministic. In that case we can obtain an explicit option pricing formula.

15.2.2 Deterministic volatilities

If we assume that both $\boldsymbol{\sigma}_t^{\text{st}}$ and $\boldsymbol{\sigma}_t^T$ are deterministic functions of time t, it follows from (15.4) and Theorem 3.3 that $\ln S_T = \ln F_T^T$ is normally distributed, that is $S_T = F_T^T$ is lognormally distributed, under the T-forward martingale measure. Theorem A.4 in Appendix A implies that the price of the stock option given in Equation (15.1) can be written in closed form as

$$
C_t = B_t^T \left\{ \mathrm{E}_t^{\mathbb{Q}^T}[F_T^T] N(d_1) - K N(d_2) \right\},
$$

where

$$
\begin{aligned}
d_1 &= \frac{1}{v_F(t,T)} \ln\left(\frac{\mathrm{E}_t^{\mathbb{Q}^T}[F_T^T]}{K}\right) + \frac{1}{2} v_F(t,T), \\
d_2 &= d_1 - v_F(t,T), \\
v_F(t,T) &= \left(\mathrm{Var}_t^{\mathbb{Q}^T}[\ln F_T^T]\right)^{1/2}.
\end{aligned}
$$

By the martingale property, $\mathrm{E}_t^{\mathbb{Q}^T}[F_T^T] = F_t^T = S_t/B_t^T$, we can compute the variance as

$$
\begin{aligned}
\nu_F(t,T)^2 &= \operatorname{Var}_t^{\mathbb{Q}^T}[\ln F_T^T] = \operatorname{Var}_t^{\mathbb{Q}^T}\left[\int_t^T \left(\boldsymbol{\sigma}^{\text{st}}(u) - \boldsymbol{\sigma}^T(u)\right)^{\top} d\boldsymbol{z}_u^T\right] \\
&= \operatorname{Var}_t^{\mathbb{Q}^T}\left[\int_t^T \sum_{j=1}^n \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u)\right) dz_{ju}^T\right] \\
&= \sum_{j=1}^n \operatorname{Var}_t^{\mathbb{Q}^T}\left[\int_t^T \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u)\right) dz_{ju}^T\right] \\
&= \sum_{j=1}^n \int_t^T \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u)\right)^2 du, \\
&= \int_t^T \sum_{j=1}^n \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u)\right)^2 du, \\
&= \int_t^T \left\|\boldsymbol{\sigma}^{\text{st}}(u) - \boldsymbol{\sigma}^T(u)\right\|^2 du \\
&= \int_t^T \|\boldsymbol{\sigma}^{\text{st}}(u)\|^2\, du + \int_t^T \|\boldsymbol{\sigma}^T(u)\|^2\, du - 2\int_t^T \boldsymbol{\sigma}^{\text{st}}(u)^{\top}\boldsymbol{\sigma}^T(u)\, du,
\end{aligned}
$$

where the third equality follows from the independence of the Brownian motions $z_1^T, \dots, z_n^T$, and the fourth equality follows from Theorem 3.3. Clearly, the first term in the final expression for the variance is due to the uncertainty about the future price of the underlying stock, the second term is due to the uncertainty about the discount factor, and the third term is due to the covariance of the stock price and the discount factor. The price of the option can be rewritten as

$$C_t = S_t N(d_1) - KB_t^T N(d_2), \tag{15.5}$$

where

$$
\begin{aligned}
d_1 &= \frac{1}{\nu_F(t,T)} \ln\left(\frac{S_t}{KB_t^T}\right) + \frac{1}{2}\nu_F(t,T), \\
d_2 &= d_1 - \nu_F(t,T).
\end{aligned}
$$

If the sensitivity vector of the stock is constant, we get

$$\nu_F(t,T)^2 = \|\boldsymbol{\sigma}^{\text{st}}\|^2(T-t) + \int_t^T \|\boldsymbol{\sigma}^T(u)\|^2\, du - 2\int_t^T \left(\boldsymbol{\sigma}^{\text{st}}\right)^{\top} \boldsymbol{\sigma}^T(u)\, du. \tag{15.6}$$

The Black–Scholes–Merton model is the special case in which the short-term interest rate r is constant, which implies a constant, flat yield curve and deterministic zero-coupon bond prices of $B_t^T = e^{-r[T-t]}$ with $\boldsymbol{\sigma}^T(u) \equiv 0$. Under these additional assumptions, the option pricing formula (15.5) reduces to the famous Black–Scholes–Merton formula

$$C(t) = S_t N(d_1) - Ke^{-r[T-t]} N(d_2), \tag{15.7}$$

where

$$d_1 = \frac{1}{\|\boldsymbol{\sigma}^{\mathrm{st}}\|\sqrt{T-t}} \ln\left(\frac{S_t}{Ke^{-r[T-t]}}\right) + \frac{1}{2}\|\boldsymbol{\sigma}^{\mathrm{st}}\|\sqrt{T-t},$$

$$d_2 = d_1 - \|\boldsymbol{\sigma}^{\mathrm{st}}\|\sqrt{T-t}.$$

The more general formula (15.5) was first shown by Merton (1973). It holds for all Gaussian term structure models, for example in the Vasicek model and the Gaussian HJM models, because the sensitivity vector and hence the volatility of the zero-coupon bonds are then deterministic functions of time. In a reduced equilibrium model such as Vasicek's, the bond price B_t^T entering the option pricing formula is given by the well-known expression for the zero-coupon bond price in the model, for example (7.36) in the one-factor Vasicek model. For the extended Vasicek model and the Gaussian HJM models the currently observed zero-coupon bond price is used in the option pricing formula. This latter approach is consistent with practitioners' use of the Black–Scholes–Merton formula since, instead of a fixed interest rate r for options of all maturities, they use the observed zero-coupon yield y_t^T until the maturity date of the option. However, the relevant variance $v_F(t,T)^2$ in Merton's formula (15.7) also differs from the Black–Scholes–Merton formula. The first of the three terms in (15.6) is exactly the variance expression that enters the Black–Scholes–Merton formula. The other two terms have to be added to take into account the variation of interest rates and the covariation of interest rates and the stock price. Practitioners seem to disregard these two terms. In Exercise 15.1 you are asked to compare the option prices for certain parameter values. In most cases the two latter variance terms will be much smaller than the first term so that the errors implied by neglecting the two last terms will be insignificant. Therefore Merton's generalization supports practitioners' use of the Black–Scholes–Merton formula. However, the assumptions underlying Merton's extension are problematic since Gaussian term structure models are highly unrealistic.

For other term structure models one must resort to numerical methods for the computation of the stock option prices. One possibility is to approximate the expected value in (15.1) by an average of payoffs generated by Monte Carlo simulations of the terminal stock price under the T-forward martingale measure, for example based on (15.4). Note that if, for example, $\boldsymbol{\sigma}_u^T$ depends on the short rate r_u, the evolution in the short rate over the time period $[t, T]$ has to be simulated together with the stock price. Alternatively, the fundamental partial differential equation can be solved numerically. See Chapter 16 for more on these numerical pricing techniques.

Apparently, the effects of stochastic interest rates on stock option pricing and hedge ratios have not been subject to much research. Based on the analysis of a model allowing for both a stochastic stock price volatility and stochastic interest rates, Bakshi et al. (1997) conclude that typical European option prices are more sensitive to fluctuations in stock price volatilities than to fluctuations in interest rates. For the pricing of short- and medium-term stock options it seems to be unimportant to incorporate interest rate uncertainty. On the other hand, taking interest rate uncertainty into account does seem to make a difference for the purposes of constructing efficient option hedging strategies and pricing long-term stock options. Whether this conclusion generalizes to other model specifications and stock options with other contractual terms, for example American options, remains an unanswered question.

15.3 OPTIONS ON FORWARDS AND FUTURES

In this section we will discuss the pricing of options on forwards and futures on a security traded at the price S_t. As before, we assume that this underlying asset has no dividend payments. We will derive explicit formulas for European call options in the case where all price volatilities are deterministic. Amin and Jarrow (1992) obtain similar results for the special case where the dynamics of the term structure is given by a Gaussian HJM model, which implies that the volatilities of the zero-coupon bonds are deterministic. We will let T denote the expiry time of the option and let $\bar{T}$ denote the time of delivery (or final settlement) of the forward or the futures contract. Here, $T \leq \bar{T}$.

15.3.1 Forward and futures prices

As shown in Section 6.2, the forward price for delivery at time $\bar{T}$ is given by $F_t^{\bar{T}} = S_t/B_t^{\bar{T}}$, while the futures price $\Phi_t^{\bar{T}}$ for final settlement at time $\bar{T}$ is characterized by

$$\Phi_t^{\bar{T}} = \mathrm{E}_t^{\mathbb{Q}}\left[S_{\bar{T}}\right] = \mathrm{E}_t^{\mathbb{Q}}\left[F_{\bar{T}}^{\bar{T}}\right].$$

As in the preceding section, we assume that the dynamics under the risk-neutral measure $\mathbb{Q}$ is

$$dS_t = S_t\left[r_t\,dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}}\right)^{\top} d\boldsymbol{z}_t^{\mathbb{Q}}\right]$$

for the price of the security underlying the forward and the futures and

$$dB_t^{\bar{T}} = B_t^{\bar{T}}\left[r_t\,dt + \left(\boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top} d\boldsymbol{z}_t^{\mathbb{Q}}\right]$$

for the price of the zero-coupon bond maturing at $\bar{T}$. Itô's Lemma yields first that

$$dF_t^{\bar{T}} = F_t^{\bar{T}}\left[-\left(\boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right) dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top} d\boldsymbol{z}_t^{\mathbb{Q}}\right], \tag{15.8}$$

and, subsequently, that

$$d(\ln F_t^{\bar{T}}) = \left[-\frac{1}{2}\|\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)\right] dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top} dz_t^{\mathbb{Q}}.$$

It follows that

$$\ln F_{\bar{T}}^{\bar{T}} = \ln F_t^{\bar{T}} + \int_t^{\bar{T}} \left[-\frac{1}{2}\|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)\right] du$$
$$+ \int_t^{\bar{T}} \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top} dz_u^{\mathbb{Q}}.$$

Since $S_{\bar{T}} = F_{\bar{T}}^{\bar{T}}$, we see that

$$S_{\bar{T}} = F_t^{\bar{T}} \exp\left\{\int_t^{\bar{T}} \left[-\frac{1}{2}\|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)\right] du \right.$$
$$\left. + \int_t^{\bar{T}} \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top} dz_u^{\mathbb{Q}}\right\}.$$

Therefore the futures price can be written as

$$\Phi_t^{\bar{T}} = F_t^{\bar{T}}\, \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left\{\int_t^{\bar{T}} \left[-\frac{1}{2}\|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)\right] du \right.\right.$$
$$\left.\left. + \int_t^{\bar{T}} \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top} dz_u^{\mathbb{Q}}\right\}\right].$$

For the case where the volatilities $\boldsymbol{\sigma}^{\mathrm{st}}$ and $\boldsymbol{\sigma}^{\bar{T}}$ are deterministic, the futures price is given in closed form as

$$\begin{aligned}
\Phi_t^{\bar{T}} &= F_t^{\bar{T}} \exp\left\{\int_t^{\bar{T}} \left[-\frac{1}{2}\|\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\|^2 - \boldsymbol{\sigma}^{\bar{T}}(u)^{\top}\left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)\right] du\right\} \\
&\quad \times \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left\{\int_t^{\bar{T}} \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)^{\top} dz_u^{\mathbb{Q}}\right\}\right] \qquad (15.9) \\
&= F_t^{\bar{T}} \exp\left\{-\int_t^{\bar{T}} \boldsymbol{\sigma}^{\bar{T}}(u)^{\top}\left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du\right\},
\end{aligned}$$

where the last equality is a consequence of Theorem 3.3 and Theorem A.2. If the volatilities are not deterministic, no explicit expression for the futures price is available.

15.3.2 Options on forwards

From the analysis in Section 4.4 we know that the price of a European call option on a forward is

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max(F_T^{\bar{T}} - K, 0)\right], \tag{15.10}$$

where T is the expiry time and K is the exercise price. Let us find the dynamics of the forward price $F_t^{\bar{T}}$ under the T-forward martingale measure $\mathbb{Q}^T$. From (4.21) we can shift the probability measure from $\mathbb{Q}$ to $\mathbb{Q}^T$ by applying the relation

$$dz_t^T = dz_t^{\mathbb{Q}} - \boldsymbol{\sigma}_t^T \, dt. \tag{15.11}$$

Substituting this into (15.8), we can write the dynamics in $F_t^{\bar{T}}$ under $\mathbb{Q}^T$ as

$$dF_t^{\bar{T}} = F_t^{\bar{T}}\left[\left(\boldsymbol{\sigma}_t^T - \boldsymbol{\sigma}_t^{\bar{T}}\right)^{\mathsf{T}}\left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right) dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)^{\mathsf{T}} dz_t^T\right].$$

Note that only if $\bar{T} = T$, the drift will be zero and $F_t^{\bar{T}}$ will be a $\mathbb{Q}^T$-martingale. It follows that

$$\begin{aligned}\ln F_T^{\bar{T}} = \ln F_t^{\bar{T}} &+ \int_t^T \left(\boldsymbol{\sigma}_u^T - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\mathsf{T}}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right) du \\ &- \frac{1}{2}\int_t^T \|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 \, du + \int_t^T \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\mathsf{T}} dz_u^T.\end{aligned}$$

Under the assumption that $\boldsymbol{\sigma}_u^{\mathrm{st}}$, $\boldsymbol{\sigma}_u^T$, and $\boldsymbol{\sigma}_u^{\bar{T}}$ are all deterministic functions of time, we have that $\ln F_T^{\bar{T}}$ (given $F_t^{\bar{T}}$) is normally distributed under $\mathbb{Q}^T$ with mean value

$$\begin{aligned}\mu_F \equiv \mathrm{E}_t^{\mathbb{Q}^T}\left[\ln F_T^{\bar{T}}\right] = \ln F_t^{\bar{T}} &+ \int_t^T \left(\boldsymbol{\sigma}^T(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)^{\mathsf{T}}\left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du \\ &- \frac{1}{2}\int_t^T \|\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\|^2 \, du\end{aligned}$$

and variance

$$v_F^2 \equiv \mathrm{Var}_t^{\mathbb{Q}^T}\left[\ln F_T^{\bar{T}}\right] = \int_t^T \left\|\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right\|^2 du.$$

Applying Theorem (A.4) in Appendix A, we can compute the option price from (15.10) as

$$C_t = B_t^T \left\{e^{\mu_F + \frac{1}{2}v_F^2} N(d_1) - K N(d_2)\right\},$$

where

$$d_1 = \frac{\mu_F - \ln K}{v_F} + v_F = \frac{\mu_F + \frac{1}{2}v_F^2 - \ln K}{v_F} + \frac{1}{2}v_F,$$

$$d_2 = d_1 - v_F.$$

Since

$$\mu_F + \frac{1}{2}v_F^2 = \ln F_t^{\bar{T}} + \int_t^T \left(\boldsymbol{\sigma}^T(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)^{\top} \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du,$$

we can replace $e^{\mu_F + \frac{1}{2}v_F^2}$ by $F_t^{\bar{T}} e^{\xi}$, where

$$\xi = \int_t^T \left(\boldsymbol{\sigma}^T(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)^{\top} \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du.$$

Hence, the option price can be rewritten as

$$C_t = B_t^T F_t^{\bar{T}} e^{\xi} N(d_1) - K B_t^T N(d_2), \tag{15.12}$$

and d_1 can be rewritten as

$$d_1 = \frac{\ln(F_t^{\bar{T}}/K) + \xi}{v_F} + \frac{1}{2}v_F.$$

15.3.3 Options on futures

A European call option on a futures has a value of

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max(\Phi_T^{\bar{T}} - K, 0)\right].$$

With deterministic volatilities we can apply (15.9) and insert $\Phi_T^{\bar{T}} = F_T^{\bar{T}} e^{-\psi(T,\bar{T},\bar{T})}$, where we have introduced the notation

$$\psi(T, U, \bar{T}) = \int_T^U \boldsymbol{\sigma}^{\bar{T}}(u)^{\top} \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du.$$

Consequently, the option price can be written as

$$\begin{aligned} C_t &= B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max(F_T^{\bar{T}} e^{-\psi(T,\bar{T},\bar{T})} - K, 0)\right] \\ &= B_t^T e^{-\psi(T,\bar{T},\bar{T})} \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max(F_T^{\bar{T}} - K e^{\psi(T,\bar{T},\bar{T})}, 0)\right]. \end{aligned}$$

We see that, under these assumptions, a call option on a futures with the exercise price K is equivalent to $e^{-\psi(T,\bar{T},\bar{T})}$ call options on a forward with the exercise price

$Ke^{\psi(T,\bar{T},\bar{T})}$. From (15.12) it follows that the price of the futures option is

$$C_t = e^{-\psi(T,\bar{T},\bar{T})} \left[B_t^T F_t^{\bar{T}} e^{\xi} N(d_1) - Ke^{\psi(T,\bar{T},\bar{T})} B_t^T N(d_2) \right],$$

which can be rewritten as

$$C_t = F_t^{\bar{T}} B_t^T e^{\xi - \psi(T,\bar{T},\bar{T})} N(d_1) - KB_t^T N(d_2),$$

where

$$d_1 = \frac{\ln(F_t^{\bar{T}}/K) + \xi - \psi(T,\bar{T},\bar{T})}{v_F} + \frac{1}{2} v_F,$$

$$d_2 = d_1 - v_F,$$

$$v_F = \left(\int_t^T \left\| \boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u) \right\|^2 du \right)^{1/2}.$$

Applying $F_t^{\bar{T}} = \Phi_t^{\bar{T}} e^{\psi(t,\bar{T},\bar{T})}$ and $\psi(t,\bar{T},\bar{T}) - \psi(T,\bar{T},\bar{T}) = \psi(t,T,\bar{T})$, we can also express the option price as

$$C_t = \Phi_t^{\bar{T}} B_t^T e^{\xi - \psi(t,T,\bar{T})} N(d_1) - KB_t^T N(d_2)$$

with

$$d_1 = \frac{\ln(\Phi_t^{\bar{T}}/K) + \xi - \psi(t,T,\bar{T})}{v_F} + \frac{1}{2} v_F.$$

In Exercise 15.2 you are asked to compute and compare prices on various options on forwards and futures on a stock assuming a constant stock volatility and the Vasicek term structure model.

15.4 CURRENCY DERIVATIVES

Corporations and individuals who operate internationally are exposed to currency risk since most foreign exchange rates fluctuate in an unpredictable manner. The exposure can be reduced or eliminated by investments in suitable financial contracts. Both on organized exchanges and in the OTC markets numerous contracts with currency dependent payoffs are traded. Some of these contracts also depend on other economic variables, for example interest rates or stock prices. However, we will focus on currency derivatives whose payments only depend on a single forward exchange rate. This is the case for standard currency forwards, futures, and options.

Before we go into the valuation of the currency derivatives, we will introduce some notation. The time t spot price of one unit of the foreign currency is denoted by ε_t. This is the number of units of the domestic currency that can be exchanged for one unit of the foreign currency. As before, r_t denotes the short-term domestic interest rate and B_t^T denotes the price (in the domestic currency) of a zero-coupon bond that delivers one unit of the domestic currency at time T. By $\check{B}_t^T$ we will denote the price in units of the foreign currency of a zero-coupon bond that delivers one unit of the foreign currency at time T. Similarly, $\check{r}_t$ and $\check{y}_t^T$ denote the foreign short rate and the foreign zero-coupon yield for maturity T, respectively.

15.4.1 Currency forwards

The simplest currency derivative is a forward contract on one unit of the foreign currency. This is a binding contract of delivery of one unit of the foreign currency at time T at a prespecified exchange rate K so that the payoff at time T is $\varepsilon_T - K$. The no-arbitrage value at time $t < T$ of this payoff is $\check{B}_t^T \varepsilon_t - B_t^T K$ since this is the value of a portfolio that provides the same payoff as the forward, namely a portfolio of one unit of the foreign zero-coupon bond maturing at T and a short position in K units of the domestic zero-coupon bond maturing at time T. The forward exchange rate at time t for delivery at time T is denoted by F_t^T and it is defined as the value of the delivery price K that makes the present value equal to zero, that is

$$F_t^T = \frac{\check{B}_t^T}{B_t^T} \varepsilon_t. \tag{15.13}$$

This relation is consistent with the results on forward prices derived in Chapter 6. The forward exchange rate can be expressed as

$$F_t^T = \varepsilon_t e^{(y_t^T - \check{y}_t^T)(T-t)},$$

where y_t^T and $\check{y}_t^T$ denote the domestic and the foreign zero-coupon rates for maturity date T, respectively. If $y_t^T > \check{y}_t^T$, the forward exchange rate will be higher than the spot exchange rate, otherwise an arbitrage will exist. Conversely, if $y_t^T < \check{y}_t^T$, the forward exchange rate will be lower than the spot exchange rate. The stated expressions for the forward exchange rate are based only on the no-arbitrage principle and hold independently of the dynamics in the spot exchange rate and the interest rates of the two countries.

15.4.2 A model for the exchange rate

In order to be able to price currency derivative securities other than currency forwards, assumptions about the evolution of the spot exchange rate are necessary. As always we focus on models where the fundamental uncertainty is represented by Brownian motions. Since we have to model the evolution in both the exchange rate and the term structures of the two countries, and these objects are not perfectly correlated, the model has to involve a multi-dimensional Brownian motion.

Foreign currency can be held in a deposit account earning the foreign short-term interest rate. Therefore, we can think of foreign currency as an asset providing a continuous dividend at a rate equal to the foreign short rate, $\check{r}_t$. Under the domestic risk-neutral measure $\mathbb{Q}$ the total expected rate of return on any asset will equal the domestic short rate. Since foreign currency provides a cash rate of return of $\breve{r}_t$, the expected percentage increase in the price of foreign currency, that is the exchange rate, must equal $r_t - \breve{r}_t$. The dynamics of the spot exchange rate will therefore be of the form

$$d\varepsilon_t = \varepsilon_t \left[(r_t - \breve{r}_t)\, dt + \left(\boldsymbol{\sigma}_t^{\varepsilon}\right)^{\top} dz_t^{\mathbb{Q}} \right], \tag{15.14}$$

where $z^{\mathbb{Q}}$ is a multi-dimensional standard Brownian motion under the risk-neutral measure $\mathbb{Q}$, and where $\boldsymbol{\sigma}_t^{\varepsilon}$ is a vector of the sensitivities of the spot exchange rate towards the changes in the individual Brownian motions. Note that, in general, r, $\breve{r}$, and $\boldsymbol{\sigma}^{\varepsilon}$ in (15.14) will be stochastic processes.

Define $Y_t = \breve{B}_t^T \varepsilon_t$, that is Y_t is the price of the foreign zero-coupon bond measured in units of the domestic currency. If we let $\breve{\boldsymbol{\sigma}}_t^T$ denote the sensitivity vector of the foreign zero-coupon bond, that is

$$d\breve{B}_t^T = \breve{B}_t^T \left[\dots, dt + \left(\breve{\boldsymbol{\sigma}}_t^T\right)^{\top} dz_t^{\mathbb{Q}} \right],$$

it follows from Itô's Lemma that the sensitivity vector for Y_t can be written as $\boldsymbol{\sigma}_t^{\varepsilon} + \breve{\boldsymbol{\sigma}}_t^T$. Furthermore, we know that, measured in the domestic currency, the expected return on any asset under the risk-neutral probability measure $\mathbb{Q}$ will equal the domestic short rate, r_t. Hence, we have

$$dY_t = Y_t \left[r_t\, dt + \left(\boldsymbol{\sigma}_t^{\varepsilon} + \breve{\boldsymbol{\sigma}}_t^T\right)^{\top} dz_t^{\mathbb{Q}} \right].$$

For the domestic zero-coupon bond the price dynamics is of the form

$$dB_t^T = B_t^T \left[r_t\, dt + \left(\boldsymbol{\sigma}_t^T\right)^{\top} dz_t^{\mathbb{Q}} \right].$$

According to (15.13), the forward exchange rate is given by $F_t^T = Y_t / B_t^T$. An application of Itô's Lemma yields that the dynamics of the forward exchange rate is

$$dF_t^T = F_t^T \left[-\left(\boldsymbol{\sigma}_t^T\right)^{\top} \left(\boldsymbol{\sigma}_t^{\varepsilon} + \breve{\boldsymbol{\sigma}}_t^T - \boldsymbol{\sigma}_t^T\right) dt + \left(\boldsymbol{\sigma}_t^{\varepsilon} + \breve{\boldsymbol{\sigma}}_t^T - \boldsymbol{\sigma}_t^T\right)^{\top} dz_t^{\mathbb{Q}} \right]. \tag{15.15}$$

This is identical to the dynamics of the forward price on a stock, except that the stock price sensitivity vector $\boldsymbol{\sigma}_t^{\mathrm{st}}$ has been replaced by the sensitivity vector for Y_t, which is $\boldsymbol{\sigma}_t^{\varepsilon} + \breve{\boldsymbol{\sigma}}_t^T$. It follows that

$$\begin{aligned} \varepsilon_T = F_T^T = F_t^T \exp\Biggl\{ & \int_t^T \left[-\frac{1}{2} \left\| \boldsymbol{\sigma}_u^{\varepsilon} + \breve{\boldsymbol{\sigma}}_u^T - \boldsymbol{\sigma}_u^T \right\|^2 - \left(\boldsymbol{\sigma}_u^T\right)^{\top} \left[\boldsymbol{\sigma}_u^{\varepsilon} + \breve{\boldsymbol{\sigma}}_u^T - \boldsymbol{\sigma}_u^T \right] \right] du \\ & + \int_t^T \left[\boldsymbol{\sigma}_u^{\varepsilon} + \breve{\boldsymbol{\sigma}}_u^T - \boldsymbol{\sigma}_u^T \right]^{\top} dz_u^{\mathbb{Q}} \Biggr\}. \end{aligned} \tag{15.16}$$

Here $\boldsymbol{\sigma}^{\varepsilon}$, $\breve{\boldsymbol{\sigma}}^T$, and $\boldsymbol{\sigma}^T$ will generally be stochastic processes.

As mentioned above, we may think of F_t^T as the forward price of a traded asset (the foreign zero-coupon bond) with no payments before maturity. We know that the forward price process (F_t^T) is a $\mathbb{Q}^T$-martingale. In particular, $\mathrm{E}_t^{\mathbb{Q}^T}[F_T^T] = F_t^T$, and the drift in F_t^T is zero under the $\mathbb{Q}^T$ measure. Consequently, the dynamics of the forward price F_t^T under the T-forward martingale measure $\mathbb{Q}^T$ is

$$dF_t^T = F_t^T \left(\boldsymbol{\sigma}_t^{\varepsilon} + \check{\boldsymbol{\sigma}}_t^T - \boldsymbol{\sigma}_t^T \right)^{\mathsf{T}} dz_t^T. \tag{15.17}$$

This can also be seen by substituting (15.11) into (15.15).

In order to obtain explicit expressions for the prices on currency derivative securities we will in the following two subsections focus on the case where $\boldsymbol{\sigma}^{\varepsilon}$, $\check{\boldsymbol{\sigma}}^T$, and $\boldsymbol{\sigma}^T$ are all deterministic functions of time. As discussed earlier, deterministic volatilities on zero-coupon bonds are obtained only in Gaussian term structure models, for example the one- or two-factor Vasicek models and Gaussian HJM models.

15.4.3 Currency futures

Let Φ_t^T denote the futures price of the foreign currency with final settlement at time T. From (6.3) we have that

$$\Phi_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[\varepsilon_T\right],$$

where we can insert (15.16). In general the expectation cannot be computed explicitly, but if we assume that $\boldsymbol{\sigma}^{\varepsilon}$, $\check{\boldsymbol{\sigma}}^T$, and $\boldsymbol{\sigma}^T$ are all deterministic, we get

$$\Phi_t^T = F_t^T \exp\left\{ - \int_t^T \boldsymbol{\sigma}^T(u)^{\mathsf{T}} \left(\boldsymbol{\sigma}^{\varepsilon}(u) + \check{\boldsymbol{\sigma}}^T(u) - \boldsymbol{\sigma}^T(u) \right) du \right\}, \tag{15.18}$$

see Exercise 15.3. Amin and Jarrow (1991) demonstrate this under the assumption that both the domestic and the foreign term structure are correctly described by Gaussian HJM models. In particular, we recover the well-known result that $\Phi_t^T = F_t^T$ when $\boldsymbol{\sigma}^T(u) = 0$, that is when the domestic term structure is non-stochastic.

15.4.4 Currency options

Let us consider a European call option on one unit of foreign currency. Let T denote the expiry date of the option and K the exercise price (expressed in the domestic currency). The option grants its owner the right to obtain one unit of the foreign currency at time T in return for a payment of K units of the domestic currency, that is the option payoff is $\max(\varepsilon_T - K, 0)$. According to the analysis in Section 4.4, the value of this option at time $t < T$ is given by

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max\left(\varepsilon_T - K, 0\right)\right],$$

where $\mathbb{Q}^T$ is the T-forward martingale measure. This relation can be used for approximating the option price by Monte Carlo simulations of the terminal exchange rate ε_T under the $\mathbb{Q}^T$ measure. Substituting the relation (15.11) into (15.14), we obtain

$$d\varepsilon_t = \varepsilon_t \left[\left(r_t - \check{r}_t + \left(\boldsymbol{\sigma}_t^{\varepsilon}\right)^{\top} \boldsymbol{\sigma}_t^T \right) dt + \left(\boldsymbol{\sigma}_t^{\varepsilon}\right)^{\top} dz_t^T \right].$$

Therefore, in the general case, we have to simulate not just the exchange rate, but also the short-term interest rates in both countries.

Let us now assume that $\boldsymbol{\sigma}_t^{\varepsilon}$, $\check{\boldsymbol{\sigma}}_t^T$, and $\boldsymbol{\sigma}_t^T$ are all deterministic functions of time. By definition, the forward price with immediate delivery is equal to the spot price so that ε_T can be replaced by F_T^T:

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max\left(F_T^T - K, 0 \right) \right].$$

It follows from (15.17) that the future forward exchange rate F_T^T is lognormally distributed with

$$v_F(t,T)^2 \equiv \mathrm{Var}_t^{\mathbb{Q}^T} \left[\ln F_T^T \right] = \int_t^T \left\| \boldsymbol{\sigma}^{\varepsilon}(u) + \check{\boldsymbol{\sigma}}^T(u) - \boldsymbol{\sigma}^T(u) \right\|^2 du.$$

Note that the future spot exchange rate under these volatility assumptions is also lognormally distributed both under the risk-neutral measure $\mathbb{Q}$ and under the T-forward martingale measure $\mathbb{Q}^T$, but not necessarily under the real-world probability measure. In line with earlier computations, the option price becomes

$$C_t = B_t^T F_t^T N(d_1) - K B_t^T N(d_2),$$

where

$$d_1 = \frac{\ln(F_t^T/K)}{v_F(t,T)} + \frac{1}{2} v_F(t,T),$$
$$d_2 = d_1 - v_F(t,T).$$

We can also insert $F_t^T = \check{B}_t^T \varepsilon_t / B_t^T$ and write the option price as

$$C_t = \varepsilon_t \check{B}_t^T N(d_1) - K B_t^T N(d_2), \tag{15.19}$$

where d_1 can be expressed as

$$d_1 = \frac{1}{v_F(t,T)} \ln\left(\frac{\varepsilon_t \check{B}_t^T}{K B_t^T} \right) + \frac{1}{2} v_F(t,T).$$

Another alternative is obtained by substituting in $B_t^T = e^{-y_t^T(T-t)}$ and $\check{B}_t^T = e^{-\check{y}_t^T(T-t)}$, which yields

$$C_t = \varepsilon_t e^{-\check{y}_t^T(T-t)} N(d_1) - K e^{-y_t^T(T-t)} N(d_2),$$

where d_1 can be written as

$$d_1 = \frac{\ln(\varepsilon_t/K) + [y_t^T - \check{y}_t^T](T-t)}{v_F(t,T)} + \frac{1}{2}v_F(t,T).$$

Similar formulas were first derived by Grabbe (1983). Amin and Jarrow (1991) demonstrate the result for the case where both the domestic and the foreign term structure of interest rates can be described by Gaussian HJM models.

In the best known model for currency option pricing, Garman and Kohlhagen (1983) assume that the short rate in both countries is constant, which implies a constant and flat yield curve in both countries. In that case we have $B_t^T = e^{-r[T-t]}$, $\check{B}_t^T = e^{-\check{r}[T-t]}$, and $\boldsymbol{\sigma}^T(t) = \check{\boldsymbol{\sigma}}^T(t) = 0$. In addition, the sensitivity vector of the exchange rate, that is $\boldsymbol{\sigma}^\varepsilon(t)$, is assumed to be a constant. Hence, the model can be viewed as a simple variation of the Black–Scholes–Merton model for stock options. Under these restrictive assumptions, the option pricing formula stated above will simplify to

$$C_t = \varepsilon_t e^{-\check{r}[T-t]}N(d_1) - Ke^{-r[T-t]}N(d_2), \tag{15.20}$$

where

$$d_1 = \frac{\ln(\varepsilon_t/K) + (r-\check{r})(T-t)}{\|\boldsymbol{\sigma}^\varepsilon\|\sqrt{T-t}} + \frac{1}{2}\|\boldsymbol{\sigma}^\varepsilon\|\sqrt{T-t},$$

$$d_2 = d_1 - \|\boldsymbol{\sigma}^\varepsilon\|\sqrt{T-t}.$$

This option pricing formula is called the **Garman–Kohlhagen formula**. If we compare with Equation (15.19), we see that the extension from constant interest rates to Gaussian interest rates implies (just as for stock options) that the interest rates r and $\check{r}$ in the Garman–Kohlhagen formula (15.20) must be replaced by the zero-coupon yields y_t^T and $\check{y}_t^T$. Furthermore, the relevant variance has to reflect the fluctuations in both the exchange rate and the discount factors. As discussed earlier, the extra terms in the variance tend to be insignificant for stock options, but for currency options the extra terms are typically not negligible.

15.4.5 Alternative exchange rate models

For exchange rates that are not freely floating, the above model for the exchange rate dynamics is inappropriate. For countries participating in a so-called target zone, the exchange rates are only allowed to fluctuate in a fixed band around some central parity. The central banks of the countries are committed to intervening in the financial markets in order to keep the exchange rate within the band. If a target zone is perfectly credible, the exchange rate model has to assign zero probability to future exchange rates outside the band.[1] Clearly, this is not the

[1] This must hold under the real-world probability measure, and since the martingale measures are equivalent to the real-world measure, it will also hold under the martingale measures.

case when the exchange rate is lognormally distributed. Krugman (1991) suggests a more appropriate model for the dynamics of exchange rates within a credible target zone. However, most target zones are not perfectly credible in the sense that the central parities and the bands may be changed by the countries involved. The possibility of these so-called *realignments* may have large effects on the pricing of currency derivatives. Christensen et al. (1997) propose a model for exchange rates in a target zone with possible realignments and show how currency options may be priced numerically within that model. See also Dumas et al. (1995) for a different, but related, model specification.

15.5 FINAL REMARKS

In this chapter we have focused on the pricing of forwards, futures, and European options on stocks and foreign exchange, when we take the stochastic nature of interest rates into account. Under rather restrictive assumptions we have derived Black–Scholes–Merton-type formulas for option prices. Explicit pricing formulas for other securities can be derived under similar assumptions. For example, Miltersen and Schwartz (1998) study the pricing of options on commodity forwards and futures under stochastic interest rates. In contrast to stocks, bonds, exchange rates, and so on, commodities will typically be valuable as consumption goods or production inputs. This value is modelled in terms of a *convenience yield*, see Hull (2009, Ch. 5). In order to be able to price options on commodity forwards and futures the dynamics of both the commodity price and the convenience yield has to be modelled. Miltersen and Schwartz obtain Black–Scholes–Merton-type pricing formulas for such options under assumptions similar to those we have applied in this chapter, for example a Gaussian process for the convenience yield of the underlying commodity.

Another class of securities traded in the international OTC markets is options on foreign securities, for example an option that pays off in euro, but the size of the payoff is determined by a U.S. stock index. The payoff is transformed into euro either by using the dollar/euro exchange rate prevailing at the expiration of the option or a prespecified exchange rate (in that case the option is called a *quanto*). Under particular assumptions on the dynamics of the relevant variables, Black–Scholes–Merton-type pricing formulas can be obtained. Consult Musiela and Rutkowski (1997, Ch. 17) for examples.

In the OTC markets some securities are traded which involve both the exchange rate between two currencies and the yield curves of both countries. A simple example is a currency swap where the two parties exchange two cash flows of interest rate payments, one cash flow determined by a floating interest rate in the first country and the other cash flow determined by a floating interest rate in the other country. Many variations of such currency swaps and also options on these swaps are traded on a large scale. Some of these securities are described in more detail in Musiela and Rutkowski (1997, Ch. 17), who also provide pricing formulas for the case of deterministic volatilities.

15.6 EXERCISES

Exercise 15.1 A stock trades at a current (time $t = 0$) price of 100 and has a constant volatility of 30%. The term structure of interest rates follows the one-factor Vasicek model with a short rate volatility $\beta = 0.03$, a mean reversion speed $\kappa = 0.3$, and a long-term average of $\theta = 0.05$. The market price of interest rate risk is $\lambda = -0.15$. The current short rate is 4%. The correlation between the short rate and the stock price is -0.2 (leading to a correlation of $+0.2$ between the stock price and the price of any bond).

(a) Compute prices of European call options on the stock for all combinations of exercise prices $K \in \{80, 90, 100, 110, 120\}$ and option maturities $T \in \{0.25, 0.5, 1, 2, 5\}$ years. The stock pays no dividends in the life of the option.
(b) Compare the prices from (a) with prices computed using the Black–Scholes formula in which the interest rate is assumed constant and set equal to the true bond yield over the remaining life of the option. How big/small are the price differences in absolute and relative terms?
(c) How sensitive are the option prices to the correlation between the stock price and the short rate? To κ? To β? To the current short rate?

Exercise 15.2 A stock trades at a current (time $t = 0$) price of 100 and has a constant volatility of 30%. The stock pays no dividends in the next 5 years. The term structure of interest rates follows the one-factor Vasicek model with a short rate volatility $\beta = 0.03$, a mean reversion speed $\kappa = 0.3$, and a long-term average of $\theta = 0.05$. The market price of interest rate risk is $\lambda = -0.15$. The current short rate is 4%. The correlation between the short rate and the stock price is -0.2 (leading to a correlation of $+0.2$ between the stock price and the price of any bond).

(a) Compute (and compare) forward prices and futures price of the stock for all delivery dates in $\bar{T} \in \{0.25, 1, 5\}$ years.
(b) Compute prices of European call options on a forward on the stock for all combinations of exercise prices $K \in \{80, 90, 100, 110, 120\}$, option maturities $T \in \{0.25, 0.5, 1\}$ years, and forward delivery dates $\bar{T} \in \{0.25, 1, 5\}$ years (with $\bar{T} \geq T$, of course).
(c) Compute prices of European call options on a futures on the stock for all combinations of exercise prices $K \in \{80, 90, 100, 110, 120\}$, option maturities $T \in \{0.25, 0.5, 1\}$ years, and futures delivery dates $\bar{T} \in \{0.25, 1, 5\}$ years (with $\bar{T} \geq T$, of course). Compare with the prices of options on forwards.
(d) How sensitive are the option prices to the correlation between the stock price and the short rate? To κ? To β? To the current short rate?

Exercise 15.3 Verify Equation (15.18).

16

Numerical Techniques

16.1 INTRODUCTION

Ideally, financial models should be simple enough to allow easily implementable and interpretable closed-form expressions for the most important quantities, that is prices and risk measures for typical assets. Indeed, throughout this book, we have continuously strived for closed-form pricing expressions. Assets with relevant early exercise features—American-style options—are, however, notoriously impossible to price in closed form, so we need to resort to numerical procedures, that is algorithms that can be implemented on a computer and will deliver an approximation of the quantity we were looking for. Some assets without early exercise features have a payoff structure too complicated to derive closed-form pricing expressions even in very simple dynamic models, so they also have to be handled by numerical procedures. Financial models should also match relevant aspects of real-life markets, which demands more complicated multi-dimensional models in which pricing in closed form is even more difficult than in the basic models. Again, one has to resort to numerical procedures.

This chapter provides an introduction to three types of numerical procedures that are useful in various pricing models. The three types of procedures have different strengths and limitations depending on the assets to be priced and the assumed underlying dynamic model. All three approaches are widely used in academic research and in practical pricing and risk management throughout the financial industry. Consequently, both researchers and industry 'quants' should master all three approaches.

Section 16.2 presents the **finite difference approach** to the numerical solution of the partial differential equations (PDEs). In a diffusion-type model the price of a financial asset can be written as a function of time and one or more state variables, and this function solves a certain second-order partial differential equation with an asset-specific terminal condition. We have demonstrated this result in a general setting in Section 4.8 and applied it in several concrete models in Chapters 7 and 8. In the class of affine or quadratic models, the relevant PDE and terminal conditions for a zero-coupon bond and many simple derivatives have closed-form solutions or can at least be reduced to solving certain ordinary differential equations. However, for other assets in the affine or quadratic models and basically all assets in other diffusion models, the appropriate PDE cannot be solved in closed form. In particular, this is true for American-style options for which the pricing function should only satisfy the PDE in the continuation region, that is the combinations of states and times where exercise is suboptimal. The continuation region is not known in advance and has to be determined jointly with the pricing function. The problem is known as a *free boundary* problem and is more complex than solving a PDE with no unknown boundaries. We will go through different variants of the finite difference approach to the standard pricing PDEs and also discuss how the techniques can be amended to cover early exercise features.

Monte Carlo simulation is the second numerical procedure covered in this chapter and is presented in Section 16.3. The starting point of this approach is the characterization of the price of an asset as an expectation of the appropriately discounted payoff under an appropriately risk-adjusted probability measure. In all of the models considered in this book, the discount rate and the payoff depend on some specified continuous-time stochastic processes. The idea of this numerical approach is to simulate a large number of sample paths of these processes. For each sample path the discounted payoff is computed, and the price of the asset is then approximated by the average of the discounted payoffs over all simulated sample paths. The standard deviation of the discounted payoffs provides information on the precision of the price approximation. This basic approach does not apply to American-style options. The continuation value of an American option at a given point depends on all the possible future paths and cannot be determined from just a single path. Consequently, it is impossible to make early exercise decisions and to compute discounted payoffs path-by-path. However, the information contained by the other simulated paths can be used to estimate the continuation value—and, hence, exercise decisions and payoffs—of an American option along a given path. Based on that idea, Monte Carlo simulation has been extended to assets with early exercise features.

Section 16.4 provides a brief introduction to the use of **approximating trees** for computing prices and risk measures in dynamic models. The basic idea is to set up a tree-structure representing a discrete-time approximation of the continuous-time dynamics of the state variables affecting the quantities we are interested in. The initial state is represented by the root node in the tree. Then a layer of nodes is added for each point in time considered in the discretization. Each layer consists of a number of nodes representing the different possible states at that point in time. Nodes at two consecutive layers are connected by branches to represent possible transitions of the state from one point in time to the next, and these branches are assigned probabilities so that the state dynamics in the tree best matches, in a certain sense, the continuous-time dynamics. An asset with a possibly state-dependent payoff at a given point in time is then priced by performing certain recursive computations backwards in time, that is through the different layers of the tree to the root node. The procedure is easily extended to handle early exercise features. While we treat trees as being approximations to continuous-time models, it should be noted that some dynamic term structure models have been formulated directly using trees, for example the model of Ho and Lee (1986), and then the limiting continuous-time model has been derived subsequently.

Section 16.5 provides a short summary and comparison of the three approaches and their applicability to fixed income pricing and risk management. The chapter serves as an introduction to numerical techniques in finance and does not cover all aspects of the different procedures. In particular, we will not go into the highly relevant formal studies of precision, convergence, and stability of the procedures. Several books are devoted solely to numerical methods in finance and contain further information on this topic. For example, Seydel (2009) and Tavella (2002) discuss several methods, whereas Glasserman (2003) focuses on Monte Carlo simulation and Tavella and Randall (2000) address only the finite difference approach. Further references on the different methods will be given throughout the chapter. Some broader presentations of numerical analysis, including many other methods useful in finance, are given, for example, by Kincaid and Cheney (2009) and Press, Teukolsky, Vetterling, and Flannery (2007).

16.2 NUMERICAL SOLUTION OF PDEs

This section gives a short presentation of the finite difference method for the numerical solution of the type of second-order partial differential equation that appears in many financial asset pricing models. We will first describe the approach for a general problem and then discuss an application to a dynamic term structure model at the end of this section. For more information about the techniques discussed here and some alternatives, the interested reader is referred to Wilmott, Dewynne, and Howison (1993), Thomas (1995), Wilmott (1998), Tavella and Randall (2000), and Seydel (2009). Schwartz (1977) and Brennan and Schwartz (1977) were the first papers applying a finite difference method to the pricing of options.

Suppose we want to find a function $f(x,t)$, with $f : \mathcal{S} \times [0,T] \to \mathbb{R}$, which solves the second-order partial differential equation (PDE)

$$\frac{\partial f}{\partial t}(x,t) + \mu(x,t)\frac{\partial f}{\partial x}(x,t) + \frac{1}{2}\sigma(x,t)^2\frac{\partial^2 f}{\partial x^2}(x,t) - r(x,t)f(x,t) = 0,$$
$$(x,t) \in \mathcal{S} \times [0,T), \tag{16.1}$$

with some terminal condition

$$f(x,T) = F(x), \quad x \in \mathcal{S}, \tag{16.2}$$

where $F : \mathcal{S} \to \mathbb{R}$ is a known function and $\mathcal{S} \subseteq \mathbb{R}$. We have seen in Section 4.8 that this is a relevant problem when pricing a financial asset in a diffusion model. Here f is the unknown price of an asset that depends on an underlying state variable x, which takes values in $\mathcal{S}$ with typically $\mathcal{S} = \mathbb{R}$ or $\mathcal{S} = \mathbb{R}_+ \equiv [0,\infty)$. The function r is the short-term continuously compounded risk-free interest rate, which is also assumed to depend on time t and the state variable x at most. In fact, in the classic one-factor diffusion models of Chapter 7, r itself is the state variable, that is $x = r$. The function F is the payoff function of the asset so that $F(x)$ is the state-dependent payoff of the asset at the maturity date T. Finally, μ is the drift rate of the state variable under a risk-neutral probability measure, while σ denotes the absolute volatility of x. In other words, the risk-neutral dynamics of the state variable is assumed to be $dx_t = \mu(x_t,t)\,dt + \sigma(x_t,t)\,dz_t$, where $z = (z_t)$ is a standard Brownian motion under the risk-neutral probability measure.

16.2.1 Discretization of the problem

The approximative solution techniques we consider here are based on a transformation of the problem (16.1)–(16.2) to a sequence of difference equations that can be solved iteratively starting with the known values at the maturity date given by (16.2). For this purpose assume that the x variable can only take on the values

$$x_{\min} \equiv x_0, x_1, x_2, \dots, x_{J-1}, x_J \equiv x_{\max},$$

where $x_{j+1} - x_j = \Delta x$ for all j, that is $x_j = x_0 + j\Delta x$. In particular, $\Delta x = (x_{\max} - x_{\min})/J$. It is intuitively clear that a good approximation to the unknown solution requires that the probability that x takes on values greater than $x_{\max}$ or smaller than $x_{\min}$ has to be negligible. In some cases the choice of one or both of the boundaries $x_{\min}$, $x_{\max}$ is natural, but in other problems a subjective choice has to be made. In many situations you care primarily about the value of the unknown function for one specific value of the state variable, for example $f(x^*, 0)$ where x^* is the current value of the state variable. In that case the boundaries should be imposed sufficiently far from x^* so that the approximate solution $f(x^*, 0)$ is fairly insensitive to small changes in those boundaries. In general, some experimentation with different values of $x_{\min}$ and $x_{\max}$ is often useful. Furthermore, assume that the time variable can only take the $N + 1$ different values

$$0, \Delta t, \dots, N\Delta t = T$$

so that $\Delta t = T/N$. Hence, the state space $\mathcal{S} \times [0, T]$ is approximated by the lattice

$$\{x_0, x_1, \dots, x_J\} \times \{0, \Delta t, \dots, N\Delta t\}.$$

The value of the function f in the lattice node (j, n) corresponding to the x-value x_j and the t-value $n\Delta t$ is denoted by $f_{j,n}$. Similarly, $\mu_{j,n}$ denotes $\mu(x_j, n\Delta t)$, $\sigma^2_{j,n}$ denotes $\sigma(x_j, n\Delta t)^2$, and $r_{j,n}$ represents $r(x_j, n\Delta t)$.

The basic idea of the finite difference approach is to replace the partial derivatives in the PDE by differences. First consider the partial derivative $\frac{\partial f}{\partial x}$. Two obvious candidates for an approximation of $\frac{\partial f}{\partial x}(x_j, n\Delta t)$ are

$$D_x^+ f_{j,n} = \frac{f_{j+1,n} - f_{j,n}}{\Delta x}, \tag{16.3}$$

$$D_x^- f_{j,n} = \frac{f_{j,n} - f_{j-1,n}}{\Delta x}, \tag{16.4}$$

where D_x^+ is called the forward-looking difference operator and D_x^- is called the backward-looking difference operator. A simple graph indicates that the central difference operator D_x given by

$$D_x f_{j,n} = \frac{f_{j+1,n} - f_{j-1,n}}{2\Delta x}, \tag{16.5}$$

will often give a more precise approximation of $\frac{\partial f}{\partial x}$. Hence, we will use that. The second-order derivative $\frac{\partial^2 f}{\partial x^2}$ is approximated by the difference operator D_x^2 given by

$$D_x^2 f_{j,n} = \frac{f_{j+1,n} - 2f_{j,n} + f_{j-1,n}}{(\Delta x)^2}. \tag{16.6}$$

Finally, we have to approximate the derivative $\frac{\partial f}{\partial t}$ appearing in the PDE (16.1). Here, the two obvious choices are a backward-looking difference and a forward-looking difference. In the following two sections we will explore both.

16.2.2 The explicit finite difference approach

First, let us apply a backward-looking difference approximation of the derivative $\frac{\partial f}{\partial t}$, that is we replace $\frac{\partial f}{\partial t}(x_j, n\Delta t)$ in (16.1) by

$$D_t^- f_{j,n} = \frac{f_{j,n} - f_{j,n-1}}{\Delta t}. \tag{16.7}$$

Substituting the approximations (16.5), (16.6), and (16.7) into the PDE (16.1) corresponding to the node (j,n) for $0 < j < J$, $0 < n \leq N$, we get

$$\frac{f_{j,n} - f_{j,n-1}}{\Delta t} + \mu_{j,n} D_x f_{j,n} + \frac{1}{2}\sigma_{j,n}^2 D_x^2 f_{j,n} - r_{j,n} f_{j,n} = 0, \tag{16.8}$$

that is

$$\frac{f_{j,n} - f_{j,n-1}}{\Delta t} + \mu_{j,n}\frac{f_{j+1,n} - f_{j-1,n}}{2\Delta x} + \frac{1}{2}\sigma_{j,n}^2 \frac{f_{j+1,n} - 2f_{j,n} + f_{j-1,n}}{(\Delta x)^2} - r_{j,n} f_{j,n} = 0,$$

which can be rewritten as

$$f_{j,n-1} = \alpha_{j,n} f_{j-1,n} + \beta_{j,n} f_{j,n} + \gamma_{j,n} f_{j+1,n}, \tag{16.9}$$

where

$$\alpha_{j,n} = \frac{1}{2}\Delta t\left(\frac{\sigma_{j,n}^2}{(\Delta x)^2} - \frac{\mu_{j,n}}{\Delta x}\right),$$

$$\beta_{j,n} = 1 - \Delta t\left(r_{j,n} + \frac{\sigma_{j,n}^2}{(\Delta x)^2}\right),$$

$$\gamma_{j,n} = \frac{1}{2}\Delta t\left(\frac{\sigma_{j,n}^2}{(\Delta x)^2} + \frac{\mu_{j,n}}{\Delta x}\right).$$

We can now compute an approximation of $f(x,0)$ by the following simple backward iterative procedure. First, put $f_{j,N} = F(x_j)$ for all j in accordance

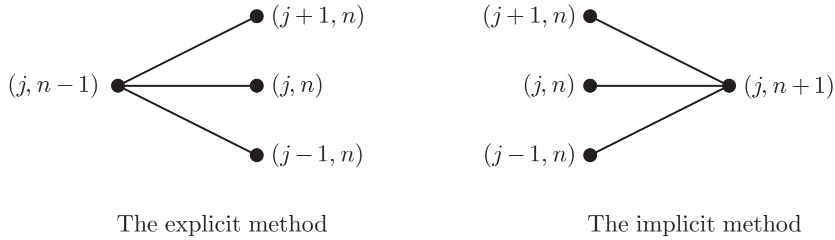

Fig. 16.1: Connections between nodes with the finite difference method.

with the terminal condition (16.2). Then use (16.9) in order to successively go backwards time step by time step until time 0 is reached. Note that the 'new' value $f_{j,n-1}$ is given explicitly in terms of the already computed values $f_{j-1,n}$, $f_{j,n}$, and $f_{j+1,n}$ by (16.9), see the left-hand side of Fig. 16.1. Hence, this method for numerical solution of the PDE is called the **explicit finite difference method.**[1]

Equation (16.9) is no good for $j = 0$ or $j = J$ since the right-hand side will then involve function values outside the lattice. In Section 16.2.5 we discuss how to determine $f_{J,n}$ and $f_{0,n}$. If we only care about a single value $f(x^*, 0) = f(x_0 + j^* \Delta x, 0)$, for example corresponding to the current value x^* of the state variable, it is sufficient to compute $f_{j,n}$ in the triangle defined by the nodes $(j^*, 0)$, $(j^* + N, N)$, and $(j^* - N, N)$. In that case the explicit finite difference method is very similar to the trinomial tree approach discussed in Section 16.4.

It seems obvious to expect and require that a finer lattice (more nodes) should produce a better approximation of the unknown function $f(x,t)$. In particular, we would like the approximate solution to converge to the exact solution when Δx and Δt approach zero. However, it can be shown that in order to ensure the convergence of the explicit finite difference method, Δt has to be very small relative to Δx. Loosely speaking, you have to take a lot of time steps to end up with a good approximation to the unknown solution (if you have implemented the method with time steps that are too large, this will often be obvious since the results will clearly be unrealistic). Although every time step only involves simple computations, see (16.9), the total procedure can thus be quite time-consuming. This conflicts with the desire to attain the approximate solution fast.

16.2.3 The implicit finite difference approach

Now let us try to approximate the partial derivative with respect to time by a forward-looking difference, that is we replace $\frac{\partial f}{\partial t}(x_j, n\Delta t)$ in (16.1) by

[1] The method is also known as the Euler method.

$$D_t^+ f_{j,n} = \frac{f_{j,n+1} - f_{j,n}}{\Delta t}. \tag{16.10}$$

Substituting (16.5), (16.6), and (16.10) into the PDE (16.1), we get the following equation corresponding to the node (j, n) for $0 < j < J$, $0 \leq n < N$:

$$\frac{f_{j,n+1} - f_{j,n}}{\Delta t} + \mu_{j,n} D_x f_{j,n} + \frac{1}{2}\sigma_{j,n}^2 D_x^2 f_{j,n} - r_{j,n} f_{j,n} = 0, \tag{16.11}$$

i.e.,

$$\frac{f_{j,n+1} - f_{j,n}}{\Delta t} + \mu_{j,n} \frac{f_{j+1,n} - f_{j-1,n}}{2\Delta x} + \frac{1}{2}\sigma_{j,n}^2 \frac{f_{j+1,n} - 2f_{j,n} + f_{j-1,n}}{(\Delta x)^2} - r_{j,n} f_{j,n} = 0,$$

which can be rewritten as

$$a_{j,n} f_{j-1,n} + b_{j,n} f_{j,n} + c_{j,n} f_{j+1,n} = f_{j,n+1}, \tag{16.12}$$

where

$$a_{j,n} = -\frac{1}{2}\Delta t \left(\frac{\sigma_{j,n}^2}{(\Delta x)^2} - \frac{\mu_{j,n}}{\Delta x} \right),$$

$$b_{j,n} = 1 + \Delta t \left(r_{j,n} + \frac{\sigma_{j,n}^2}{(\Delta x)^2} \right),$$

$$c_{j,n} = -\frac{1}{2}\Delta t \left(\frac{\sigma_{j,n}^2}{(\Delta x)^2} + \frac{\mu_{j,n}}{\Delta x} \right).$$

Again we want to apply a backward iterative procedure beginning with the known function values at the terminal date, that is $f_{j,N} = F(j\Delta x)$. Suppose we know $f_{j,n+1}$ for all j and thus want to compute $f_{j,n}$ for all j. In contrast to the explicit approach we cannot directly compute $f_{j,n}$ from values already known, see Fig. 16.1. But since (16.12) has to hold for all $j = 1, \ldots, J-1$, we have a system of linked equations involving the unknown function values $f_{j,n}$. More precisely, we have $J-1$ equations in the $J+1$ unknowns $f_{0,n}, \ldots, f_{J,n}$. Therefore we need either to add two more equations (linearly independent of the other equations) or to fix the values of two of the unknowns. In particular, if we add two equations of the form

$$a_{J,n} f_{J-1,n} + b_{J,n} f_{J,n} = d_{J,n+1} \tag{16.13}$$

and

$$b_{0,n} f_{0,n} + c_{0,n} f_{1,n} = d_{0,n+1}, \tag{16.14}$$

the full system of equations will have a particularly nice structure, which simplifies the solution. In order to see this, write Equations (16.12), (16.13), and (16.14) in matrix form as

$$
\begin{pmatrix}
b_{0,n} & c_{0,n} & 0 & 0 & 0 & \dots & 0 \\
a_{1,n} & b_{1,n} & c_{1,n} & 0 & 0 & \dots & 0 \\
0 & a_{2,n} & b_{2,n} & c_{2,n} & 0 & \dots & 0 \\
\vdots & & \ddots & \ddots & \ddots & & \vdots \\
\vdots & & & \ddots & \ddots & \ddots & \vdots \\
0 & \dots & 0 & 0 & a_{J-1,n} & b_{J-1,n} & c_{J-1,n} \\
0 & \dots & 0 & 0 & 0 & a_{J,n} & b_{J,n}
\end{pmatrix}
\begin{pmatrix}
f_{0,n} \\ f_{1,n} \\ f_{2,n} \\ \vdots \\ \vdots \\ f_{J-1,n} \\ f_{J,n}
\end{pmatrix}
=
\begin{pmatrix}
d_{0,n+1} \\ d_{1,n+1} \\ d_{2,n+1} \\ \vdots \\ \vdots \\ d_{J-1,n+1} \\ d_{J,n+1}
\end{pmatrix}. \tag{16.15}
$$

Here $d_{j,n+1} = f_{j,n+1}$ for $j = 1, \dots, J-1$. Given $f_{j,n+1}$ for all j, we can compute $f_{j,n}$ for all j by solving the equation system (16.15). The matrix is *tridiagonal*, which simplifies the solution of the system, as will be discussed in Section 16.2.6. In Section 16.2.5 we discuss how to come up with equations of the form (16.13) and (16.14) at the boundaries of the lattice.

The method described above is called the **implicit finite difference method**.[2] Since it involves solving a system of equations in every time step, it is more complicated to implement than the explicit finite difference method. On the other hand, it can be shown that it is not necessary to use very small time steps relative to Δx to ensure convergence of the approximate solution generated by the implicit method to the true solution. Hence, the implicit method is considered superior to the explicit method.

16.2.4 The Crank–Nicolson approach

Equation (16.8) is the key equation in the explicit finite difference method. If we replace n by $n+1$, it reads

$$\frac{f_{j,n+1} - f_{j,n}}{\Delta t} + \mu_{j,n+1} D_x f_{j,n+1} + \frac{1}{2}\sigma_{j,n+1}^2 D_x^2 f_{j,n+1} - r_{j,n+1} f_{j,n+1} = 0. \tag{16.16}$$

In the implicit finite difference method the key equation is given by (16.11), that is

$$\frac{f_{j,n+1} - f_{j,n}}{\Delta t} + \mu_{j,n} D_x f_{j,n} + \frac{1}{2}\sigma_{j,n}^2 D_x^2 f_{j,n} - r_{j,n} f_{j,n} = 0. \tag{16.17}$$

[2] The method is also known as the backward Euler method.

The so-called **Crank–Nicolson method** is based on taking 'the average' of the two equations (16.16) and (16.17), which gives

$$\frac{f_{j,n+1}-f_{j,n}}{\Delta t}+\frac{1}{2}\left\{\mu_{j,n+1}D_x f_{j,n+1}+\mu_{j,n}D_x f_{j,n}\right\}+\frac{1}{2}\left\{\frac{1}{2}\sigma^2_{j,n+1}D^2_x f_{j,n+1}+\frac{1}{2}\sigma^2_{j,n}D^2_x f_{j,n}\right\}$$
$$-\frac{1}{2}\left\{r_{j,n+1}f_{j,n+1}+r_{j,n}f_{j,n}\right\}=0.$$

After substitution of the difference operators and some manipulations, we arrive at the equation

$$A_{j,n}f_{j-1,n}+B_{j,n}f_{j,n}+C_{j,n}f_{j+1,n}=-A_{j,n+1}f_{j-1,n+1}+B^*_{j,n+1}f_{j,n+1}-C_{j,n+1}f_{j+1,n+1}, \tag{16.18}$$

where

$$A_{j,n}=\frac{1}{4}\Delta t\left(\frac{\sigma^2_{j,n}}{(\Delta x)^2}-\frac{\mu_{j,n}}{\Delta x}\right),$$

$$B_{j,n}=-1-\frac{1}{2}\Delta t\left(\frac{\sigma^2_{j,n}}{(\Delta x)^2}+r_{j,n}\right),$$

$$C_{j,n}=\frac{1}{4}\Delta t\left(\frac{\sigma^2_{j,n}}{(\Delta x)^2}+\frac{\mu_{j,n}}{\Delta x}\right),$$

$$B^*_{j,n}=-1+\frac{1}{2}\Delta t\left(\frac{\sigma^2_{j,n}}{(\Delta x)^2}+r_{j,n}\right).$$

Again we can perform successive backwards iterations, where the right-hand side of (16.18) is known and the left-hand side involves the unknowns $f_{j-1,n}$, $f_{j,n}$, and $f_{j+1,n}$. Again we need to add two equations of the form

$$A_{J,n}f_{J-1,n}+B_{J,n}f_{J,n}=d_{J,n+1}$$

and

$$B_{0,n}f_{0,n}+C_{0,n}f_{1,n}=d_{0,n+1}$$

in order to 'complete the system'. The resulting system of equations has a tridiagonal structure as for the implicit method.

In line with the implicit method, convergence of the Crank–Nicolson method does not require particularly small time steps relative to Δx. The Crank–Nicolson method will typically converge faster than the implicit method and will thus typically give a more precise approximation of the unknown solution for the same lattice.

16.2.5 How to handle the boundaries

In some cases it seems reasonable to fix the value of the asset at the boundaries of the lattice, that is in the nodes (J, n) and $(0, n)$. We then directly obtain equations like (16.13) and (16.14). Say we fix $f_{J,n} = k$, then simply put $a_{J,n} = 0$, $b_{J,n} = 1$, and $d_{J,n+1} = k$. As an example, consider the pricing of a European put option on a stock in the Black–Scholes–Merton model where the state variable x is the price of the underlying stock, and the interest rate r is assumed constant. Since the stock price is lognormally distributed in that model, the obvious choice is to set the lower bound at $x_{\min} = 0$ and to let $f_{0,n} = Ke^{-r[T-n\Delta t]}$, where K is the exercise price of the put option. If the stock price should reach zero, it will stay at zero, so that the payoff of the put option will be K. Conversely, we can set $f_{J,n} = 0$ since the put option will be close to worthless for very high values of the stock.

In many situations, however, it seems difficult to fix values of the unknown function $f(x, t)$ on the boundaries, but it may be reasonable to impose conditions on the derivatives of the function at the boundaries. For some assets it may be a reasonable approximation to impose the condition that the second-order derivative $\frac{\partial^2 f}{\partial x^2}$ is zero at the upper bound. If, furthermore, the first-order derivative $\frac{\partial f}{\partial x}$ at the upper bound is approximated using a one-sided, backward-looking difference as in (16.4), the implicit finite difference method results in

$$\frac{f_{J,n+1} - f_{J,n}}{\Delta t} + \mu_{J,n}\frac{f_{J,n} - f_{J-1,n}}{\Delta x} - r_{J,n}f_{J,n} = 0,$$

which can be rewritten as

$$\frac{\Delta t}{\Delta x}\mu_{J,n}f_{J-1,n} + \left(1 + r_{J,n}\Delta t - \frac{\Delta t}{\Delta x}\mu_{J,n}\right)f_{J,n} = f_{J,n+1}.$$

Note that this equation is of the form (16.13).

The appropriate boundary conditions are dependent on the nature of the underlying model and on the specific asset to be priced. Often both thoughtful consideration and practical experiments are needed to find some appropriate conditions.

16.2.6 How to solve a tridiagonal system of equations

With the implicit finite difference method and the Crank–Nicolson method a number of tridiagonal systems of equations like (16.15) have to be solved. While a mathematician would attack the problem by computing the inverse of the matrix, this is not the most efficient way to solve the system of equations on the computer. A more efficient solution can be implemented as follows. First, perform a **Gauss elimination**:[3]

$$\begin{aligned}
&\text{for } j = 0, 1, \ldots, J-1: \\
&\qquad a_{j+1,n} := a_{j+1,n}/b_{j,n}, \\
&\qquad b_{j+1,n} := b_{j+1,n} - a_{j+1,n}c_{j,n}.
\end{aligned}$$

[3] The symbol := should be interpreted as the assignment operator as it would be understood by a computer. As an example, the declaration $a_{j+1,n} := a_{j+1,n}/b_{j,n}$ means that the variable $a_{j+1,n}$ is assigned a new value equal to its current value divided by the value of the variable $b_{j,n}$. Hence, this new value of $a_{j+1,n}$ should be used in all later calculations until a new value is assigned to the variable.

Second, perform a **forward substitution**:

$$\text{for } j = 0, 1, \ldots, J-1: \quad d_{j+1,n+1} := d_{j+1,n+1} - a_{j+1,n} d_{j,n+1}.$$

Third, perform a **backward substitution**:

$$\begin{aligned}
&f_{J,n} := d_{J,n+1}/b_{J,n},\\
&\text{for } j = J-1, J-2, \ldots, 1, 0: \\
&\qquad f_{j,n} := \left[d_{j,n+1} - c_{j,n} f_{j+1,n}\right]/b_{j,n}.
\end{aligned}$$

Note that if the matrix is the same at all time steps, that is the coefficients $a_{j,n}$, $b_{j,n}$, and $c_{j,n}$ are independent of n, it is only necessary to do the Gauss elimination once. Also note when several systems of equations involving the same matrix—but different right-hand sides—have to be solved, a forward and a backward substitution have to be carried out for each right-hand side, while the Gauss elimination is the same in all cases. This is relevant if several assets are to be priced in the same model.

16.2.7 American-style derivatives

For a derivative asset with American-style exercise features the pricing function $f(x,t)$ has to satisfy the PDE (16.1) only in the continuation region, that is the subset of $\mathcal{S} \times [0,T]$ in which it is not optimal to exercise the derivative. If the continuation region was known, it would be easy to adapt the finite difference technique to this case, but the continuation region is unknown and has to be determined jointly with the pricing function. For standard derivative securities and diffusion models, the continuation region is separated from the exercise region by a security-specific curve $\tilde{x}(t)$, $t \in [0,T]$, the so-called exercise boundary. Some derivatives should be exercised for sufficiently high values of the state variable, and exercise is then optimal at time t if and only if x_t is above $\tilde{x}(t)$. Other derivatives are optimally exercised for sufficiently low values of the state variable, that is if and only if x_t is below $\tilde{x}_t$. In a sense we add a boundary condition to the PDE of the form

$$f(\tilde{x}(t), t) = F(\tilde{x}(t), t),$$

where $F(x,t)$ gives the payoff if the derivative is exercised at time t in a situation where the state variable has the value x. An approximation of the exercise boundary $\tilde{x}(t)$ can be found simultaneously with the pricing function in the backward iterative procedure by checking whether early exercise is optimal. Suppose we know the value of the American-style derivative at time $(n+1)\Delta t$, that is we know $f_{j,n+1}$ for all j. First, compute a value $f_{j,n}$ for all j as explained in the previous sections. Then replace that value with $\max\{f_{j,n}, F(x_j, n\Delta t)\}$. Next, we can go backwards another time step. This simple approach was suggested by Brennan and Schwartz (1977).

The procedure just outlined is first to solve the system of equations for $f_{j,n}$ for all j and then adjust each of the solution values $f_{j,n}$. For the implicit method and the Crank–Nicolson method the early exercise feature can be included in a more efficient way by checking for early exercise during the solution of the equation system. This can be obtained by adjusting the backward substitution slightly:

$$
\begin{aligned}
&f_{J,n} := \max\left\{d_{J,n+1}/b_{J,n}, F(x_J, n\Delta t)\right\},\\
&\text{for } j = J-1, J-2, \dots, 1, 0:\\
&\qquad f_{j,n} := \max\left\{\left[d_{j,n+1} - c_{j,n} f_{j+1,n}\right]/b_{j,n}, F(x_j, n\Delta t)\right\}.
\end{aligned}
$$

Now, when the value of $f_{j,n}$ is computed, the term $f_{j+1,n}$ on the right-hand side has already been modified to take into account the possibility of early exercise in the node $(j+1, n)$. The computed value of $f_{j,n}$ is thus based on a better approximation of $f_{j+1,n}$.

The above procedures are simple and produce results of reasonable accuracy. There are other and computationally more efficient methods, but they are considerably more complex to understand and implement. See Tavella and Randall (2000) and Tangman, Gopaul, and Bhuruth (2008) for more information.

16.2.8 Assets with intermediate payments

As explained in Section 4.8, the price $f(x,t)$ of an asset giving a continuous dividend yield $q(x,t)$ has to satisfy a PDE of the form

$$
\frac{\partial f}{\partial t}(x,t) + \mu(x,t)\frac{\partial f}{\partial x}(x,t) + \frac{1}{2}\sigma(x,t)^2\frac{\partial^2 f}{\partial x^2}(x,t) - (r(x,t) - q(x,t))f(x,t) = 0,
$$
$$
(x,t) \in \mathcal{S} \times [0,T)
$$

with appropriate boundary conditions. Here the new term with $q(x,t)$ does not cause any problems in the finite difference methods discussed above.

Next consider an asset giving discrete payments at predetermined points in time, say $F(x(\tau_i), \tau_i)$ at time τ_i, $i = 1, 2, \dots, n$, where $0 < \tau_1 < \dots < \tau_n = T$. This is, for example, relevant for a coupon bond. The price $f(x,t)$ of such an asset has to satisfy the PDE in between payment dates, that is in all intervals (τ_i, τ_{i+1}). At a given payment date the price of the asset will drop by an amount exactly equal to the size of the payment. We can price such an asset using one of the backward iterative procedures described in earlier sections with the following small adjustment for each intermediate payment $F(x, \tau_i)$. Suppose that $\tau_i \in [n\Delta t, (n+1)\Delta t)$ and that the values $f_{j,n+1}$ for all j have already been computed. Then the values at time $n\Delta t$ are determined by first computing $f_{j,n}$ as described earlier and then replacing $f_{j,n}$ with $f_{j,n} + F(x_j, \tau_i)e^{-r_{j,n}[\tau_i - n\Delta t]}$. Preferably, the grid should be spaced so that the intermediate payment dates are on the grid.

16.2.9 Risk measures

Any finite difference method delivers an approximation of the full pricing function, including current prices for different values of the state variables. While the main interest, of course, is the price at the currently observable value of the state variable, the price approximations for neighbouring grid points are useful for computing risk measures relevant for hedging and risk management purposes. The grid should be constructed so that the current value of the state variable is located on the grid, say equal to x_j, which in general should be relatively far from the boundaries. Then $f_{j,0} = f(x_j, 0)$ denotes the finite difference based approximation of the current price of the asset considered. The most relevant risk measure is the derivative of the price with respect to the state variable, sometimes called the 'Delta' of the asset, which is closely related to the asset's duration as defined in Equation (12.3). This derivative is then approximated by the two-sided difference (16.5), that is

$$\frac{\partial f}{\partial x}(x_j, 0) \approx \frac{f_{j+1,0} - f_{j-1,0}}{2\Delta x}.$$

Another frequently used risk measure is the second-order derivative with respect to the current price. This is called the 'Gamma' of the asset and is closely related to the asset's convexity which was also defined in Chapter 12. We can approximate this using (16.6), that is

$$\frac{\partial^2 f}{\partial x^2}(x_j, 0) \approx \frac{f_{j+1,0} - 2f_{j,0} + f_{j-1,0}}{(\Delta x)^2}.$$

The derivative with respect to time indicates how the price evolves if the state variable should remain unchanged and is approximated by

$$\frac{\partial f}{\partial t}(x_j, 0) \approx \frac{f_{j,1} - f_{j,0}}{\Delta t}.$$

In order to assess the sensitivity of the computed price and risk measures with respect to a parameter of the underlying model, the finite difference routine has to be re-run for the new set of parameters.

16.2.10 Multi-dimensional problems

In multi-factor diffusion models, pricing functions solve a multi-dimensional PDE, compare the general setting in Section 4.8 and the specific models in Chapter 8. For simplicity and concreteness, let us take a dimension of two and look for a function $f(x, y, t)$ with $f : \mathbb{R}_+ \times \mathbb{R}_+ \times [0, T] \to \mathbb{R}$ solving the second-order partial differential equation

$$\begin{aligned}
&\frac{\partial f}{\partial t}(x,y,t) + \mu_1(x,y,t)\frac{\partial f}{\partial x}(x,y,t) + \mu_2(x,y,t)\frac{\partial f}{\partial y}(x,y,t) \\
&+ \frac{1}{2}\sigma_1(x,y,t)^2\frac{\partial^2 f}{\partial x^2}(x,y,t) + \frac{1}{2}\sigma_2(x,y,t)^2\frac{\partial^2 f}{\partial y^2}(x,y,t) \\
&+ \sigma_{12}(x,y,t)\frac{\partial^2 f}{\partial x \partial y}(x,y,t) - r(x,y,t)f(x,y,t) = 0, \\
&\quad (x,y,t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times [0,T),
\end{aligned} \tag{16.19}$$

with the terminal condition, representing the payoff of the asset,

$$f(x,y,T) = F(x,y), \quad \forall (x,y) \in \mathbb{R}_+ \times \mathbb{R}_+.$$

To ease the notation somewhat, we have assumed that $x, y \in \mathbb{R}_+$. We discretize the problem by assuming that (x, y, t) can only take values in

$$\{0, \Delta x, 2\Delta x, \dots, J\Delta x \equiv x_{\max}\} \times \left\{0, \Delta y, 2\Delta y, \dots, K\Delta y \equiv y_{\max}\right\}$$
$$\times \{0, \Delta t, 2\Delta t, \dots, N\Delta t \equiv T\}$$

and define $f_{j,k,n} = f(j\Delta x, k\Delta y, n\Delta t)$. Except for $\partial^2 f/\partial x \partial y$, the partial derivatives can be approximated by differences as in the one-dimensional case. The mixed second-order derivative in $(j\Delta x, k\Delta y, n\Delta t)$ is well approximated by

$$D^2_{xy} f_{j,k,n} = \frac{f_{j+1,k+1,n} - f_{j+1,k-1,n} - f_{j-1,k+1,n} + f_{j-1,k-1,n}}{4\Delta x \Delta y}.$$

Substituting the approximations into the PDE (16.19), we once again obtain either an explicit or implicit system of related equations depending on the approximation of the time derivative. The Crank–Nicolson approach comes from 'averaging' the relevant equations from the implicit and the explicit method. Again, the explicit method cannot be recommended due to bad stability and convergence properties.

The implicit method and the Crank–Nicolson method lead to a set of equations of the form

$$\begin{aligned}
&a_{j,k,n} f_{j-1,k-1,n} + b_{j,k,n} f_{j-1,k,n} + c_{j,k,n} f_{j-1,k+1,n} \\
&+ a'_{j,k,n} f_{j,k-1,n} + b'_{j,k,n} f_{j,k,n} + c'_{j,k,n} f_{j,k+1,n} \\
&+ a''_{j,k,n} f_{j+1,k-1,n} + b''_{j,k,n} f_{j+1,k,n} + c''_{j,k,n} f_{j+1,k+1,n} = d_{j,k,n+1},
\end{aligned}$$

where the right-hand side does not involve values of f at time $n\Delta t$ or earlier. Before collecting these equations—and appropriate equations for the boundaries—in a matrix system $M_n \boldsymbol{f}_n = \boldsymbol{d}_{n+1}$, we have to decide on the ordering of the points (x_j, y_k), that is of the indices (j, k). One obvious choice is $(0, 0), (0, 1), \dots, (0, K), (1, 0), (1, 1), \dots, (1, K), \dots, (J, 0), (J, 1), \dots, (J, K)$. In this case the typical row of the matrix M_n looks like

$$0, \dots, 0, a_{j,k,n}, b_{j,k,n}, c_{j,k,n}, 0, \dots, 0, a'_{j,k,n}, b'_{j,k,n}, c'_{j,k,n},$$
$$0, \dots, 0, a''_{j,k,n}, b''_{j,k,n}, c''_{j,k,n}, 0, \dots, 0$$

with $K - 2$ zeros both between $c_{j,k,n}$ and $a'_{j,k,n}$ and between $c'_{j,k,n}$ and $a''_{j,k,n}$. In total, the most narrow band including all non-zero elements has a width of $3 + (K - 2) + 3 + (K - 2) + 3 = 2K + 5$ elements. Performing the Gauss elimination, this band will typically be full of non-zero elements that have to be computed and stored. The Gauss elimination and forward and backward substitutions are now considerably more time-consuming than for the tridiagonal matrix in the one-dimensional setting. The higher the band-width of the matrix, the more computations are needed. Analogously, ordering the (j, k) indices as $(0,0), (1,0), \dots, (J,0), (0,1), (1,1), \dots, (J,1), \dots, (0,K), (1,K), \dots, (J,K)$ leads to a matrix with a band width of $2J + 5$. For some problems, a high precision may require a finer grid along one dimension than along the other dimension, so that, for example, J is much larger than K. In such cases the ordering minimizing the band width of the matrix should be chosen. Instead of solving the matrix system directly by Gauss elimination and forward/backward substitutions, various iterative solution techniques can be used, for example successive overrelaxation. These are particularly relevant when the matrix has a high bandwidth (or no band structure at all). Note that early exercise features and intermediate payments can be included exactly as in the one-dimensional case.

In principle, the implicit method and the Crank–Nicolson method easily extend to higher-dimensional problems, but the number of computations and the required computer memory grow exponentially, as do the number of boundaries that demand special care. Problems with two or three state variables can be handled, but higher dimensions seem intractable unless the PDE has a specific structure that simplifies computations.

A good alternative to the implicit and Crank–Nicolson methods for multi-dimensional PDEs is the **alternating direction implicit finite difference method** or **ADI method** for short. The basic idea is to use an implicit approximation in only one dimension and explicit approximations in the other dimensions, which leads to a fairly simple system of equations to be solved, and then alternate between which dimension is treated implicitly. We will illustrate how to implement this idea on the two-dimensional PDE

$$\begin{aligned}
\frac{\partial f}{\partial t}(x,y,t) + \mu_1(x)\frac{\partial f}{\partial x}(x,y,t) + \mu_2(y)\frac{\partial f}{\partial y}(x,y,t) + \frac{1}{2}\sigma_1(x)^2\frac{\partial^2 f}{\partial x^2}(x,y,t)& \\
+ \frac{1}{2}\sigma_2(y)^2\frac{\partial^2 f}{\partial y^2}(x,y,t) - r(x,y)f(x,y,t) = 0,& \\
(x,y,t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times [0,T),& \qquad (16.20)
\end{aligned}$$

which is a special case of (16.19). Note that a PDE like (16.20) is relevant, for example, for the Longstaff–Schwartz two-factor model of Section 8.4.2.

Suppose we know $f_{j,k,n+1}$ for all (j,k) and wish to determine $f_{j,k,n}$ for all (j,k). We insert an extra time point $n + \frac{1}{2}$ and go from $n+1$ to $n + \frac{1}{2}$ by treating x implicitly and y explicitly. That is, we approximate $\partial^2 f/\partial x^2(x_j, y_k, (n+1/2)\Delta t)$ and $\partial f/\partial x(x_j, y_k, (n+1/2)\Delta t)$ by $D_x^2 f_{j,k,n+1/2}$ and $D_x f_{j,k,n+1/2}$, while $\partial^2 f/\partial y^2(x_j, y_k, (n+1/2)\Delta t)$ and $\partial f/\partial y(x_j, y_k, (n+1/2)\Delta t)$ are approximated by $\bar{D}_y^2 f_{j,k,n+1}$ and $\bar{D}_y f_{j,k,n+1}$.[4] The PDE (16.20) is thus approximated in the point $(j, k, n + \frac{1}{2})$ for $0 < j < J$ and $0 < k < K$ by

[4] If the PDE involves a mixed derivative $\partial^2 f/\partial x \partial y$, it will also have to be approximated explicitly, that is using known function values only.

$$\frac{f_{j,k,n+1} - f_{j,k,n+\frac{1}{2}}}{\frac{1}{2}\Delta t} + \mu_1(j\Delta x)\frac{f_{j+1,k,n+\frac{1}{2}} - f_{j-1,k,n+\frac{1}{2}}}{2\Delta x}$$
$$+ \mu_2(k\Delta y)\frac{f_{j,k+1,n+1} - f_{j,k-1,n+1}}{2\Delta y}$$
$$+ \frac{1}{2}\sigma_1(j\Delta x)^2\frac{f_{j+1,k,n+\frac{1}{2}} - 2f_{j,k,n+\frac{1}{2}} + f_{j-1,k,n+\frac{1}{2}}}{(\Delta x)^2}$$
$$+ \frac{1}{2}\sigma_2(k\Delta y)^2\frac{f_{j,k+1,n+1} - 2f_{j,k,n+1} + f_{j,k-1,n+1}}{(\Delta y)^2} - r(j\Delta x, k\Delta y)f_{j,k,n+\frac{1}{2}} = 0,$$

which can be rewritten as

$$a_{jk}f_{j-1,k,n+\frac{1}{2}} + b_{jk}f_{j,k,n+\frac{1}{2}} + c_{jk}f_{j+1,k,n+\frac{1}{2}} = A_{jk}f_{j,k-1,n+1} + B_{jk}f_{j,k,n+1} + C_{jk}f_{j,k+1,n+1},$$

where

$$a_{jk} = \frac{\sigma_1(j\Delta x)^2\Delta t}{4(\Delta x)^2} - \frac{\mu_1(j\Delta x)\Delta t}{4\Delta x},$$
$$b_{jk} = -\frac{\sigma_1(j\Delta x)^2\Delta t}{2(\Delta x)^2} - \frac{1}{2}r(j\Delta x, k\Delta y)\Delta t - 1,$$
$$c_{jk} = \frac{\sigma_1(j\Delta x)^2\Delta t}{4(\Delta x)^2} + \frac{\mu_1(j\Delta x)\Delta t}{4\Delta x},$$
$$A_{jk} = -\frac{\sigma_2(k\Delta y)^2\Delta t}{4(\Delta y)^2} + \frac{\mu_2(k\Delta y)\Delta t}{4\Delta y},$$
$$B_{jk} = \frac{\sigma_2(k\Delta y)^2\Delta t}{2(\Delta y)^2} - 1,$$
$$C_{jk} = -\frac{\sigma_2(k\Delta y)^2\Delta t}{4(\Delta y)^2} - \frac{\mu_2(k\Delta y)\Delta t}{4\Delta y}.$$

Again, we have to handle boundary points separately. Suppose that $\sigma_1(0) = \sigma_2(0) = 0$ and $f(\infty, y, t) = f(x, \infty, t) = 0$. Then we can put $f_{J,k,n+\frac{1}{2}} = 0$ and $f_{j,K,n+\frac{1}{2}} = 0$. For $j = 0$, approximate $\partial f/\partial x$ by the one-sided difference $\frac{f_{1,k,n+\frac{1}{2}} - f_{0,k,n+\frac{1}{2}}}{\Delta x}$. Analogously, for $k = 0$, approximate $\partial f/\partial y$ by $\frac{f_{j,1,n+1} - f_{j,0,n+1}}{\Delta y}$. We can then set up simple equations at the boundaries. As in the one-dimensional case, all the equations can be gathered in matrix form

$$M_k f_{\cdot,k,n+\frac{1}{2}} = d_{k,n+1},$$

where $f_{\cdot,k,n+\frac{1}{2}}$ is the vector $(f_{0,k,n+\frac{1}{2}}, f_{1,k,n+\frac{1}{2}}, \ldots, f_{J,k,n+\frac{1}{2}})^{\mathsf{T}}$, M_k is a tridiagonal matrix, and $d_{k,n+1}$ is a vector that can be determined from known values. We have such a system of equations for each k. By solving this system of equations, we find $f_{j,k,n+\frac{1}{2}}$ for all j and this particular value of k. Since we have assumed that $f_{j,K,n+\frac{1}{2}} = 0$, we have to solve such a system of equations for k equal to $0, 1, 2, \dots, K-1$, that is K systems of equations, each of dimension $J+1$, have to be solved. Then we have determined $f_{j,k,n+\frac{1}{2}}$ for all (j,k).

Next we go from time point $n+\frac{1}{2}$ to n by treating x explicitly and y implicitly so that the PDE (16.20) in $(x_j, y_k, n\Delta t)$ is approximated by

$$\begin{aligned}
&\frac{f_{j,k,n+\frac{1}{2}} - f_{j,k,n}}{\frac{1}{2}\Delta t} + \mu_1(j\Delta x)\frac{f_{j+1,k,n+\frac{1}{2}} - f_{j-1,k,n+\frac{1}{2}}}{2\Delta x} \\
&+ \mu_2(k\Delta y)\frac{f_{j,k+1,n} - f_{j,k-1,n}}{2\Delta y} + \frac{1}{2}\sigma_1(j\Delta x)^2\frac{f_{j+1,k,n+\frac{1}{2}} - 2f_{j,k,n+\frac{1}{2}} + f_{j-1,k,n+\frac{1}{2}}}{(\Delta x)^2} \\
&+ \frac{1}{2}\sigma_2(k\Delta y)^2\frac{f_{j,k+1,n} - 2f_{j,k,n} + f_{j,k-1,n}}{(\Delta y)^2} - r(j\Delta x, k\Delta y)f_{j,k,n} = 0,
\end{aligned}$$

which can be rewritten as

$$A_{jk}f_{j,k-1,n} + \hat{B}_{jk}f_{j,k,n} + C_{jk}f_{j,k+1,n} = a_{jk}f_{j-1,k,n+\frac{1}{2}} + \hat{b}_{jk}f_{j,k,n+\frac{1}{2}} + c_{jk}f_{j+1,k,n+\frac{1}{2}},$$

where

$$\hat{B}_{jk} = \frac{\sigma_2(k\Delta y)^2\Delta t}{2(\Delta y)^2} + \frac{1}{2}r(j\Delta x, k\Delta y)\Delta t + 1,$$

$$\hat{b}_{jk} = -\frac{\sigma_1(j\Delta x)^2\Delta t}{2(\Delta x)^2} + 1.$$

Boundary points are handled as in the previous step. For every j, we can thus find $f_{j,k,n}$ for all k by solving a system of equations of the form

$$Q_j\, f_{j,\cdot,n} = e_{j,n+\frac{1}{2}},$$

where Q_j is a tridiagonal matrix, $f_{j,\cdot,n}$ is the vector $(f_{j,0,n}, f_{j,1,n}, \dots, f_{j,K,n})^{\top}$, and $e_{j,n+\frac{1}{2}}$ is a vector depending only on known values. Since we have assumed $f_{J,k,n} = 0$, we have to solve such a system of equations for j equal to $0, 1, 2, \dots, J-1$, that is we have to solve J systems of equations, each of dimension $K+1$, in this step. Then we will finally have found $f_{j,k,n}$ for all (j,k).

Clearly the ADI method involves numerical solution of a high number of systems of equations, but they all have a very simple structure and can be solved very fast. In spite of its relative simplicity, the ADI method does have nice convergence properties, see Thomas (1995). The procedure can be adapted to American-style securities in the same way as the other finite difference method. The ADI method can be extended to higher dimensions but, of course, computational complexity grows fast.

16.2.11 An application to the one-factor CIR model

We will illustrate how the implicit finite difference method is implemented for the one-factor dynamic term structure model that was introduced by Cox, Ingersoll, and Ross (1985b) and thoroughly discussed in Section 7.5. In that model the short-term interest rate itself is the state variable, that is $r = x$, and the relevant PDE is

$$\frac{\partial f}{\partial t}(r,t) + (\hat{\varphi} - \hat{\kappa} r)\frac{\partial f}{\partial r}(r,t) + \frac{1}{2}\beta^2 r \frac{\partial^2 f}{\partial r^2}(r,t) - rf(r,t) = 0, \quad (r,t) \in \mathbb{R}_+ \times [0,T),$$

as can be seen by combining (7.2) and (7.45). As always, the PDE is accompanied by an asset-specific terminal condition, $f(r,T) = F(r)$. The state space $\mathbb{R}_+ \times [0,T)$ is replaced by the grid $\{0, \Delta t, 2\Delta r, \dots, J\Delta r \equiv r_{\max}\} \times \{0, \Delta t, 2\Delta t, \dots, N\Delta t \equiv T\}$ so that an artificial upper bound on the short rate is imposed. In the interior of the grid, the implicit finite difference method leads to Equation (16.12), where the coefficients in this specific model are

$$\begin{aligned}
a_{j,n} &= -\frac{1}{2}\frac{\Delta t}{\Delta r}\left(\beta^2 j - (\hat{\varphi} - \hat{\kappa} j\Delta r)\right),\\
b_{j,n} &= 1 + \Delta t\left(j\Delta r + \frac{\beta^2 j}{\Delta r}\right),\\
c_{j,n} &= -\frac{1}{2}\frac{\Delta t}{\Delta r}\left(\beta^2 j + (\hat{\varphi} - \hat{\kappa} j\Delta r)\right).
\end{aligned}$$

The coefficients are independent of n, because neither the risk-neutral drift nor the volatility of the short rate depend on time.

At the lower boundary $r = 0$, the PDE simplifies to

$$\frac{\partial f}{\partial t}(0,t) + \hat{\varphi}\frac{\partial f}{\partial r}(0,t) = 0.$$

By approximating $\partial f/\partial r(0, n\Delta t)$ by the one-sided difference $(f_{1,n} - f_{0,n})/\Delta t$ and, as before, the derivative with respect to time by $(f_{0,n+1} - f_{0,n})/\Delta t$, we obtain the equation

$$\left(1 + \hat{\varphi}\frac{\Delta t}{\Delta r}\right) f_{0,n} - \hat{\varphi}\frac{\Delta t}{\Delta r} f_{1,n} = f_{0,n+1},$$

which is on the form (16.14) as desired. This can be used for any asset. Concerning the upper bound, first recall that it should be chosen so that the probability of exceeding that bound is negligible. For reasonable parameters in the CIR model, an $r_{\max}$ three or four times the long-term (risk-neutral) average $\hat{\varphi}/\hat{\kappa}$ should be sufficiently high. For interest rates that high, it is often reasonable to assume that the pricing function is roughly linear so that its second-order derivative is close to zero. The PDE then simplifies to

$$\frac{\partial f}{\partial t}(r_J,t) + (\hat{\varphi} - \hat{\kappa} r_J)\frac{\partial f}{\partial r}(r_J,t) - r_J f(r_J,t) = 0.$$

By applying the one-sided approximation $(f_{J,n} - f_{J-1,n})/\Delta r$ of $\frac{\partial f}{\partial r}(r_J,t)$ and the usual approximation of the derivative with respect to time, we arrive at the equation

$$\frac{\Delta t}{\Delta r}\left(\hat{\varphi} - \hat{\kappa} J\Delta r\right) f_{J-1,n} + \left(1 + J\Delta r\Delta t - \frac{\Delta t}{\Delta r}\left(\hat{\varphi} - \hat{\kappa} J\Delta r\right)\right) f_{J,n} = f_{J,n+1},$$

which is of the desired form (16.13). We can now stack up all the equations and obtain the matrix equation $Mf_n = d_{n+1}$, where the matrix M is tridiagonal and is the same for all time steps so that the Gauss elimination has only to be performed once.

Assume the parameter values $\beta = 0.2$, $\hat{\kappa} = 0.3$, and $\hat{\varphi} = 0.02$. We consider a 5-year zero-coupon bond with a face value of 100. For illustrative purposes, we first use a very coarse grid with (i) $\Delta r = 0.005 = 0.5\%$ and $r_{\max} = 0.15 = 15\%$ so that $J = 30$ and (ii) $\Delta t = 0.25$ so that only $N = 20$ time steps are taken. All the necessary calculations can then be handled in a simple spreadsheet. Table 16.1 presents part of the results from an application of the implicit finite difference method. The table shows that the method delivers time 0 prices very close to those calculated using the closed-form expression (7.46). Even with this very coarse grid, the finite difference based price deviates by less than 0.1% for the most relevant levels of the interest rate. Near the upper bound the prices are less precise, but if you are interested in prices at such high interest rate levels, you should pick a higher $r_{\max}$. Figure 16.2 shows the percentage deviation of the bond price computed with the implicit finite difference method for various combinations of J, N, and $r_{\max}$ and confirms the intuition that finer grids generally lead to more accurate prices. However, holding N fixed, increasing J might not lead to better results. Conversely, for a fixed J, increasing N might lead to less accurate price estimates. The best convergence is obtained when J and N are increased simultaneously according to a certain relation. Note that even for the relatively fine grid with $J = 20,000$ and $N = 100$, all the prices can be calculated within a couple of seconds on any modern computer.

Next, we consider 2-year options on a 5-year zero-coupon bond. To illustrate early exercise, we take a put option with an exercise price of 82. As before, the face value of the underlying bond is 100, and we use the same model parameters. Table 16.2 shows prices of both European and American puts computed by the implicit finite difference method using $J = 2000$, $r_{\max} = 20\%$, and $N = 100$ time steps over the life of the option. The known closed-form expression for the zero-coupon bond price in the CIR model is used to determine the option payoff upon exercise. The table also lists the price of the European put computed using the closed-form expression and the percentage deviation of the price computed by the implicit method. The price deviations are fairly small, although it should be possible to achieve better accuracy than 2%. However, note that the closed-form expression also involves an approximation to implement the cumulative distribution function of the non-central χ^2-distribution. Moreover, we have not tried to fine-tune the finite difference implementation. It is not surprising that the method is more accurate for the zero-coupon bond with a state-independent terminal value than for options with a state-dependent payoff with a kink at the exercise price.

With the stated parameter values, the table shows that the price of the American put is considerably higher than the price of the corresponding European put, so the early exercise feature is highly valuable. It is easy to compute the critical short rate over the life of the option (that is the exercise boundary) when checking for early exercise while going backwards in time according to the finite difference method. The critical short rate at a specific point in time is the value of the short rate where you are indifferent between exercising the American option and keeping it

Table 16.1: The implicit finite difference method applied to a 5-year zero-coupon bond in the one-factor CIR model. The model parameter values are $\beta = 0.2$, $\hat{\kappa} = 0.3$, and $\hat{\varphi} = 0.02$. The face value of the bond is 100. The approximating grid is based on $J = 30$, $r_{\max} = 0.15 = 15\%$, and $N = 20$ so that $\Delta r = 0.5\%$ and $\Delta t = 0.25$.

r (%)	Analytic price	Numerical error (%)	Time						
			0	0.25	0.5	...	4.5	4.75	5
0.0	85.707	−0.10	85.622	86.637	87.644	...	99.662	99.884	100
0.5	84.684	−0.09	84.606	85.630	86.648	...	99.440	99.769	100
1.0	83.673	−0.08	83.605	84.638	85.666	...	99.219	99.653	100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
4.0	77.855	−0.04	77.823	78.900	79.986	...	97.906	98.967	100
4.5	76.925	−0.04	76.894	77.977	79.073	...	97.690	98.853	100
5.0	76.007	−0.04	75.973	77.064	78.168	...	97.474	98.740	100
5.5	75.099	−0.05	75.062	76.158	77.271	...	97.259	98.627	100
6.0	74.203	−0.06	74.159	75.262	76.382	...	97.044	98.514	100
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
14.0	61.229	−1.02	60.604	61.790	63.025	...	93.690	96.737	100
14.5	60.498	−1.17	59.791	60.982	62.223	...	93.484	96.627	100
15.0	59.776	−1.33	58.978	60.174	61.421	...	93.277	96.518	100

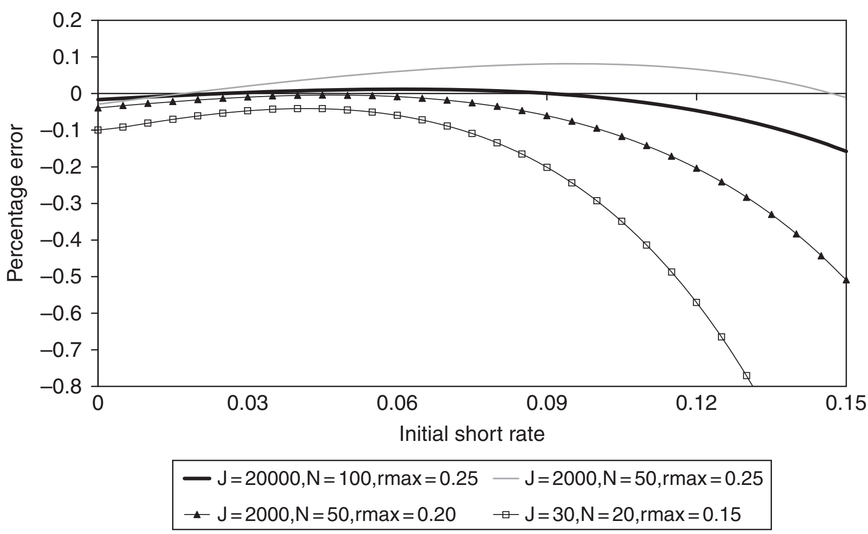

Fig. 16.2: The implicit finite difference method applied to a 5-year zero-coupon bond in the one-factor CIR model. The model parameter values are $\beta = 0.2$, $\hat{\kappa} = 0.3$, and $\hat{\varphi} = 0.02$. The face value of the bond is 100. The graphs show the percentage deviation of the prices calculated with the implicit finite difference method relative to the prices calculated with the closed-form solution for different approximating grids.

Table 16.2: Bond option prices in the one-factor CIR model. The model parameter values are $\beta = 0.2$, $\hat{\kappa} = 0.3$, and $\hat{\varphi} = 0.02$. The bond is a zero-coupon bond with 5 years to maturity and a face value of 100. The options expire in 2 years and have an exercise price of 82. The numerical prices are computed by the implicit finite difference method with $J = 2000$, $r_{\max} = 20\%$, and $N = 100$.

r (%)	European put			American put
	Analytic price	Numerical price	pct. deviation	Numerical price
1	0.6555	0.6448	−1.6353	1.3412
2	0.9267	0.9053	−2.3099	2.1382
3	1.2152	1.1867	−2.3457	3.1465
4	1.5185	1.4866	−2.0995	4.4049
5	1.8342	1.8026	−1.7228	5.9930
6	2.1601	2.1324	−1.2836	7.7972
7	2.4943	2.4740	−0.8143	9.5585
8	2.8348	2.8254	−0.3305	11.2780
9	3.1798	3.1849	0.1604	12.9567
10	3.5279	3.5510	0.6559	14.5955

alive. You should exercise a put on a bond when the bond price is sufficiently low, that is when the short rate is sufficiently high. Figure 16.3 shows the critical short rates for the 2-year American put on the 5-year zero-coupon bond as computed in the finite difference implementation. Early exercise is optimal above the curve. In particular, if the current short rate is above 4.8%, the American put should be exercised immediately.

In Exercise 16.6, you are asked to implement the implicit finite difference method for the one-factor Vasicek model and apply it to the pricing of bonds and bond options, as done above for the CIR model.

16.3 MONTE CARLO SIMULATION

This section introduces the Monte Carlo simulation approach to the pricing of assets in financial models. The approach builds on the result that the no-arbitrage price of an asset equals the expectation (under an appropriate probability measure) of the (appropriately discounted) payoff of the asset. The procedure is to simulate a large number of sample paths of the stochastic processes affecting the (discounted) payoff. The average (discounted) payoff over all sample paths leads to an estimate of the price of the asset. In finance, the approach was first applied to the pricing of European options by Boyle (1977) and has since been improved and generalized in various directions and applied to many different assets and models. The basic Monte Carlo approach is very intuitive and easy to implement with standard software tools. The Monte Carlo approach is becoming increasingly popular in the financial industry for, at least, the following two reasons. First, the computational complexity of the Monte Carlo approach grows only linearly in the dimension of the problem (that is the number of standard Brownian motions generating the underlying uncertainty), in contrast to the finite difference approach

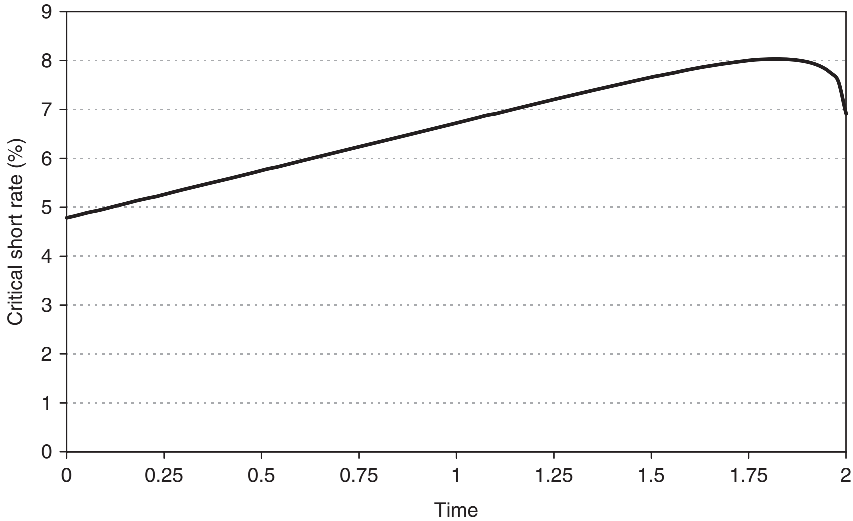

Fig. 16.3: The critical short rate for the American put on a zero-coupon bond in the one-factor CIR model. The model parameter values are $\beta = 0.2$, $\hat{\kappa} = 0.3$, and $\hat{\varphi} = 0.02$. The bond is a zero-coupon bond with 5 years to maturity and a face value of 100. The option expires in 2 years and has an exercise price of 82. The critical short rate at a specific point in time is the value of the short rate where the value of keeping the option alive (and exercising it later when optimal) equals the payoff from immediate exercise. A put on a bond is optimally exercised whenever the short rate is above the critical rate. The critical short rates are determined by the implicit finite difference method with $J = 2000$, $r_{\max} = 20\%$, and $N = 100$.

and the approximating tree approach. Monte Carlo simulation is thus well-suited for high-dimensional models, which are gaining attention and support both among researchers and practitioners. Second, while the basic Monte Carlo approach does not apply to American-style options, the approach has been successfully extended to these assets in recent years. Since many fixed income securities have early exercise features, and numerical methods are almost always required to price them, the extension of Monte Carlo to American-style options is highly valuable for practitioners. Furthermore, the Monte Carlo approach can handle assets with path-dependent payoffs (such as the so-called lookback options) and models in which the relevant stochastic processes are not diffusions (such as some models in the Heath-Jarrow-Morton class considered in Chapter 10). The main problematic feature of the approach is the computation time which is long because a large number of sample paths has to be simulated to produce highly accurate price approximations, but there are various ways to speed up the computations.

We will first describe the basic approach and its applications in some dynamic term structure models. Then we discuss various ways of improving the precision of the price approximations produced. We present some numerical results for the one-factor CIR model. Finally, the extension to American-style options is briefly presented. For more information about Monte Carlo and its applications in finance, we refer the interested reader to the books of Glasserman (2003) and Asmussen and Glynn (2007). Further references are given below.

16.3.1 The basics

Suppose you want to know the expectation (under some probability measure) of a function F of a random variable x, that is $\mathrm{E}[F(x)]$, but you cannot compute it explicitly. By simulating a number of samples $x^1, \dots, x^M$ of the random variable, you can approximate the expectation by a simple average

$$\mathrm{E}[F(x)] \approx \bar{F}_M \equiv \frac{1}{M} \sum_{m=1}^{M} F(x^m). \tag{16.21}$$

Intuitively, this will only be a good approximation if the distribution of the samples is representative of the true distribution of the random variable x, which typically requires M to be large. Here x might be the value of a stochastic process (x_t) at a given time, say time T, where the dynamics of the process is known. For a diffusion process with dynamics

$$dx_t = \mu(x_t, t)\, dt + \sigma(x_t, t)\, dz_t, \tag{16.22}$$

the exact distribution of x_T is only known for some functions μ and σ. If the exact distribution of x_T is unknown, it is necessary to simulate entire sample paths in order to end up with samples of x_T.

The generation of a sample path involves splitting the relevant time interval, say $[0, T]$, into a number of subintervals of length Δt, and setting up an iterative procedure for computing $x_{t+\Delta t}$ from x_t and a sample value of the random shock $z_{t+\Delta t} - z_t$ to the process over the subinterval. The simplest such procedure is based on the Euler approximation of the above dynamics, which is

$$x_{t+\Delta t} = x_t + \mu(x_t, t)\Delta t + \sigma(x_t, t)\varepsilon\sqrt{\Delta t}, \tag{16.23}$$

where ε is a sample from the standard normal $N(0, 1)$ distribution so that $\varepsilon\sqrt{\Delta t}$ has the same distribution as $z_{t+\Delta t} - z_t$. Iterating forward time step by time step, using independent samples of ε, will produce (an approximation of) a sample path of (x_t) and, in particular, one sample value of x_T and an associated value $F(x_T)$. After M repetitions of this procedure, the approximation in (16.21) can be computed. Note that F can be allowed to be path-dependent, that is dependent on values of the process (x_t) at more than one point in time. If entire sample paths are simulated anyway, this will not significantly complicate the procedure.

As explained in Chapter 4 and applied throughout the book, the no-arbitrage price of a financial asset is the expectation of the appropriately discounted payoff under an appropriately risk-adjusted probability measure. The Monte Carlo simulation approach is thus directly applicable to the pricing of financial assets. Let us illustrate the idea in a one-factor diffusion model in which the risk-neutral dynamics of the short rate is of the form

$$dr_t = \mu^{\mathbb{Q}}(r_t, t)\, dt + \beta(r_t, t)\, dz_t^{\mathbb{Q}}, \tag{16.24}$$

where $z^{\mathbb{Q}} = (z_t^{\mathbb{Q}})$ is a one-dimensional standard Brownian motion under the risk-neutral probability measure $\mathbb{Q}$. Suppose we want to price an asset with a payoff at time T given by $H(r_T)$ for some function H. Then we know that the price P_t at time t is given by

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} H(r_T)\right]. \tag{16.25}$$

The idea is then to generate M samples of r_T and $I_T = \int_t^T r_u\,du$ from their joint $\mathbb{Q}$-distribution. Let r_T^m and I_T^m denote the values in the m'th sample, where $m = 1, 2, \dots, M$. The Monte Carlo approximation of the price is then

$$P_t \approx P_t^{\mathrm{MC}} \equiv \frac{1}{M}\sum_{m=1}^{M} e^{-I_T^m} H(r_T^m). \tag{16.26}$$

How are the samples (r_T^m, I_T^m) generated? In some models, the short-rate process is so simple that the joint $\mathbb{Q}$-distribution of r_T and I_T is known and one can then draw samples directly from that distribution. More generally, the joint distribution is unknown and it is necessary to simulate entire sample paths of the short rate, for example by the Euler approximation scheme

$$r_{t+\Delta t} = r_t + \mu^{\mathbb{Q}}(r_t, t)\Delta t + \beta(r_t, t)\varepsilon\sqrt{\Delta t}. \tag{16.27}$$

The time period $[t, T]$ is divided into N sub-intervals of length Δt separated by the time points $t_n = t + n\Delta t$, $n = 0, 1, \dots, N$. In particular, $t_N = T$ and $N = (T - t)/\Delta t$. We can now generate M discretized sample paths enumerated by $m = 1, 2, \dots, M$ of the short-rate process by iteratively computing

$$r_{t_{n+1}}^m = r_{t_n}^m + \mu^{\mathbb{Q}}(r_{t_n}^m, t_n)\Delta t + \beta(r_{t_n}^m, t_n)\varepsilon_{n+1}^m\sqrt{\Delta t}, \quad n = 0, 1, \dots, N-1. \tag{16.28}$$

Here the $M \cdot N$ different values ε_n^m are independent draws from a standard normal distribution. In particular, this gives us a sample of the terminal short rate r_T^m, and we can approximate the integral in (16.25) by a simple sum $I_T^m = \sum_{n=0}^{N-1} r_{t_n}^m\,\Delta t$. Again, the price is then approximated by (16.26).

Monte Carlo simulation is often implemented in a programming language using a 'for $m = 1$ to M do ...' algorithm so that the sum in (16.26) is iteratively updated as additional samples are generated. As discussed later, a high value of M (in many cases 10,000 or even higher) is often needed to obtain a good price approximation. In principle, all the computations can be done in a simple spreadsheet, for example, by simultaneously generating all the M paths in a spreadsheet with M rows and N columns, but if M is very high this becomes highly impractical and time-consuming.

If you want to use Monte Carlo simulation to approximate prices of several different assets depending only on the short rate process, you might as well use the same set of simulated short rate paths for all assets. It is then computationally efficient to price the assets simultaneously since then there is no need to store paths for later use. When a given short rate path, say path number m, is generated, then the (discounted) payoff is computed for each asset and added to the sum in (16.26). Then you can move on to path number $m + 1$.

As we have seen in earlier chapters, it is sometimes advantageous to perform a change of numeraire and an associated change of measure so that the price of the asset considered can be expressed on the form

$$P_t = S_t \, \mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{H(r_T)}{S(r_T)}\right],$$

where $S = (S_t)$ is the price process of the numeraire asset and $\mathbb{Q}^S$ is the associated risk-adjusted probability measure. Here we have assumed that both the numeraire and the asset be priced depending solely on the short rate. A Monte Carlo based approximation of the price P_t now involves simulating draws of r_T under the $\mathbb{Q}^S$-measure in line with the $\mathbb{Q}^S$-dynamics

$$dr_t = \mu^{\mathbb{Q}^S}(r_t, t)\, dt + \beta(r_t, t)\, dz_t^{\mathbb{Q}^S},$$

which may again necessitate simulating entire sample paths of the short rate. In this case, only draws of the terminal short rate are needed, not draws of the integral $\int_t^T r_u \, du$. Both the payoff function of the asset to be priced and the terminal value of the numeraire can be allowed to depend on the entire path of the short rate. Recall that pricing under the risk-neutral measure $\mathbb{Q}$ is the special case where the numeraire is the balance of the bank account, that is $S_t = \exp\left(\int_0^t r_u \, du\right)$, which is obviously path-dependent. Although the Monte Carlo approach can be used to approximate the price of virtually any interest rate dependent asset via (16.26), it is worthwhile thinking about a change of measure that will simplify the computations.

In Gaussian models, the joint risk-neutral distribution of $I_T = \int_t^T r_u \, du$ and r_T, or the distribution of the short rate r_T under another relevant measure $\mathbb{Q}^S$, is known, so if only assets maturing on the same date T are to be priced, there is no need to simulate the entire short rate path over $[t, T]$. However, if various assets maturing at many different dates are to be priced, then it may be computationally more efficient to simulate a set of M paths up to the latest maturity date and use that set of paths to value all the assets, rather than performing separate simulations for each maturity date or even each single asset. As it can easily become very time-consuming to apply the Monte Carlo approach, one should think carefully about what information is needed and how the algorithm is best designed and implemented before beginning the computer implementation.

The Monte Carlo approach easily extends to settings where the underlying uncertainty is generated by a multi-dimensional standard Brownian motion. Suppose z in (16.22) is of dimension d. Then ε in the Euler approximation (16.23) must also have dimension d, that is ε should be a draw from the d-dimensional standard normal distribution, which is just a vector of d independent draws from the one-dimensional standard normal distribution. The work load grows roughly linearly with the dimension of the problem, in contrast to, for example, the finite difference methods where the number of grid points and computations grows exponentially. While finite difference methods therefore are virtually impossible to implement for problems of a dimension higher than three, the Monte Carlo approach is applicable for high-dimensional problems.

16.3.2 Approximation schemes for continuous-time processes

A key part of the Monte Carlo approach is the discrete-time approximation of the continuous-time dynamics. Above, the short-rate process defined by the dynamics (16.24) was approximated by (16.27). This is the so-called Euler approximation. For some continuous-time processes we can easily come up with better approximations, which should then also lead to more precise price approximations. As a first example, consider the Ornstein–Uhlenbeck process introduced in Section 3.8.2 and used in the one-factor Vasicek model in which the short rate has risk-neutral dynamics

$$dr_t = \kappa[\hat{\theta} - r_t]\, dt + \beta\, dz_t^{\mathbb{Q}},$$

see (7.35). Here the exact risk-neutral distribution of the future short rate is well-known,

$$r_{t+\Delta t} | r_t \sim N\left(\hat{\theta} + (r_t - \hat{\theta})e^{-\kappa \Delta t}, \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa \Delta t}\right)\right).$$

Equivalently, we can write

$$r_{t+\Delta t} = e^{-\kappa \Delta t} r_t + \hat{\theta}(1 - e^{-\kappa \Delta t}) + \varepsilon\beta\sqrt{\frac{1}{2\kappa}(1 - e^{-2\kappa \Delta t})}, \tag{16.29}$$

where $\varepsilon \sim N(0,1)$. In contrast, by using (16.27), we obtain the approximation

$$r_{t+\Delta t} = r_t + \kappa(\hat{\theta} - r_t)\Delta t + \beta\varepsilon\sqrt{\Delta t} = (1 - \kappa\Delta t)r_t + \kappa\hat{\theta}\Delta t + \beta\varepsilon\sqrt{\Delta t}, \tag{16.30}$$

which also leads to a normally distributed $r_{t+\Delta t}$, but with the wrong variance and mean. Note that the Euler approximation (16.30) can also be seen as the result of applying the first-order Taylor approximation $e^x \approx 1 + x$ to the exponential terms in the exact relation (16.29). The difference between the two recursive schemes will be small when Δt is small, but a small Δt requires many time steps and thus many draws of random numbers, which will increase the computation time. In any case, one might as well use the exact relation (16.29) instead of the Euler approximation (16.30) as the computational effort per time step is similar.

As another example, consider a geometric Brownian motion $x = (x_t)$ with dynamics of the form

$$dx_t = \mu x_t\, dt + \sigma x_t\, dz_t,$$

where $z = (z_t)$ is a standard Brownian motion under the relevant pricing measure. Here, the Euler approximation is

$$x_{t+\Delta t} = x_t + \mu x_t \Delta t + \sigma x_t \varepsilon\sqrt{\Delta t}, \tag{16.31}$$

where $\varepsilon \sim N(0,1)$. As for any other process, the Euler approximation leads to a normally distributed $x_{t+\Delta t}$ conditional on x_t. However, the exact solution is known to imply that

$$x_{t+\Delta t} = x_t \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma(z_{t+\Delta t} - z_t)\right\},$$

see Section 3.8.1, so simulations should use

$$x_{t+\Delta t} = x_t \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)\Delta t + \sigma\varepsilon\sqrt{\Delta t}\right\}. \tag{16.32}$$

The relation (16.32) produces the correct lognormal distribution. Not only is (16.32) more precise than (16.31), the Euler approximation (16.31) also induces the risk of obtaining negative values of x, although the continuous-time process stays non-negative. If x_t is close to zero, a large negative draw of ε may indeed lead to a negative $x_{t+\Delta t}$. The exact discretization (16.32) avoids negative values of x. Geometric Brownian motions are widely used in the modelling of stock prices, but are also used in the market models for the pricing of caps, floors, and swaptions studied in Chapter 11.

Another stochastic process often used is the square-root process introduced in Section 3.8.3. For example, in the Cox–Ingersoll–Ross model of Section 7.5, the short-term interest rate is assumed to have dynamics of the form

$$dr_t = \kappa(\theta - r_t)\,dt + \beta\sqrt{r_t}\,dz_t \tag{16.33}$$

both under the real-world probability measure and under the risk-neutral probability measure (with different values of the constants κ and θ under the two measures). Recall that future values of the short rate are then non-centrally χ^2 distributed. The 'brute force' Euler approximation is

$$r_{t+\Delta t} = r_t + \kappa(\theta - r_t)\Delta t + \beta\sqrt{r_t}\,\varepsilon\sqrt{\Delta t},$$

which again implies that $r_{t+\Delta t}$ is normally distributed conditional on r_t. In particular, the Euler approximation involves the risk of obtaining negative rates, which is inconsistent with the original square-root process and, moreover, makes it impossible to compute the square-root in the next time step so that the simulation procedure is not well-defined. There are various ad hoc fixes. With a minor adjustment of the square-root term,

$$r_{t+\Delta t} = r_t + \kappa(\theta - r_t)\Delta t + \beta\sqrt{\max(r_t, 0)}\,\varepsilon\sqrt{\Delta t}, \tag{16.34}$$

we can implement the procedure, but we may still run into negative rates. We can avoid negative rates by using

$$r_{t+\Delta t} = \max\left(0, r_t + \kappa(\theta - r_t)\Delta t + \beta\sqrt{r_t}\,\varepsilon\sqrt{\Delta t}\right), \tag{16.35}$$

or by introducing the absolute value

$$r_{t+\Delta t} = \left|r_t + \kappa(\theta - r_t)\Delta t + \beta\sqrt{r_t}\,\varepsilon\sqrt{\Delta t}\right|. \tag{16.36}$$

With any of the three alternatives we miss the correct distribution and moments of $r_{t+\Delta t}$.

It is, in fact, possible to simulate $r_{t+\Delta t}$ in a way that produces the exact non-central χ^2 distribution of the square-root process. However, it is quite time-consuming and involves sampling from a central χ^2 and a Poisson distribution, which is somewhat more complicated than just sampling from a standard normal distribution as in the other simulation procedures discussed in this presentation. The exact simulation procedure is described in detail in Glasserman (2003, Sec. 3.4). Andersen (2008) presents two interesting alternative simulation schemes of medium complexity which, when compared to the simpler alternatives outlined above, perform well in his numerical tests. For small-scale simulations involving square-root processes where high precision and high computational speed are not so important, the coarser simulation procedures given by (16.34)–(16.36) generally do a decent job as can be seen in Section 16.3.6 below.

16.3.3 Normally distributed random numbers

The Monte Carlo simulation schemes given above require a sequence of independent draws (or samples) from the standard normal distribution, $N(0,1)$. Some spreadsheet applications, programming environments, and other software tools may have a built-in procedure for generating such draws, but not all of them are of a good quality, that is if you use the procedure for generating a number of such draws, the distribution of these draws may be quite different from the standard normal distribution. If your favorite computer tool does not have a built-in generator of normally distributed random numbers, or you do not trust the output from it, you can generate draws from the $N(0,1)$ distribution by transforming draws from a uniform distribution on the unit interval, a distribution we will denote by $U[0,1]$. Most computer tools used for financial applications have a built-in generator of random numbers from the $U[0,1]$ distribution, but there are also algorithms for generating these draws that can easily be implemented in any programming environment, see for example Press, Teukolsky, Vetterling, and Flannery (2007, Ch. 7).

There are several ways to transform independent draws from the uniform $U[0,1]$ distribution into independent draws from the normal $N(0,1)$ distribution. A popular choice is the so-called Box–Muller transformation suggested by Box and Muller (1958). Given two draws U_1 and U_2 from the uniform $U[0,1]$ distribution, ε_1 and ε_2 defined by

$$\varepsilon_1 = \sqrt{-2\ln U_1}\,\cos(2\pi U_2), \qquad \varepsilon_2 = \sqrt{-2\ln U_1}\,\sin(2\pi U_2)$$

are two independent draws from the standard normal distribution. An alternative approach is to transform a draw U from the $U[0,1]$ distribution into a draw ε from the $N(0,1)$ distribution by

$$\varepsilon = N^{-1}(U),$$

where $N^{-1}(\cdot)$ denotes the inverse of the probability distribution function $N(\cdot)$ associated with the standard normal distribution, that is $N(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp(-z^2/2)\,dz$. This follows from the fact that $\mathbb{P}(\varepsilon < a) = \mathbb{P}(N^{-1}(U) < a) = \mathbb{P}(U < N(a)) = N(a)$. Of course, this approach requires an implementation of the inverse normal distribution $N^{-1}(\cdot)$, which is not known in closed form. Again, some software tools (such as Microsoft Excel) have a built-in algorithm for computing the inverse normal distribution, but the precision of the algorithm is generally unknown to the user, and the computation is bound to be more time-consuming than when using the Box–Muller transformation.[5]

16.3.4 Risk measures

In addition to the price of an asset, traders and analysts are also interested in various risk measures, such as the sensitivity of the price with respect to small changes in state variables or key parameters. For example, in one-factor diffusion models, the duration and convexity measures defined and discussed in Section 12.3 contain important information about the market risks associated with fixed income assets. Suppose the price of the asset is given as a function $B(r,t)$ of the short-term interest rate and time, then the duration $D(r,t)$ and convexity $K(r,t)$ are defined as

$$D(r,t) = -\frac{1}{B(r,t)}\frac{\partial B}{\partial r}(r,t), \qquad K(r,t) = \frac{1}{2B(r,t)}\frac{\partial^2 B}{\partial r^2}(r,t),$$

compare (12.3) and (12.4). It should be clear by now how the price $B(r,t)$ can be approximated by Monte Carlo simulation of the short rate process under an appropriate pricing measure. As all the paths used for computing the price have initial values equal to the current short rate r, they contain no information about the price of the asset at slightly different current short rates, which is needed to approximate the derivatives $\frac{\partial B}{\partial r}(r,t)$ and $\frac{\partial^2 B}{\partial r^2}(r,t)$. Therefore, additional simulations are needed.

As explained when developing the finite difference techniques in Section 16.2, we can approximate the first-order derivative either by the one-sided upward difference (16.3), the one-sided downward difference (16.4), or the more precise central difference (16.5). With our current notation, the central difference is

$$\frac{\partial B}{\partial r}(r,t) \approx \frac{B(r+\Delta r,t) - B(r-\Delta r,t)}{2\Delta r},$$

[5] Here is another procedure mentioned in some finance textbooks. Given 12 independent draws $U_1,\dots,U_{12}$ from the $U[0,1]$ distribution, the variable

$$\varepsilon = \left(\sum_{i=1}^{12} U_i\right) - 6$$

is taken to represent a draw from the $N(0,1)$ distribution. Each U_i has mean zero and variance 1/12, so ε will have the correct first two moments, namely mean zero and variance one. However, it is clear that this ε can only take values in the interval $(-6,6)$ and is therefore not normally distributed, although the probability that a draw from the $N(0,1)$ distribution falls outside this interval is extremely small. In contrast, the Box–Muller transformation produces draws from the correct distribution. Moreover, the Box–Muller transformation is much more parsimonious as it delivers two normal variates from two uniform variates, where the above alternative delivers one normal variate from 12 uniform variates.

where $\Delta r > 0$ represents a small deviation from the current short rate. Consequently, we have to compute Monte Carlo estimates of both $B(r + \Delta r, t)$ and $B(r - \Delta r, t)$. Then we will also have the necessary values to compute the approximation of the second-order derivative

$$\frac{\partial^2 B}{\partial r^2}(r,t) \approx \frac{B(r + \Delta r, t) - 2B(r,t) + B(r - \Delta r, t)}{(\Delta r)^2},$$

see (16.6). For this purpose it is necessary to simulate a new set of paths with the initial value $r + \Delta r$ and a new set of paths with the initial value $r - \Delta r$. However, to reduce a potential bias in the approximation of $\partial B/\partial r$ (and reduce the computation time), exactly the same sequence of draws of random numbers—such as the ε_n^m, $m = 1, \dots, M$, $n = 1, \dots, N$ in (16.28)—should be used in the generation of all three sets of short rate paths with the different initial values r, $r + \Delta r$, and $r - \Delta r$. Should the simulations for some reason be biased, say upwards, in the sense that the price approximation $B^{\mathrm{MC}}(r,t)$ exceeds the true (but generally unknown price) $B(r,t)$, then it is likely that the price approximation $B^{\mathrm{MC}}(r + \Delta r, t)$ (respectively $B^{\mathrm{MC}}(r - \Delta r, t)$ will exceed the true price $B(r + \Delta r, t)$ (respectively $B(r - \Delta r, t)$ by nearly the same amount. Hence, the approximation of the derivatives $\partial B/\partial r$ and $\partial^2 B/\partial r^2$ will be virtually unbiased. To avoid storage of the many random numbers, the Monte Carlo approximations of the three prices $B(r,t)$, $B(r + \Delta r, t)$, and $B(r - \Delta r, t)$ should be computed simultaneously so that, for any $m = 1, \dots, M$, three different short rate paths corresponding to different initial values are generated using the same sequence of random numbers $\varepsilon_1^m, \dots, \varepsilon_N^m$.

The price sensitivity with respect to the value of any given model parameter can be computed in a similar way. For example, in the Vasicek model we can compute an estimate of the sensitivity of the price of a specific asset with respect to the interest rate volatility parameter around a given value β by computing two Monte Carlo approximations of the price, one based on simulated paths using an interest rate volatility of $\beta + \Delta\beta$ and another based on simulated paths using the volatility $\beta - \Delta\beta$, where $\Delta\beta$ is positive and small (compared to β). Then we can approximate the derivative of the price with respect to β in the same way as we approximated $\partial B/\partial r$ above. Again, the same sequence of random numbers ε_n^m should be used when computing the two price approximations.

Monte Carlo simulation also applies to the estimation of other risk measures used in practical risk management such as the so-called Value-at-Risk (VaR) measure. The VaR of a certain portfolio of assets is defined as a threshold loss such that the loss on the portfolio over a pre-specified time horizon (typically 1 day or 10 days) exceeds this threshold value only with a certain pre-specified probability (often taken to be 1% or 5%). Obviously, the VaR is determined by the probability distribution of the value of the portfolio at the end of the specified time horizon. Monte Carlo simulations of the uncertain factors that influence the portfolio value lead to an approximation of the distribution. The VaR can be estimated from the lower tail of the approximated distribution. The Monte Carlo simulation approach to pricing is based on computing an average as an approximation to an expectation over some distribution. It should be noted that tail probabilities are more difficult to estimate precisely than means. Again, we refer to Asmussen and Glynn (2007) and Glasserman (2003) for more information.

16.3.5 Precision and improvements

The prices (and risk measures) delivered by the Monte Carlo simulation procedure are only approximations of the true, typically unknown, prices, so an interesting question is how precise these approximations are. Consider the approximation

$$\bar{F} \equiv \mathrm{E}_t[F_T] \approx \bar{F}_M \equiv \frac{1}{M}\sum_{m=1}^{M} F_T^m,$$

where the realization of F_T depends on the time T value (and maybe also on earlier values) of some underlying stochastic process x. Each m corresponds to a sample of x_T (and maybe relevant earlier values) and F_T^m denotes the associated realization of F_T. Suppose that each F_T^m is an unbiased estimator of F_T. Then the average $(1/M)\sum_m F_T^m$ is indeed the best estimate of the expectation. Suppose $\mathrm{Var}_t[F_T^m] = \sigma_F^2 < \infty$. Then, according to the central limit theorem, $\sqrt{M}(\bar{F}_M - \bar{F})$ will converge in distribution to the normal distribution $N(0, \sigma_F^2)$ as $M \to \infty$. Consequently, for large M, the approximation error $\bar{F}_M - \bar{F}$ is roughly $N(0, \sigma_F^2/M)$ distributed. In particular, a 95% confidence interval for the true $\bar{F}$ is given by

$$\left[\bar{F}_M - 1.96\frac{\sigma_F}{\sqrt{M}}, \bar{F}_M - 1.96\frac{\sigma_F}{\sqrt{M}}\right],$$

where the true standard deviation σ_F in practice has to be replaced by the sample standard deviation, that is

$$\hat{\sigma}_F = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(F_T^m - \bar{F}_M)^2} = \sqrt{\frac{1}{M-1}\sum_{m=1}^{M}(F_T^m)^2 - \frac{M}{M-1}(\bar{F}_M)^2}.$$

These results give a rough link between the number of samples, M, and the precision of the Monte Carlo approximation of the expectation. To double the accuracy of the approximation—that is to halve the standard deviation of the approximation error, $\sigma_F/\sqrt{M}$—we must quadruple the number of samples. To improve the accuracy by a factor of 10 (giving an extra digit of precision), we must increase the number of samples by a factor of 100. As the time required for computing the Monte Carlo approximation is roughly proportional to M, it is clear that a high precision can become very costly with respect to computation time.

It is not always possible to ensure that each simulated terminal value F_T^m is an unbiased estimator of the true unknown value. As discussed above, if the underlying process x is Gaussian or otherwise relatively simple, it is sometimes possible to ensure that the simulated values are draws from the exact distribution. But generally, the continuous-time process has to be discretized in a way that does not replicate the exact distribution. The resulting values F_T^m may therefore have a bias. Most discretization methods lead to the correct distribution when the number of discretization intervals, N in the notation introduced earlier, goes to infinity. In these cases, a high N may also be needed to obtain high precision which also increases the computation time. However, experimental experience indicates that for the diffusion processes typically used in financial models, a fairly low N

(corresponding to, say, 10–100 time steps per year) leads to a satisfactory precision for standard financial assets, at least when M is sufficiently high. In any concrete case, experimentation with different values of N and M is recommended so as to obtain an acceptable precision with an acceptable computational time.

The above discussion motivates the search for variance reduction techniques, and various such techniques have been suggested and are frequently used in financial applications of Monte Carlo simulation. We will only give a brief introduction. Readers seeking more details should consult Glasserman (2003, Ch. 4) and the references therein.

Control variates

Suppose you use Monte Carlo simulation to estimate an expectation,

$$\mathrm{E}[F] \approx \bar{F}_M \equiv \frac{1}{M} \sum_{m=1}^{M} F^m,$$

as in (16.21). In some cases you might be able to compute in closed form the expectation of a different random variable, say G, depending on the same underlying uncertainty as F, that is you know the exact value $\mathrm{E}[G]$. By computing a Monte Carlo approximation $\bar{G}_M$ of the same expectation, you can observe the approximation error $\bar{G}_M - \mathrm{E}[G]$. This information is often useful in estimating the approximation error for $\mathrm{E}[F]$, which then leads to a better approximation. To see how this works, let b be some constant, and define an adjusted Monte Carlo estimate of $\mathrm{E}[F]$ by

$$\bar{F}_M(b) = \bar{F}_M - b\left(\bar{G}_M - \mathrm{E}[G]\right),$$

which basically assumes that the Monte Carlo error for F equals b times the Monte Carlo error for G. Equivalently, we can adjust each of the sample values F^m to

$$F^m(b) = F^m - b\left(G^m - \mathrm{E}[G]\right)$$

and then take the average

$$\begin{aligned}
\frac{1}{M} \sum_{m=1}^{M} F^m(b) &= \frac{1}{M} \sum_{m=1}^{M} \left(F^m - b\left(G^m - \mathrm{E}[G]\right)\right) \\
&= \frac{1}{M} \sum_{m=1}^{M} F^m - b\left(\frac{1}{M} \sum_{m=1}^{M} G^m - \mathrm{E}[G]\right) = \bar{F}_M(b),
\end{aligned}$$

which results in the same estimate. G is referred to as a control variate (for F).

The purpose of the adjustment is to ensure that the adjusted $F^m(b)$ values have a lower variance than the original F^m values so that the adjusted average $\bar{F}_M(b)$ is more precise than the original $\bar{F}_M$. The variance of any adjusted value $F(b) = F^m(b)$ is

$$\mathrm{Var}[F(b)] = \mathrm{Var}\left[F - b\left(G - \mathrm{E}[G]\right)\right] = \mathrm{Var}[F] + b^2\, \mathrm{Var}[G] - 2b\, \mathrm{Cov}[F, G].$$

The variance is minimized for b equal to

$$b^* \equiv \frac{\mathrm{Cov}[F,G]}{\mathrm{Var}[G]} = \rho[F,G]\frac{\sigma[F]}{\sigma[G]}, \tag{16.37}$$

which defines the optimally adjusted Monte Carlo estimate $\bar{F}_M(b^*)$. Not surprisingly, the optimal adjustment depends on the correlation between F and G. The variance of the optimally adjusted estimate is

$$\mathrm{Var}[F(b^*)] = \left(1 - \rho[F,G]^2\right)\mathrm{Var}[F], \tag{16.38}$$

see Exercise 16.4. Hence, the squared correlation determines the variance reduction. A correlation of ± 0.95 reduces the variance by roughly 90%, while a correlation of ± 0.7 leads to a 50% variance reduction. Recall that in order to reduce the variance by 90% (50%) by simply increasing the number of samples, M, requires ten (four) times as many samples with a similar increase in computation time. The control variate technique offers the same variance reduction with a very small increase in computation time.[6]

In practice, $\sigma[F]$ and $\rho[F,G]$ (and maybe even $\sigma[G]$) are unknown, but they can be estimated from the simulated samples leading to the following estimate of b^*,

$$\hat{b}_M = \frac{\sum_{m=1}^{M} F^m G^m - M\bar{F}_M\bar{G}_M}{\sum_{m=1}^{M}(G^m)^2 - M(\bar{G}_M)^2},$$

which is then used in the adjustments, that is the adjusted Monte Carlo estimate of $\mathrm{E}[F]$ is $\bar{F}_M - \hat{b}_M\left(\bar{G}_M - \mathrm{E}[G]\right)$, where both $\bar{F}_M$, $\bar{G}_M$, and $\hat{b}_M$ are computed from the simulated samples.

The control variate technique can be used to enhance the Monte Carlo approximation of the price of an asset by identifying a related asset with a price known in closed form. If you want to price a derivative asset, the underlying asset itself is often a good and easily implemented control variate. In many dynamic term structure models bond prices are known in closed form and can be used as control variates for various derivatives. This lines up with the idea of calibrating models to observed bond prices discussed in Chapter 9: why trust the computed derivative prices if the underlying assets are not priced correctly? Alternatively, for specific derivatives, it may be possible to find a related derivative that can be priced in closed form.

Above, the control variate was assumed to be the price of a different asset in the same model, but we can also use the price of the same asset in a different model, namely a model which is nested by the model you really care about. For example, in a two-factor diffusion model of the term structure (such as a model in which the short rate has stochastic volatility) it might be impossible to price various fixed income assets in closed form, but if you turn off one source of uncertainty by assuming that the second factor is constant, you may be able to derive an explicit pricing expression. Then you can use the price of such an asset in the simple model

[6] Some texts describe the control variate technique only for $b = 1$, see for example (Hull, 2009, Sec. 19.7), but unless the true b^* is close to one, implementing the control variate technique assuming $b = 1$ is inefficient and may in some cases even lead to higher variance.

as a control variate when computing the price of the same asset in the general model by Monte Carlo. Of course, the simulation-based price approximation in the simple model will only involve samples of the first factor, so it is unlikely to obtain a dramatic variance reduction in this way. But even a modest variance reduction at almost no extra computational cost should be appreciated.

In principle, the control variate technique can be used in combination with other numerical procedures than Monte Carlo simulation. If you compute a price F^{num} of an asset with a finite difference approach, you can easily miss the true, unknown price F^{exact}. But you may be able to find a related asset for which an exact price G^{exact} is available so that pricing that asset with the same finite difference grid leads to an observation of the numerical pricing error, $G^{\text{num}} - G^{\text{exact}}$. If you know the appropriate value of b^* stated in (16.37), you can adjust the finite difference based price of the first asset to the more precise approximation

$$\tilde{F} = F^{\text{num}} - b^* \left(G^{\text{num}} - G^{\text{exact}}\right).$$

But b^* is generally unknown and—in contrast to Monte Carlo simulation—the finite difference approach cannot produce an estimate of b^*. If the two assets are close to being perfectly positively correlated, applying the above adjustment with $b^* = 1$ will probably produce a better price approximation than the unadjusted price F^{num} coming out of the finite difference algorithm, but no guarantees can be given.

Antithetic variates

Consider again the basic Monte Carlo approximation

$$\mathrm{E}[F(x)] \approx \bar{F}_M \equiv \frac{1}{M} \sum_{m=1}^{M} F(x^m).$$

Here each sample x^m is based on a sequence of independent draws from the standard normal distribution, $\varepsilon_1^m, \dots, \varepsilon_N^m$ (sometimes $N = 1$ is sufficient). The antithetic variate technique is based on the simple observation that when ε is $N(0,1)$-distributed, then $-\varepsilon$ is also $N(0,1)$-distributed. Applying the sequence $-\varepsilon_1^m, \dots, -\varepsilon_N^m$ instead of the original sequence will lead to a different sample $\tilde{x}^m$ of the underlying random quantity x. Doing this for all M samples, we can compute the antithetic variates estimator

$$\bar{F}_M^{\text{AV}} \equiv \frac{1}{2M} \left(\sum_{m=1}^{M} F(x^m) + \sum_{m=1}^{M} F(\tilde{x}^m) \right) = \frac{1}{M} \sum_{m=1}^{M} \left(\frac{F(x^m) + F(\tilde{x}^m)}{2} \right).$$

Since $F(x^m)$ and $F(\tilde{x}^m)$ are not independent, estimates of the variance of $\bar{F}_M^{\text{AV}}$ and the associated confidence intervals should *not* be derived from the sample variance of all $2M$ values $F(x^1), F(\tilde{x}^1), \dots, F(x^M), F(\tilde{x}^M)$. Instead, note that the M pairwise averages

$$\frac{F(x^1) + F(\tilde{x}^1)}{2}, \dots, \frac{F(x^M) + F(\tilde{x}^M)}{2}$$

will be independent, so their sample standard deviation $\hat{\sigma}_M^{\mathrm{AV}}$ leads to a 95% confidence interval

$$\left[\bar{F}_M^{\mathrm{AV}} - 1.96\frac{\hat{\sigma}_M^{\mathrm{AV}}}{\sqrt{M}}, \bar{F}_M^{\mathrm{AV}} - 1.96\frac{\hat{\sigma}_M^{\mathrm{AV}}}{\sqrt{M}}\right]$$

for the true expectation $\mathrm{E}[F(x)]$.

Generating M antithetic pairs of samples $(x^m, \tilde{x}^m)$ is computationally slightly simpler than generating $2M$ independent samples $x^1, \dots, x^{2M}$, and leads to a tighter confidence interval. The antithetic variate technique is intuitively appealing. If the sample values ε_n^m for some reason are biased upwards and this leads to an upwards bias in $F(x^m)$, then it is likely that the 'antithetic samples' $F(\tilde{x}^m)$ are biased in the opposite direction. The overall average $\bar{F}_M^{\mathrm{AV}}$ is then closer to being unbiased.

Moment matching

The idea of moment matching is to adjust the samples ε_n^m from the standard normal distribution in order to make sure that they do in fact have a zero average and a unit sample variance. The first step is to generate and store all the $M \cdot N$ samples ε_n^m, then compute the sample average $\hat{\mu}$ and sample variance $\hat{\sigma}^2$, and finally replace each ε_n^m by

$$\tilde{\varepsilon}_n^m = \frac{\varepsilon_n^m - \hat{\mu}}{\hat{\sigma}}.$$

Now the adjusted $\tilde{\varepsilon}_n^m$ are used in the simulation of the relevant stochastic processes and random variables. While moment matching has some immediate appeal, it comes with two major problems. First, the storage and handling of all the $M \cdot N$ samples is cumbersome, time-consuming, and challenges the memory capacity of some computers. Second, since all the adjusted values $\tilde{\varepsilon}_n^m$ are linked via $\hat{\mu}$ and $\hat{\sigma}$, they are not independent, so the standard convergence results and construction of confidence intervals are not justified. Furthermore, note that the antithetic variate technique automatically ensures that the first moment is matched.

Stratified sampling

If we simulate a fairly low number M of samples from a given distribution, the distribution of the sampled values may be very unlike the true distribution. For example, when sampling from the standard normal distribution, there might be many fewer extremely high or low sample values than in the true distribution, either by chance or a due to a bad random number generator. In some cases, for example for the valuation of certain exotic options, the extreme values have a high influence on the asset value, that is the expectation to be estimated with the simulation procedure. Stratified sampling is, loosely speaking, a way to ensure that the distribution of the sample values is in accordance with the true distribution.

As a brief illustration, suppose that you want to estimate an expectation $\mathrm{E}[F(x)]$, where x is a standard normal random variable. The basic Monte Carlo approach is to generate M samples $x^1, \dots, x^M$ from the $N(0,1)$ distribution and compute

$(1/M)\sum_{m=1}^{M} F(x^m)$. The samples $x^1, \ldots, x^M$ can, in principle, be any set of real numbers. With stratified sampling, the real line is divided into a number of intervals, say K intervals, of the form $A_k = (a_{k-1}, a_k]$ for $k = 1, \ldots, K$, where $-\infty \equiv a_0 < a_1 < \cdots < a_{K-1} < a_K \equiv \infty$. If $p_1, \ldots, p_K$ are positive numbers summing up to one, we can define

$$a_k = N^{-1}(p_1 + \cdots + p_k), \quad k = 1, \ldots, K-1,$$

where $N^{-1}(\cdot)$ is the inverse of the probability distribution function associated with the standard normal distribution. This implies that the probability that x falls in A_k is given by

$$\begin{aligned}\mathbb{P}(x \in A_k) &= \mathbb{P}(a_{k-1} < x \leq a_k) = \mathbb{P}(x \leq a_k) - \mathbb{P}(x \leq a_{k-1}) \\ &= (p_1 + \cdots + p_k) - (p_1 + \cdots + p_{k-1}) = p_k.\end{aligned}$$

By Bayes' rule, we have

$$\mathrm{E}[F(x)] = \sum_{k=1}^{K} \mathbb{P}(x \in A_k)\, \mathrm{E}[F(x)|x \in A_k] = \sum_{k=1}^{K} p_k\, \mathrm{E}[F(x)|x \in A_k].$$

If, for each k, we generate M_k samples $x^{k1}, \ldots, x^{kM_k}$ from the conditional distribution of x given $x \in A_k$, we can approximate each of the conditional expectations by

$$\mathrm{E}[F(x)|x \in A_k] \approx \frac{1}{M_k} \sum_{m=1}^{M_k} F(x^{km})$$

and, consequently, the stratified sampling Monte Carlo approximation is

$$\mathrm{E}[F(x)] \approx \sum_{k=1}^{K} \frac{p_k}{M_k} \sum_{m=1}^{M_k} F(x^{km}). \tag{16.39}$$

It remains to be explained how samples x^{km} can be generated from the conditional distribution of x given $x \in A_k = (a_{k-1}, a_k]$. Let U be a draw from the uniform $U(0,1)$ distribution. Then

$$x = N^{-1}\big(N(a_{k-1}) + [N(a_k) - N(a_{k-1})]\, U\big)$$

has the desired conditional distribution. This follows from the fact that

$$\begin{aligned}&\mathbb{P}(x < a) = \mathbb{P}\left(N(x) < N(a)\right) = \mathbb{P}\big(N(a_{k-1}) + [N(a_k) - N(a_{k-1})]\, U < N(a)\big) \\ &\quad = \mathbb{P}\left(U < \frac{N(a) - N(a_{k-1})}{N(a_k) - N(a_{k-1})}\right),\end{aligned}$$

which is equal to $[N(a) - N(a_{k-1})]/[N(a_k) - N(a_{k-1})]$ for $a \in (a_{k-1}, a_k]$ and equal to zero for $a \notin (a_{k-1}, a_k]$.

Stratification induces dependence among the generated samples. If you need to simulate entire sample paths of some process, you should not use stratification along each path, as it is important to respect the independence of the increments to a standard Brownian motion. However, you can use stratification *across* different paths. For more information on how to compute a variance estimate for the stratified sampling approximation (16.39), how to optimally select the subsets A_k, and other relevant issues, the reader should consult Glasserman (2003, Sec. 4.3).

Importance sampling

For some functions $F(\cdot)$, many of the samples $x^1, \dots, x^M$ might have little effect on the estimate of $\mathrm{E}[F(x)]$, if F is zero or very small in certain regions compared to other regions, which means that precious computation time has been wasted. For example, let x represent the price of an underlying asset and $F(x) = \max(x - K, 0)$ the payoff of a call option with an exercise price of K. Suppose that the call option is deep out-of-the-money in the sense that the initial price of the underlying asset is much smaller than K. Then it is highly likely that most of the simulations will lead to a zero payoff and thus no contribution to the estimate of the option price. The idea of importance sampling is to put more weight on samples that contribute more to the expectation. The distribution is shifted to the area where 'most of the action is'. For details and examples, see Glasserman (2003, Sec. 4.6) or similar material.

Quasi-Monte Carlo

The variance reduction techniques outlined above all try to improve the random sampling from the relevant uncertain objects, that is random variables or stochastic processes capturing the underlying uncertainty. Quasi-Monte Carlo methods (also known as low-discrepancy methods) are very different. They do not try to mimic the inherent randomness. Instead of drawing random numbers from some distribution, they use purely deterministic sequences of numbers, sometimes called quasi-random sequences, although this is a not a very descriptive term. Suppose that the goal is to compute some expectation $\mathrm{E}[F(x)]$ and x is somehow related to d independent, uniformly $U[0,1]$ distributed random variables, for example $x = h(U_1, \dots, U_d)$. If we let $f = F \circ h$ denote the composite function, then we are looking for

$$\mathrm{E}[f(U_1, \dots, U_d)] = \int_{[0,1]^d} f(u)\, du,$$

where the integration is over the d-dimensional unit cube $[0,1]^d$. The quasi-Monte Carlo approximation is then

$$\int_{[0,1]^d} f(u)\, du \approx \frac{1}{N} \sum_{n=1}^{N} f(u_n),$$

where $u_1, \dots, u_N$ is a carefully and deterministically chosen sequence of points in the cube $[0,1]^d$. By increasing N and thus adding points, the cube is being filled in a specific way. There are many different ways of choosing the sequence of points.

Glasserman (2003, Ch. 5) presents a good overview and comparison of various methods, and concludes that the so-called Sobol' sequences are the most effective for financial applications. Quasi-Monte Carlo methods can potentially improve the convergence rate of the approximation compared to the genuine Monte Carlo methods. On the other hand, the genuine Monte Carlo methods involving truly random numbers are very intuitive, accessible to a broad range of people, and easy to implement with a variety of software tools.

16.3.6 An application to the one-factor CIR model

We illustrate some of the aspects of Monte Carlo simulation in the one-factor CIR model, in which the risk-neutral dynamics of the short-term interest rate is of the form (16.33). As in our finite difference example in Section 16.2.11, the initial short rate is 5%, the parameter values are assumed to be $\beta = 0.2$, $\kappa = 0.3$, and $\theta = 0.02/0.3 \approx 0.0667$, and we focus on the pricing of a 5-year zero-coupon bond with a face value of 100. The exact price of the bond is 76.007. Figure 16.4 depicts ten simulated paths of the short rate (thin erratic curves) produced by the discretization (16.34) as well as the average short rate over 10,000 simulated paths (thicker and smoother curve). Note that some of the paths dip slightly below zero as explained below (16.34).

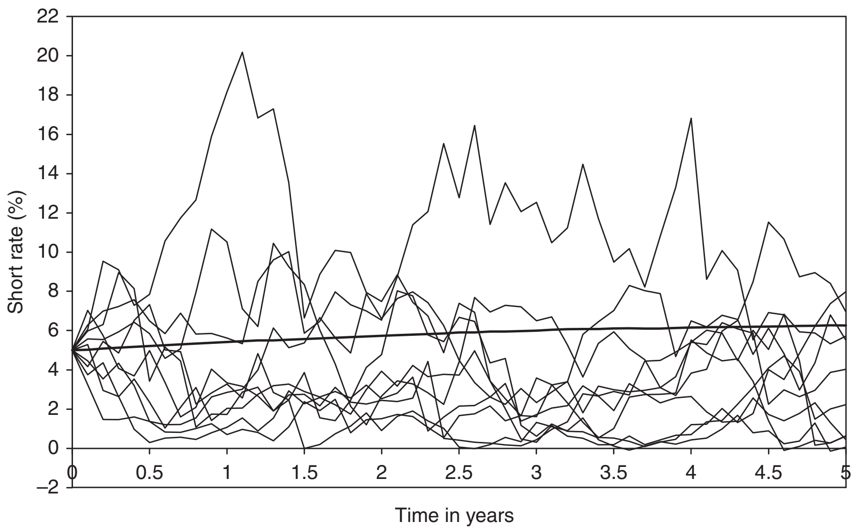

Fig. 16.4: Monte Carlo simulation of the short rate in the one-factor CIR model. Each of the ten thin curves shows a simulated path of the short rate using the discretization (16.34) with a time interval of $\Delta t = 0.1$ years. The initial short rate is 5%, the (risk-neutral) parameters are $\beta = 0.2$, $\kappa = 0.3$, and $\theta = 0.02/0.3 \approx 0.0667$. The thick curve shows the average short rate over 10,000 simulated paths.

The price of a zero-coupon bond maturing at time T with a unit payoff is approximated as

$$B^T(r_t, t) = \mathrm{E}^{\mathbb{Q}}\left[\exp\left\{-\int_t^T r_u\, du\right\}\right] \approx \frac{1}{M}\sum_{m=1}^{M} \exp\left\{-\sum_{n=0}^{N-1} r^m_{t_n}\right\},$$

using the notation introduced earlier. To illustrate the procedure, Table 16.3 shows a part of the simulated short rates and the discount factor $\exp\{-\sum_{n=0}^{N-1} r^m_{t_n}\}$, scaled by the face value of 100, along the ten paths illustrated in Fig. 16.4. The average of the exponential discount factor over those ten paths is 81.58, which is far from the exact price of the bond. Table 16.4 shows the simulation-based estimates of the 5-year zero-coupon bond price and the associated 95% confidence interval for 100, 1000, and 10000 paths, both for $\Delta t = 0.25$ and $\Delta t = 0.1$ without including antithetic variates as well as for $\Delta t = 0.1$ with antithetic variates. Note that the price estimate comes closer to the exact price and the confidence interval narrows as the number of paths is increased. The price estimates based on a discretization time interval of $\Delta t = 0.1$ are closer to the exact price than the estimates based on $\Delta t = 0.25$, but the improvement is modest compared to the fact that computational effort is roughly $0.25/0.1 = 2.5$ times harder. For $\Delta t = 0.1$ and $M = 10,000$, the estimate based on the antithetic variates technique is very close to the exact price. All the confidence intervals in Table 16.4 embrace the exact price.

Table 16.5 compares the bond price approximations generated by Monte Carlo simulation using the three different discretizations (16.34), (16.35), and (16.36). The same sequence of random numbers is used in all cases. In this experiment the discretization (16.34) produces the best results. In fact, for the two other

Table 16.3: Illustration of the Monte Carlo approach for pricing a zero-coupon bond in the one-factor CIR model. The bond matures in 5 years and has a face value of 100. The table shows ten paths of the short rate simulated using the discretization (16.34) with $\Delta t = 0.1$ and the associated approximation of the discounted payoff $100\exp\left\{-\sum_{n=0}^{N-1} r^m_{t_n}\right\}$. The average discounted payoff is 81.58, which is the approximation of the bond price based on these ten paths. The initial short rate is 5%, the (risk-neutral) parameters are $\beta = 0.2$, $\kappa = 0.3$, and $\theta = 0.02/0.3 \approx 0.0667$. The exact price of the bond is 76.007.

Path number	Time in years						Discounted payoff
	0	0.1	0.2	0.3	…	5	
1	0.05	0.06280	0.06988	0.07225	…	0.00620	64.45
2	0.05	0.05285	0.02949	0.02642	…	0.07982	88.64
3	0.05	0.03747	0.04358	0.02625	…	0.06954	83.05
4	0.05	0.03241	0.01472	0.01461	…	0.05514	85.35
5	0.05	0.04162	0.05442	0.04857	…	0.00063	91.47
6	0.05	0.05980	0.06307	0.08953	…	0.00634	84.23
7	0.05	0.06054	0.09535	0.09098	…	0.02227	73.29
8	0.05	0.05575	0.05543	0.05932	…	0.05712	75.35
9	0.05	0.07027	0.05748	0.04069	…	0.00442	88.61
10	0.05	0.04454	0.03535	0.04337	…	0.04030	81.39

Table 16.4: Price approximations of a 5-year zero-coupon bond with a face value of 100 in the one-factor CIR model. The table shows estimated prices and 95% confidence intervals using the discretization (16.34) of the short rate process. The initial short rate is 5%, the (risk-neutral) parameters are $\beta = 0.2$, $\kappa = 0.3$, and $\theta = 0.02/0.3 \approx 0.0667$. The exact price of the bond is 76.007.

# paths	$\Delta t = 0.25$		$\Delta t = 0.1$		$\Delta t = 0.1$, antithetic	
M	price	95% interval	price	95% interval	price	95% interval
100	73.09	[69.93;76.24]	74.13	[70.94;77.32]	75.54	[74.65;76.42]
1,000	76.32	[75.51;77.13]	76.29	[75.48;77.10]	75.92	[75.68;76.15]
10,000	76.20	[75.96;76.45]	76.12	[75.87;76.37]	76.02	[75.94;76.09]

Table 16.5: Price approximations of a 5-year zero-coupon bond with a face value of 100 in the one-factor CIR model. The table shows estimated prices and 95% confidence intervals using three different discretizations of the short rate process. The discretization time step is $\Delta t = 0.1$, and the antithetic variate technique is applied. The initial short rate is 5%, the (risk-neutral) parameters are $\beta = 0.2$, $\kappa = 0.3$, and $\theta = 0.02/0.3 \approx 0.0667$. The exact price of the bond is 76.007.

# paths	Discretization (16.34)		Discretization (16.35)		Discretization (16.36)	
M	price	95% interval	price	95% interval	price	95% interval
100	75.54	[74.65;76.42]	75.44	[74.54;76.33]	75.32	[74.41;76.22]
1,000	75.92	[75.68;76.15]	75.80	[75.56;76.04]	75.69	[75.45;75.93]
10,000	76.02	[75.94;76.09]	75.91	[75.83;75.98]	75.79	[75.72;75.87]

discretizations, the exact bond price falls outside the 95% confidence interval obtained with 10, 000 paths.

In Exercise 16.3, you are asked to implement the Monte Carlo approach to the pricing of bonds and bond options when the short rate dynamics are given by the one-factor Vasicek model.

16.3.7 Early exercise features

The Monte Carlo pricing approach explained above assumes that the payoff of the asset to be priced is received at fixed and known points in time. Hence, the approach is not directly applicable to derivatives with early exercise features such as American options (which, for now, we take to include Bermudan options). At first, you might think the following naive approach will work: once a full path has been simulated, look for the point in time at which the option payoff along that path is maximized, and discount that payoff back to the initial date. The average of the discounted payoffs over all paths may seem to give a reasonable estimate of the price of the American option. However, this estimate will be biased upwards and very often significantly overstate the true option price. The problem is that the exercise decision for each path is determined assuming perfect foresight about the future part of the path. In reality, of course, the exercise decision at a given point in time can only depend on the available information, which means the path of the relevant variables up to that point in time. The exercise decision at a given point in time and a given state of the world is based on a comparison between the payoff from immediately exercising the option and the continuation value of the option, that is the value of the option if it is not exercised now, and then optimally exercised in the future, up to and including the maturity date of the option. This continuation value is itself an expectation of future discounted payoffs and you cannot estimate an expectation from a single path. Various methods for incorporating early exercise decisions in the Monte Carlo approach have been explored in the literature. Here we will introduce the Least-Squares Monte Carlo approach and subsequently mention a few alternatives. A common feature of all the known methods is that they involve substantial computational effort.

Least-squares Monte Carlo

The idea of the least-squares Monte Carlo approach is to use regressions to estimate continuation values from the cross-section of simulated paths. Different variants of the approach have been suggested by Carriére (1996), Tsitsiklis and Van Roy (2001), and Longstaff and Schwartz (2001), with the latter being the apparently most popular version among practitioners. Further theoretical and experimental studies of the convergence of the methods can be found in, for example Clément et al. (2002) and Stentoft (2004a, 2004b).

The method can only handle a finite number of possible exercise dates, so if the option is truly American, this involves an approximation. In that case, it is natural, although not necessary, to check for exercise at every time step in the discretization of the underlying stochastic processes. If the option is Bermudan, the discretization should ensure that each of the known possible exercise dates coincides with a discretization time point, that is equals $t_i = t_0 + i\Delta t$ for some i.

Let $T_1 < \cdots < T_k < \cdots < T_K = T$ denote the future dates at which we want to check for exercise. Let $x = (x_t)$ denote the underlying stochastic process describing the dynamics of interest rates and other state variables. Finally, let $P(x_t, t)$ denote the price of the option at time t. The option value can then be determined recursively starting at the last possible exercise date with

$$P\left(x_{T_K}, T_K\right) = H\left(x_{T_K}, T_K\right),$$

where H gives the payoff as a function of state and time. The payoff can be zero for a range of values of the state variable. At every earlier possible exercise date T_k, the option price satisfies the relation

$$P\left(x_{T_k}, T_k\right) = \max\left(H\left(x_{T_k}, T_k\right), \mathrm{E}^{\mathbb{Q}}_{T_k}\left[\exp\left\{-\int_{T_k}^{T_{k+1}} r(x_u)\, du\right\} P\left(x_{T_{k+1}, T_{k+1}}\right)\right]\right),$$

where $H(x_{T_k}, T_k)$ is the payoff if the option is exercised at time T_k. The term

$$C\left(x_{T_k}, T_k\right) = \mathrm{E}^{\mathbb{Q}}_{T_k}\left[\exp\left\{-\int_{T_k}^{T_{k+1}} r(x_u)\, du\right\} P\left(x_{T_{k+1}, T_{k+1}}\right)\right]$$

is the continuation value of the option. Note that it involves the price of the option at the following possible exercise date and, due to the backwards recursive structure, this price incorporates the best possible exercise decisions in all future paths. Simulating forward the x process, it is no problem to compute the payoffs from exercise at any date, but the continuation value depends on all the possible future paths of the process. In principle, when considering exercise at time T_k, one could simulate a new set of paths starting from X_{T_k} to estimate the continuation value, but the work load will be incredibly high unless the number of exercise dates is very low.[7]

The continuation value at any potential exercise date T_k is an unknown function of the value of the state variable at that date. The least-squares Monte Carlo approach assumes that the continuation value is a linear combination of certain basis functions denoted by $\psi_1, \dots, \psi_L$, that is

$$C(x, T_k) = \sum_{\ell=1}^{L} \beta_\ell^k \psi_\ell(x) = (\beta^k)^\top \psi(x),$$

where $\psi(x) = (\psi_1(x), \dots, \psi_L(x))^\top$ collects the basis function in a vector and $\beta^k = (\beta_1^k, \dots, \beta_L^k)^\top$ collects the coefficients. The basis functions are pre-determined and will typically be various polynomials of the state vector x.

The challenge is to find the 'best' values of the coefficient vectors $\beta^1, \dots, \beta^{K-1}$ corresponding to all potential early exercise dates. This is done in the following manner. A set of M paths of the state variable are simulated up to time $T_K = T$, the latest possible exercise date. First, the coefficient vector β^{K-1} corresponding to the penultimate potential exercise date T_{K-1} is to be determined. For each path m, we can observe at time T_{K-1} whether the option is in-the-money, that is $H\left(x^m_{T_{K-1}}, T_{K-1}\right) > 0$, in which case early exercise might be relevant. For each of these 'in-the-money paths' we observe the option value at the terminal date T_K and discount it back to time T_{K-1}, which gives us an observed continuation value

$$\begin{aligned} C^m_{K-1} \equiv C^m\left(x^m_{T_{K-1}}, T_{K-1}\right) &= \exp\left\{-\int_{T_{K-1}}^{T_K} r(x^m_u)\, du\right\} P\left(x^m_{T_K, T_K}\right) \\ &= \exp\left\{-\int_{T_{K-1}}^{T_K} r(x^m_u)\, du\right\} H\left(x^m_{T_K, T_K}\right). \end{aligned}$$

In some models, the integral of the short rate has to be approximated by a sum. Let $\psi^m_{K-1} = \psi(x^m_{T_{K-1}})$. The coefficient vector β^{K-1} is then chosen to minimize the squared differences

$$\sum_m \left[C^m_{K-1} - (\beta^{K-1})^\top \psi^m_{K-1}\right]^2,$$

that is as the coefficient estimate in the regression

$$C^m_{K-1} = (\beta^{K-1})^\top \psi^m_{K-1} + \varepsilon,$$

[7] The method of nested sets of simulations is closely related to the so-called random tree method of Broadie and Glasserman (1997) and Broadie et al. (1997).

where ε is a mean-zero residual. Let $\hat{\beta}^{K-1}$ denote the least-squares estimate of β^{K-1}. The estimated continuation value at time T_{K-1} as a function of the state variable is then

$$\hat{C}(x, T_{K-1}) = \left(\hat{\beta}^{K-1}\right)^{\top} \psi(x).$$

Note that only the 'in-the-money paths' are used in the regression. According to Longstaff and Schwartz (2001), this improves the accuracy of the method. For each of the 'in-the-money paths' m at time T_{K-1}, the decision is to exercise at time T_{K-1} if and only if the exercise payoff $H\left(x^m_{T_{K-1}}, T_{K-1}\right)$ exceeds the estimated continuation value $\hat{C}\left(x^m_{T_{K-1}}, T_{K-1}\right)$. In that case, the possible payoff from exercise at the later date T_K along path m becomes irrelevant as the option will not live longer than T_{K-1} if path m should be realized. We need to keep track of the chosen exercise date along each path. Let τ^m denote the optimal exercise date along path m when exercise decisions are made as just described. For each path m where $H\left(x^m_{T_{K-1}}, T_{K-1}\right) > \hat{C}\left(x^m_{T_{K-1}}, T_{K-1}\right)$, put $\tau^m = T_{K-1}$. For the other paths, put $\tau^m = T_K$, indicating that the option is exercised at time T_K or expires worthless. These pathwise exercise dates are updated as we go backwards in time.

At the previous potential exercise date T_{K-2}, the coefficient vector β^{K-2} is estimated in a similar way by regressing observed continuation values $C^m_{K-2} \equiv C^m\left(x^m_{T_{K-2}}, T_{K-2}\right)$ on the basis function values $\psi^m_{K-2} \equiv \psi^m(x^m_{T_{K-2}})$. Again, only paths m for which the option is in-the-money at time T_{K-2} are included in the regressions. Note that the continuation value for path m is now to be computed as

$$C^m_{K-2} \equiv C^m\left(x^m_{T_{K-2}}, T_{K-2}\right) = \exp\left\{-\int_{T_{K-2}}^{\tau^m} r(x^m_u)\, du\right\} H\left(x^m_{\tau^m, \tau^m}\right)$$

using the best of the future possible exercise dates and the associated payoff discounted back to T_{K-2}. Given the least-squares estimate $\hat{\beta}^{K-2}$ and the corresponding continuation value function $\hat{C}(x, T_{K-2})$, the option is now being exercised at time T_{K-2} along path m if the immediate payoff exceeds the estimated continuation value. So if the option is in-the-money at time T_{K-2} along path m, and $H\left(x^m_{T_{K-2}}, T_{K-2}\right) > \hat{C}\left(x^m_{T_{K-2}}, T_{K-2}\right)$, put $\tau^m = T_{K-2}$.

Proceeding backwards in this way, we produce for each path an optimal exercise date and an associated payoff, and the value of the American option is then approximated by the average over all the paths:

$$P_t \approx P_t^{\mathrm{LSM}} \equiv \frac{1}{M} \sum_{m=1}^{M} \exp\left\{-\int_t^{\tau^m} r(x^m_u)\, du\right\} H\left(x^m_{\tau^m}, \tau^m\right).$$

Along some paths, exercise is never optimal so that $\tau^m = T_K$ and $H\left(x^m_{\tau^m}, \tau^m\right) = 0$. These paths can be ignored when computing the average.

By using the information on the simulated paths to estimate the exercise decisions, some element of foresight is present in the approach that induces a high bias in the estimated option price. Yet, this is far from the perfect foresight of the naive approach. The high bias can be avoided by using one (fairly small) set of simulated paths to estimate the continuation value function at all potential exercise dates and another set of simulated paths (larger and independent of the first set of paths) to compute an 'out-of-sample' option value. Experiments show virtually no difference between the 'in-sample' and the 'out-of-sample' option prices indicating that the high bias is not a practical concern.

The exercise rule chosen is based on the approximation of the continuation value and thus unlikely to be the truly optimal exercise rule, which induces a low bias. The magnitude of the low bias is determined by the type and number of basis functions used in the estimation of continuation values. Other things being equal, adding basis functions that are not spanned by the basis functions already included will lead to a better approximation of the continuation value function and the estimated option price, in particular when the current number of basis functions is low. If you are already applying a high number of basis functions, adding another is not likely to have much effect on the accuracy. However, increasing the number of basis functions will definitely increase the complexity of each regression and the computation time, so a tradeoff between accuracy and computational effort has to be found from a mix of experiences with related problems and experiments with the specific problem to be handled.

The basis functions typically considered are standard polynomials[8] or various forms of non-standard polynomials referred to as Legendre, Laguerre, Hermite, or Chebyshev polynomials, see for example Moreno and Navas (2003) for the definitions and interrelations. It is likely that the best choice of basis functions and the number of basis functions to include depend on the dimension and complexity of the model as well as the precise type of option to be priced. Longstaff and Schwartz (2001) consider various tests in which the prices computed with the least-squares Monte Carlo approach are compared to prices computed with the finite difference approach and report that even a low number of basis polynomials produces good results. For example, in a simple Black–Scholes–Merton framework, they show that with just a constant and the first three Laguerre polynomials, American put options can be very accurately priced. In higher-dimensional problems, a higher number of basis functions (up to around 20 in their examples) is required. Moreno and Navas (2003) provide additional experimental evidence that the method is robust to the type of polynomials used. Moreover, adding higher-order polynomials does not significantly improve the performance when polynomials up to order four are included. In fact, adding many high-order polynomials seems to cause numerical problems in the regressions.

Alternative methods

Various other methods for handling early exercise decisions in a simulation-based framework have been suggested. A general idea which has been explored in different settings, in various versions, is the search for a parametric approximation of the exercise boundary, that is the bound that separates the contin-

[8] Standard polynomials are $\psi^\ell(x) = x^\ell$ if x is one-dimensional, $\psi^\ell(x_1, x_2) = x_1^j x_2^{\ell-j}$ for $j = 0, 1, \dots, \ell$ if $x = (x_1, x_2)$ is two-dimensional, and so on.

uation region from the exercise region. In a problem with a one-dimensional state variable, the exercise boundary is just a curve, see Fig. 16.3, which shows the exercise boundary for an American put option on a bond in the one-factor CIR model. If the exercise boundary is postulated to be some parameterized function—maybe a linear combination of various basis functions—one can search for the parameters maximizing the price of the American option. The exercise boundary does not have to be estimated with high precision, as the option value is not so sensitive to the choice of exercise boundary around the optimal one. This suggests a two-step procedure. The first step is to simulate a fairly low number of paths of the underlying state variable. For any given set of parameters and thus a given exercise boundary, the option payoff along each path is determined by the first time the simulated state variable crosses the exercise boundary, and the option value is estimated as the average of the appropriately discounted payoffs. An optimization routine is added to find the set of parameters that maximizes the option value and thus defines the approximate exercise boundary. The second step involves simulating a higher number of paths. The option payoff along each path is determined from the approximate exercise boundary estimated in the first step. Again the estimated option price is the average discounted payoff over all paths. The paths used in the second step must be independent of the paths used in the first step to avoid any 'in-sample high bias'. The estimated option price is biased low since it is based on a specific exercise strategy which is not the truly optimal one, although we hope that it comes close.

For one-dimensional problems, we can easily price American options with finite differences or a tree-based method, so a simulation-based method for American options is mostly needed for high-dimensional problems. However, the exercise region and its boundary can be much more complex than in the one-dimensional case described above. Therefore, it is more difficult to come up with a parameterized form of the boundary. The boundary will probably involve many parameters to optimize over in the first step of the procedure with a corresponding increase in complexity and computation time. There is still hope, though. Andersen (2000) considers the valuation of a Bermudan swaption in a multi-factor LIBOR market model. In such a model, the true exercise boundary will depend on all the underlying forward LIBOR rates and is therefore very complex. Andersen suggests that one lets the exercise decision of the option depend only on the value of the underlying swap, which leads to a significant simplification of the parameterized exercise boundary and the maximization needed to identify the best parameterized boundary. In various tests he demonstrates that his approach leads to very good approximations of the option price. Again this indicates that a detailed and precise estimate of the exercise boundary is not needed in order to obtain useful simulation-based price approximations for typical derivatives with early exercise features. The same idea is likely to be useful in multi-factor diffusion models such as those considered in Chapter 8. In a study on the pricing of American options involving multiple stocks, Garcia (2003) also implements simple but carefully formulated low-dimensional exercise decision rules in high-dimensional settings and finds that the estimated option prices are very accurate.

For other simulation-based methods and techniques for American-style options, see Tilley (1993), Barraquand and Martineau (1995), Boyle et al. (1997), Carr and Yang (2001), Andersen and Broadie (2004), and Haugh and Kogan (2004).

16.4 APPROXIMATING TREES

The use of trees for derivatives pricing dates back at least to Cox et al. (1979) and Rendleman and Bartter (1979), and binomial trees are now presented in most introductory finance textbooks and widely used. A tree-structure can be used to describe the possible sample paths of a discrete-time stochastic process or to approximate the sample paths of a continuous-time process. The root node represents the initial value of the process, and the different layers of nodes in the tree contain the possible values of the process at different future dates. Nodes at two consecutive layers are connected by branches to represent possible transitions of the process value from one point in time to the next. These branches are associated with transition probabilities. Assets depending only on that stochastic process can be valued by a backward iterative procedure through the tree, from the final layer of nodes (corresponding to the maturity date of the asset) to the root node (corresponding to the current date). Early exercise features are easily handled by comparing in each node the payoff from immediate exercise to the continuation value, which is known due to the backward iterative procedure.

The Cox–Ross–Rubinstein binomial tree approximates the stock price dynamics in the Black–Scholes–Merton model. For term structure modelling, the tree must provide the dynamics of the state variables assumed relevant for pricing the fixed income securities. Some dynamic term structure models have been formulated directly using trees, see Ho and Lee (1986), Pedersen et al. (1989), and Black et al. (1990).[9] As reflected by this book, the modern and most popular models are set in continuous time. For such models, approximating trees might be useful for pricing assets that cannot be priced in closed form. For the one-factor continuous-time models presented in Chapter 7, an approximating tree must reflect the dynamics of the short-term interest rate. The construction and application of trees for one-factor models have been discussed by, among others, Nelson and Ramaswamy (1990), Tian (1993), Hull and White (1990b, 1993, 1994a, 1996), and Hull (2009, Ch. 30). Hull and White (1994b) present an extension of the approach to some two-factor models. If a binomial process is used for each of the two factors, it is natural to use a quadronomial tree, that is a tree with four branches from each node, to represent the possible transitions of both factors. Unfortunately, this approach implies that the number of nodes and branches in the tree grows exponentially with the number of factors, and trees seem to be rarely used for models with three or more state variables. It should be mentioned, however, that He (1990) introduces a general approach to approximating an n-factor diffusion model by an $(n+1)$-nomial discrete-time model, that is a tree with $n+1$ branches from each node, and this approach should thus be applicable to even high-dimensional models.

16.4.1 Constructing trinomial trees for general one-factor models

We will focus on the construction of trees for one-factor short rate models like those presented in Chapter 7. In contrast to the binomial tree normally used for

[9] For some of these models, the limiting continuous-time model has been derived subsequently. For example, the continuous-time version of the binomial model introduced by Ho and Lee (1986) is discussed in Section 9.3.

approximating the stock price in the Black–Scholes–Merton model, the trees used for approximating a one-factor interest rate model are typically trinomial, that is there are three branches leaving each node. Going from two to three branches from each node increases the flexibility and makes it easier to replicate mean reversion, see the discussion in Hull and White (1990b, 1993), which is a typical feature of interest rate models.

The branching procedure is also more flexible than in the Cox–Ross–Rubinstein stock price tree. In the binomial stock price tree the branching is the same throughout the tree. One branch represents multiplication of the stock price by u and the other multiplication by $d < u$. The probability of following the u-branch is the same in all parts of the tree, as is thus also the case for the d-branches. This makes sense since the distribution of the relative change in the stock price of a future time interval is independent of the current or past prices. In any reasonable interest rate model, the drift (and maybe also the volatility) of the short rate over the next time period depends on its current level. This feature is accommodated by letting transition probabilities vary over the nodes and by letting the set of accessible nodes depend on the current node, that is the current interest rate level. Intuitively, assuming that the short rate exhibits mean reversion, if the short rate is already very low corresponding to the lowest possible node in the tree at a given date, it is very unlikely that it will decrease further over the next time step, so it would be inefficient to add another node corresponding to an even lower short rate in the next layer of the tree. The opposite is the case if the short rate is already very high. In other words, non-standard branching is possible at the boundaries of the tree.

First, we discuss how to construct a trinomial tree approximating some state variable x having diffusion dynamics

$$dx_t = \mu(x_t, t)\, dt + \sigma(x_t, t)\, dz_t.$$

Afterwards, we will link this procedure to specific one-factor short rate models. Let $t = t_0$ denote the current date and Δt the time step applied in the tree so that the time points considered in the tree are $t_n = t_0 + n\Delta t$, $n = 1, 2, \dots, N$, where the final point in time $t_N = T$ is determined from the desired applications of the tree, for example the maturity date of some asset to be priced. The tree we construct here is equidistant in time, but that can be generalized to any time spacing $t_{n+1} = t_n + \Delta t_n$ with varying time steps Δt_n, and such generalizations may be valuable for the pricing of some assets. At any time t_n, we will construct the tree so that the distance between any two neighbouring nodes is the same, $\Delta x_n > 0$, but we can generally allow for different distances at different points in time. The choice of Δx_n is discussed later.

The value of the approximate process at time t_n in node number j is denoted by $x_{j,n}$. Let x_0 denote the initial (time $t = t_0$) state. The only node in the tree corresponding to time t_0 is referred to as node 0, so $x_{0,0} = x_0$ is the value of the state variable in the root node. Now consider the branching from a general node (j, n) in the tree with an associated state $x_{j,n} = x_0 + j\Delta x_n$ so that a positive (negative) j refers to a value of the state variable higher (lower) than the initial state. The three possible subsequent nodes are the 'up'-node $(k + 1, n + 1)$, the 'middle' node $(k, n + 1)$, and the 'down'-node $(k - 1, n + 1)$, so that the three possible values of the state variable are $x_{k+1,n+1} = x_0 + (k+1)\Delta x_{n+1}$, $x_{k,n} = x_0 + k\Delta x_{n+1}$, and $x_{k-1,n} = x_0 + (k-1)\Delta x_{n+1}$. The probabilities associated with the transitions are denoted by $q^u_{j,n}$, $q^m_{j,n}$, and $q^d_{j,n}$, respectively, with the natural requirement that

$$q^u_{k,n} + q^m_{k,n} + q^d_{k,n} = 1. \tag{16.40}$$

The choice of k is also discussed later. Note that even if $k = j$, the middle of the possible states $x_{k,n+1}$ will differ from the current state $x_{j,n}$ whenever $\Delta x_n \neq \Delta x_{n+1}$.

The probabilities are chosen such that the first two moments of the state variable after a time step in the tree coincide with the same moments of the continuous-time process. Together with (16.40), we then have a system of three equations for determining the three probabilities.[10] Let

$$M_{j,n} = \mathrm{E}^{\mathbb{Q}}\left[x_{t_{n+1}} | x_{t_n} = x_{j,n}\right], \qquad V^2_{j,n} = \mathrm{Var}^{\mathbb{Q}}\,\mathrm{E}\left[x_{t_{n+1}} | x_{t_n} = x_{j,n}\right]$$

denote the conditional mean and variance of the state variable after the time step according to the true continuous-time process. These moments are known in closed form for some functions μ and σ, or they can be approximated by

$$M_{j,n} \approx x_{j,n} + \mu_{j,n}\Delta t, \qquad V^2_{j,n} \approx \sigma^2_{j,n}\Delta t, \tag{16.41}$$

where we have introduced the notation

$$\mu_{j,n} = \mu(x_{j,n}, t_n), \qquad \sigma_{j,n} = \sigma(x_{j,n}, t_n).$$

In order to match the mean, we need to have

$$M_{j,n} = q^u_{j,n} x_{k+1,n+1} + q^m_{j,n} x_{k,n+1} + q^d_{j,n} x_{k-1,n+1} = x_{k,n+1} + \left(q^u_{j,n} - q^d_{j,n}\right)\Delta x_{n+1},$$

which implies that

$$\left(q^u_{j,n} - q^d_{j,n}\right)\Delta x_{n+1} = M_{j,n} - x_{k,n+1}. \tag{16.42}$$

To match the variance, we need

$$\begin{aligned} V^2_{j,n} + (M_{j,n})^2 &= q^u_{j,n} x^2_{k+1,n+1} + q^m_{j,n} x^2_{k,n+1} + q^d_{j,n} x^2_{k-1,n+1} \\ &= x^2_{k,n+1} + 2\left(q^u_{j,n} - q^d_{j,n}\right) x_{k,n+1}\Delta x_{n+1} + \left(q^u_{j,n} + q^d_{j,n}\right)(\Delta x_{n+1})^2, \end{aligned}$$

and applying (16.42), we get

$$\left(q^u_{j,n} + q^d_{j,n}\right)(\Delta x_{n+1})^2 = V^2_{j,n} + \left(M_{j,n} - x_{k,n+1}\right)^2. \tag{16.43}$$

The solution to the system of equations given by (16.40), (16.42), and (16.43) is

[10] The two moment matching conditions are central for the determination of the transition probabilities in the tree and for the convergence of the discrete-time tree process to the true continuous-time process. For theoretical results on convergence, see, for example Hubalek and Schachermayer (1998) and Lesne et al. (2000).

$$q^u_{j,n} = \frac{V^2_{j,n} + \left(M_{j,n} - x_{k,n+1}\right)^2}{2(\Delta x_{n+1})^2} + \frac{M_{j,n} - x_{k,n+1}}{2\Delta x_{n+1}}, \tag{16.44}$$

$$q^m_{j,n} = 1 - \frac{V^2_{j,n} + \left(M_{j,n} - x_{k,n+1}\right)^2}{(\Delta x_{n+1})^2},$$

$$q^d_{j,n} = \frac{V^2_{j,n} + \left(M_{j,n} - x_{k,n+1}\right)^2}{2(\Delta x_{n+1})^2} - \frac{M_{j,n} - x_{k,n+1}}{2\Delta x_{n+1}}. \tag{16.45}$$

At this point, we cannot be sure that these numbers are non-negative and thus really represent probabilities, but we still have flexibility with respect to the choice of k and Δx_{n+1}.

A high volatility makes it more likely that the process moves far away from the expected value, so it makes sense to let the spacing Δx_{n+1} in the state dimension reflect the volatility of the process over the next time step. Let us assume that the diffusion coefficient $\sigma(x, t)$ is independent of x. Then the approximation of $V_{j,n}$ in (16.41) is independent of j, that is $V_{j,n} = V_n$ for all j.[11] Suppose that we use a spacing of

$$\Delta x_{n+1} = \delta V_n$$

where $\delta > 0$ is a coefficient to be fixed. Then

$$q^u_{j,n} = \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{M_{j,n} - x_{k,n+1}}{V_n}\right)^2 + \frac{1}{2\delta}\left(\frac{M_{j,n} - x_{k,n+1}}{V_n}\right),$$

$$q^m_{j,n} = 1 - \frac{1}{\delta^2} - \frac{1}{\delta^2}\left(\frac{M_{j,n} - x_{k,n+1}}{V_n}\right)^2,$$

$$q^d_{j,n} = \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{M_{j,n} - x_{k,n+1}}{V_n}\right)^2 - \frac{1}{2\delta}\left(\frac{M_{j,n} - x_{k,n+1}}{V_n}\right).$$

Furthermore, pick k so that the state in the middle node, $x_{k,n+1}$, comes as close as possible to the expected state $M_{j,n}$. This implies that

$$|M_{j,n} - x_{k,n+1}| \le \frac{1}{2}\Delta x_{n+1} = \frac{\delta}{2} V_n \qquad \Rightarrow \qquad \frac{M_{j,n} - x_{k,n+1}}{V_n} \in \left[-\frac{\delta}{2}, \frac{\delta}{2}\right].$$

The up-probability will always be non-negative if δ is chosen such that

$$\frac{1}{2\delta^2} y^2 + \frac{1}{2\delta} y + \frac{1}{2\delta^2} \ge 0 \qquad \Leftrightarrow \qquad y^2 + \delta y + 1 \ge 0$$

for all $y \in [-\delta/2, \delta/2]$. The maximum of $y^2 + \delta y + 1$ is reached for $y = -\delta/2$ with a value of $(-\delta/2)^2 + \delta(-\delta/2) + 1 = 1 - \delta^2/4$. Hence we need $\delta \le 2$. To ensure a non-negative down-probability, we also need $\delta \le 2$. The mid-probability will be non-negative, whenever $\delta^2 \ge 1 + y^2$ for all $y \in [-\delta/2, \delta/2]$. This requires $\delta \ge 2/\sqrt{3} \approx 1.155$. So if we pick a δ between $2/\sqrt{3}$ and 2, we can really interpret

[11] In fact, even if σ is independent of x, the exact variance of the change $x_{t_{n+1}} - x_{t_n}$ over the time interval $[t_n, t_{n+1}]$ may vary with the value $x_{t_n} = x_{j,n}$; that depends on the drift $\mu(x, t)$.

$q^u_{j,n}$, $q^m_{j,n}$, and $q^d_{j,n}$ as probabilities. The standard choice for the spacing coefficient is $\delta = \sqrt{3} \approx 1.732$ which, according to Hull and White (1990b), reduces the approximation error and thus should improve convergence.

In the following we will make the stronger assumption that the variance V_n^2 is also independent of n, that is it is constant. This will be 'approximately true' if the diffusion coefficient $\sigma(x, t)$ is constant. In this case, the spacing of the tree is the same throughout,

$$\Delta x = \delta V \approx \delta\sigma\sqrt{\Delta t}.$$

The branching of the tree from any node (j, n) is defined by the distance between j and k, where $(k, n+1)$ is the middle of the three nodes that can be reached from (j, n). This distance is determined by the drift of the process. In many applications, the drift is such that for most nodes (j, n) in the tree, the middle of the three subsequent nodes is $(j, n+1)$ so that k is equal to j. This is referred to as 'standard branching' and is illustrated in the left part of Fig. 16.5. If we use the approximations (16.41), we have

$$M_{j,n} - x_{k,n+1} = M_{j,n} - x_{j,n+1} = \mu_{j,n}\Delta t,$$

and the probabilities (16.44)–(16.45) with a general Δx can be rewritten as

$$\begin{aligned} q^u_{j,n} &= \frac{\sigma^2\Delta t}{2(\Delta x)^2} + \frac{\mu_{j,n}^2(\Delta t)^2}{2(\Delta x)^2} + \frac{\mu_{j,n}\Delta t}{2\Delta x}, \\ q^m_{j,n} &= 1 - \frac{\sigma^2\Delta t}{(\Delta x)^2} - \frac{\mu_{j,n}^2(\Delta t)^2}{(\Delta x)^2}, \\ q^d_{j,n} &= \frac{\sigma^2\Delta t}{2(\Delta x)^2} + \frac{\mu_{j,n}^2(\Delta t)^2}{2(\Delta x)^2} - \frac{\mu_{j,n}\Delta t}{2\Delta x}. \end{aligned} \tag{16.46}$$

For $\Delta x = \delta\sigma\sqrt{\Delta t}$, we get

$$q^u_{j,n} = \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 + \frac{1}{2\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}, \tag{16.47}$$

$$q^m_{j,n} = 1 - \frac{1}{\delta^2} - \frac{1}{\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2,$$

$$q^d_{j,n} = \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 - \frac{1}{2\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}. \tag{16.48}$$

If Δt and the absolute value of the drift are small, the terms with $\left(\mu_{j,n}\sqrt{\Delta t}/\sigma\right)^2$ are considerably smaller than the other terms and are sometimes ignored, but in general it cannot be recommended to leave them out.

Intuitively, if the ratio of the expectation to the standard deviation of the change in the state variable over the next time step, that is $\mu_{j,n}\sqrt{\Delta t}/\sigma$, is very big, $q^d_{j,n}$ in (16.48) might become negative, and this is exactly when $k > j$, so that the

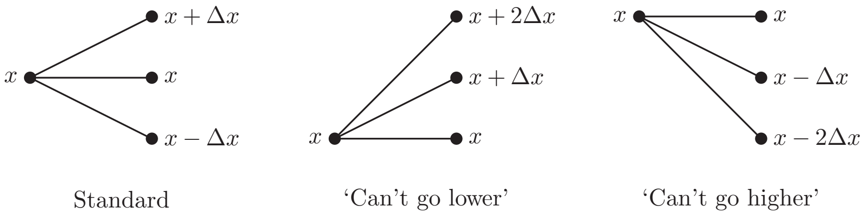

Fig. 16.5: The three possible branching methods in trinomial trees.

standard branching procedure should not be applied. If x is a process with mean reversion, high drifts occur for low state values, and the tree is not allowed to develop to even lower values of the state. In this case we will suppose that $k = j + 1$, so that the state will either remain unchanged, increase by Δx, or increase by $2\Delta x$, see Figure 16.5. Hence,

$$M_{j,n} - x_{k,n+1} = M_{j,n} - x_{j+1,n+1} = \mu_{j,n}\Delta t - \Delta x,$$

which leads to the probabilities

$$q^u_{j,n} = \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 - \frac{1}{2\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma},$$

$$q^m_{j,n} = -\frac{1}{\delta^2} - \frac{1}{\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 + \frac{2}{\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma},$$

$$q^d_{j,n} = 1 + \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 - \frac{3}{2\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}$$

for the non-standard 'can't go lower' branching. Conversely, if $\mu_{j,n}\sqrt{\Delta t}/\sigma$ is highly negative, $q^u_{j,n}$ in (16.47) might become negative. For a mean-reverting process this occurs when the state is very high and $k < j$. Assuming $k = j - 1$, the state will either remain unchanged, decrease by Δx, or decrease by $2\Delta x$ over the next time step, see Fig. 16.5. The tree is not allowed to develop to even higher values of the state. In this case,

$$M_{j,n} - x_{k,n+1} = M_{j,n} - x_{j-1,n+1} = \mu_{j,n}\Delta t + \Delta x,$$

which leads to the probabilities

$$q^u_{j,n} = 1 + \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 + \frac{3}{2\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma},$$

$$q^m_{j,n} = -\frac{1}{\delta^2} - \frac{1}{\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 - \frac{2}{\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma},$$

$$q^d_{j,n} = \frac{1}{2\delta^2} + \frac{1}{2\delta^2}\left(\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}\right)^2 + \frac{1}{2\delta}\frac{\mu_{j,n}\sqrt{\Delta t}}{\sigma}$$

for the non-standard 'can't go higher' branching.

In the original Vasicek model of Section 7.4 and its Hull–White extension of Section 9.4, the short-term interest rate r_t follows a mean-reverting diffusion process with a constant volatility, so the above procedure immediately applies with r_t replacing x_t. Note that when constructing and applying interest rate trees for pricing assets, it is the interest rate dynamics under the appropriate pricing measure that has to be approximated. A tree for the short-term interest rate approximating its risk-neutral dynamics contains all the necessary information to price assets with payoffs that can be stated as functions of the short rate at the payment date. For the Vasicek model, the relevant first and second (risk-neutral) moments are known in closed form, see (7.33) and (7.34) with risk-neutral parameters, and the variance of the short rate after a Δt period is independent of the current value of the short rate.

We cannot directly apply the recipe above to constructing an approximating tree for the short rate in the CIR model since the short rate volatility depends on the level of the short rate. However, following Nelson and Ramaswamy (1990), we can transform the short rate r into a state variable x with constant volatility. When

$$dr_t = \kappa(\theta - r_t)\,dt + \beta\sqrt{r_t}\,dz_t \tag{16.49}$$

and $x_t = f(r_t)$, Itô's Lemma implies that

$$dx_t = \left(f'(r_t)\kappa(\theta - r_t) + \frac{1}{2}f''(r_t)\beta^2 r_t\right)\,dt + f'(r_t)\beta\sqrt{r_t}\,dz_t.$$

A deterministic volatility is obtained when $f'(r)$ is proportional to $1/\sqrt{r}$. In particular, with $x = f(r) = 2\sqrt{r}/\beta$, we have $f'(r) = 1/(\beta\sqrt{r}) = \beta^{-1}r^{-1/2}$ and $f''(r) = -\frac{1}{2}\beta^{-1}r^{-3/2}$. Hence, the dynamics of the state variable becomes

$$dx_t = \left[\left(\frac{2\kappa\theta}{\beta^2} - \frac{1}{2}\right)x_t^{-1} - \frac{\kappa}{2}x_t\right]\,dt + dz_t \tag{16.50}$$

with a unit volatility. A tree approximating the dynamics of x can now be constructed and the value $x_{j,n}$ at any node transforms into a short rate of $r_{j,n} = f^{-1}(x_{j,n}) = (\beta x_{j,n}/2)^2$. In the construction of the x-tree in this case, the approximate expressions for the conditional mean and variance have to be invoked due to the complicated drift term in (16.50).

The zero lower bound on the short rate in the CIR model constitutes another challenge. A zero short rate translates into a zero value of the process x. We should construct the tree so that it does not allow negative values of the state and so that it respects the special behaviour of the short rate at zero. One way out is to choose the spacing coefficient δ so that the initial state $x_0 = 2\sqrt{r_0}/\beta$ is a multiple of $\Delta x = \delta\sqrt{\Delta t}$. Then zero is a possible state value in the approximating tree, although it is not necessarily ever reached if the non-standard 'can't go lower' branching applies at a level above Δx. Of course, if a node with $x = 0$ is reached, we will not allow for transitions to lower, negative states. The drift of x is not well-defined at zero, so instead of matching the moments of x, we will match the moments of the short rate itself. For a zero short rate, the volatility is zero, hence we only have to match the expectation which can be achieved with only two branches. The lower branch simply leaves x (and thus the short rate) unchanged at zero. Suppose that the upper branch leads to a node with a state value of $x = k\Delta x$ corresponding to a short rate of $r = (\beta k \Delta x/2)^2$. If q_0^u and $q_0^d = 1 - q_0^u$ denote the probabilities of the nodes, we want to ensure that

$$\kappa\theta\Delta t = q_0^u \left(\frac{k\Delta x}{2}\right)^2 = q_0^u \frac{k^2}{4}\delta^2\beta^2\Delta t,$$

so the up-probability is set to

$$q_0^u = \frac{4\kappa\theta}{\delta^2 k^2 \beta^2}. \tag{16.51}$$

Of course, we want $q_0^u \leq 1$, but we want q^u to be as high as possible so that the probability of being stuck at zero is as low as possible. Therefore k is chosen as the smallest integer exceeding $2\sqrt{\kappa\theta}/(\delta\beta)$.

16.4.2 An application to the one-factor CIR model

We illustrate the construction and use of trinomial trees in the one-factor CIR model, in which the risk-neutral dynamics of the short-term interest rate is of the form (16.49). Just as in our finite difference and Monte Carlo examples in Sections 16.2.11 and 16.3.6, the initial short rate is 5%, the parameter values are assumed to be $\beta = 0.2$, $\kappa = 0.3$, and $\theta = 0.02/0.3 \approx 0.0667$, and we focus on the pricing of a 5-year zero-coupon bond with a face value of 100. We will construct a tree with a time step of $\Delta t = 0.1$ years and thus $N = 50$ time intervals. The initial short rate translates to $x_0 \approx 2.2361$. With $\delta = \sqrt{3}$, we would have $\Delta x \approx 0.5477$ and $x_0/\Delta x \approx 4.0825$. To ensure that x_0 is a multiple of Δx, we decrease Δx to 0.4472, corresponding to $\delta \approx 1.4142$, since then $x_0/\Delta x = 5$. Now we can let the tree grow. As long as the probabilities (16.47)–(16.48) associated with the standard branching method are positive, the tree grows by standard branching. As the dynamics of the short rate in the original CIR model has no direct time-dependence, the transition probabilities are independent of n and depend only on j. With the given parameters, the shift to the 'can't go lower' branching occurs at the lower bound of zero, where the state can subsequently either go to $2\Delta x$, with a probability given by (16.51), or stay at zero. In the top of the tree, the shift to the 'can't go higher' branching occurs at $j = 24$ corresponding to $x \approx 12.97$ or $r \approx 1.68 = 168\%$. The geometry of the tree is illustrated in Fig. 16.6, and the associated transition probabilities are listed in Table 16.6.

The tree is now ready to be used for valuation of interest rate dependent assets by going backwards through the tree from the maturity date of the asset. Let $f_{j,n}$ denote the value of the asset in node (j, n). For the 5-year zero-coupon bond, we begin by assigning a value of 100 to each of the nodes at the maturity date, that is $f_{j,N} = 100$ for all relevant j, which in our case is $j = -5, -4, \dots, 24$. In any node (j, n) with standard branching, we compute the value by

$$f_{j,n} = e^{-r_{j,n}\Delta t}\left(q_j^u f_{j+1,n+1} + q_j^m f_{j,n+1} + q_j^d f_{j-1,n+1}\right) \qquad \text{(standard branching)} \tag{16.52}$$

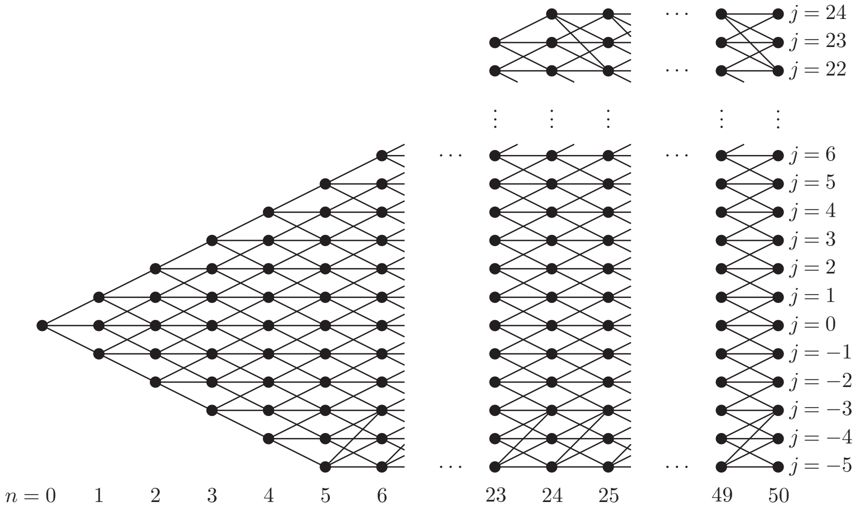

Fig. 16.6: The geometry of the approximating tree for the one-factor CIR model with the parameters stated in the text.

Table 16.6: Transition probabilities in the approximating tree for the one-factor CIR model with the parameters stated in the text.

j	r_j	q_j^u	q_j^m	q_j^d	j	r_j	q_j^u	q_j^m	q_j^d
24	1.682	0.7959	0.0666	0.1375	9	0.392	0.1422	0.4012	0.4565
23	1.568	0.1416	0.0347	0.8237	8	0.338	0.1478	0.4179	0.4343
22	1.458	0.1372	0.0690	0.7937	7	0.288	0.1542	0.4333	0.4125
21	1.352	0.1336	0.1021	0.7643	6	0.242	0.1616	0.4473	0.3911
20	1.250	0.1305	0.1340	0.7355	5	0.200	0.1700	0.4600	0.3700
19	1.152	0.1281	0.1646	0.7073	4	0.162	0.1796	0.4713	0.3491
18	1.058	0.1264	0.1939	0.6797	3	0.128	0.1907	0.4811	0.3282
17	0.968	0.1254	0.2220	0.6526	2	0.098	0.2036	0.4893	0.3071
16	0.882	0.1250	0.2488	0.6262	1	0.072	0.2189	0.4956	0.2856
15	0.800	0.1253	0.2744	0.6003	0	0.050	0.2378	0.4994	0.2628
14	0.722	0.1263	0.2987	0.5750	−1	0.032	0.2628	0.4994	0.2378
13	0.648	0.1280	0.3217	0.5502	−2	0.018	0.3000	0.4916	0.2084
12	0.578	0.1305	0.3435	0.5260	−3	0.008	0.3700	0.4600	0.1700
11	0.512	0.1336	0.3640	0.5024	−4	0.002	0.6003	0.2744	0.1253
10	0.450	0.1375	0.3833	0.4792	−5	0.000	0.3750	0.0000	0.6250

using the already computed values of the asset at the subsequent point in time. At the top of the tree, this is replaced by

$$f_{j,n} = e^{-r_{j,n}\Delta t}\left(q_j^u f_{j,n+1} + q_j^m f_{j-1,n+1} + q_j^d f_{j-2,n+1}\right). \text{ ('can't go higher' branching)}$$

At the bottom of the tree, use

$$f_{j,n} = e^{-r_{j,n}\Delta t}\left(q_j^u f_{j+2,n+1} + q_j^m f_{j+1,n+1} + q_j^d f_{j,n+1}\right), \text{ ('can't go lower' branching)}$$

where we have put $q_j^m = 0$ so the middle term can be ignored. We end up with an initial zero-coupon bond price of 76.082, not far from the exact price of 76.007. The same tree can be used to price other assets depending only on the risk-neutral behaviour of the short rate over the 5-year period. Early exercise is easily checked for along the way. For example, if $F_{j,n}$ denotes the payoff from exercising the asset in node (j, n), the recursive valuation relation (16.52) has to be modified to

$$f_{j,n} = \max\left\{F_{j,n}; e^{-r_{j,n}\Delta t}\left(q_j^u f_{j+1,n+1} + q_j^m f_{j,n+1} + q_j^d f_{j-1,n+1}\right)\right\}.$$

In Exercise 16.5, you are asked to use the tree for valuing 2-year European and American put options on the 5-year zero-coupon bond. Exercise 16.6 deals with the construction and implementation of a trinomial tree approximating the original Vasicek model.

16.4.3 More on interest rate trees

Hull and White (1994a, 1996) have developed a two-step procedure for constructing trinomial trees matching observed bond prices. The procedure applies directly to the Hull–White extension of the original Vasicek model in which the short rate dynamics is

$$dr_t = \left(\hat{\varphi}(t) - \kappa r_t\right)\, dt + \beta\, dz_t$$

under the risk-neutral pricing measure. Moreover, it is easily generalized to models in which some function $f(r_t)$ of the short rate follows a similar process, which is true for the Black–Karasinski model where $f(r) = \ln r$, cf. (7.54) in Section 7.7. The procedure does not apply to the Cox–Ingersoll–Ross model or similar models. The first step is to build a tree approximating an auxiliary process $R^* = (R_t^*)$ with dynamics

$$dR_t^* = -\kappa R_t^*\, dt + \beta\, dz_t,$$

which is symmetrical around zero. The tree is constructed very similarly to the general recipe described above, but the terms corresponding to $(\mu(x, t)\sqrt{\Delta t}/\beta)^2$ are ignored in the probabilities. These terms are typically small, at least near the centre of the tree. The second step is to compute the short rate values $r_{j,n}$ that are to be assigned to the different nodes in a way that ensures matching of the observed prices of zero-coupon bonds maturing at time $\Delta t, 2\Delta t, \dots, N\Delta t$. This is done by a forward induction procedure through the tree. Now we have a short rate tree with probabilities computed in the first step and node values computed in the second step. Interest rate derivatives can then be valued by the usual backward induction technique through the tree.

Leippold and Wiener (2004) discuss how to generalize the Hull–White two-step procedure to other short rate models and present an alternative way to match observed bond prices which avoids the cumbersome and time-consuming forward induction step.

The trees constructed by the above procedure are recombining, which is natural for approximating a diffusion process, since the increment over a future period is independent of past values. However, the payoffs of some assets depend on the path of the underlying process. In order to price such assets, we have to use a non-recombining tree in which we distinguish, for example, between a 'first up, then down' transition and a 'first down, then up' transition. The number of nodes in a non-recombining tree grows exponentially with the number of time steps, which makes applications very computationally intensive and accurate price approximations hard to obtain within a reasonable time limit.

A comparison of the left-hand side graphs in Fig. 16.1 and 16.5 indicates a close relation between the explicit finite difference approach and the trinomial tree approach. If we 'backdate' and slightly rewrite (16.52), we get

$$f_{j,n-1} = e^{-r_{j,n-1}\Delta t} q_j^d f_{j-1,n} + e^{-r_{j,n-1}\Delta t} q_j^m f_{j,n} + e^{-r_{j,n-1}\Delta t} q_j^u f_{j+1,n},$$

which is exactly of the same form as the key recursive valuation equation (16.9) in the explicit finite difference approach. The coefficients on the right-hand side of those two expressions are closely related, but not completely identical. For example, compare the expression for $\alpha_{j,n}$ below (16.9) with $e^{-r_{j,n-1}\Delta t} q_j^d$ where q_j^d is given by (16.46). With the trinomial tree approach, the lower and upper bounds on the state space are determined endogenously (based on the drift and volatility of the underlying state variable), whereas the bounds are preset with the finite difference approach. Intuitively, the trinomial tree approach should thus exhibit somewhat better efficiency and accuracy than the explicit finite difference approach.

Finally, note that estimates of relevant risk measures can also be derived from a trinomial tree. For example, when an asset has been priced using a trinomial tree for some state variable x, the initial sensitivity of the price with respect to changes in the state is approximated by

$$\frac{\partial f}{\partial x}(x_0, 0) \approx \frac{f_{1,1} - f_{-1,1}}{2\Delta x}.$$

Other derivatives can be approximated in a similar way.

16.5 CONCLUDING REMARKS

This chapter has introduced the three most widely used numerical methods for computing prices and risk measures in interest rate models. Various interesting extensions and improvements of computational efficiency have been suggested for all three methods, and more specialized and detailed textbook presentations should be consulted before setting up large-scale implementations for real-life applications in the financial industry. Let us briefly summarize the strengths and weaknesses of the three general methods. The finite difference approach and the approximating tree approach require that the underlying uncertainty can be described by a diffusion process of low dimension, whereas the Monte Carlo approach can handle non-diffusions and higher-dimensional processes. The finite difference approach and the approximating tree approach are easily applied to assets with early exercise opportunities. Early exercise features have long been considered impossible to incorporate in the Monte Carlo approach, but recent developments have shown that this is manageable with a tolerable increase in computational complexity. Furthermore, assets with path-dependent payoffs are easily priced with the Monte Carlo approach, but not with the two alternative approaches.

16.6 EXERCISES

Exercise 16.1 (Finite difference, Vasicek) Develop a spreadsheet or a computer program in which you apply the implicit finite difference approach to the one-factor Vasicek model. In particular, you should be able to price a 5-year zero-coupon bond with a face value of 100 and to price 2-year European and American put options on that bond with an exercise price of 82. The risk-neutral dynamics of the short rate is of the form

$$dr_t = \kappa \left[\theta - r_t\right] \, dt + \beta \, dz_t,$$

where the parameter values are $\kappa = 0.3$, $\beta = 0.03$, and $\theta = 0.065$. The initial short rate is $r_0 = 0.05$. Use a grid spacing defined by $\Delta t = 0.1$ years and $\Delta r = 0.0025 = 0.25\%$.

(a) Explain how you set the upper and lower bound on the grid.

(b) What assumptions do you make about the pricing function at the boundaries of the grid? As suggested in the text, the term with the second-order derivative in the PDE is often ignored at the upper and lower bound. Use the closed-form pricing formula for zero-coupon bonds to investigate whether the term with the second-order derivative is really small at the boundaries, compared to the other terms in the PDE.

(c) What are the prices of the bond and the European put option according to the closed-form pricing formulas developed in Chapter 7?

(d) What are the prices of the same assets according to your trinomial tree?

(e) Vary the grid spacing parameters Δt and Δx and discuss the effects on the accuracy of the method.

(f) Use the finite difference method to price the American put option and to find the associated exercise boundary.

(g) Can you use the control variate technique to fine-tune your price estimate of the American put?

Exercise 16.2 (Finite difference, CIR, mortgage-backed bonds) The purpose of this exercise is to reproduce the results on the valuation of mortgages and mortgage-backed bonds presented in Section 14.5.4. The exercise overlaps somewhat with Exercise 14.1.

Assume that the one-factor CIR model is true, that is

$$dr_t = \kappa[\theta - r_t] \, dt + \beta\sqrt{r_t} \, dz_t, \qquad \lambda(r, t) = \lambda\sqrt{r}/\beta,$$

where $z = (z_t)$ is a standard Brownian motion under the real-world probability measure $\mathbb{P}$. Assume that $\kappa = 0.3, \theta = 0.045$, and $\beta = 0.15$, and determine the value of the constant λ such that the long rate y_∞ equals 0.05.

Consider a 30-year mortgage with quarterly payments. The mortgage is an annuity loan in the sense that the sum of the scheduled interest payment and scheduled repayment is the same for all 120 payment dates. The mortgage is issued by a financial institution and financed by the issuance of 'pass-through' mortgage-backed bonds. The annualized coupon rate is 5% and the face value is set at 100. At each of the payment dates, the borrower must also pay a fee to the financial institution. The fee equals 0.125% of the outstanding debt after the previous payment date (corresponding to an extra 0.5% on an annual basis). This fee is not passed through to bond investors.

(a) Compute and illustrate graphically the different components of the scheduled payments from the borrower and the scheduled payments to the bond investor over the 30-year period.

The mortgage borrower has the right to prepay the loan at any payment date. In order to value the mortgage and the bond and to find the optimal prepayment strategy, you are asked to solve the relevant partial differential equation numerically using a finite difference technique. First, however, assume that the borrower cannot choose to prepay so that all payments are as scheduled.

(b) Explain how you can apply and implement the finite difference technique for the CIR model. In particular, explain how you will take into account intermediate payments in your finite difference valuation, for example the quarterly payments of the bond introduced above.

(c) Use your finite difference scheme to compute the price of the bond *ignoring possible prepayments*, that is compute the present value of the scheduled bond payments. Compare with the price you can compute analytically using the well-known pricing formula for zero-coupon bonds in the CIR model. Do this for various selected values of the current short-term interest rate, for example 2%, 3%, 4%, 5%, 6%, and 7%. This will serve as a check of your finite difference implementation and can also give you some guidance concerning the choice of the grid spacing parameters Δt and Δr.

Next, you have to include the possibility of prepayment. The borrower can only prepay by paying the outstanding debt (and prepayment costs, if there are any such costs), not by buying back the bond at market price. The optimal prepayment strategy can be represented by a critical short rate at each payment date such that prepayment is optimal below the critical rate and suboptimal above the critical rate. Your implementation should allow for prepayment costs of the form

$$X(t_n) = x_0 + x_1 D(t_n),$$

where $D(t_n)$ is the outstanding debt immediately after t_n. Your implementation should also allow for the case in which there is some fixed probability Π^e of prepaying although it is suboptimal from an option-theory perspective (when prepayment is optimal from an option-theory perspective, the prepayment probability is set to 1).

(d) Explain how your finite difference algorithm handles the possible prepayments and how it finds the critical short rates.

(e) First assume no costs and no suboptimal prepayments (i.e. $x_0 = x_1 = \Pi^e = 0$). Find the optimal prepayment strategy. Find the present value of the mortgage and the bond for selected values of the current short-term interest rate. Discuss your results and the value of the prepayment option. How do the results depend on the parameters of the interest rate process?

(f) Discuss and illustrate the effects of prepayment costs and the possibility of suboptimal prepayments on the optimal prepayment strategy and the valuation of the mortgage and the bond by varying the parameters x_0, x_1, and Π^e. For example, consider $x_0 \in \{0, 2, 4\}$, $x_1 \in \{0, 0.02, 0.04\}$, and $\Pi^e \in \{0, 0.05, 0.1\}$.

Exercise 16.3 (Monte Carlo, Vasicek) Develop a spreadsheet or a computer program in which you apply the Monte Carlo simulation approach to the one-factor Vasicek model. In particular, you should be able to price a 5-year zero-coupon bond with a face value of 100 and a 2-year European put option on that bond with an exercise price of 82. The risk-neutral dynamics of the short rate is of the form

$$dr_t = \kappa \left[\theta - r_t\right] dt + \beta \, dz_t,$$

where the parameter values are $\kappa = 0.3$, $\beta = 0.03$, and $\theta = 0.065$. The initial short rate is $r_0 = 0.05$.

(a) What are the prices of the bond and the European put option according to the closed-form pricing formulas developed in Chapter 7?

(b) Apply the Monte Carlo simulation approach to approximate the price of the bond in three different ways using $M = 10,000$ samples:

   (i) No discretization: simulate samples directly from the known exact distribution of the short rate at the maturity date of the bond.

   (ii) Exact discretization: split the 5-year period into subperiods of length $\Delta t = 0.1$ year and simulate discrete sample paths using the exact distribution of $r_{t+\Delta t}$ conditional on r_t over each subperiod $[t, t + \Delta t]$.

   (iii) Euler discretization: split the 5-year period into subperiods of length $\Delta t = 0.1$ year and simulate discrete sample paths using an Euler approximation of $r_{t+\Delta t}$ given r_t over each subperiod $[t, t + \Delta t]$.

   For each of the three alternatives, find an approximate 95% confidence interval for the bond price and discuss the accuracy of the pricing approximations.

(c) For one of the three alternatives (preferably the most accurate one) considered in the previous question, vary the number of samples, M, and discuss the effect on the accuracy of the price approximations for the bond.

(d) Derive a Monte Carlo based price approximation for the European put option on the bond. Can you implement all three alternative methods discussed above to price the put? Compare the price approximation with the price computed with the explicit pricing formula.

Exercise 16.4 (Control variates) For the control variate technique, show the Equations (16.37) and (16.38).

Exercise 16.5 (Trinomial tree, CIR) Develop a spreadsheet or a computer program for constructing and applying a trinomial tree for the one-factor CIR model as explained in Section 16.4.2. Confirm that the transition probabilities are as stipulated in Table 16.6 and that the tree produces the bond price stated in the text. Apply the tree for valuing a European put option and an American put option written on this 5-year zero-coupon bond. The options mature in 2 years and have an exercise price of 82 (you can compare the computed prices with the finite difference results in Section 16.2.11).

Exercise 16.6 (Trinomial tree, Vasicek) Consider the construction of a trinomial tree approximating the dynamics of the short-term interest rate in the Vasicek model, that is,

$$dr_t = \kappa(\theta - r_t)\,dt + \beta\,dz_t.$$

(a) Write down the transition probabilities for nodes in which the standard branching procedure applies. Assume $\delta = \sqrt{3}$.

(b) For what value of j do you have to switch to the non-standard 'can't go higher' branching? Write down the transition probabilities for that type of branching.

(c) For what value of j do you have to switch to the non-standard 'can't go lower' branching? Write down the transition probabilities for that type of branching.

Develop a spreadsheet or a computer program in which you can construct and apply the trinomial tree for the Vasicek model. Assume that the above equation shows the risk-neutral dynamics of the short rate and that the parameter values are $\kappa = 0.3$, $\beta = 0.03$, and $\theta = 0.065$. The initial short rate is $r_0 = 0.05$. Construct the tree so that it can be applied to the pricing of a 5-year zero-coupon bond with a face value of 100 and 2-year European and American put options on that bond with an exercise price of 82. Let $\Delta t = 0.1$ year to begin with.

(d) What are the prices of the bond and the European put option according to the closed-form pricing formulas developed in Chapter 7?

(e) What are the prices of the same assets according to your trinomial tree?

(f) Vary Δt (and thus also $\Delta r = \beta\sqrt{3\Delta t}$) and discuss how the accuracy of the prices determined in the tree depends on Δt.

(g) Illustrate the probability distribution of the short rate in 5 years according to the approximating tree and compare this distribution graphically with the exact probability distribution.

(h) Use the trinomial tree to price the American put option and to find the associated exercise boundary.

(i) Can you use the control variate technique to fine-tune your price estimate of the American put?

APPENDIX A

Results on the Lognormal Distribution

A random variable Y is said to be lognormally distributed if the random variable $X = \ln Y$ is normally distributed. In the following we let m be the mean of X and s^2 be the variance of X, so that

$$X = \ln Y \sim N(m, s^2).$$

The probability density function for X is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi s^2}} \exp\left\{-\frac{(x-m)^2}{2s^2}\right\}, \qquad x \in \mathbb{R}.$$

Theorem A.1 *The probability density function for Y is given by*

$$f_Y(y) = \frac{1}{\sqrt{2\pi s^2 y}} \exp\left\{-\frac{(\ln y - m)^2}{2s^2}\right\}, \qquad y > 0,$$

and $f_Y(y) = 0$ for $y \leq 0$.

This result follows from the general result on the distribution of a random variable which is given as a function of another random variable; see any introductory text book on probability theory and distributions.

Theorem A.2 *For $X \sim N(m, s^2)$ and $\gamma \in \mathbb{R}$ we have*

$$\mathrm{E}\left[e^{-\gamma X}\right] = \exp\left\{-\gamma m + \frac{1}{2}\gamma^2 s^2\right\}.$$

Proof: Per definition we have

$$\mathrm{E}\left[e^{-\gamma X}\right] = \int_{-\infty}^{+\infty} e^{-\gamma x} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}}\, dx.$$

Manipulating the exponent we get

$$\begin{aligned}
\mathrm{E}\left[e^{-\gamma X}\right] &= e^{-\gamma m + \frac{1}{2}\gamma^2 s^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{1}{2s^2}\left[(x-m)^2 + 2\gamma(x-m)s^2 + \gamma^2 s^4\right]}\, dx \\
&= e^{-\gamma m + \frac{1}{2}\gamma^2 s^2} \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-[m-\gamma s^2])^2}{2s^2}}\, dx \\
&= e^{-\gamma m + \frac{1}{2}\gamma^2 s^2},
\end{aligned}$$

where the last equality is due to the fact that the function

$$x \mapsto \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-[m-\gamma s^2])^2}{2s^2}}$$

is a probability density function, namely the density function for an $N(m - \gamma s^2, s^2)$ distributed random variable. $\square$

Using this theorem, we can easily compute the mean and the variance of the lognormally distributed random variable $Y = e^X$. The mean is (let $\gamma = -1$)

$$\mathrm{E}[Y] = \mathrm{E}\left[e^X\right] = \exp\left\{m + \frac{1}{2}s^2\right\}.$$

With $\gamma = -2$ we get

$$\mathrm{E}\left[Y^2\right] = \mathrm{E}\left[e^{2X}\right] = e^{2(m+s^2)},$$

so that the variance of Y is

$$\begin{aligned}\mathrm{Var}[Y] &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 \\ &= e^{2(m+s^2)} - e^{2m+s^2} \\ &= e^{2m+s^2}\left(e^{s^2} - 1\right).\end{aligned}$$

The next theorem provides an expression for the truncated mean of a lognormally distributed random variable, that is the mean of the part of the distribution that lies above some level. We define the indicator variable $\mathbf{1}_{\{Y>K\}}$ to be equal to 1 if the outcome of the random variable Y is greater than the constant K and equal to 0 otherwise.

Theorem A.3 *If* $X = \ln Y \sim N(m, s^2)$ *and* $K > 0$, *then we have*

$$\begin{aligned}\mathrm{E}\left[Y\mathbf{1}_{\{Y>K\}}\right] &= e^{m+\frac{1}{2}s^2} N\left(\frac{m - \ln K}{s} + s\right) \\ &= \mathrm{E}\left[Y\right] N\left(\frac{m - \ln K}{s} + s\right).\end{aligned}$$

Proof: Because $Y > K \Leftrightarrow X > \ln K$, it follows from the definition of the expectation of a random variable that

$$\begin{aligned}\mathrm{E}\left[Y\mathbf{1}_{\{Y>K\}}\right] &= \mathrm{E}\left[e^X \mathbf{1}_{\{X>\ln K\}}\right] \\ &= \int_{\ln K}^{+\infty} e^x \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-m)^2}{2s^2}}\, dx \\ &= \int_{\ln K}^{+\infty} \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-[m+s^2])^2}{2s^2}} e^{\frac{2ms^2+s^4}{2s^2}}\, dx \\ &= e^{m+\frac{1}{2}s^2} \int_{\ln K}^{+\infty} f_{\bar{X}}(x)\, dx,\end{aligned}$$

where

$$f_{\bar{X}}(x) = \frac{1}{\sqrt{2\pi s^2}} e^{-\frac{(x-[m+s^2])^2}{2s^2}}$$

is the probability density function for an $N(m+s^2, s^2)$ distributed random variable. The calculations

$$\begin{aligned}
\int_{\ln K}^{+\infty} f_{\bar{X}}(x)\,dx &= \mathrm{Prob}(\bar{X} > \ln K) \\
&= \mathrm{Prob}\left(\frac{\bar{X}-[m+s^2]}{s} > \frac{\ln K-[m+s^2]}{s}\right) \\
&= \mathrm{Prob}\left(\frac{\bar{X}-[m+s^2]}{s} < -\frac{\ln K-[m+s^2]}{s}\right) \\
&= N\left(-\frac{\ln K-[m+s^2]}{s}\right) \\
&= N\left(\frac{m-\ln K}{s}+s\right)
\end{aligned}$$

complete the proof. □

Theorem A.4 *If $X = \ln Y \sim N(m, s^2)$ and $K > 0$, we have*

$$\begin{aligned}
\mathrm{E}\left[\max\left(0, Y-K\right)\right] &= e^{m+\frac{1}{2}s^2} N\left(\frac{m-\ln K}{s}+s\right) - KN\left(\frac{m-\ln K}{s}\right) \\
&= \mathrm{E}\left[Y\right] N\left(\frac{m-\ln K}{s}+s\right) - KN\left(\frac{m-\ln K}{s}\right).
\end{aligned}$$

Proof: Note that

$$\begin{aligned}
\mathrm{E}\left[\max\left(0, Y-K\right)\right] &= \mathrm{E}\left[(Y-K)\mathbf{1}_{\{Y>K\}}\right] \\
&= \mathrm{E}\left[Y\mathbf{1}_{\{Y>K\}}\right] - K\mathrm{Prob}\left(Y>K\right).
\end{aligned}$$

The first term is known from Theorem A.3. The second term can be rewritten as

$$\begin{aligned}
\mathrm{Prob}\left(Y>K\right) &= \mathrm{Prob}\left(X>\ln K\right) \\
&= \mathrm{Prob}\left(\frac{X-m}{s} > \frac{\ln K-m}{s}\right) \\
&= \mathrm{Prob}\left(\frac{X-m}{s} < -\frac{\ln K-m}{s}\right) \\
&= N\left(-\frac{\ln K-m}{s}\right) \\
&= N\left(\frac{m-\ln K}{s}\right).
\end{aligned}$$

The claim now follows immediately. □

References

Abramowitz, M. and I. Stegun (1972). *Handbook of Mathematical Functions*. Dover Publications.

Acharya, V. V., S. R. Das, and R. K. Sundaram (2002). Pricing credit derivatives with rating transitions. *Financial Analysts Journal 58*(3), 28–44.

Aiyagari, S. R. (1994). Frictions in asset pricing and macroeconomics. *European Economic Review 38*(3–4), 932–939.

Altman, E., B. Brady, A. Resti, and A. Sironi (2005). The link between default and recovery rates: Theory, empirical evidence, and implications. *Journal of Business 78*(6), 2203–2228.

Amihud, Y., H. Mendelson, and L. H. Pedersen (2005). Liquidity and asset prices. *Foundations and Trends in Finance 1*(4), 269–364.

Amin, K. I. and R. A. Jarrow (1991). Pricing foreign currency options under stochastic interest rates. *Journal of International Money and Finance 10*(3), 310–329.

Amin, K. I. and R. A. Jarrow (1992). Pricing options on risky assets in a stochastic interest rate economy. *Mathematical Finance 2*(4), 217–237.

Amin, K. I. and A. J. Morton (1994). Implied volatility functions in arbitrage-free term structure models. *Journal of Financial Economics 35*(2), 141–180.

Andersen, L. (2000). A simple approach to the pricing of Bermudan swaptions in the multifactor LIBOR market model. *Journal of Computational Finance 3*(2), 1–32.

Andersen, L. (2007). Discount curve construction with tension splines. *Review of Derivatives Research 10*(3), 227–267.

Andersen, L. (2008). Simple and efficient simulation of the Heston stochastic volatility model. *Journal of Computational Finance 11*(3), 1–42.

Andersen, L. and J. Andreasen (2000). Volatility skews and extensions of the LIBOR market model. *Applied Mathematical Finance 7*(1), 1–32.

Andersen, L. and M. Broadie (2004). A primal-dual simulation algorithm for pricing multidimensional American options. *Management Science 50*(9), 1222–1234.

Andersen, L. and R. Brotherton-Ratcliffe (2001, December). Extended LIBOR market models with stochastic volatility. Working paper available at SSRN: http://ssrn.com/abstract=294853.

Andersen, L. and J. Sidenius (2005). Extensions of the Gaussian copula model. *Journal of Credit Risk 1*(1), 29–70.

Andersen, T. G. and J. Lund (1997). Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics 77*(2), 343–377.

Anderson, N., F. Breedon, M. Deacon, A. Derry, and G. Murphy (1996). *Estimating and Interpreting the Yield Curve*. John Wiley & Sons, Inc.

Arrow, K. (1951). An extension of the basic theorems of classical welfare economics. In J. Neyman (Ed.), *Proceedings of the Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 507–532. University of California Press.

Arrow, K. (1953). Le rôle des valeurs boursières pour la repartition la meillure des risques. *Econometrie 40*, 41–47. English translation: Arrow (1964).

Arrow, K. (1964). The role of securities in the optimal allocation of risk-bearing. *Review of Economic Studies 31*(2), 91–96.

Arrow, K. (1971). *Essays in the Theory of Risk Bearing*. North-Holland.

Asmussen, S. and P. W. Glynn (2007). *Stochastic Simulation: Algorithms and Analysis*. Springer Verlag.

Azevedo-Pereira, J. A., D. P. Newton, and D. A. Paxson (2003). Fixed-rate endowment mortgage and mortgage indemnity valuation. *Journal of Real Estate Finance and Economics 26*(2–3), 197–221.

Babbs, S. H. and N. J. Webber (1994, February). A theory of the term structure with an official short rate. Working paper, Warwick Business School, University of Warwick, Coventry CV4 7AL, UK.

Bachelier, L. (1900). *Théorie de la Spéculation*, Volume 3 of *Annales de l'Ecole Normale Supérieure*. Gauthier-Villars. English translation in Cootner (1964).

Bakshi, G., D. Madan, and F. Zhang (2006). Investigating the role of systematic and firm-specific factors in default risk: Lessons from empirically evaluating credit risk models. *Journal of Business 79*(4), 1955–1988.

Bakshi, G. S. and Z. Chen (1996). Inflation, asset prices and the term structure of interest rates in monetary economies. *Review of Financial Studies 9*(1), 241–275.

Bakshi, G. S. and Z. Chen (1997). An alternative valuation model for contingent claims. *Journal of Financial Economics 44*(1), 123–165.

Bakshi, G. S., C. Cao, and Z. Chen (1997). Empirical performance of alternative option pricing models. *Journal of Finance 52*(5), 2003–2049.

Balduzzi, P., G. Bertola, and S. Foresi (1997). A model of target changes and the term structure of interest rates. *Journal of Monetary Economics 39*(2), 223–249.

Balduzzi, P., S. R. Das, and S. Foresi (1998). The central tendency: A second factor in bond yields. *Review of Economics and Statistics 80*(1), 62–72.

Balduzzi, P., S. R. Das, S. Foresi, and R. Sundaram (1996). A simple approach to three-factor affine term structure models. *Journal of Fixed Income 6*(3), 43–53.

Bank for International Settlements (2009). *BIS Semiannual OTC derivatives statistics at end-March 2009*. Bank for International Settlements. Available at http://www.bis.org/statistics/derstats.htm.

Bank for International Settlements (2010, March). *BIS Quarterly Review: International Banking and Financial Market Developments. Statistical Annex*. Bank for International Settlements. Available at http://www.bis.org/publ.

Bansal, R. and I. Shaliastovich (2009, July). A long-run risks explanation of predictability puzzles in bond and currency markets. Working paper, Duke University.

Bansal, R. and A. Yaron (2004). Risks for the long run: A potential resolution of asset pricing puzzles. *Journal of Finance 59*(4), 1481–1509.

Barraquand, J. and D. Martineau (1995). Numerical valuation of high dimensional multivariate American securities. *Journal of Financial and Quantitative Analysis 30*(3), 383–405.

Batten, J. A., T. A. Fetherston, and P. G. Szilagyi (Eds.) (2004). *European Fixed Income Markets*. Wiley.

Beaglehole, D. R. and M. S. Tenney (1991). General solutions of some interest rate-contingent claim pricing equations. *Journal of Fixed Income 1*(2), 69–83.

Beaglehole, D. R. and M. S. Tenney (1992). Corrections and additions to "A nonlinear equilibrium model of the term structure of interest rates". *Journal of Financial Economics 32*(3), 345–353.

Bebchuk, L. A. (2002). Ex ante costs of violating absolute priority in bankruptcy. *Journal of Finance 57*(1), 445–460.

Berndt, A., B. Hollifield, and P. Sandås (2010, May). The role of mortgage brokers in the subprime crisis. Available at SSRN: http://ssrn.com/abstract=1573312.

Bhansali, V., R. Gingrich, and F. A. Longstaff (2008). Systemic credit risk: What is the market telling us? *Financial Analysts Journal 64*(4), 16–24.

Bhar, R. and C. Chiarella (1997). Transformation of Heath-Jarrow-Morton models to Markovian systems. *European Journal of Finance 3*(3), 1–26.

Bhar, R., C. Chiarella, N. El-Hassan, and X. Zheng (2000). The reduction of forward rate dependent volatility HJM models to Markovian form: Pricing European bond options. *Journal of Computational Finance 3*(3), 47–72.

Bharath, S. T. and T. Shumway (2008). Forecasting default with the Merton distance to default model. *Review of Financial Studies 21*(3), 1339–1369.

Bielecki, T. R. and M. Rutkowski (2002). *Credit Risk: Modeling, Valuation and Hedging*. Springer.

Björk, T. (2009). *Arbitrage Theory in Continuous Time* (Third ed.). Oxford University Press.

Björk, T. and B. J. Christensen (1999). Interest rate dynamics and consistent forward rate curves. *Mathematical Finance 9*(4), 323–348.

Björk, T. and C. Landén (2002). On the construction of finite dimensional realizations for nonlinear forward rate models. *Finance and Stochastics 6*(3), 303–331.

Black, F. (1976). The pricing of commodity contracts. *Journal of Financial Economics 3*(1–2), 167–179.

Black, F. and J. Cox (1976). Valuing corporate securities: Some effects of bond indenture provisions. *Journal of Finance 31*(2), 351–367.

Black, F. and P. Karasinski (1991). Bond and option pricing when short rates are lognormal. *Financial Analysts Journal 47*(4), 52–59.

Black, F. and M. Scholes (1973). The pricing of options and corporate liabilities. *Journal of Political Economy 81*(3), 637–654.

Black, F., E. Derman, and W. Toy (1990). A one-factor model of interest rates and its application to treasury bond options. *Financial Analysts Journal 46*(1), 33–39.

Bliss, R. R. and D. C. Smith (1997, November). The stability of interest rate processes. Working paper 97-13, Federal Reserve Bank of Atlanta.

Bongaerts, D., F. de Jong, and J. Driessen (2009). Derivative pricing with liquidity risk: Theory and evidence from the credit default swap market. Forthcoming in *Journal of Finance*.

Boudoukh, J., M. Richardson, R. Stanton, and R. F. Whitelaw (1999, June). A multifactor, nonlinear, continuous-time model of interest rate volatility. Working paper, Stern School of Business, NYU and Haas School of Business, UC Berkeley. Forthcoming in “Volatility and Time Series Econometrics: Essays in Honor of Robert F. Engle”, Oxford University Press.

Box, G. E. P. and M. E. Muller (1958). A note on the generation of random normal deviates. *The Annals of Mathematical Statistics 29*(2), 610–611.

Boyle, P., M. Broadie, and P. Glasserman (1997). Monte Carlo methods for security pricing. *Journal of Economic Dynamics and Control 21*(8–9), 1267–1321.

Boyle, P. P. (1977). Options: A Monte Carlo approach. *Journal of Financial Economics 4*(3), 323–338.

Brace, A., D. Gatarek, and M. Musiela (1997). The market model of interest rate dynamics. *Mathematical Finance 7*(2), 127–155.

Breeden, D. T. (1986). Consumption, production, inflation and interest rates. *Journal of Financial Economics 16*(1), 3–39.

Breeden, D. T., M. R. Gibbons, and R. H. Litzenberger (1989). Empirical tests of the consumption-oriented CAPM. *Journal of Finance 44*(2), 231–262.

Brennan, M. J. and E. S. Schwartz (1977). The valuation of American put options. *Journal of Finance 32*(2), 449–462.

Brennan, M. J. and E. S. Schwartz (1979). A continuous time approach to the pricing of bonds. *Journal of Banking and Finance 3*(2), 133–155.

Brennan, M. J. and E. S. Schwartz (1980). Analyzing convertible bonds. *Journal of Financial and Quantitative Analysis 15*(4), 907–929.

Brennan, M. J. and E. S. Schwartz (1984). Valuation of corporate claims: Optimal financial policy and firm vaulation. *Journal of Finance 39*(3), 593–609.

Brennan, M. J. and E. S. Schwartz (1985). Determinants of GNMA mortgage prices. *Journal of the American Real Estate and Urban Economics Association 13*(3), 209–228.

Brenner, R., R. Harjes, and K. Kroner (1996). Another look at alternative models of the short-term interest rate. *Journal of Financial and Quantitative Analysis 31*(1), 85–107.

Brigo, D. and F. Mercurio (2006). *Interest Rate Models – Theory and Practice* (Second ed.). Springer-Verlag.

Briys, E. and F. de Varenne (1997). Valuing risky fixed rate debt: An extension. *Journal of Financial and Quantitative Analysis 32*(2), 239–248.

Broadie, M. and P. Glasserman (1997). Pricing American-style securities using simulation. *Journal of Economic Dynamics and Control 21*(8–9), 1323–1352.

Broadie, M., P. Glasserman, and G. Jain (1997). Enhanced Monte Carlo estimates for American option prices. *Journal of Derivatives 5*(1), 25–44.

Brown, R. H. and S. M. Schaefer (1994). The term structure of real interest rates and the Cox, Ingersoll, and Ross model. *Journal of Financial Economics 35*(1), 3–42.

Brown, S. J. and P. H. Dybvig (1986). The empirical implications of the Cox, Ingersoll, and Ross theory of the term structure of interest rates. *Journal of Finance 41*(3), 617–630.

Brunnermeier, M. K. (2009). Deciphering the liquidity and credit crunch 2007–2008. *Journal of Economic Perspectives 23*(1), 77–100.

Burtschell, X., J. Gregory, and J.-P. Laurent (2009). A comparative analysis of CDO pricing models. *Journal of Derivatives 16*(4), 9–37.

Buser, S. A., P. H. Hendershott, and A. B. Sanders (1990). Determinants of the value of call options on default-free bonds. *Journal of Business 63*(1), S33–S50.

Campbell, J. Y. (1986). A defense of traditional hypotheses about the term structure of interest rates. *Journal of Finance 41*(1), 183–193.

Campbell, J. Y. and J. F. Cocco (2003). Household risk management and optimal mortgage choice. *The Quarterly Journal of Economics 118*(4), 1449–1494.

Campbell, J. Y., A. W. Lo, and A. C. MacKinlay (1997). *The Econometrics of Financial Markets*. Princeton University Press.

Canabarro, E. (1995). Where do one-factor interest rate models fail? *Journal of Fixed Income 5*(2), 31–52.

Carr, P. and G. Yang (2001). Simulating Bermudan interest rate derivatives. In M. Avellaneda (Ed.), *Quantitative Analysis in Financial Markets*, Volume 2, pp. 295–316. Springer-Verlag.

Carriére, J. F. (1996). Valuation of the early-exercise price for options using simulations and nonparametric regression. *Insurance: Mathematics and Economics 19*(1), 19–30.

Carverhill, A. (1994). When is the short rate Markovian? *Mathematical Finance 4*(4), 305–312.

Carverhill, A. (1995). A note on the models of Hull and White for pricing options on the term structure. *Journal of Fixed Income 5*(2), 89–96.

Casassus, J., P. Collin-Dufresne, and R. S. Goldstein (2005). Unspanned stochastic volatility and fixed income derivatives pricing. *Journal of Banking and Finance 29*(11), 2723–2749.

Cathcart, L. and L. El-Jahel (1998). Valuation of defaultable bonds. *Journal of Fixed Income 8*(1), 65–78.

Cetin, U., R. Jarrow, P. Protter, and Y. Yildirim (2004). Modeling credit risk with partial information. *Annals of Applied Probability 14*(3), 1167–1178.

Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. Sanders (1992). An empirical comparison of alternative models of the short-term interest rate. *Journal of Finance 47*(3), 1209–1227.

Chapman, D. A., J. B. Long, Jr., and N. D. Pearson (1999). Using proxies for the short rate: When are three months like an instant? *Review of Financial Studies 12*(4), 763–806.

Chen, L. (1996). Stochastic mean and stochastic volatility—a three-factor model of the term structure of interest rates and its applications in derivatives pricing and risk management. *Financial Markets, Institutions & Instruments 5*(1), 1–88.

Chen, L., D. Lesmond, and J. Wei (2007). Corporate yield spreads and bond liquidity. *Journal of Finance 62*(1), 119–149.

Chen, N.-F., R. Roll, and S. A. Ross (1986). Economic forces and the stock market. *Journal of Business 59*(3), 383–403.

Chen, R.-R. and L. Scott (1992). Pricing interest rate options in a two-factor Cox-Ingersoll-Ross model of the term structure. *Review of Financial Studies 5*(4), 613–636.

Chen, R.-R. and L. Scott (1993). Maximum likelihood estimation for a multifactor equilibrium model of the term structure of interest rates. *Journal of Fixed Income 3*(3), 14–31.

Cheridito, P., D. Filipovic, and R. L. Kimmel (2007). Market price of risk specifications for affine models: Theory and evidence. *Journal of Financial Economics 83*(1), 123–170.

Cherubini, U., E. Luciano, and W. Vecchiato (2004). *Copula Methods in Finance*. Wiley.

Cheyette, O. (1996, August). Markov representation of the Heath–Jarrow–Morton model. Working paper, BARRA Inc.

Childs, P. D., S. H. Ott, and T. J. Riddiough (1996). The pricing of multiclass commercial mortgage-backed securities. *Journal of Financial and Quantitative Analysis 31*(4), 581–603.

Christensen, B. J., R. Poulsen, and M. Sørensen (2001, November). Optimal inference in diffusion models of the short rate of interest. Working paper 102, Centre for Analytical Finance, Aarhus.

Christensen, J. H. E., F. X. Diebold, and G. D. Rudebusch (2009). An arbitrage-free generalized Nelson-Siegel term structure model. *Econometrics Journal 12*(3), C33–C64.

Christensen, P. O. and B. G. Sørensen (1994). Duration, convexity, and time value. *Journal of Portfolio Management 20*(2), 51–60.

Christensen, P. O., K. Larsen, and C. Munk (2010, May). Equilibrium in securities markets with heterogeneous investors and unspanned income risk. Available at SSRN: http://ssrn.com/abstract=1364237.

Christensen, P. O., D. Lando, and K. R. Miltersen (1997). State-dependent realignments in target zone currency regimes. *Review of Derivatives Research 1*(4), 295–323.

Christensen, P. O., C. R. Flor, D. Lando, and K. R. Miltersen (2002). Dynamic capital structure with callable debt and debt renegotiation. Available at SSRN: http://ssrn.com/abstract=320161.

Clément, E., D. Lamberton, and P. Protter (2002). An analysis of a least squares regression method for American option pricing. *Finance and Stochastics 6*(4), 449–471.

Clewlow, L. and C. Strickland (1994). A note on parameter estimation in the two-factor Longstaff and Schwartz interest rate model. *Journal of Fixed Income 3*(4), 95–100.

Cochrane, J. H. (2005). *Asset Pricing* (Revised ed.). Princeton University Press.

Collin-Dufresne, P. and R. S. Goldstein (2001). Do credit spreads reflect stationary leverage? Reconciling structural and reduced-form frameworks. *Journal of Finance 56*(5), 1929–1957.

Collin-Dufresne, P. and R. S. Goldstein (2002a). Do bonds span the fixed income markets? Theory and evidence for unspanned stochastic volatility. *Journal of Finance 57*(4), 1685–1730.

Collin-Dufresne, P. and R. S. Goldstein (2002b). Pricing swaptions within the affine framework. *Journal of Derivatives 10*(1), 1–18.

Collin-Dufresne, P., R. S. Goldstein, and J. S. Martin (2001). The determinants of credit spread changes. *Journal of Finance 56*(6), 2177–2207.

Constantinides, G. M. (1992). A theory of the nominal term structure of interest rates. *Review of Financial Studies 5*(4), 531–552.

Constantinides, G. M. and D. Duffie (1996). Asset pricing with heterogeneous consumers. *Journal of Political Economy 104*(2), 219–240.

Cootner, P. H. (Ed.) (1964). *The Random Character of Stock Market Prices*. MIT Press.

Courtadon, G. (1982). The pricing of options on default-free bonds. *Journal of Financial and Quantitative Analysis 17*(1), 75–100.

Cox, J. C. and S. A. Ross (1976). The valuation of options for alternative stochastic processes. *Journal of Financial Economics 3*(1-2), 145–166.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1979). Duration and the measurement of basis risk. *Journal of Business 52*(1), 51–61.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1981a). A re-examination of traditional hypotheses about the term structure of interest rates. *Journal of Finance 36*(4), 769–799.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1981b). The relation between forward prices and futures prices. *Journal of Financial Economics 9*(4), 321–346.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1985a). An intertemporal general equilibrium model of asset prices. *Econometrica 53*(2), 363–384.

Cox, J. C., J. E. Ingersoll, Jr., and S. A. Ross (1985b). A theory of the term structure of interest rates. *Econometrica 53*(2), 385–407.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979). Option pricing: A simplified approach. *Journal of Financial Economics 7*(3), 229–263.

Crouhy, M., D. Galai, and R. Mark (2000). A comparative analysis of current credit risk models. *Journal of Banking and Finance 24*(1–2), 59–117.

Crouhy, M. G., R. A. Jarrow, and S. M. Turnbull (2008). The subprime credit crisis of 07. *Journal of Derivatives 16*(1), 81–110.

Culbertson, J. M. (1957). The term structure of interest rates. *Quarterly Journal of Economics 71*(4), 489–504.

Dai, Q. and K. J. Singleton (2000). Specification analysis of affine term structure models. *Journal of Finance 55*(5), 1943–1978.

Das, S. R. and P. Tufano (1996). Pricing credit-sensitive debt when interest rates, credit ratings and credit spreads are stochastic. *Journal of Financial Engineering 5*(2), 161–198.

Daves, P. R. and M. C. Ehrhardt (1993). Joint cross-section/time-series maximum likelihood estimation for the parameters of the Cox-Ingersoll-Ross bond pricing model. *The Financial Review 28*(2), 203–237.

De Jong, F., J. Driessen, and A. Pelsser (2001). LIBOR market models versus swap market models for pricing interest rate derivatives: An empirical analysis. *European Finance Review 5*(3), 201–237.

Debreu, G. (1953). Une economie l'incertain. Working paper, Electricité de France.

Debreu, G. (1954). Valuation equilibrium and pareto optimum. *Proceedings of the National Academy of Sciences 40*(7), 588–592.

Debreu, G. (1959). *Theory of Value*. Yale University Press.

Delbaen, F. and W. Schachermayer (1994). A general version of the fundamental theorem of asset pricing. *Mathematische Annalen 300*(1), 463–520.

Delbaen, F. and W. Schachermayer (1999). The fundamental theorem of asset pricing for unbounded stochastic processes. *Mathematische Annalen 312*(2), 215–250.

Dell' Aquila, R., E. Ronchetti, and F. Trojani (2003). Robust GMM analysis of models for the short rate process. *Journal of Empirical Finance 10*(3), 373–397.

Demyanyk, Y. S. and O. Van Hemert (2008, December). Understanding the subprime mortgage crisis. Available at SSRN: http://ssrn.com/abstract=1020396. Forthcoming in *Review of Financial Studies*.

Deng, Y., J. M. Quigley, and R. Van Order (2000). Mortgage terminations, heterogeneity and the exercise of mortgage options. *Econometrica 68*(2), 275–307.

Dick-Nielsen, J., P. Feldhütter, and D. Lando (2009, March). Corporate bond liquidity before and after the onset of the subprime crisis. Available at http://ssrn.com/ abstract=1364635.

Ding, C. G. (1992). Algorithm AS 275: Computing the non-central χ^2 distribution function. *Applied Statistics 41*(2), 478–482.

Dorn, J. (2007). Modeling of CDO squareds: Capturing the second dimension. *Journal of Fixed Income 17*(2), 27–45.

Dothan, M. U. (1978). On the term structure of interest rates. *Journal of Financial Economics 6*(1), 59–69.

Dothan, M. U. (1990). *Prices in Financial Markets*. Oxford University Press.

Driessen, J. (2005). Is default event risk priced in corporate bonds? *Review of Financial Studies 18*(1), 165–195.

Duarte, J. (2004). Evaluating an alternative risk preference in affine term structure models. *Review of Financial Studies 17*(2), 379–404.

Duffee, G. R. (1996). Idiosyncratic variation of Treasury Bill yields. *Journal of Finance 51*(2), 527–551.

Duffee, G. R. (1998). The relation between treasury yields and corporate bond yield spreads. *Journal of Finance 53*(6), 2225–2241.

Duffee, G. R. (1999). Estimating the price of default risk. *Review of Financial Studies 12*(1), 197–226.

Duffee, G. R. (2002). Term premia and interest rate forecasts in affine models. *Journal of Finance 57*(1), 405–443.

Duffie, D. (2001). *Dynamic Asset Pricing Theory* (Third ed.). Princeton University Press.

Duffie, D. and L. G. Epstein (1992). Asset pricing with stochastic differential utility. *Review of Financial Studies 5*(3), 411–436.

Duffie, D. and N. Gârleanu (2001). Risk and valuation of collateralized debt obligations. *Financial Analysts Journal 57*(1), 41–59.

Duffie, D. and M. Huang (1996). Swap rates and credit quality. *Journal of Finance 51*(3), 921–949.

Duffie, D. and R. Kan (1996). A yield-factor model of interest rates. *Mathematical Finance 6*(4), 379–406.

Duffie, D. and D. Lando (2001). Term structures of credit spreads with incomplete accounting information. *Econometrica 69*(3), 633–664.

Duffie, D. and J. Liu (2001). Floating-fixed credit spreads. *Financial Analysts Journal 57*(3), 76–87.

Duffie, D. and K. Singleton (1999). Modeling term structures of defaultable bonds. *Review of Financial Studies 12*(4), 687–720.

Duffie, D. and K. Singleton (2003). *Credit Risk: Pricing, Measurement, and Management*. Princeton University Press.

Duffie, D. and R. Stanton (1992). Pricing continuously resettled contingent claims. *Journal of Economic Dynamics and Control 16*(3–4), 561–574.

Duffie, D. and H. Zhu (2009, May). Does a central clearing counterparty reduce counterparty risk? Available at SSRN: http://ssrn.com/abstract=1348343.

Duffie, D., J. Ma, and J. Yong (1995). Black's consol rate conjecture. *Annals of Applied Probability 5*(2), 356–382.

Duffie, D., J. Pan, and K. Singleton (2000). Transform analysis and asset pricing for affine jump-diffusions. *Econometrica 68*(6), 1343–1376.

Duffie, D., L. Saita, and K. Wang (2007). Multi-period corporate default prediction with stochastic covariates. *Journal of Financial Economics 83*(3), 635–665.

Dumas, B., L. P. Jennergren, and B. Näslund (1995). Realignment risk and currency option pricing in target zones. *European Economic Review 39*(8), 1523–1544.

Dunn, K. B. and J. J. McConnell (1981a). A comparison of alternative models for pricing GNMA mortage-backed securities. *Journal of Finance 36*(2), 471–484.

Dunn, K. B. and J. J. McConnell (1981b). Valuation of GNMA mortage-backed securities. *Journal of Finance 36*(3), 599–616.

Dybvig, P. H. (1988). Bond and bond option pricing based on the current term structure. Working paper, School of Business, Washington University, St. Louis.

Eberhart, A. C. and L. W. Senbet (1993). Absolute priority rule violations and risk incentives for financially distressed firms. *Financial Management 22*(3), 101–116.

Eberhart, A. C. and L. A. Weiss (1998). The importance of deviations from the absolute priority rule in Chapter 11 bankruptcy proceedings. *Financial Management 27*(4), 106–110.

Eckner, A. (2009). Computational techniques for basic affine models of portfolio credit risk. *Journal of Computational Finance 13*(1), 63–97.
Elton, E. J., M. J. Gruber, D. Agrawal, and C. Mann (2001). Explaining the rate spread on corporate bonds. *Journal of Finance 56*(1), 247–277.
Emery, K., S. Ou, J. Tennant, A. Matos, and R. Cantor (2009, February). Corporate default and recovery rates, 1920-2008. Special comment, Moody's Investors Service.
Eom, Y. H., J. Helwege, and J.-Z. Huang (2004). Structural models of corporate bond pricing: An empirical analysis. *Review of Financial Studies 17*(2), 499–544.
Fabozzi, F. J. (2010). *Bond Markets, Analysis and Strategies* (Seventh ed.). Pearson Prentice-Hall.
Fama, E. F. (1981). Stock returns, real activity, inflation, and money. *American Economic Review 71*(4), 545–565.
Fama, E. F. and M. Gibbons (1982). Inflation, real returns and capital investment. *Journal of Monetary Economics 9*(3), 297–323.
Feldhütter, P. (2008a, April). Can affine models match the moments in bond yields? Available at SSRN: http://ssrn.com/abstract=915323.
Feldhütter, P. (2008b). An empirical investigation of an intensity-based model for pricing CDO tranches. Available at SSRN: http://ssrn.com/abstract=984836.
Filipović, D. (1999). A note on the Nelson-Siegel family. *Mathematical Finance 9*(4), 349–359.
Fischer, E., R. Heinkel, and J. Zechner (1989). Dynamic capital structure choice: Theory and tests. *Journal of Finance 44*(1), 19–40.
Fisher, I. (1896). Appreciation and interest. *Publications of the American Economic Association*, 23–29 and 88–92.
Fisher, I. (1907). *The Rate of Interest*. Macmillan.
Fisher, L. and R. L. Weil (1971). Coping with the risk of interest rate fluctuations: Returns to bondholders from naive and optimal strategies. *Journal of Business 44*(4), 408–431.
Fisher, M. and C. Gilles (1998). Around and around: The expectations hypothesis. *Journal of Finance 52*(1), 365–383.
Flesaker, B. (1993). Testing the Heath-Jarrow-Morton/Ho-Lee model of interest rate contingent claims pricing. *Journal of Financial and Quantitative Analysis 28*(4), 483–485.
Fons, J. S. (1994). Using default rates to model the term structure of credit risk. *Financial Analysts Journal 50*(5), 25–32.
Garcia, D. (2003). Convergence and biases of Monte Carlo estimates of American option prices using a parametric exercise rule. *Journal of Economic Dynamics and Control 27*(10), 1855–1879.
Garman, M. B. and S. W. Kohlhagen (1983). Foreign currency option values. *Journal of International Money and Finance 2*(3), 231–237.
Geman, H. (1989). The importance of the forward neutral probability in a stochastic approach of interest rates. Working paper, ESSEC.
Geske, R. (1979). The valuation of compound options. *Journal of Financial Economics 7*(1), 63–81.
Gibbons, M. R. and K. Ramaswamy (1993). A test of the Cox, Ingersoll, and Ross model of the term structure. *Review of Financial Studies 6*(3), 619–658.
Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer Verlag.
Glasserman, P. and S. G. Kou (2003). The term structure of simple forward rates with jump risk. *Mathematical Finance 13*(3), 383–410.
Goldstein, R. and F. Zapatero (1996). General equilibrium with constant relative risk aversion and Vasicek interest rates. *Mathematical Finance 6*(3), 331–340.
Goldstein, R., N. Ju, and H. Leland (2001). An EBIT-based model of dynamic capital structure. *Journal of Business 74*(4), 483–512.

Gorovoy, V. and V. Linetsky (2007). Intensity-based valuation of residential mortgages: An analytically tractable model. *Mathematical Finance 17*(4), 541–573.

Gorton, G. (2009). The subprime panic. *European Financial Management 15*(1), 10–46.

Grabbe, J. O. (1983). The pricing of call and put options on foreign exchange. *Journal of International Money and Finance 2*(3), 239–253.

Green, J. and J. B. Shoven (1986). The effects of interest rates on mortgage prepayments. *Journal of Money, Credit, and Banking 18*(1), 41–59.

Gregory, J. and J.-P. Laurent (2005). Basket default swaps, CDOs and factor copulas. *Journal of Risk 7*(4), 103–122.

Grossman, S. J., A. Melino, and R. J. Shiller (1987). Estimating the continuous-time consumption-based asset-pricing model. *Journal of Business & Economic Statistics 5*(3), 315–327.

Gupta, A. and M. G. Subrahmanyam (2005). Pricing and hedging interest rate options: Evidence from cap floor markets. *Journal of Banking and Finance 29*(3), 701–733.

Hagan, P. S. and G. West (2006). Interpolation methods for curve construction. *Applied Mathematical Finance 13*(2), 89–129.

Hagan, P. S., D. Kumar, A. S. Lesniewski, and D. E. Woodward (2002). Managing smile risk. *Wilmott Magazine* (September), 84–108.

Hansen, A. T. and P. L. Jørgensen (2000). Fast and accurate approximation of bond prices when short interest rates are log-normal. *Journal of Computational Finance 3*(3), 27–46.

Harrison, J. M. and D. M. Kreps (1979). Martingales and arbitrage in multiperiod securities markets. *Journal of Economic Theory 20*(3), 381–408.

Harrison, J. M. and S. R. Pliska (1981). Martingales and stochastic integrals in the theory of continuous trading. *Stochastic Processes and their Applications 11*(3), 215–260.

Harrison, J. M. and S. R. Pliska (1983). A stochastic calculus model of continuous trading: Complete markets. *Stochastic Processes and their Applications 15*(3), 313–316.

Haugh, M. B. and L. Kogan (2004). Pricing American options: A duality approach. *Operations Research 52*(2), 258–270.

He, H. (1990). Convergence from discrete- to continuous-time contingent claims prices. *Review of Financial Studies 3*(4), 523–546.

Heath, D., R. Jarrow, and A. Morton (1990). Contingent claim valuation with a random evolution of interest rates. *Review of Future Markets 9*(1), 54–82.

Heath, D., R. Jarrow, and A. Morton (1992). Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation. *Econometrica 60*(1), 77–105.

Heidari, M. and L. Wu (2003). Are interest rate derivatives spanned by the term structure of interest rates? *Journal of Fixed Income 13*(1), 75–86.

Helwege, J. and C. M. Turner (1999). The slope of the credit yield curve for speculative-grade issuers. *Journal of Finance 54*(5), 1869–1884.

Hicks, J. R. (1939). *Value and Capital*. Oxford: Clarendon Press.

Ho, T. S. Y. (1992). Key rate durations: A measure of interest rate risks. *Journal of Fixed Income 2*(2), 29–44.

Ho, T. S. Y. and S.-B. Lee (1986). Term structure movements and pricing interest rate contingent claims. *Journal of Finance 41*(5), 1011–1029.

Hogan, M. (1993). Problems in certain two-factor term structure models. *Annals of Applied Probability 3*(2), 576–581.

Hogan, M. and K. Weintraub (1993). The log-normal interest rate model and Eurodollar futures. Working paper, Citibank, New York.

Honoré, P. (1998, January). Maturity induced bias in estimating spot-rate diffusion models. Working paper, Department of Finance, Aarhus School of Business.

Huang, C.-F. and R. H. Litzenberger (1988). *Foundations for Financial Economics*. Prentice-Hall.

Huang, J.-Z. and M. Huang (2003, May). How much of corporate-treasury yield spread is due to credit risk? A new calibration approach. Working paper. Available at SSRN: http://ssrn.com/abstract=307360.

Hubalek, F. and W. Schachermayer (1998). When does convergence of asset price processes imply convergence of option prices? *Mathematical Finance 8*(4), 385–403.

Hübner, G. (2001). The analytic pricing of asymmetric defaultable swaps. *Journal of Banking and Finance 25*(2), 295–316.

Huge, B. and D. Lando (1999). Swap pricing with two-sided default risk in a rating-based model. *European Finance Review 3*(3), 239–268.

Hull, J. and A. White (1990a). Pricing interest-rate-derivative securities. *Review of Financial Studies 3*(4), 573–592.

Hull, J. and A. White (1990b). Valuing derivative securities using the explicit finite difference method. *Journal of Financial and Quantitative Analysis 25*(1), 87–100.

Hull, J. and A. White (1993). One-factor interest-rate models and the valuation of interest-rate derivative securities. *Journal of Financial and Quantitative Analysis 28*(2), 235–254.

Hull, J. and A. White (1994a). Numerical procedures for implementing term structure models I: Single-factor models. *Journal of Derivatives 2*(1), 7–16.

Hull, J. and A. White (1994b). Numerical procedures for implementing term structure models II: Two-factor models. *Journal of Derivatives 2*(2), 37–48.

Hull, J. and A. White (1995). "A note on the models of Hull and White for pricing options on the term structure": Response. *Journal of Fixed Income 5*(2), 97–102.

Hull, J. and A. White (1996). Using Hull-White interest rate trees. *Journal of Derivatives 3*(3), 26–36.

Hull, J. and A. White (2003). The valuation of credit default swap options. *Journal of Derivatives 10*(5), 40–50.

Hull, J. and A. White (2004). Valuation of a CDO and an n-th to default CDS without Monte Carlo simulation. *Journal of Derivatives 12*(2), 8–23.

Hull, J. and A. White (2006). Valuing credit derivatives using an implied copula approach. *Journal of Derivatives 14*(2), 8–28.

Hull, J., M. Predescu, and A. White (2006, November). The valuation of correlation-dependent credit derivatives using a structural model. Working paper, University of Toronto and Oxford University.

Hull, J. C. (2009). *Options, Futures, and Other Derivatives* (7th edn). Prentice-Hall, Inc.

Ingersoll, Jr., J. E. (1987). *Theory of Financial Decision Making*. Lanham, MD: Rowman & Littlefield.

Ingersoll, Jr., J. E., J. Skelton, and R. Weil (1978). Duration forty years later. *Journal of Financial and Quantitative Analysis 13*(4), 627–650.

Inui, K. and M. Kijima (1998). A Markovian framework in multi-factor Heath-Jarrow-Morton models. *Journal of Financial and Quantitative Analysis 33*(3), 423–440.

James, J. and N. Webber (2000). *Interest Rate Modelling*. Wiley.

Jamshidian, F. (1987). Pricing of contingent claims in the one factor term structure model. Working paper, Merrill Lynch Capital Markets.

Jamshidian, F. (1989). An exact bond option formula. *Journal of Finance 44*(1), 205–209.

Jamshidian, F. (1991, June). Forward induction and construction of yield curve diffusion models. *Journal of Fixed Income 1*(1), 62–74.

Jamshidian, F. (1996). Bond, futures and option valuation in the quadratic interest rate model. *Applied Mathematical Finance 3*, 93–115.

Jamshidian, F. (1997). LIBOR and swap market models and measures. *Finance and Stochastics 1*(4), 293–330.

Jarrow, R. A. (2004). Risky coupon bonds as a portfolio of zero-coupon bonds. *Finance Research Letters 1*(2), 10–15.

Jarrow, R. A. and S. M. Turnbull (1995). Pricing derivatives on financial securities subject to credit risk. *Journal of Finance 50*(1), 53–85.

Jarrow, R. A. and Y. Yildirim (2002). Valuing default swaps under market and credit risk correlation. *Journal of Fixed Income 11*(4), 7–19.

Jarrow, R. A. and F. Yu (2001). Counterparty risk and the pricing of defaultable securities. *Journal of Finance 56*(5), 1765–1799.

Jarrow, R. A., D. Lando, and S. M. Turnbull (1997). A Markov model for the term structure of credit risk spreads. *Review of Financial Studies 10*(2), 481–523.

Jarrow, R. A., D. Lando, and F. Yu (2005). Default risk and diversification: Theory and empirical implications. *Mathematical Finance 15*(1), 1–26.

Jarrow, R. A., H. Li, and F. Zhao (2007). Interest rate caps "smile" too! But can the LIBOR market models capture the smile? *Journal of Finance 62*(1), 345–382.

Jarrow, R. A. and P. Protter (2004). Structural versus reduced form models: A new information based perspective. *Journal of Investment Management 2*(2), 1–10.

Jarrow, R. A., H. Li, S. Liu, and C. Wu (2010). Reduced-form valuation of callable corporate bonds: Theory and evidence. *Journal of Financial Economics 95*(2), 227–248.

Jaschke, S. R. (1998). Arbitrage bounds for the term structure of interest rates. *Finance and Stochastics 2*(1), 29–40.

Jeffrey, A. (1995). Single factor Heath-Jarrow-Morton term structure models based on Markov spot interest rate dynamics. *Journal of Financial and Quantitative Analysis 30*(4), 619–642.

Jensen, M. C. and W. H. Meckling (1976). Theory of the firm: Managerial behavior, agency costs and ownership structure. *Journal of Financial Economics 3*(4), 305–360.

Joint Center for Housing Studies of Harvard University (2008). *The State of the Nation's Housing 2008*. Joint Center for Housing Studies of Harvard University. Available at http://www.jchs.harvard.edu /publications/publications_by-year. htm.

Jones, E. P., S. P. Mason, and E. Rosenfeld (1984). Contingent claims analysis of corporate capital strucutres: An empirical investigation. *Journal of Finance 39*(3), 611–627.

Joshi, M. and R. Rebonato (2003). A displaced-diffusion stochastic volatility LIBOR market model: Motivation, definition and implementation. *Quantitative Finance 3*(6), 458–469.

Kan, R. (1992, June). Shape of the yield curve under CIR single factor model: A note. Working paper, University of Chicago.

Karatzas, I. and S. E. Shreve (1988). *Brownian Motion and Stochastic Calculus*, Volume 113 of *Graduate Texts in Mathematics*. New York: Springer-Verlag.

Karatzas, I. and S. E. Shreve (1998). *Methods of Mathematical Finance*, Volume 39 of *Applications of Mathematics*. New York: Springer-Verlag.

Karlin, S. and H. M. Taylor (1981). *A Second Course in Stochastic Processes*. Academic Press, Inc.

Kau, J. B., D. C. Keenan, W. J. Muller, III, and J. F. Epperson (1992). A generalized valuation model for fixed-rate residential mortgages. *Journal of Money, Credit, and Banking 24*(3), 279–299.

Kau, J. B., D. C. Keenan, W. J. Muller, III, and J. F. Epperson (1995). The value at origination of fixed-rate mortgages with default and prepayment. *Journal of Real Estate Finance and Economics 11*(1), 5–36.

Kim, J., K. Ramaswamy, and S. Sundaresan (1993). Does default risk in coupons affect the valuation of corporate bonds? *Financial Management 22*(3), 117–131.

Kincaid, D. and W. Cheney (2009). *Numerical Analysis: Mathematics of Scientific Computing* (3rd revised edn). American Mathematical Society.

Knez, P. J., R. Litterman, and J. Scheinkman (1994). Explorations into factors explaining money market returns. *Journal of Finance 49*(5), 1861–1882.

Kraft, H. and C. Munk (2007). Bond durations: Corporates vs. Treasuries. *Journal of Banking and Finance 31*(12), 3720–3741.

Kraft, H. and M. Steffensen (2007). Bankruptcy, counterparty risk, and contagion. *Review of Finance 11*(2), 209–252.

Krugman, P. R. (1991). Target zones and exchange rate dynamics. *The Quarterly Journal of Economics 106*(3), 669–682.

Lando, D. (1994). *Three Essays on Contingent Claims Pricing*. Ph. D. thesis, Cornell University.

Lando, D. (1998). On Cox processes and credit risky securities. *Review of Derivatives Research 2*(2-3), 99–120.

Lando, D. (2004). *Credit Risk Modeling*. Princeton University Press.

Lando, D. and A. Mortensen (2005). On the pricing of step-up bonds in the European telecom sector. *Journal of Credit Risk 1*(1), 71–110.

Langetieg, T. C. (1980). A multivariate model of the term structure. *Journal of Finance 35*(1), 71–97.

Leippold, M. and Z. Wiener (2004). Efficient calibration of trinomial trees for one-factor short rate models. *Review of Derivatives Research 7*(3), 213–239.

Leippold, M. and L. Wu (2002). Asset pricing under the quadratic class. *Journal of Financial and Quantitative Analysis 37*(2), 271–295.

Leland, H. (1994). Corporate debt value, bond covenants, and optimal capital structure. *Journal of Finance 49*(4), 1213–1252.

Leland, H. E. and K. B. Toft (1996). Optimal capital structure, endogenous bankruptcy, and the term structure of credit spreads. *Journal of Finance 51*(3), 987–1019.

LeRoy, S. F. (1996). Mortgage valuation under optimal prepayment. *Review of Financial Studies 9*(3), 817–844.

LeRoy, S. F. and J. Werner (2001). *Principles of Financial Economics*. Cambridge University Press.

Lesne, J.-P., J.-L. Prigent, and O. Scaillet (2000). Convergence of discrete time option pricing models under stochastic interest rates. *Finance and Stochastics 4*(1), 81–93.

Li, A., P. Ritchken, and L. Sankarasubramanian (1995). Lattice models for pricing American interest rate claims. *Journal of Finance 50*(2), 719–737.

Li, D. X. (2000). On default correlation: A copula approach. *Journal of Fixed Income 9*(4), 43–54.

Li, H. and F. Zhao (2006). Unspanned stochastic volatility: Evidence from hedging interest rate derivatives. *Journal of Finance 61*(1), 341–378.

Liao, S.-L., M.-S. Tsai, and S.-L. Chiang (2008). Closed-form mortgage valuation using reduced-form model. *Real Estate Economics 36*(2), 313–347.

Linton, O., E. Mammen, J. Nielsen, and C. Tanggaard (2001). Yield curve estimation by kernel smoothing methods. *Journal of Econometrics 105*(1), 185–223.

Litterman, R. and J. Scheinkman (1991, June). Common factors affecting bond returns. *Journal of Fixed Income 1*(1), 54–61.

Litzenberger, R. H. and R. Rolfo (1984). An international study of tax effects on government bonds. *Journal of Finance 39*(1), 1–22.

Longstaff, F. A. (1989). A nonlinear general equilibrium model of the term structure of interest rates. *Journal of Financial Economics 23*(2), 195–224.

Longstaff, F. A. (1993). The valuation of options on coupon bonds. *Journal of Banking and Finance 17*(1), 27–42.

Longstaff, F. A. (2005). Borrower credit and the valuation of mortgage-backed securities. *Real Estate Economics 33*(4), 619–661.

Longstaff, F. A. and E. S. Schwartz (1992a). Interest rate volatility and the term structure: A two-factor general equilibrium model. *Journal of Finance 47*(4), 1259–1282.

Longstaff, F. A. and E. S. Schwartz (1992b). A two-factor interest rate model and contingent claims valuation. *Journal of Fixed Income 2*(3), 16–23.

Longstaff, F. A. and E. S. Schwartz (1993a). Implementation of the Longstaff-Schwartz interest rate model. *Journal of Fixed Income 3*(2), 7–14.

Longstaff, F. A. and E. S. Schwartz (1993b). Interest rate volatility and bond prices. *Financial Analysts Journal 49*(4), 70–74.

Longstaff, F. A. and E. S. Schwartz (1994). Comments on "a note on parameter estimation in the two-factor Longstaff and Schwartz interest rate model". *Journal of Fixed Income 3*(4), 101–102.

Longstaff, F. A. and E. S. Schwartz (1995a). A simple approach to valuing risky fixed and floating rate debt. *Journal of Finance 50*(3), 789–819.

Longstaff, F. A. and E. S. Schwartz (1995b). Valuing credit derivatives. *Journal of Fixed Income 5*(1), 6–12.

Longstaff, F. A. and E. S. Schwartz (2001). Valuing American options by simulation: A simple least-squares approach. *Review of Financial Studies* 14(1), 113–147.

Longstaff, F. A., S. Mithal, and E. Neis (2005). Corporate yield spreads: Default risk or liquidity? New evidence from the credit-default swap market. *Journal of Finance 60*(5), 2213–2253.

Lutz, F. (1940). The structure of interest rates. *Quarterly Journal of Economics 55*(1), 36–63.

Macaulay, F. R. (1938). *Some Theoretical Problems Suggested by the Movements of Interest Rates, Bond Yields, and Stock Prices in the United States since 1856*. New York: Columbia University Press.

Madan, D. B. and H. Unal (1998). Pricing the risks of default. *Review of Derivatives Research 2*(2–3), 121–160.

Madan, D. B. and H. Unal (2000). A two-factor hazard-rate model for pricing risky debt and the term structure of credit spreads. *Journal of Financial and Quantitative Analysis 35*(1), 43–65.

Marsh, T. A. and E. R. Rosenfeld (1983). Stochastic processes for interest rates and equilibrium bond prices. *Journal of Finance 38*(2), 635–646.

Marshall, D. (1992). Inflation and asset returns in a monetary economy. *Journal of Finance 47*(4), 1315–1342.

McConnell, J. J. and M. Singh (1994). Rational prepayments and the valuation of collateralized mortgage obligations. *Journal of Finance 49*(3), 891–921.

McCulloch, J. H. (1971). Measuring the term structure of interest rates. *Journal of Business 44*(1), 19–31.

McCulloch, J. H. (1975). The tax-adjusted yield curve. *Journal of Finance 30*(3), 811–830.

McCulloch, J. H. (1993). A reexamination of traditional hypotheses about the term structure of interest rates: A comment. *Journal of Finance 48*(2), 779–789.

Mehra, R. and E. C. Prescott (1985). The equity premium: A puzzle. *Journal of Monetary Economics 15*(2), 145–162.

Mella-Barral, P. and W. Perraudin (1997). Strategic debt service. *Journal of Finance 52*(2), 531–556.

Mengle, D. (2007). Credit derivatives: An overview. Economic Review Fourth Quarter, Federal Reserve Bank of Atlanta.

Mercurio, F. and J. M. Moraleda (2000). An analytically tractable interest rate model with humped volatility. *European Journal of Operational Research 120*(1), 205–214.

Merton, R. C. (1970). A dynamic general equilibrium model of the asset market and its application to the pricing of the capital structure of the firm. Working paper 497–70, Sloan School of Management, MIT. Reprinted as Chapter 11 in Merton (1992).

Merton, R. C. (1973). Theory of rational option pricing. *Bell Journal of Economics and Management Science 4*(1), 141–183. Reprinted as Chapter 8 in Merton (1992).

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *Journal of Finance 29*(2), 449–470. Reprinted as Chapter 12 in Merton (1992).

Merton, R. C. (1976). Option pricing when underlying stock returns are discontinuous. *Journal of Financial Economics 3*(1–2), 125–144. Reprinted as Chapter 9 in Merton (1992).

Merton, R. C. (1992). *Continuous-Time Finance*. Padston, UK: Basil Blackwell Inc.

Miltersen, K. R. (1994). An arbitrage theory of the term structure of interest rates. *Annals of Applied Probability 4*(4), 953–967.

Miltersen, K. R. (1998). Pricing of interest rate contingent claims: Implementing a simulation approach. *Journal of Computational Finance 1*(3), 37–62.

Miltersen, K. R. and E. S. Schwartz (1998). Pricing of options on commodity futures with stochastic term structures of convenience yields and interest rates. *Journal of Financial and Quantitative Analysis 33*(1), 33–59.

Miltersen, K. R., K. Sandmann, and D. Sondermann (1997). Closed form solutions for term structure derivatives with log-normal interest rates. *Journal of Finance 52*(1), 409–430.

Modiagliani, F. and M. H. Miller (1958). The cost of capital, corporation finance, and the theory of investment. *American Economic Review 48*(3), 261–297. Errata: Modiagliani and Miller (1963).

Modiagliani, F. and M. H. Miller (1963). Corporate income taxes and the cost of capital: A correction. *American Economic Review 53*(3), 433–443.

Modigliani, F. and R. Sutch (1966). Innovations in interest rate policy. *American Economic Review 56*(1-2), 178–197.

Moreno, M. and J. F. Navas (2003). On the robustness of least-squares Monte Carlo for pricing American derivatives. *Review of Derivatives Research 6*(2), 107–128.

Mortensen, A. (2006). Semi-analytical valuation of basket credit derivatives in intensity-based models. *Journal of Derivatives 13*(4), 8–26.

Morton, A. J. (1988). Arbitrage and martingales. Technical Report 821, Cornell University, New York.

Munk, C. (1999). Stochastic duration and fast coupon bond option pricing in multi-factor models. *Review of Derivatives Research 3*(2), 157–181.

Munk, C. (2002). Price bounds on bond options, swaptions, caps, and floors assuming only nonnegative interest rates. *International Review of Economics and Finance 11*(4), 335–347.

Munk, C. (2010). *Financial asset pricing theory*. Lecture notes, Aarhus University. To be published by Oxford University Press.

Musiela, M. and M. Rutkowski (1997). *Martingale Methods in Financial Modelling*, Volume 36 of *Applications of Mathematics*. Springer-Verlag.

Negishi, T. (1960). Welfare economics and existence of an equilibrium for a competetive economy. *Metroeconometrica 12*(2-3), 92–97.

Nelsen, R. (1999). *An Introduction to Copulas*. New York: Springer.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious modeling of yield curves. *Journal of Business 60*(4), 473–489.

Nelson, D. B. and K. Ramaswamy (1990). Simple binomial processes as diffusion approximations in financial models. *Review of Financial Studies 3*(3), 393–430.

Nielsen, L. T., J. Saa-Requejo, and P. Santa-Clara (1993). Default risk and interest rate risk: The term structure of default spreads. Working paper, INSEAD.

Ogden, J. (1987). An analysis of yield curve notes. *Journal of Finance 42*(1), 99–110.

Øksendal, B. (2003). *Stochastic Differential Equations* (6th edn). Springer-Verlag.

Papageorgiou, N. and F. S. Skinner (2006). Credit spreads and the zero-coupon treasury spot curve. *Journal of Financial Research 29*(3), 421–439.

Parwani, K., K. Emery, and R. Cantor (2009, May). European corporate default and recovery rates, 1985-2008. Special comment, Moody's Investors Service.

Pearson, N. D. and T.-S. Sun (1991, June). An empirical examination of the Cox, Ingersoll, and Ross model of the term structure of interest rates. Technical report, Graduate School of Business, Columbia University, New York, NY 10027, USA.

Pearson, N. D. and A. Zhou (1999, October). A nonparametric analysis of the forward rate volatilities. Working paper, University of Illinois and SUNY Binghamton.

Pedersen, H. W., E. S. W. Shiu, and A. E. Thorlacius (1989). Arbitrage-free pricing of interest-rate contingent claims. *Transactions of Society of Actuaries 41*, 231–279.

Pennacchi, G. (2008). *Theory of Asset Pricing*. Pearson Education.

Phoa, W. and M. Shearer (1997). A note on arbitrary yield curve reshaping sensitivities using key rate durations. *Journal of Fixed Income 7*(3), 67–71.

Piazzesi, M. (2005). Bond yields and the federal reserve. *Journal of Political Economy 113*(2), 311–344.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (2007). *Numerical Recipes: The Art of Scientific Computing* (Third ed.). Cambridge University Press.

Rebonato, R. (1996). *Interest-Rate Option Models*. John Wiley & Sons, Inc.

Rendleman, R. and B. Bartter (1979). Two state option pricing. *Journal of Finance 34*(5), 1093–1110.

Rendleman, R. and B. Bartter (1980). The pricing of options on debt securities. *Journal of Financial and Quantitative Analysis 15*(1), 11–24.

Riedel, F. (2000). Decreasing yield curves in a model with an unknown constant growth rate. *European Finance Review 4*(1), 51–67.

Riedel, F. (2004). Heterogeneous time preferences and humps in the yield curve–the preferred habitat theory revisited. *European Journal of Finance 10*(1), 3–22.

Ritchken, P. and L. Sankarasubramanian (1995). Volatility structures of forward rates and the dynamics of the term structure. *Mathematical Finance 5*(1), 55–72.

Rogers, L. C. G. (1995). Which model for term-structure of interest rates should one use? In M. H. A. Davis, D. Duffie, W. H. Fleming, and S. E. Shreve (Eds), *Mathematical Finance*, Volume 65 of *The IMA Volumes in Mathematics and Its Applications*, pp. 93–115. Springer-Verlag.

Ross, S. A. (1978). A simple approach to the valuation of risky streams. *Journal of Business 51*(3), 453–475.

Rubinstein, M. and E. Reiner (1991). Breaking down the barriers. *Risk 4*(8), 28–35.

Saa-Requejo, J. and P. Santa-Clara (1999). Bond pricing with default risk. Working paper, UCLA.

Sanders, A. (2008). The subprime crisis and its role in the financial crisis. *Journal of Housing Economics 17*(4), 254–261.

Sandmann, K. and D. Sondermann (1997). A note on the stability of lognormal interest rate models and the pricing of Eurodollar futures. *Mathematical Finance 7*(2), 119–125.

Sankaran, M. (1963). Approximations to the non-central χ^2 distribution. *Biometrika 50* (1–2), 199–204.

Sarig, O. and A. Warga (1989). Some empirical estimates of the risk structure of interest rates. *Journal of Finance 44*(5), 1351–1360.

Schrager, D. F. and A. A. J. Pelsser (2006). Pricing swaptions in affine term structure models. *Mathematical Finance 16*(4), 673–694.

Schwartz, E. S. (1977). The valuation of warrants: Implementing a new approach. *Journal of Financial Economics 4*(1), 79–93.

Schwartz, E. S. and W. N. Torous (1989). Prepayment and the valuation of mortgage-backed securities. *Journal of Finance 44*(2), 375–392.

Schwartz, E. S. and W. N. Torous (1992). Prepayment, default, and the valuation of mortgage pass-through securities. *Journal of Business 65*(2), 221–239.

Seydel, R. U. (2009). *Tools for Computational Finance* (4th edn). Springer Verlag.

Sharp, N. J., P. V. Johnson, D. P. Newton, and P. W. Duck (2009). A new prepayment model (with default): An occupation-time derivative approach. *Journal of Real Estate Finance and Economics 39*(2), 118–145.

Shea, G. S. (1984). Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations. *Journal of Financial and Quantitative Analysis 19*(3), 253–269.

Shea, G. S. (1985). Interest rate term structure estimation with exponential splines: A note. *Journal of Finance 40*(1), 319–325.

Shiller, R. J. (2008). *The Subprime Solution*. Princeton, NJ: Princeton University Press.

Shimko, D. C., M. Tejima, and D. R. van Deventer (1993). The pricing of risky debt when interest rates are stochastic. *Journal of Fixed Income 3*(2), 58–65.

Singleton, K. J. and L. Umantsev (2002). Pricing coupon-bond options and swaptions in affine term structure models. *Mathematical Finance 12*(4), 427–446.

Skiadas, C. (2009). *Asset Pricing Theory*. Princeton University Press.

Stambaugh, R. F. (1988). The information in forward rates: Implications for models of the term structure. *Journal of Financial Economics 21*(1), 41–70.

Stanton, R. (1995). Rational prepayment and the valuation of mortgage-backed securities. *Review of Financial Studies 8*(3), 677–708.

Stanton, R. and N. Wallace (1998). Mortgage choice: What's the point? *Real Estate Economics 26*(2), 173–205.

Stentoft, L. (2004a). Assessing the least squares Monte-Carlo approach to American option valuation. *Review of Derivatives Research 7*(2), 129–168.

Stentoft, L. (2004b). Convergence of the least squares Monte-Carlo approach to American option valuation. *Management Science 50*(9), 1193–1203.

Svensson, L. E. O. (1995). Estimating forward interest rates with the extended Nelson-Siegel method. *Sveriges Riksbank Quarterly Review 3*, 13–26.

Swan, P. L. (2009). The political economy of the subprime crisis: Why subprime was so attractive to its creators. *European Journal of Political Economy 25*(1), 124–132.

Tanggaard, C. (1997). Nonparametric smoothing of yield curves. *Review of Quantitative Finance and Accounting 9*(3), 251–267.

Tangman, D. Y., A. Gopaul, and M. Bhuruth (2008). A fast high-order finite difference algorithm for pricing American options. *Journal of Computational and Applied Mathematics 222*(1), 17–29.

Tavella, D. (2002). *Quantitative Methods in Derivatives Pricing*. Wiley.

Tavella, D. and C. Randall (2000). *Pricing Financial Instruments: The Finite Difference Method*. Wiley.

Telmer, C. I. (1993). Asset-pricing puzzles and incomplete markets. *Journal of Finance 48*(5), 1803–1832.

Thomas, J. W. (1995). *Numerical Partial Differential Equations*, Volume 22 of *Texts in Applied Mathematics*. Springer-Verlag.

Tian, Y. (1993). A simplified approach to the pricing of interest-rate contingent claims. *Journal of Financial Engineering 1*(1), 285–314.

Tilley, J. A. (1993). Valuing American options in a path simulation model. *Transactions of the Society of Actuaries 45*, 83–104.

Tomz, M. and M. L. J. Wright (2007). Do countries default in "bad times"? *Journal of the European Economic Association 5*(2–3), 352–360.

Trolle, A. B. and E. S. Schwartz (2009). A general stochastic volatility model for the pricing of interest rate derivatives. *Review of Financial Studies 22*(5), 2007–2057.

Tsitsiklis, J. and B. Van Roy (2001). Regression methods for pricing complex American-style options. *IEEE Transactions on Neural Networks 12*(4), 694–703.

Van Hemert, O. (2009, February). Household interest rate risk management. Unpublished working paper, NYU Stern. Available at SSRN: http://ssrn.com/abstract=891034.

Vasicek, O. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics 5*(2), 177–188.

Vetzal, K. R. (1997). Stochastic volatility, movements in the short term interest rates, and bond option values. *Journal of Banking and Finance 21*(2), 169–196.

Wachter, J. A. (2006). A consumption-based model of the term structure of interest rates. *Journal of Financial Economics 79*(2), 365–399.

Wang, J. (1996). The term structure of interest rates in a pure exchange economy with heterogeneous investors. *Journal of Financial Economics 41*(1), 75–110.

Wei, J. Z. (1997). A simple approach to bond option pricing. *Journal of Futures Markets 17*(2), 131–160.

Weil, P. (1989). The equity premium puzzle and the risk-free rate puzzle. *Journal of Monetary Economics 24*(3), 401–421.

Wilcox, D. W. (1992). The construction of U.S. consumption data: Some facts and their implications for empirical work. *American Economic Review 82*(4), 922–941.

Willner, R. (1996). A new tool for portfolio managers: Level, slope, and curvature durations. *Journal of Fixed Income 6*(1), 48–59.

Wilmott, P. (1998). *Derivatives: The Theory and Practice of Financial Engineering*. Oxford University Press.

Wilmott, P., J. Dewynne, and S. Howison (1993). *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press.

Wu, S. (2008, August). Long-run consumption risk and the real yield curve. Available at SSRN: http://ssrn.com/abstract=1130087.

Yu, F. (2002). Modeling expected return on defaultable bonds. *Journal of Fixed Income 12*(2), 69–81.

Zhou, C. (2001a). An analysis of default correlations and multiple defaults. *Review of Financial Studies 14*(2), 555–576.

Zhou, C. (2001b). The term structure of credit spreads with jump risk. *Journal of Banking and Finance 25*(11), 2015–2040.

This page intentionally left blank

Index