Fixed Income Analysis:
Securities, Pricing, and Risk Management

Claus Munk*

This version: March 15, 2004

*Department of Accounting and Finance, University of Southern Denmark, Campusvej 55, DK-5230 Odense M, Denmark. Phone: ++45 6550 3257. Fax: ++45 6593 0726. E-mail: cmu@sam.sdu.dk. Internet homepage: http://www.sam.sdu.dk/~cmu

Contents

Preface

This book provides an introduction to the markets for fixed-income securities and the models and methods that are used to analyze such securities. The class of fixed-income securities covers 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, several related securities with payments that are tied to the development in some particular index, interest rate, or asset price are typically also termed fixed-income securities. In the broadest sense of the term, the many different interest rate and bond derivatives are also considered fixed-income products. Maybe a more descriptive term for this broad class of securities is "interest rate securities", since the values of these financial contracts are derived from current interest rates and expectations and uncertainty about future interest rates. The key concept in the analysis of fixed-income securities is the term structure of interest rates, which is loosely defined as the dependence between interest rates and maturities.

The outline of this book is as follows. The first two chapters deal with the most common fixed-income securities, fix much of the notation and terminology, and discuss basic relations between key concepts.

The main part of the book discusses models of the evolution of the term structure of interest rates over time. Chapter 3 introduces much of the mathematics needed for developing and analyzing modern dynamic models of interest rates. Chapter 4 reviews some of the important general results on asset pricing. In particular, we define and relate the key concepts of arbitrage, state-price deflators, and risk-neutral probability measures. The connection to market completeness and individual investors' behavior is also addressed, just as the implications of the general asset pricing theory for the modeling of the term structure are discussed. Chapter 5 applies the general asset pricing tools to explore the economics of the term structure of interest rates. For example we discuss the relation between the term structure of interest rates and macro-economic variables such as aggregate consumption, production, and inflation. We will also review some of the traditional hypotheses on the shape of the yield curve, e.g. the expectation hypotheses. Chapter 6 further develops the consequences of the general asset pricing theory for the modeling of the term structure of interest rates and the pricing of derivatives.

Chapters 7 to 12 develop models for the pricing of fixed income securities and the management of interest rate risk. Chapter 7 goes through so-called one-factor models. This type of models 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. Chapter 8 explores multi-factor models which have several advantages over one-factor models, but are also more complicated to analyze and apply. In Chapter 9 we discuss how one-and multi-factor models can be extended to be consistent with current market information, such as bond prices and volatilities. Chapter 10 introduces and analyzes so-called Heath-Jarrow-Morton models, which are characterized by taking the current market term structure of interest rates as given and then modeling 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. Chapter 12 discusses how the different interest rate models can be applied for interest rate risk management.

The subject of Chapter 13 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. In Chapter 14 (only some references are listed in the current version) we discuss the pricing of corporate bonds and other fixed-income securities where the default risk of the issuer cannot be ignored. 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 (only some references are listed in the current version) describes several numerical techniques that can be applied in cases where explicit pricing and hedging formulas are not available.

Style...

Prerequisites...

There are several other books that cover much of the same material or focus on particular elements discussed in this book. Books emphasizing descriptions of markets and products: Fabozzi (2000), van Horne (2001). Books emphasizing modern interest rate modeling: James and Webber (2000), Pelsser (2000), Rebonato (1996),...

I appreciate comments and corrections from Rasmus H. Andersen, Lennart Damgaard, Hans Frimor, Mette Hansen, Stig Secher Hesselberg, Frank Emil Jensen, Kasper Larsen, Morten Mosegaard, Per Plotnikoff, and other people. I also appreciate the excellent secretarial assistance of Lene Holbæk.

Chapter 1

Basic interest rate markets, concepts, and relations

1.1 Introduction

The value of an asset equals the value of its future cash flow. Traditionally, the value of a bond is computed by discounting all its future payments using the same interest rate, namely the yield of the bond. If we have two bonds with the same payment dates but different yields, we will therefore discount the payments from the bonds with different interest rates. This is clearly illogical. The present value of a given payment at a given future point in time is independent of which asset the payment stems from. All sure payments at the same date should be discounted with the same rate. On the other hand, there is no reason to discount payments at different dates with the same discount rate. The interest rate on a loan will normally depend on the maturity of the loan, and on the bond markets there will often be differences between the yields on short-term bonds and long-term bonds. **The term structure of interest rates** is the relation between interest rates and their maturity.

In this chapter we first provide a brief introduction to the most basic markets for borrowing and lending and discuss different types of interest rates and discount factors as well as the relation between them. Finally, we will discuss how information about the term structure can be extracted from market data.

1.2 Markets for bonds and interest rates

This section gives a brief description of markets for bonds and other debt contracts. A bond is simply a tradable loan agreement. Most bonds that are issued and subsequently traded at either organized exchanges or over-the-counter (OTC) represent medium- or long-term loans with maturities in excess of one year and often between 10 and 30 years. Some short-term bonds are also issued and traded, but much of the short-term borrowing activity takes place in the so-called money market, where large financial institutions (including the central bank) and large corporations form several types of loan agreements with maturities ranging from a few hours up to around one year. These agreements are usually not traded after the original contracting. The interest rates set in the money market directly affect the interest rates that banks offer and charge their commercial and household customers. Small investors may participate in the money markets through money market funds. Below, we introduce the most important types of bonds that are traded. More details on bond markets can be found in e.g. Fabozzi (2000). We also look at some of the debt contracts available in the money market. In Chapter 2 we will discuss many other interest rate related securities, such as futures and options on bonds and interest rates.

The issuer of the bond (the debtor or borrower) issues a contract in which he is obligated to pay certain payments at certain future points in time. Typically, a bond issue consists of a series of identical bonds. The simplest possible bond is a **zero-coupon bond**, which is a bond promising a single payment at a given future date, the maturity date of the bond, and no other payments. Bonds which promise more than one payment when issued are referred to as **coupon bonds**.

Typically, the payments of coupon bonds follow a regular schedule so that the payments occur at regular intervals (quarterly, semi-annually, or annually) and 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. 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. 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. 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. Finally, few bonds are *perpetuities* or *consols* that last forever and only pay interest. 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, e.g. to the 6-month interest rate for a floating rate bond with semi-annual payments.

Bond markets can be divided into the national markets of different countries and the international market (also known as the Eurobond market). The largest national bond markets are those of the U.S., Japan, Germany, Italy, and France followed by other Western European countries and Australia.

In the national market of a country, primarily bonds issued by domestic issuers are traded, but often some bonds issued by certain foreign governments or corporations or international associations are also traded. The bonds issued in a given national market must comply with the regulation of that particular country. Bonds issued in the less regulated Eurobond market are usually underwritten by an international syndicate and offered to investors in several countries simultaneously. Many Eurobonds are listed on one national exchange, often in Luxembourg or London, but most of the trading in these bonds takes place OTC (over-the-counter). Other Eurobonds are issued as a private placement with financial institutions. Most Eurobonds are issued in U.S. dollars ("Eurodollar bonds"), the common European currency Euro, Pound Sterling, Swiss francs, or Japanese yen.

Divided according to the type of issuer, national bond markets have two or three major categories: government(-related) bonds, corporate bonds, and – in some countries – mortgage-backed bonds. In addition, some bonds issued by international institutions or foreign governments are often traded. Eurobonds are typically issued by international institutions, governments, or large corporations.

In most national bond markets, the major part of bond trading is in **government bonds**, which are simply 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 and politically unstable countries, the default risk on government bonds cannot be ignored. In the U.S., government bonds are issued by the Department of the Treasury and called Treasury securities. 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.[1] 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.[2] 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.[3]

In some countries including the U.S., bonds issued by various public institutions, e.g. utility companies, railway companies, export support funds, etc., 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 borrowing costs for e.g. farmers, homeowners, and students. However, these bonds are typically not backed

[1] 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.

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

[3] More information on Treasury securities can be found on the homepage of the Bureau of the Public Debt at the Department of the Treasury, see `www.publicdebt.treas.gov`.

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 some countries, corporations will traditionally borrow funds by issuing bonds, so-called **corporate bonds**. This is the case in the U.S., where there is a large market for such bonds. In other countries, e.g. Germany, corporations tend to borrow funds through bank loans, so that the market for corporate bonds is very 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, e.g. 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.

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.

Mortgage-backed bonds constitute a large part of some bond markets, e.g. in the U.S., Germany, Sweden, and Denmark. 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, etc.) or non-residential (corporations, farms, etc.). 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 have 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, e.g. pass-throughs issued on the basis of a pool of fixed-rate annuity mortgage loans have a payment schedule equal to that of annuity bond. 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 are 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"). Ginnie Mae pass-throughs are even guaranteed by the U.S. government, but the securities issued by the two other institutions are also considered virtually free of default risk.

Finally, let us take a brief look at the prevailing debt contracts in the **money market**. While we focus on the U.S. market, similar contracts exist in many other countries and many of the contracts are also made in the Euromarket. The debt contracts in the money market are mainly zero-coupon loans, which have a single repayment date. Financial institutions borrow large amounts over short periods from each other by issuing *certificates of deposit*, also known in the market as CDs. In the Euromarket deposits are negotiated for various terms and currencies, but most deposits are in U.S. dollars and for a period of one, three, or six months. Interest rates set on deposits at the London interbank market are called LIBOR rates (LIBOR is short for London Interbank offered rate).

To manage very short-term liquidity, financial institutions often agree on overnight loans, so-called *federal funds*. The interest rate charged on such loans is called the Fed funds rate. The Federal Reserve has a target Fed funds rate and buys and sells securities in open market operations to manage the liquidity in the market, thereby also affecting the Fed funds rate. 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.

Large corporations, both financial corporations and others, often borrow short-term 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, e.g. 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 their monetary policy. The interest rate on repos is called the repo rate.

1.3 Discount factors and zero-coupon bonds

We will assume throughout that the face value is equal to 1 (dollar) unless stated otherwise. 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 must be decreasing, i.e.

$$1 \geq B_t^T \geq B_t^S \geq 0, \quad T < S. \tag{1.1}$$

An example of an estimated market discount function is shown in Figure 1.1 on page 16.

Next, consider a coupon bond with payment dates $T_1, T_2, \dots, T_n$, where we assume without loss of generality that $T_1 < T_2 < \dots < T_n$. The payment at date T_i is denoted by Y_i. Such 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, etc. If all these zero-coupon bonds are traded in the market, the price of the coupon bond at any time t must be

$$B_t = \sum_{T_i > t} Y_i B_t^{T_i}, \tag{1.2}$$

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.

Example 1.1 Consider a bullet bond with a face value of 100, a coupon rate of 7%, annual payments, and exactly three 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 (1.2), 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, riskfree profits can be locked in by buying the bullet bond and selling 7 one-year, 7 two-year, and 107 three-year zero-coupon bonds. If the price of the bullet bond is higher than 105.97, sell the bullet bond and buy 7 one-year, 7 two-year, and 107 three-year zero-coupon bonds. □

If not all the relevant zero-coupon bonds are traded, we cannot justify the relation (1.2) as a result of the no-arbitrage principle. Still, it is a valuable relation. Suppose that an investor has determined (from private or macro economic 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 (1.2).

The market prices of all bonds reflect a market discount function, which is the result of the supply 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 Sections 1.5 and 1.6.

1.4 Zero-coupon rates and forward 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 five-year interest rate is 5%. Interest rates are always quoted on an annual basis, i.e. 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. Spot rates applies to a period beginning at the time the rate is set, whereas forward rates applies to a future period of time. The precise definitions follow below.

1.4.1 Annual compounding

Given the price B_t^T at time t on a zero-coupon bond maturing at time T, the relevant discount rate between time t and time T is the yield on the zero-coupon bond, the so-called **zero-coupon rate** or **spot rate** for date T. Let $\hat{y}_t^T$ denote this rate computed using annual compounding. We then have the following relationship:

$$B_t^T = (1 + \hat{y}_t^T)^{-(T-t)} \tag{1.3}$$

or

$$\hat{y}_t^T = \left(B_t^T\right)^{-1/(T-t)} - 1. \tag{1.4}$$

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. An example of a zero-coupon yield curve is shown in Figure 1.2 on page 17.

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 $t \leq T < S$. 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

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

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 (1.5) in terms of zero-coupon bond prices as

$$B_t^S = B_t^T \left(1 + \hat{f}_t^{T,S}\right)^{-(S-T)}, \tag{1.6}$$

so that the forward rate is given by

$$\hat{f}_t^{T,S} = \left(\frac{B_t^T}{B_t^S}\right)^{1/(S-T)} - 1. \tag{1.7}$$

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

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

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

1.4.2 Compounding over other discrete periods – LIBOR rates

In practice, many interest rates are quoted using semi-annually, quarterly, or monthly compounding. An interest rate or 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.

As discussed earlier, interest rates are set for loans with various maturities and currencies at the international money markets, the most commonly used being the LIBOR rates that are fixed in London. Traditionally, these rates are quoted using a compounding period equal to the maturity of the interest rate. If, for example, the three-month interest rate is $l_t^{t+0.25}$ per year, it means that the present value of one dollar paid three months from now is

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

Conversely, the three-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

$$B_t^T = \frac{1}{1 + l_t^T (T - t)} \tag{1.8}$$

and

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

Similarly, a six-month forward rate of $L_t^{T,T+0.5}$ valid for the period $[T, T + 0.5]$ means that

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

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,

$$L_t^{T,S} = \frac{1}{S - T}\left(\frac{B_t^T}{B_t^S} - 1\right). \tag{1.9}$$

Although such spot and forward rates are quoted on many different money markets, we shall use the term (spot/forward) LIBOR rate for all such money market interest rates computed with discrete compounding.

1.4.3 Continuous compounding

Increasing the compounding frequency m, the effective annual return of one 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

$$B_t^T = e^{-y_t^T(T-t)}, \tag{1.10}$$

so that

$$y_t^T = -\frac{1}{T-t} \ln B_t^T. \tag{1.11}$$

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 $T \mapsto \hat{y}_t^T$. We have the following relation between the continuously compounded and the annually compounded zero-coupon rates:

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

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

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

in analogy with (1.6). Consequently,

$$f_t^{T,S} = -\frac{\ln B_t^S - \ln B_t^T}{S-T}. \tag{1.12}$$

Using (1.10), we get the following relation between zero-coupon rates and forward rates under continuous compounding:

$$f_t^{T,S} = \frac{y_t^S(S-t) - y_t^T(T-t)}{S-T}. \tag{1.13}$$

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**. Letting $S \to T$ in the expression (1.12), we get

$$f_t^T = -\frac{\partial \ln B_t^T}{\partial T} = -\frac{\partial B_t^T / \partial T}{B_t^T}, \tag{1.14}$$

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

$$B_t^T = e^{-\int_t^T f_t^u \, du}. \tag{1.15}$$

Applying (1.13), the relation between the infinitesimal forward rate and the spot rates can be written as

$$f_t^T = \frac{\partial [y_t^T(T-t)]}{\partial T} = y_t^T + \frac{\partial y_t^T}{\partial T}(T-t) \tag{1.16}$$

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 (1.15) and (1.10) that

$$y_t^T = \frac{1}{T-t} \int_t^T f_t^u \, du, \tag{1.17}$$

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

1.4.4 Different ways to represent the term structure of interest rates

It is important to realize 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, etc. Academics frequently apply continuous compounding since the mathematics involved in many relevant computations is more elegant when exponentials are used.

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, cf. Section 2.9.

1.5 Determining the zero-coupon yield curve: 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.

Example 1.2 Consider a market where two bullet bonds are traded, a 10% bond expiring in one year and a 5% bond expiring in two years. Both have annual payments and a face value of 100. The one-year bond has the payment structure of a zero-coupon bond: 110 dollars in one 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 one dollar in a year. If the price of the one-year bullet bond is 100, the one-year discount factor is given by

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

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

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

cf. (1.2). Isolating B_t^{t+2}, we get

$$B_t^{t+2} = \frac{1}{105} B_{2,t} - \frac{5}{105} B_t^{t+1}. \tag{1.18}$$

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

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

From (1.18) we see that we can construct a two-year zero-coupon bond as a portfolio of 1/105 units of the two-year bullet bond and −5/105 units of the one-year zero-coupon bond. This is equivalent to a portfolio of 1/105 units of the two-year bullet bond and $-5/(105 \cdot 110)$ units of the one-year bullet bond. Given the discount factors, zero-coupon rates and forward rates can be calculated as shown in Section 1.4.1–1.4.3. □

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 date. 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, e.g. if the bond matures before time $t + M$. Let $B_{i,t}$ denote the price of bond i. From (1.2) 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

$$\begin{pmatrix} B_{1,t} \\ B_{2,t} \\ \vdots \\ B_{M,t} \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{12} & \ldots & Y_{1M} \\ Y_{21} & Y_{22} & \ldots & Y_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ Y_{M1} & Y_{M2} & \ldots & Y_{MM} \end{pmatrix} \begin{pmatrix} B_t^{t+1} \\ B_t^{t+2} \\ \vdots \\ B_t^{t+M} \end{pmatrix}. \tag{1.19}$$

The conditions in 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 enters the portfolio replicating the zero-coupon bond maturing at $t + j$.

Then we must have that

$$\begin{pmatrix} 0 \\ 0 \\ \vdots \\ 1 \\ \vdots \\ 0 \end{pmatrix} = \begin{pmatrix} Y_{11} & Y_{21} & \dots & \dots & \dots & Y_{M1} \\ Y_{12} & Y_{22} & \dots & \dots & \dots & Y_{M2} \\ \vdots & \vdots & \ddots & & & \vdots \\ Y_{1j} & Y_{2j} & \dots & \dots & \dots & Y_{Mj} \\ \vdots & \vdots & & & \ddots & \vdots \\ Y_{1M} & Y_{2M} & \dots & \dots & \dots & Y_{MM} \end{pmatrix} \begin{pmatrix} x_1(j) \\ x_2(j) \\ \vdots \\ x_j(j) \\ \vdots \\ x_M(j) \end{pmatrix}, \tag{1.20}$$

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 (1.19) and the solution $(x_1(j), \dots, x_M(j))$ to (1.20):[4]

$$\sum_{i=1}^{M} x_i(j) B_{i,t} = B_t^{t+j}. \tag{1.21}$$

Thus, first the zero-coupon bonds can be constructed, i.e. (1.20) is solved for each $j = 1, \dots, M$, and next (1.21) can be applied to compute the discount factors.

Example 1.3 In Example 1.2 we considered a two-year 5% bullet bond. Assume now that a two-year 8% serial bond with the same payment dates is traded. The payments from this bond are 58 dollars in one year and 54 dollars in two 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 (1.19) 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 1.4 In the Examples 1.2 and 1.3 we have considered three bonds: a one-year bullet bond, a two-year bullet bond, and a two-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:

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

[4] In matrix notation, Equation (1.19) can be written as $\boldsymbol{B}_{\text{cpn}} = \mathbf{Y} \boldsymbol{B}_{\text{zero}}$ and Equation (1.20) can be written as $\boldsymbol{e}_j = \mathbf{Y}^\top \boldsymbol{x}(j)$, where $\boldsymbol{e}_j$ is the vector on the left hand side of (1.20), and the other symbols are self-explanatory (the symbol $^\top$ indicates transposition). Hence,

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

which is equivalent to (1.21).

No solution exists. In Example 1.2 we found that the solution to the first two equations is

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

In contrast, we found in Example 1.3 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

$$58 \cdot 0.9091 + 54 \cdot 0.8139 \approx 96.68, \tag{1.22}$$

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, i.e. provides the same cash flow. We know that the serial bond is equivalent to a portfolio of 58 one-year zero-coupon bonds and 54 two-year zero-coupon bonds, all with a face value of 1 dollar. In Example 1.2 we found that the one-year zero-coupon bond is equivalent to 1/110 units of the one-year bullet bond, and that the two-year zero-coupon bond is equivalent to a portfolio of $-5/(105 \cdot 110)$ units of the one-year bullet bond and 1/105 units of the two-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 one-year bullet bond and

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

units of the two-year bullet bond. This portfolio will give exactly the same cash flow as the serial bond, i.e. 58 dollars in one year and 54 dollars in two 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 (1.22). □

In some markets, the government bonds are issued with many different payment dates. The system (1.19) will then typically have fewer equations than unknowns. In that case there are many solutions to the equation system, i.e. many sets of discount factors can be consistent both with observed prices and the no-arbitrage pricing principle.

1.6 Determining the zero-coupon yield curve: Parameterized forms

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. We will next consider methods to estimate the entire discount function $T \mapsto B_t^T$ (at least up to some large T). The basic assumption is that the discount function is of a given functional form with some unknown parameters. The value of these parameters are then estimated 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. The unknown parameters can be estimated by least-squares methods.

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 enhance flexibility, some of the procedures suggested in the literature divide the maturity axis into subintervals and use separate functions (of the same type) in each subinterval. To ensure a continuous and smooth term structure of interest rates, one must impose certain conditions for the maturities separating the subintervals. Procedures of this type are called *spline* methods.

In this section we will focus on two of the most frequently applied parameterization techniques, namely cubic splines and the Nelson-Siegel parameterization. An overview of some of the many other approaches suggested in the literature can be seen in Anderson, Breedon, Deacon, Derry, and Murphy (1996, Ch. 2). For some recent procedures, see Jaschke (1998) and Linton, Mammen, Nielsen, and Tanggaard (2001).

The purpose of the procedures is to estimate the term structure of interest rate at a given date t from the market prices of bonds at that date. To simplify the notation in what follows, let $\bar{B}(\tau)$ denote the discount factor for the next τ periods, i.e. $\bar{B}(\tau) = B_t^{t+\tau}$. 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 $\bar{f}(\tau) = f_t^{t+\tau}$.

1.6.1 Cubic splines

In this subsection we will discuss a version of the cubic splines approach introduced by McCulloch (1971) and later modified by McCulloch (1975) and Litzenberger and Rolfo (1984). Given prices for M bonds with time-to-maturities of $T_1 \leq T_2 \leq \dots \leq T_M$. Divide the maturity axis into subintervals defined by the "knot points" $0 = \tau_0 < \tau_1 < \dots < \tau_k = T_M$. 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

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

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 $\alpha_j, \beta_j, \gamma_j$, and δ_j are constants.

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

$$\bar{B}(\tau) = \alpha_0 + \beta_0 \tau + \gamma_0 \tau^2 + \delta_0 \tau^3. \tag{1.23}$$

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

$$\bar{B}(\tau) = \left(1 + \beta_0 \tau + \gamma_0 \tau^2 + \delta_0 \tau^3\right) + \left(\alpha_1 + \beta_1(\tau - \tau_1) + \gamma_1(\tau - \tau_1)^2 + \delta_1(\tau - \tau_1)^3\right). \tag{1.24}$$

To get a smooth transition between (1.23) and (1.24) in the point $\tau = \tau_1$, we demand that

$$\bar{B}(\tau_1-) = \bar{B}(\tau_1+), \tag{1.25}$$

$$\bar{B}'(\tau_1-) = \bar{B}'(\tau_1+), \tag{1.26}$$

$$\bar{B}''(\tau_1-) = \bar{B}''(\tau_1+), \tag{1.27}$$

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 (1.25) ensures that the discount function is continuous in the knot point τ_1. The condition (1.26) 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 (1.27) requires the same to be true for the second-order derivative, which ensures an even smoother behavior of the graph around the knot point τ_1.

The condition (1.25) implies $\alpha_1 = 0$. Differentiating (1.23) and (1.24), 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 (1.26) 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 (1.27) 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

$$\bar{B}(\tau) = 1 + \beta_0 \tau + \gamma_0 \tau^2 + \delta_0 \tau^3 + \sum_{j=1}^{k-1} \delta_j (\tau - \tau_j)^3 I_j(\tau). \tag{1.28}$$

Let $t_1, t_2, \dots, t_N$ denote the time distance from today (date t) to the 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 (1.2), 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

$$B_i = \sum_{n=1}^{N} Y_{in} \bar{B}(t_n) + \varepsilon_i. \tag{1.29}$$

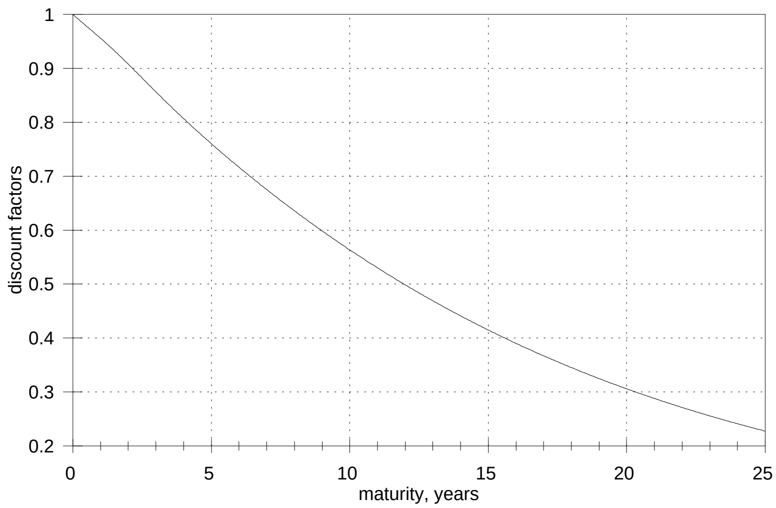


Figure 1.1: The discount function, $\tau \mapsto \bar{B}(\tau)$, estimated using cubic splines and prices of Danish government bonds February 14, 2000.

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 (1.28) into (1.29) yields

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

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.[5] Substituting the estimated parameters into (1.28), we get an estimated discount function, from which estimated zero-coupon yield curves and forward rate curves can be derived as explained earlier in the chapter.

It remains to describe how the number of subintervals k and the knot points τ_j are to be chosen. McCulloch suggested to let k be the nearest integer to $\sqrt{M}$ and to define the knot points by

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

where $h_j = [j \cdot M/k]$ (here the square 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, cf. the preferred habitats hypothesis.

Figure 1.1 shows the discount function on the Danish government bond markets on February 14, 2000 estimated using cubic splines and data from 14 bonds with maturities up to 25 years. Figure 1.2 shows the associated zero-coupon yield curve and the term structure of forward rates.

Discount functions estimated using cubic splines will usually have a credible form for maturities less than the longest maturity in the data set. Although there is nothing in the approach that

[5] See, for example, Johnston (1984).

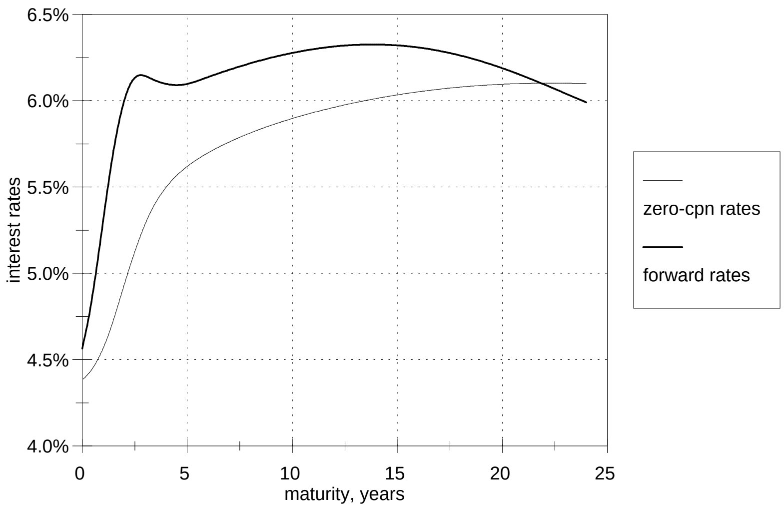


Figure 1.2: The zero-coupon yield curve, $\tau \mapsto \bar{y}(\tau)$, and the term structure of forward rates, $\tau \mapsto \bar{f}(\tau)$, estimated using cubic splines and prices of Danish government bonds February 14, 2000.

ensures that the resulting discount function is positive and decreasing, as it should be according to (1.1), this will almost always be the case. 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, and the method cannot be expected to provide reasonable values beyond the longest maturity T_M, since none of the bonds are affected by that very long end of the term structure.

Two other properties of the cubic splines approach are more disturbing. First, the derived zero-coupon rates will often increase or decrease significantly for maturities approaching T_M, cf. Shea (1984, 1985). Second, 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 bond prices and the precise location of the knot points. Therefore, forward rate curves estimated using cubic splines should only be applied with great caution.

1.6.2 The Nelson-Siegel parameterization

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

$$\bar{f}(\tau) = \beta_0 + \beta_1 e^{-\tau/\theta} + \beta_2 \frac{\tau}{\theta} e^{-\tau/\theta}, \tag{1.30}$$

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 1.3 shows the graphs of the three functions that constitutes (1.30). 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

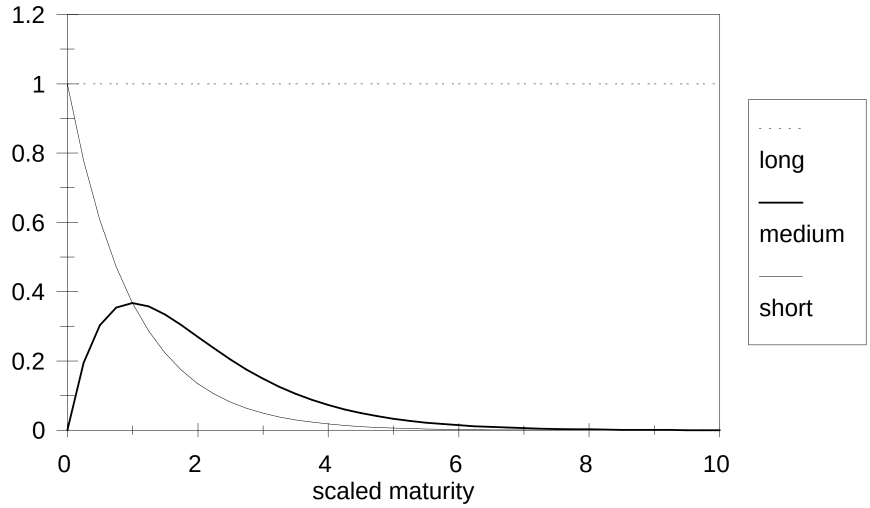


Figure 1.3: The three curves which the Nelson-Siegel parameterization combines.

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 (1.17) on page 9, the term structure of zero-coupon rates is given by

$$\bar{y}(\tau) = \beta_0 + (\beta_1 + \beta_2)\frac{1 - e^{-\tau/\theta}}{\tau/\theta} - \beta_2 e^{-\tau/\theta},$$

which we will rewrite as

$$\bar{y}(\tau) = a + b\frac{1 - e^{-\tau/\theta}}{\tau/\theta} + ce^{-\tau/\theta}. \tag{1.31}$$

Figure 1.4 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.

If we could directly observe zero-coupon rates $\bar{y}(T_i)$ for different maturities T_i, $i = 1, \dots, M$, we could, given θ, 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 $\varepsilon_i \sim N(0, \sigma^2)$, $i = 1, \dots, M$, 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, i.e. 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 (1.31) is given by

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

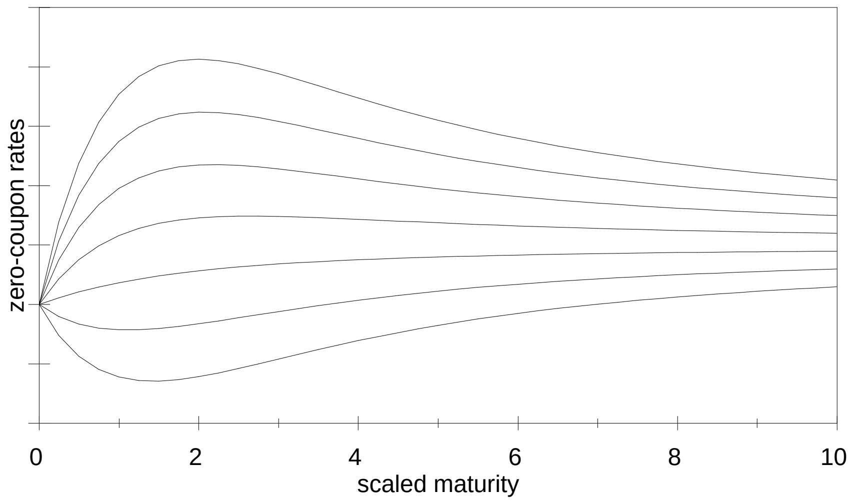


Figure 1.4: Possible forms of the zero-coupon yield curve using the Nelson-Siegel parameterization.

Substituting this into (1.29), we get

$$B_i = \sum_{n=1}^{N} Y_{in} \exp\left\{-at_n - b\theta\left(1 - e^{-t_n/\theta}\right) - ct_n e^{-t_n/\theta}\right\} + \varepsilon_i. \tag{1.32}$$

Since this is a non-linear expression in the unknown parameters, the estimation must be based on generalized least squares, i.e. non-linear regression techniques. See e.g. Gallant (1987).

1.6.3 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 no 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 (1.29) 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 riskless 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.

1.7 Exercises

EXERCISE 1.1 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.

EXERCISE 1.2 Consider two annuity bonds with identical time-to-maturity and payment dates. Let n be the number of remaining payments on each bond, and let R_1 and R_2 be the coupon rates of the two bonds. In addition, let B_1 and B_2 denote the prices (quoted price plus accrued interests) of the bonds. Show that in the absence of arbitrage, the following relation must hold:

$$\frac{B_1}{B_2} = \frac{R_1}{R_2} \frac{1-(1+R_2)^{-n}}{1-(1+R_1)^{-n}}.$$

In some markets callable mortgage-backed annuity bonds with different coupon rates, but the same maturity and payment dates, are traded. Should you expect the relation above to hold for such bonds? Explain!

EXERCISE 1.3 Show that if the discount function does *not* satisfy the condition

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

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

EXERCISE 1.4 Consider two bullet bonds, both with annual payments and exactly four 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 four-year discount factor? What is the four-year zero-coupon interest rate?

Chapter 2

Fixed income securities

2.1 Introduction

This chapter provides a description of the most important financial securities whose cash flows and values only depend on the term structure of interest rates. In Chapter 1 we already considered bonds and the relation between bond prices and the term structure of interest rates. In this chapter we will describe and discuss floating-rate bonds, forwards, futures and options written on bonds, interest rate futures, caps, floors, interest rate swaps and swaptions.

We will also derive numerous relations between the prices of these securities that will hold in the absence of arbitrage and hence independently of the preferences of investors and the future dynamics of the term structure. For example we will show

- a put-call parity for 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
- 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 options on bonds. In later chapters we can therefore focus on the pricing of these "basic" securities.

2.2 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.

Suppose that the payment dates of the bond are $T_1 < \cdots < T_n$, where $T_i - T_{i-1} = \delta$ for all i. In practice, δ will typically equal 0.25, 0.5, or 1 year, corresponding to quarterly, semi-annual, or annual payments. 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 $H(1 + \delta l_{T_{i-1}}^{T_i})$. 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.

Let us look at the valuation of a floating rate 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.

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

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

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

$$B_t^{\text{fl}} = H(1 + \delta l_{T_{i(t)-1}}^{T_{i(t)}}) B_t^{T_{i(t)}}, \quad T_0 \le t < T_n. \tag{2.1}$$

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.

While few floating rate bonds are traded, the results above are also very useful for the analysis of interest rate swaps studied in Section 2.9.

2.3 Forwards on bonds

A forward contract is an agreement between two parties on a trade of a given asset or security at a given future point in time and at a price that is already fixed when the agreement is made. We will refer to this price as the delivery price. Usually, the delivery price is set so that the net present value of the future transaction (computed at the agreement date) is equal to zero. This value of the delivery price is called the **forward price**. Note that the forward price will depend on the time of delivery and the underlying asset to be transacted. In the following we will take a closer look at forwards on zero-coupon bonds and coupon bonds. For a general introduction to forwards, see Hull (2003).

Consider an agreement on a delivery at time T of a zero-coupon bond maturing at time S (where $S > T$) for a price of K. We assume that the face value of zero-coupon bonds is 1 (dollar). The following theorem gives formulas for the time t value $V_t^{T,S}$ of a long position in such a forward contract and for the forward price of the zero-coupon bond. Here and throughout the text, B_t^T denotes the time t price of a zero-coupon bond that pays 1 (dollar) at time T.

Theorem 2.1 *The unique no-arbitrage value at time t of a forward with delivery at time T of a zero-coupon bond maturing at time S at the delivery price K is given by*

$$V_t^{T,S} = B_t^S - K B_t^T. \tag{2.2}$$

The unique no-arbitrage forward price on the zero-coupon bond is

$$F_t^{T,S} = \frac{B_t^S}{B_t^T}. \tag{2.3}$$

Proof: Consider two portfolios, namely

(I) a long position in the forward contract and K zero-coupon bonds maturing at time T,

(II) a zero-coupon bond maturing at time S (i.e. the underlying bond).

At time T the forward yields a payoff of $B_T^S - K$, so that the total time T value of portfolio I is $B_T^S - K + K = B_T^S$, which is identical to the value of portfolio II. Since none of the portfolios yield payments before time T, the no-arbitrage principle implies that they must have the same current price, i.e.

$$V_t^{T,S} + K B_t^T = B_t^S,$$

from which (2.2) follows. Since the forward price $F_t^{T,S}$ is that value of K that makes $V_t^{T,S} = 0$, equation (2.3) follows directly from (2.2). □

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 (1.8), (1.9), and (2.3) that the gain is equal to

$$\begin{aligned} H\left(B_T^S - F_t^{T,S}\right) &= H\left(\frac{1}{1+(S-T)l_T^S} - \frac{1}{1+(S-T)L_t^{T,S}}\right) \\ &= \frac{(S-T)(L_t^{T,S} - l_T^S)H}{\left(1+(S-T)l_T^S\right)\left(1+(S-T)L_t^{T,S}\right)}. \end{aligned} \tag{2.4}$$

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, i.e. 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, let us consider a forward on a coupon bond. As before, let T be the delivery date and K be the delivery price. The underlying coupon bond is assumed to yield payments at the points in time $T_1 < T_2 < \cdots < T_n$, where $T < T_n$. The time T_i payment is denoted by Y_i, $i = 1, 2, \ldots, n$. The time t value of the bond is therefore given by

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

where the sum is over all future payment dates; cf. (1.2) on page 6. Let $V_t^{T,\text{cpn}}$ denote the time t value of this forward contract. We then have the following result:

Theorem 2.2 *The unique no-arbitrage value of a forward on a coupon bond is given by*

$$V_t^{T,cpn} = \sum_{T_i > T} Y_i B_t^{T_i} - K B_t^T = B_t - \sum_{t<T_i<T} Y_i B_t^{T_i} - K B_t^T, \tag{2.5}$$

and the unique no-arbitrage forward price of a coupon bond is

$$F_t^{T,cpn} = \frac{\sum_{T_i>T} Y_i B_t^{T_i}}{B_t^T} = \frac{B_t - \sum_{t<T_i<T} Y_i B_t^{T_i}}{B_t^T} = \sum_{T_i>T} Y_i F_t^{T,T_i}. \tag{2.6}$$

Proof: Consider the portfolios

(I) a forward contract, K zero-coupon bonds maturing at T and, for each T_i with $t < T_i < T$, Y_i zero-coupon bonds maturing at T_i,

(II) the underlying coupon bond.

Between time t and the delivery date T, the two portfolios have exactly the same payments. At time T the value of portfolio I equals $B_T - K + K = B_T$, which is identical to the value of portfolio II. Absence of arbitrage implies that

$$V_t^{T,\text{cpn}} + K B_t^T + \sum_{t<T_i<T} Y_i B_t^{T_i} = B_t,$$

from which (2.5) follows. The forward price follows immediately. □

2.4 Interest rate forwards – forward rate agreements

As discussed in Section 1.4 forward interest rates are rates for a future period relative to the time where 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 so-called 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

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

cf. (1.9), i.e. 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

$$B_T^S H[1+(S-T)K] - H = H\left(\frac{1+(S-T)L_t^{T,S}}{1+(S-T)l_T^S} - 1\right) = \frac{(S-T)(L_t^{T,S} - l_T^S)H}{1+(S-T)l_T^S}. \tag{2.7}$$

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 (2.7) is closely related, but not identical, to the gain/loss on a forward on a zero-coupon bond, cf. (2.4).

2.5 Futures on bonds

As a forward contract, a futures contract is also an agreement of a future transaction of a given asset or security. The distinct characteristic 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 (i.e. the value of the future payments) returns to zero immediately after each 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.

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. At maturity of the contract the futures 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.

Due to the marking-to-market settlement procedure, it is far more difficult to price futures than forwards. In particular, the no-arbitrage principle is not sufficient to derive a unique futures price. As shown by Cox, Ingersoll, and Ross (1981b), the futures price and the forward price will be identical at all points in time if there is no uncertainty about future interest rates, since in that case the timing of the payoffs does not matter.[1] Such an assumption is of course unacceptable in the case of futures on assets that depend on the term structure of interest rates, such as futures on bonds. We will return to the valuation of futures and the relation between forward prices and futures prices in Chapter 6.

2.6 Interest rate futures – Eurodollar futures

Interest rate futures is a class of fixed income instruments that trade with a very high volume at several international exchanges, e.g. CME (Chicago Mercantile Exchange), LIFFE (London International Financial Futures & Options Exchange), and MATIF (Marché à Terme International

[1]A proof of this result can also be found in Appendix 3A in Hull (2003).

de France). The CME interest rate futures involve the three-month Eurodollar deposit rate and are called **Eurodollar futures**. The interest rate involved in the futures contracts traded at LIFFE and MATIF is the three-month LIBOR rate on the Euro currency. We shall simply refer to all these contracts as Eurodollar futures and refer to the underlying interest rate as the three-month LIBOR rate, whose value at time t we denote by $l_t^{t+0.25}$.

The price quotation of Eurodollar futures is a bit 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 three-month LIBOR rate according to the relation

$$\tilde{\mathcal{E}}_T^T = 100\left(1 - l_T^{T+0.25}\right), \tag{2.8}$$

which using (1.8) on page 8 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 is defined 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 (2.8) that the LIBOR futures rate converges to the three-month LIBOR spot rate, as the maturity of the futures contract approaches.

The actual Eurodollar futures price is given by

$$\mathcal{E}_t^T = 100 - 0.25(100 - \tilde{\mathcal{E}}_t^T) = 100 - 25\varphi_t^T$$

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, i.e. 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 (2.7). 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.

2.7 Options on bonds

Options on government bonds are traded at several exchanges and also on the OTC-markets (OTC: Over-the-counter). 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.

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 2.8. 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.

It is well-known that the no-arbitrage principle in itself does not yield a unique price for stock options, but only upper and lower boundaries for the price, cf. Merton (1973) or Hull (2003). This is also the case for options on bonds. 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 2.1. Just as for stock options, the absence of arbitrage is sufficient to derive a precise relation between prices on European call and put options with the same underlying asset, exercise price, and maturity date. This relation, the so-called put-call parity, is stated below both for options on zero-coupon bonds and options on coupon bonds.

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).

2.7.1 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.

A European **call option** on this zero-coupon bond gives the owner the right, but not the obligation, to buy the zero-coupon bond at time T for a price of K. We let $C_t^{K,T,S}$ denote the time t price of the call option. At maturity the value of the call equals its payoff:

$$C_T^{K,T,S} = \max\left(B_T^S - K, 0\right).$$

A European **put option** on the zero-coupon bond gives the owner the right, but not the obligation, to sell the zero-coupon bond at time T for a price of K. We let $\pi_t^{K,T,S}$ denote the time t price of the put option. The value at maturity is equal to

$$\pi_T^{K,T,S} = \max\left(K - B_T^S, 0\right).$$

If the owner of the option makes use of his right to buy/sell the underlying asset, the option is said to be exercised. 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.

The put-call parity gives a precise relation between the prices of European call and put options with the same underlying asset, exercise price and maturity date. For options on zero-coupon bonds the relation is as follows:

Theorem 2.3 *In absence of arbitrage, the prices of European call and put options on zero-coupon bonds satisfy the relation*

$$C_t^{K,T,S} + KB_t^T = \pi_t^{K,T,S} + B_t^S. \tag{2.9}$$

Proof: 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

$$\max\left(B_T^S - K, 0\right) + K = \max\left(B_T^S, K\right)$$

and will have a current time t price given by the left-hand side of (2.9). Another portfolio consisting of a put option and one unit of the underlying zero-coupon bond has a time T value of

$$\max\left(K - B_T^S, 0\right) + B_T^S = \max\left(K, B_T^S\right)$$

and a time t price corresponding to the right-hand side of (2.9). None of the portfolios provide payments before time T. Therefore, there will be an obvious arbitrage opportunity unless (2.9) is satisfied. □

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.

American options can be exercised at the time of maturity T or any point in time before time T. 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; cf. Merton (1973) and Hull (2003). 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, i.e. sufficiently high interest rates.

2.7.2 Options on coupon bonds

Consider a coupon bond with payments Y_i at time T_i $(i = 1, 2, \ldots, n)$, where $T_1 < T_2 < \cdots < T_n$. Let B_t denote the time t price of this bond, i.e.

$$B_t = \sum_{T_i > t} Y_i B_t^{T_i}.$$

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 value of the options is 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

(1) only the payments of the bonds after maturity of the option are relevant for the payoff and the value of the option;[2]

(2) 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 put-call parity for European options on coupon bonds is as follows:

[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.

Theorem 2.4 *Absence of arbitrage implies that*

$$C_t^{K,T,cpn} + KB_t^T = \pi_t^{K,T,cpn} + B_t - \sum_{t<T_i\leq T} Y_i B_t^{T_i}. \tag{2.10}$$

The proof of this result is left for the reader in Exercise 2.2.

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 e.g. in Hull (2003, Chap. 12). 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 $K(1-B_{T_l}^T)$. 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 $K(1 - B_{T_i}^{T_{i+1}})$. 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.[4] 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 may be optimal for sufficiently low bond prices, i.e. high interest rates.

2.8 Caps, floors, and collars

2.8.1 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.[5] The interest rate to be paid at time T_i is determined by the δ-period money market interest rate prevailing at time $T_{i-1} = T_i - \delta$, i.e. 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 $T_0, T_1, \ldots, T_{n-1}$ 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 ($i = 1, \ldots, n$) 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 $i = 1, 2, \ldots, n$. If a borrower buys such a cap, the net payment at time T_i cannot exceed $H\delta K$. The period length δ is often referred to as the **frequency** or the **tenor** of the cap.[6] 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.

[4] 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.

[5] 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*.

[6] The word tenor is sometimes used for the set of payment dates $T_1, \ldots, T_n$.

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

$$\mathcal{C}^i_{T_i} = H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right) \tag{2.11}$$

and no other payments. A caplet is a call option on the zero-coupon yield 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, i.e.

$$\mathcal{C}^i_t = B^{T_i}_t H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right), \quad T_i - \delta \leq t \leq T_i.$$

In particular,

$$\mathcal{C}^i_{T_i-\delta} = B^{T_i}_{T_i-\delta} H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right).$$

Applying (1.8) on page 8, we can rewrite this value 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. Accordingly, 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 can write this as

$$\mathcal{C}^i_t = H(1 + \delta K)\pi_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}. \tag{2.12}$$

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

$$\mathcal{C}_t = \sum_{i=1}^{n} \mathcal{C}^i_t = H(1 + \delta K) \sum_{i=1}^{n} \pi_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad t < T_0. \tag{2.13}$$

At all dates after the first reset date, the next payment of the cap will already be known. If we again use the notation $T_{i(t)}$ for the nearest following payment date after time t, 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

$$\begin{aligned}
\mathcal{C}_t = {} & H B^{T_{i(t)}}_t \delta \max\left(l^{T_{i(t)}}_{T_{i(t)}-\delta} - K, 0\right) \\
& + (1 + \delta K) H \sum_{i=i(t)+1}^{n} \pi_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad T_0 \leq t \leq T_n.
\end{aligned} \tag{2.14}$$

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.

2.8.2 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

$$\mathcal{F}^i_{T_i} = H\delta \max\left(K - l^{T_i}_{T_i-\delta}, 0\right), \tag{2.15}$$

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.

The (hypothetical) contracts that only yield one of the payments in (2.15) 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 also think of a 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

$$\mathcal{F}^i_{T_i-\delta} = H(1+\delta K) \max\left(B^{T_i}_{T_i-\delta} - \frac{1}{1+\delta K}, 0\right). \tag{2.16}$$

The total value of the floor contract at any time $t < T_0$ is therefore given by

$$\mathcal{F}_t = H(1+\delta K) \sum_{i=1}^{n} C_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad t < T_0, \tag{2.17}$$

and later the value is

$$\begin{aligned} \mathcal{F}_t &= H B_t^{T_{i(t)}} \delta \max\left(K - l^{T_{i(t)}}_{T_{i(t)}-\delta}, 0\right) \\ &\quad + (1+\delta K) H \sum_{i=i(t)+1}^{n} C_t^{(1+\delta K)^{-1}, T_i-\delta, T_i}, \quad T_0 \le t \le T_n. \end{aligned} \tag{2.18}$$

2.8.3 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.

2.8.4 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. The examples are taken from Musiela and Rutkowski (1997, Ch. 16).

- 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 due to the contract does not exceed a certain pre-specified level. Consequently, the ordinary cap payments in (2.11) 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 (2.11) 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.

2.9 Swaps and swaptions

2.9.1 Swaps

Many different types of swaps are traded on the OTC markets, e.g. currency swaps, credit swaps, 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, e.g. 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.

The first swap was contracted in 1981, and nowadays the international swap markets are enormous, both in terms of transactions and outstanding contracts. The organization ISDA (International Swaps and Derivatives Association) publishes key figures showing the size and development of the markets for swaps and interest rate options traded OTC. After 1997 the published figures are for all products together and are not informative for the markets of the different securities. At the end of 1997, the total notional principal (face value) of all reported, active interest rate swaps amounted to 22.3 trillion U.S. dollars (i.e. 22 300 000 million), and the swap market was indisputably the largest OTC derivative market.[7] The total OTC derivative market more than doubled in size from 1997 to 2000, and there is no reason to believe that the growth rate of the swap market itself is significantly different.

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, i.e. 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, i.e. 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

[7] For interest rate swaps denoted in the Danish currency the corresponding number is 133 billion U.S. dollars (i.e. 133 000 million), an increase of 16% relative to the year before.

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 (2003, Ch. 6).

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 $T_1, \dots, T_n$, 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 $T_0, T_1, \dots, T_{n-1}$ 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

$$V_t^{\text{fix}} = \sum_{i=i(t)}^{n} HK\delta B_t^{T_i} = HK\delta \sum_{i=i(t)}^{n} B_t^{T_i}. \tag{2.19}$$

The floating rate payments are exactly the same as the coupon payments on a floating rate bond, which was discussed in Section 2.2, i.e. at time T_i $(i = 1, 2, \dots, n)$ the payment is $H\delta l_{T_i-\delta}^{T_i}$. Note that this payment is already known at time $T_i - \delta$. According to (2.1), the value of such a floating bond at any time $t \in [T_0, T_n)$ is given by $H(1 + \delta l_{T_{i(t)}-\delta}^{T_{i(t)}})B_t^{T_{i(t)}}$. 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

$$\begin{aligned} V_t^{\text{fl}} &= H(1+\delta l_{T_{i(t)}-\delta}^{T_{i(t)}})B_t^{T_{i(t)}} - HB_t^{T_n} \\ &= H\delta l_{T_{i(t)}-\delta}^{T_{i(t)}} B_t^{T_{i(t)}} + H\left[B_t^{T_{i(t)}} - B_t^{T_n}\right], \quad T_0 \le t < T_n. \end{aligned}$$

At and before time T_0, the first term is not present, so the value of the floating rate payments is simply

$$V_t^{\text{fl}} = H\left[B_t^{T_0} - B_t^{T_n}\right], \quad t \le T_0. \tag{2.20}$$

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

$$H\delta l_{T_i-\delta}^{T_i} B_{T_i-\delta}^{T_i} = H\delta \frac{l_{T_i-\delta}^{T_i}}{1+\delta l_{T_i-\delta}^{T_i}},$$

where we have applied (1.8) on page 8. 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

$$HB_{T_i-\delta}^{T_i-\delta} - HB_{T_i-\delta}^{T_i} = H - \frac{H}{1+\delta l_{T_i-\delta}^{T_i}} = H\delta \frac{l_{T_i-\delta}^{T_i}}{1+\delta l_{T_i-\delta}^{T_i}},$$

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

$$H\left(B_t^{T_i-\delta} - B_t^{T_i}\right) = H\delta B_t^{T_i} \frac{\frac{B_t^{T_i-\delta}}{B_t^{T_i}} - 1}{\delta} = H\delta B_t^{T_i} L_t^{T_i-\delta,T_i}, \tag{2.21}$$

where we have applied (1.9) on page 8. Thus, the value at time $t \le T_i - \delta$ of getting $H\delta l_{T_i-\delta}^{T_i}$ at time T_i is equal to $H\delta B_t^{T_i} L_t^{T_i-\delta,T_i}$, i.e. the unknown future spot rate $l_{T_i-\delta}^{T_i}$ in the payoff is replaced by the current forward rate for $L_t^{T_i-\delta,T_i}$ and then discounted by the current riskfree 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

$$V_t^{\text{fl}} = H\delta B_t^{T_{i(t)}} l_{T_{i(t)}-\delta}^{T_{i(t)}} + H\delta \sum_{i=i(t)+1}^{n} B_t^{T_i} L_t^{T_i-\delta,T_i}, \quad t > T_0.$$

At or before time T_0, the first term is not present, so we get

$$V_t^{\text{fl}} = H\delta \sum_{i=1}^{n} B_t^{T_i} L_t^{T_i-\delta,T_i}, \quad t \le T_0. \tag{2.22}$$

The value of a payer swap is

$$\mathbf{P}_t = V_t^{\text{fl}} - V_t^{\text{fix}},$$

while the value of a receiver swap is

$$\mathbf{R}_t = V_t^{\text{fix}} - V_t^{\text{fl}}.$$

In particular, the value of a payer swap at or before its starting date T_0 can be written as

$$\mathbf{P}_t = H\delta \sum_{i=1}^{n} B_t^{T_i}\left(L_t^{T_i-\delta,T_i} - K\right), \quad t \le T_0, \tag{2.23}$$

using (2.19) and (2.22), or as

$$\mathbf{P}_t = H\left(\left[B_t^{T_0} - B_t^{T_n}\right] - \sum_{i=1}^{n} K\delta B_t^{T_i}\right), \quad t \le T_0, \tag{2.24}$$

using (2.19) and (2.20). If we let $Y_i = K\delta$ for $i = 1, \ldots, n-1$ and $Y_n = 1+K\delta$, we can rewrite (2.24) as

$$\mathbf{P}_t = H\left(B_t^{T_0} - \sum_{i=1}^{n} Y_i B_t^{T_i}\right), \quad t \le T_0. \tag{2.25}$$

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:

$$\mathcal{C}_t = \mathcal{F}_t + \mathbf{P}_t. \tag{2.26}$$

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** $\tilde{l}^{\delta}_{T_0}$ prevailing at time T_0 for a swap with frequency δ and payments dates $T_i = T_0 + i\delta$, $i = 1, 2, \ldots, n$, 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, i.e. $\mathbf{P}_{T_0} = \mathbf{R}_{T_0} = 0$. The swap rate is sometimes called the equilibrium swap rate or the par swap rate. Applying (2.23), we can write the swap rate as

$$\tilde{l}^{\delta}_{T_0} = \frac{\sum_{i=1}^{n} L_{T_0}^{T_i-\delta,T_i} B_{T_0}^{T_i}}{\sum_{i=1}^{n} B_{T_0}^{T_i}},$$

which can also be written as a weighted average of the relevant forward rates:

$$\tilde{l}^{\delta}_{T_0} = \sum_{i=1}^{n} w_i L_{T_0}^{T_i-\delta,T_i}, \tag{2.27}$$

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 (2.24) 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

$$\tilde{l}^{\delta}_{T_0} = \frac{1 - B_{T_0}^{T_n}}{\delta \sum_{i=1}^{n} B_{T_0}^{T_i}}. \tag{2.28}$$

Substituting (2.28) 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 $T_1, \ldots, T_n$, can be written in terms of the current swap rate as

$$\begin{aligned}\mathbf{P}_{T_0} &= H\left(\tilde{l}^{\delta}_{T_0}\delta\left(\sum_{i=1}^{n} B_{T_0}^{T_i}\right) - K\delta\left(\sum_{i=1}^{n} B_{T_0}^{T_i}\right)\right) \\ &= \left(\sum_{i=1}^{n} B_{T_0}^{T_i}\right) H\delta\left(\tilde{l}^{\delta}_{T_0} - K\right).\end{aligned} \tag{2.29}$$

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 (2.23)–(2.25). 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

$$\tilde{L}_t^{\delta,T_0} = \frac{B_t^{T_0} - B_t^{T_n}}{\delta \sum_{i=1}^n B_t^{T_i}} = \frac{\sum_{i=1}^n L_t^{T_i-\delta,T_i} B_t^{T_i}}{\sum_{i=1}^n B_t^{T_i}}. \tag{2.30}$$

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, i.e. 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 2.3, 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.

2.9.2 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.

Let us first focus on a European receiver swaption. At time T_0, the value of a receiver swap with payment dates $T_i = T_0 + i\delta$, $i = 1, 2, \dots, n$, and a fixed rate K 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 $Y_n = 1 + K\delta$; cf. (2.25). Hence, the time T_0 payoff of a receiver swaption is

$$\mathcal{R}_{T_0} = \max\left(\mathbf{R}_{T_0} - 0, 0\right) = H \max\left(\sum_{i=1}^n Y_i B_{T_0}^{T_i} - 1, 0\right), \tag{2.31}$$

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 these options 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

$$\mathcal{P}_{T_0} = \max\left(\mathbf{P}_{T_0} - 0, 0\right) = \max\left(-\mathbf{R}_{T_0}, 0\right) = H \max\left(1 - \sum_{i=1}^n Y_i B_{T_0}^{T_i}, 0\right). \tag{2.32}$$

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 (2.29) to express the payoff of a European payer swaption as

$$\mathcal{P}_{T_0} = \left(\sum_{i=1}^{n} B_{T_0}^{T_i} \right) H\delta \max\left(\tilde{l}_{T_0}^{\delta} - K, 0 \right). \tag{2.33}$$

The interpretation of this payoff is that the payer swaption gives the right to pay a fixed annual rate of K in a swap beginning at T_0 instead of the equilibrium swap rate $\tilde{l}_{T_0}^{\delta}$ that prevails at that date. Hence, at all the payment dates of the swap $T_1, \dots, T_n$, the payer swaption allows the holder to reduce the fixed rate payment by $H\delta\left(\tilde{l}_{T_0}^{\delta} - K\right)$, which the holder will do whenever this difference is positive. Discounting back to T_0 and summing up, we get (2.33). Similarly, the payoff for a European receiver swaption can be written as

$$\mathcal{R}_{T_0} = \left(\sum_{i=1}^{n} B_{T_0}^{T_i} \right) H\delta \max\left(K - \tilde{l}_{T_0}^{\delta}, 0 \right).$$

Also note that the following **payer-receiver parity** holds for European swaptions having the same underlying swap and the same exercise rate:

$$\mathcal{P}_t - \mathcal{R}_t = \mathbf{P}_t, \quad t \leq T_0, \tag{2.34}$$

cf. Exercise 2.4. In words, a payer swaption minus a receiver swaption is indistinguishable form a forward payer swap.

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, i.e. 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 option holder upon exercise receives a swap with the same time to maturity no matter when the option is exercised.

2.9.3 Exotic swap instruments

The following examples of exotic swap market products are adapted from Musiela and Rutkowski (1997) and Hull (2003):

- **Float-for-floating swap:** Two floating interest rates are swapped, e.g. the three-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, e.g. 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, i.e. 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, i.e. 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 $T_1, \ldots, T_m$, 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 $T_1, \ldots, T_L$. 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.

2.10 Exercises

EXERCISE 2.1 Show that the no-arbitrage price of a European call on a zero-coupon bond will satisfy

$$\max\left(0, B_t^S - KB_t^T\right) \le C_t^{K,T,S} \le B_t^S(1-K)$$

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, cf. Hull (2003, Ch. 8). Derive similar bounds for a European call on a coupon bond.

EXERCISE 2.2 Give a proof of the put-call parity for options on coupon bonds in Theorem 2.4.

EXERCISE 2.3 Let $\tilde{l}^{\delta}_{T_0}(k)$ be the equilibrium swap rate for a swap with payment dates $T_1, T_2, \dots, T_k$, where $T_i = T_0 + i\delta$ as usual. Suppose that $\tilde{l}^{\delta}_{T_0}(1), \dots, \tilde{l}^{\delta}_{T_0}(n)$ are known. Find a recursive procedure for deriving the associated discount factors $B^{T_1}_{T_0}, B^{T_2}_{T_0}, \dots, B^{T_n}_{T_0}$.

EXERCISE 2.4 Show the parity (2.34). 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 2.5 Consider a swap with starting date T_0 and a fixed rate K. For $t \le T_0$, show that $V_t^{\text{fl}}/V_t^{\text{fix}} = \tilde{L}_t^{\delta,T_0}/K$, where $\tilde{L}_t^{\delta,T_0}$ is the forward swap rate.

Chapter 3

Stochastic processes and stochastic calculus

In the previous chapter we saw that many interest rate dependent securities cannot be priced uniquely just by appealing to no-arbitrage arguments. To derive prices and hedging strategies we have to model the uncertainty about the term structure of interest rates at relevant future dates. In order to analyze the relation between interest rates and other macroeconomic variables such as aggregate consumption or production, we also have to take the uncertainty about the future values of these variables into account. For example, the uncertainty about future consumption will affect individuals' supply and demand for bonds and, hence, affect the interest rates set today. In modern finance, stochastic processes are used to model the evolution of uncertain variables over time. Therefore, a basic knowledge of stochastic processes and how to do computations involving stochastic processes is needed in order to understand, evaluate, and develop models of the term structure of interest rates. This chapter is devoted to a relatively brief 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 the textbooks of Øksendal (1998) and Karatzas and Shreve (1988).

3.1 Probability spaces

The basic object for studies of uncertain events is a **probability space**, which is a triple $(\Omega, \mathcal{F}, \mathbb{P})$. Here, 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 dice (meaning the number of "eyes" on top of the dice), the state space is $\Omega = \{1, 2, 3, 4, 5, 6\}$. An event is a set of possible outcomes, i.e. a subset of the state space. In the example with the dice, some events are $\{1, 2, 3\}$, $\{4, 5\}$, $\{1, 3, 5\}$, $\{6\}$, and $\{1, 2, 3, 4, 5, 6\}$. The second component of a probability space, $\mathcal{F}$, is the set of events to which a probability can be assigned, i.e. the set of "probabilizable" events. Hence, $\mathcal{F}$ is a set of subsets of the state space! It is required that

(i) the entire state space can be assigned a probability, i.e. $\Omega \in \mathcal{F}$;

(ii) if some event $F \subseteq \Omega$ can be assigned a probability, so can its complement $F^c \equiv \Omega \setminus F$, i.e.

$F \in \mathcal{F} \Rightarrow F^c \in \mathcal{F}$; and

(iii) given a sequence of probabilizable events, the union is also probabilizable, i.e. $F_1, F_2, \dots \in \mathcal{F} \Rightarrow \cup_{i=1}^{\infty} F_i \in \mathcal{F}$.

Often $\mathcal{F}$ is referred to as a **sigma-field**. The final component of a probability space is a **probability measure** $\mathbb{P}$, which formally is a function from the sigma-field $\mathcal{F}$ 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, i.e. 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-field, $\mathcal{F}$, of events. In the example of the dice, a probability measure $\mathbb{P}$ corresponding to the idea that the dice 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 $\mathbb{Q}(\{1\}) = 1/12$, $\mathbb{Q}(\{2\}) = \dots = \mathbb{Q}(\{5\}) = 1/6$, and $\mathbb{Q}(\{6\}) = 3/12$, which may be appropriate if the dice is believed to be "unfair".

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

3.2 Stochastic processes

The state of many systems or objects changes over time in a manner that cannot be predicted with certainty. This is also true for many economic objects such as stock prices, interest rates, and exchange rates. Such an object can be described by a stochastic process, which is a family of random variables with one random variable for each time we observe the state of the object. We will denote a generic stochastic process by the symbol x, which is then given as a collection $(x_t)_{t \in \mathcal{T}}$ of random variables defined on a common probability space $(\Omega, \mathcal{F}, \mathbb{P})$. We will only consider real-valued stochastic process, i.e. all the random variables take values in (a subset of) $\mathbb{R}^K$ for some integer $K \geq 1$. The set $\mathcal{T}$ consists of all the points in time at which we care about the state of the object, which is represented by the value of the process. Associated with a stochastic process is information about exactly how the state can change over time.

3.2.1 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 $\mathcal{T} = \{0, 1, 2, \dots, T\}$ or $\{0, 1, 2, \dots\}$. If the state can take on all values in a given interval (e.g. all real numbers), the process is called a **continuous-variable stochastic process**. On the other hand, if the state can take on countably many separated values, the process is called a **discrete-variable stochastic process**.

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, with many investors there will be some trades at almost any point in time, so that prices and interest rates etc. 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. In practice, these economic variables can only take on countably many values, e.g. 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 modeling 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. In sum, we will use continuous-time, continuous-variable stochastic processes throughout to describe the evolution in prices and rates. Therefore the remaining section of this chapter will be devoted to that type of stochastic processes.

3.2.2 Basic concepts

Let us consider a continuous-time, continuous-variable stochastic process $x = (x_t)_{t \in \mathbb{R}_+}$, where the random variable x_t represents the state or value of the object at time t. Here we assume that we are interested in the state at every point in time in $\mathbb{R}_+ = [0, \infty)$, where time is measured relative to some given starting time, "time 0." In this case an outcome is an entire set of values $\{x_t | t \geq 0\}$, which we will call a **(sample) path**. A path of a stochastic process is a possible realization of the evolution of the process over time. The state space Ω is the set of all paths. Events are subsets of Ω, i.e. sets of paths. The following are examples of some events: $\{x_t \leq 10 \mid \text{ for all } t \leq 1\}$, $\{x_5 \geq 0\}$, and $\{x_1 \leq a_1, x_{3/2} \geq a_2\}$. Attached to all possible events is a probability given by a probability measure $\mathbb{P}$.

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.

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 $(x_t)_{t \in [0, t')}$, where $x_t \in \mathcal{S}$, will be known (at least in the models we consider). These values constitute the **history** of the process up to time t'. The future values are still stochastic.

3.2.3 Markov processes and martingales

As time passes 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 subsets A of the value space $\mathcal{S}$. If for all $t, t' \in \mathbb{R}_+$ with $t < t'$ and all $A \subseteq \mathcal{S}$, 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 presence, 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 it, 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 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 $t \geq 0$ we have that

$$\mathrm{E}_t^{\mathbb{P}}\left[x_s\right] = x_t, \qquad \text{for all } s > t.$$

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.

3.2.4 Continuous or discontinuous paths

We will only consider stochastic processes having paths that are continuous functions of time, so that one can depict the evolution of the process by a continuous curve. The most fundamental process with this property is the so-called standard Brownian motion or Wiener process, which we will describe in detail in the next section. From the standard Brownian motion many other interesting continuous-path processes can be constructed as we will see in later sections.

Stochastic processes which have paths with discontinuities (jumps) also exist. The jumps of such processes are often modeled 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

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

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 break-through. 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 (which does not necessarily conflict with the usual assumption of rational investors). 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.

There are numerous financial models of stock markets that allow for jumps in stock prices, e.g. Merton (1976) discusses the pricing of stock options in such a framework. On the other hand, there are only very few models allowing for jumps in interest rates.[2] This can be justified empirically by the observation that sudden, large movements are not nearly as frequent in the bond markets as in the stock markets. There are also theoretical arguments supporting these findings. In a general equilibrium model of the economy, Wu (1999) shows among other things that jumps in the overall productivity of the economy will cause jumps in stock prices, but not in bond prices or interest rates. 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; see e.g. Lando (1998). In the main part of the text we will focus on default-free contracts and use continuous-path processes.

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. A standard Brownian motion is defined relative to a probability measure $\mathbb{P}$, under which the increments have the properties above. For example, for all $t < t'$ and all $h \in \mathbb{R}$ we have that

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

where $N(\cdot)$ denotes the cumulative distribution function for an $N(0, 1)$-distributed random stochastic variable. To be precise, we should use the term $\mathbb{P}$-standard Brownian motion, but the probability

[2]For an example see Babbs and Webber (1994).

measure is often clear from the context. 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 in 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.

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

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

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.[3] Note that since a normally distributed random variable can take on infinitely many values, a standard Brownian motion has infinitely many paths that each has 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

[3] Most spreadsheets and programming tools have a built-in procedure that generates uniformly distributed numbers over the interval $[0, 1]$. Such uniformly distributed random numbers can be transformed into standard normally distributed numbers in several ways. One example: Given uniformly distributed numbers U_1 and U_2, the numbers ε_1 and ε_2 defined by

$$\varepsilon_1 = \sqrt{-2 \ln U_1} \sin(2\pi U_2), \qquad \varepsilon_2 = \sqrt{-2 \ln U_1} \cos(2\pi U_2)$$

will be independent standard normally distributed random numbers. This is the so-called Box-Muller transformation. See e.g. Press, Teukolsky, Vetterling, and Flannery (1992, Sec. 7.2).

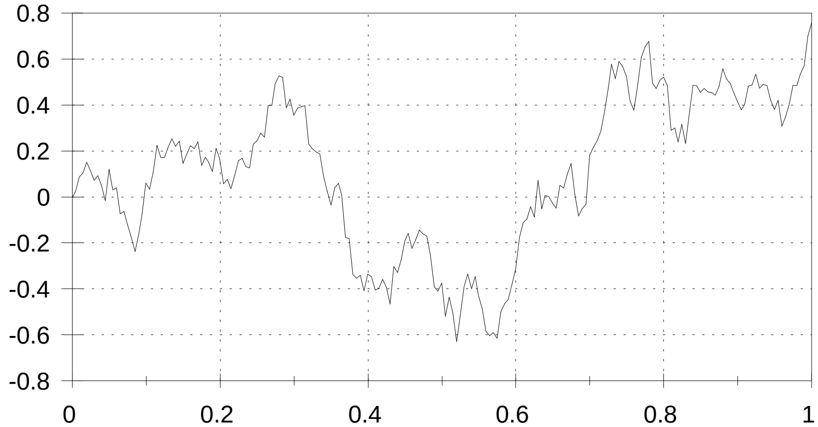


Figure 3.1: A simulated path of a standard Brownian motion based on 200 subintervals.

and define a new stochastic process $x = (x_t)_{t \geq 0}$ by

$$x_t = x_0 + \mu t + \sigma z_t, \quad t \geq 0, \tag{3.1}$$

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 μt and variance $\sigma^2 t$, i.e. $x_t \sim N(x_0 + \mu t, \sigma^2 t)$.

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

$$dx_t = \mu\, dt + \sigma\, dz_t, \tag{3.2}$$

where dz_t can loosely be interpreted as a $N(0, dt)$-distributed random variable. To give this a precise mathematical meaning, it must 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 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, also a generalized Brownian motion is a Markov process, and the

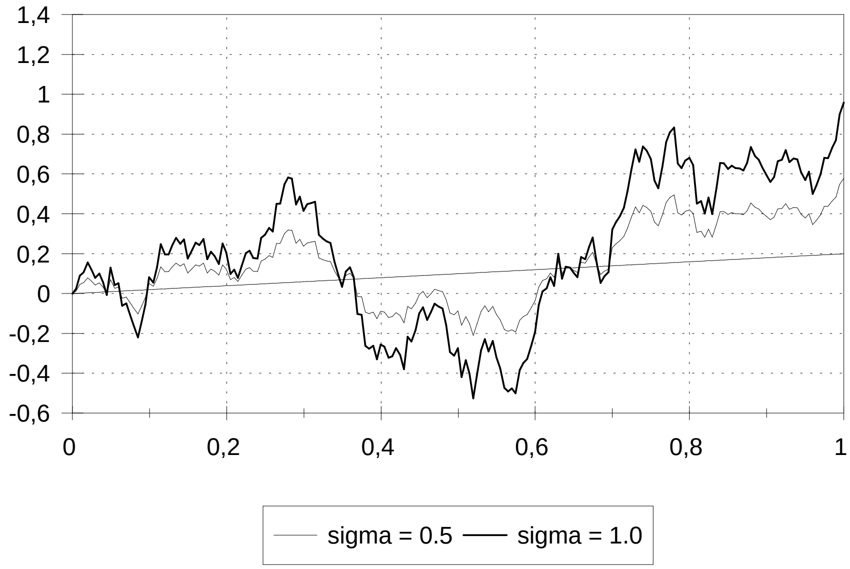


Figure 3.2: Simulation of a generalized Brownian motion with $\mu = 0.2$ and $\sigma = 0.5$ or $\sigma = 1.0$. The straight line shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals.

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

$$dx_t = \mu(t)\, dt + \sigma(t)\, dz_t. \tag{3.3}$$

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

$$x_{t'} - x_t = \int_t^{t'} \mu(u)\, du + \int_t^{t'} \sigma(u)\, dz_u. \tag{3.4}$$

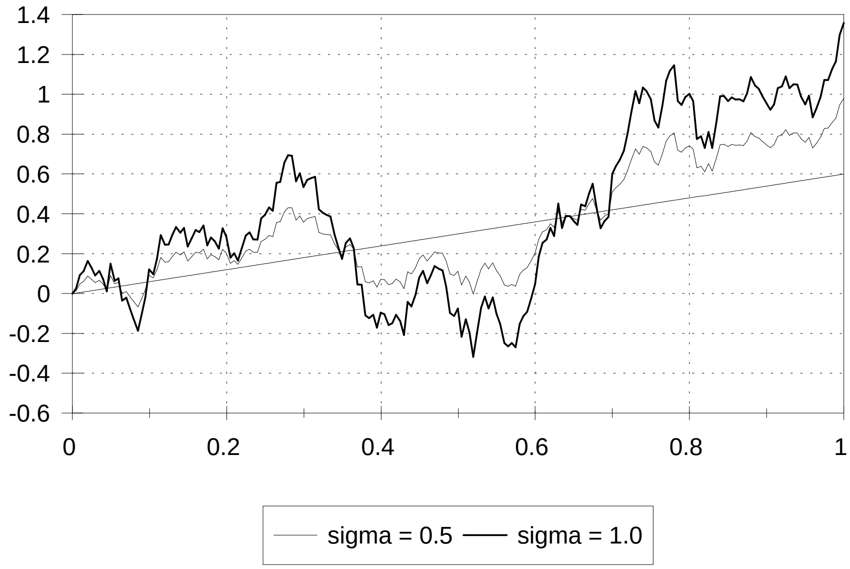


Figure 3.3: Simulation of a generalized Brownian motion with $\mu = 0.6$ and $\sigma = 0.5$ or $\sigma = 1.0$. The straight line shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals.

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, i.e. 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

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

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.[4] This expression generalizes (3.2), where μ and σ were assumed to be constants, and (3.3), where μ and σ were functions of time only. An equation

[4]For the process x to be mathematically meaningful, the functions $\mu(x, t)$ and $\sigma(x, t)$ must satisfy certain conditions. See e.g. Øksendal (1998, Ch. 7) and Duffie (2001, App. E).

like (3.5), 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.

In the expression (3.5) 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 (the history up to and including time t). To be more precise, the change in a diffusion process over any interval $[t, t']$ is

$$x_{t'} - x_t = \int_t^{t'} \mu(x_u, u)\, du + \int_t^{t'} \sigma(x_u, u)\, dz_u, \tag{3.6}$$

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 (3.5). The drift rate $\mu(x_t, t)$ and the variance rate $\sigma(x_t, t)^2$ are really the limits

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

A diffusion process is a Markov process as can be seen from (3.5), 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. 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 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 on the form

$$dx_t = \mu_t\, dt + \sigma_t\, dz_t, \tag{3.7}$$

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. 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 (1998, Ch. 4). The expression (3.7) gives an intuitive understanding of the evolution of an Itô process, but it is more precise to state the evolution in the integral form

$$x_{t'} - x_t = \int_t^{t'} \mu_u \, du + \int_t^{t'} \sigma_u \, dz_u, \tag{3.8}$$

where the last term again is a stochastic integral.

3.6 Stochastic integrals

3.6.1 Definition and properties of stochastic integrals

In (3.6) and (3.8) 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 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 $t \equiv t_0 < t_1 < \dots < t_n \equiv t'$, so that σ_u is constant on each subinterval $[t_i, t_{i+1})$. The stochastic integral is then defined by

$$\int_t^{t'} \sigma_u \, dz_u = \sum_{i=0}^{n-1} \sigma_{t_i} \left(z_{t_{i+1}} - z_{t_i} \right). \tag{3.9}$$

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:

$$\int_t^{t'} \sigma_u \, dz_u = \lim_{m \to \infty} \int_t^{t'} \sigma_u^{(m)} \, dz_u. \tag{3.10}$$

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 (1998). 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 3.1 *The stochastic integral $\int_t^{t'} \sigma_u \, dz_u$ has the following properties:*

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

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

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

Theorem 3.2 *If z is a Brownian motion, and $\sigma(u)$ 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 the 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, i.e.

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

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, 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.

3.6.2 The martingale representation theorem

As discussed above any process $y = (y_t)$ of the form $y_t = \int_0^t \sigma_u\, dz_u$, or 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 an Itô integral. This is the so-called martingale representation theorem:

Theorem 3.3 *Suppose the process $M = (M_t)$ is a martingale with respect to a probability measure under which $z = (z_t)$ is a standard Brownian motion. Then a unique process $\theta = (\theta_t)$ exists such that*

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

The integrand process θ is such that for any t the value θ_t is known at time t. This result is used in the chapter on general asset pricing results. For a mathematically precise statement of the result and a proof, see Øksendal (1998, Thm. 4.3.4).

3.6.3 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 3.4 *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 and β 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, Jarrow, and Morton (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 3.5 *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

$$dy_t = \left(\frac{\partial g}{\partial t}(x_t, t) + \frac{\partial g}{\partial x}(x_t, t)\mu_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(x_t, t)\sigma_t^2 \right) dt + \frac{\partial g}{\partial x}(x_t, t)\sigma_t \, dz_t. \tag{3.11}$$

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 (1998, Ch. 4) and similar textbooks. The result can also be written in the following way, which may be easier to remember:

$$dy_t = \frac{\partial g}{\partial t}(x_t, t)\,dt + \frac{\partial g}{\partial x}(x_t, t)\,dx_t + \frac{1}{2}\frac{\partial^2 g}{\partial x^2}(x_t, t)(dx_t)^2. \tag{3.12}$$

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[5] $\mathrm{Var}[(dz_t)^2] = 2(dt)^2$. 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 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. It is therefore important to be familiar with the way it works. It is a good idea to train yourself by doing the exercises at the end of this chapter.

3.8 Important diffusion processes

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

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

$$dx_t = \mu x_t\,dt + \sigma x_t\,dz_t, \tag{3.13}$$

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 (3.5) 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.

The expression (3.13) 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

[5] 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.

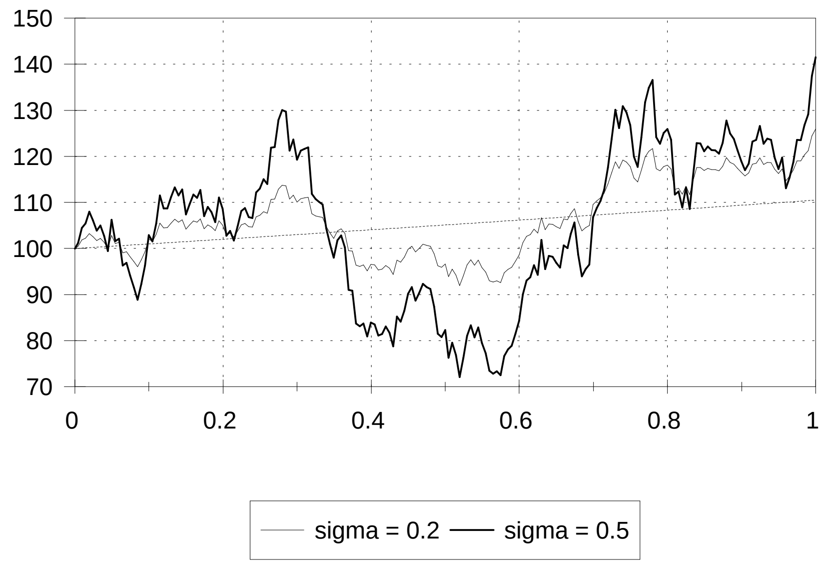


Figure 3.4: Simulation of a geometric Brownian motion with initial value $x_0 = 100$, relative drift rate $\mu = 0.1$, and a relative volatility of $\sigma = 0.2$ and $\sigma = 0.5$, respectively. The smooth curve shows the trend corresponding to $\sigma = 0$. The simulations are based on 200 subintervals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

on the asset over the next instant. The constant μ is the 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, cf. Section 6.6, 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, cf. Section 7.6.

Next, we will find an explicit expression for x_t, i.e. we will find a solution to the stochastic differential equation (3.13). 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 3.5 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

$$x_{t'} = x_t \exp\left\{\left(\mu - \frac{1}{2}\sigma^2\right)(t' - t) + \sigma(z_{t'} - z_t)\right\}. \tag{3.14}$$

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 (3.14) that the future value $x_{t'}$ (given x_t) will be lognormally distributed. The probability density function for $x_{t'}$ (given x_t) is given by

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

and the mean and variance are

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

cf. Appendix A.

The geometric Brownian motion in (3.13) 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

$$dx_t = \mu(t)x_t\, dt + \sigma(t)x_t\, dz_t, \tag{3.15}$$

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

$$x_{t'} = x_t \exp\left\{\int_t^{t'} \left(\mu(u) - \frac{1}{2}\sigma(u)^2\right) du + \int_t^{t'} \sigma(u)\, dz_u\right\}. \tag{3.16}$$

According to Theorem 3.2, $\int_t^{t'} \sigma(u)\, dz_u$ 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 rate 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

$$dx_t = [\varphi - \kappa x_t]\, dt + \beta\, dz_t, \tag{3.17}$$

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

$$dx_t = \kappa\, [\theta - x_t]\, dt + \beta\, dz_t, \tag{3.18}$$

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

$$\begin{aligned} x_{t'} &= e^{-\kappa(t'-t)} x_t + \int_t^{t'} \varphi e^{-\kappa(t'-u)}\, du + \int_t^{t'} \beta e^{-\kappa(t'-u)}\, dz_u \\ &= e^{-\kappa(t'-t)} x_t + \theta\left(1 - e^{-\kappa(t'-t)}\right) + \int_t^{t'} \beta e^{-\kappa(t'-u)}\, dz_u. \end{aligned} \tag{3.19}$$

This holds for all $t' > t \geq 0$. In particular, we get that the solution to the stochastic differential equation (3.17) can be written as

$$x_t = e^{-\kappa t} x_0 + \theta\left(1 - e^{-\kappa t}\right) + \int_0^t \beta e^{-\kappa(t-u)}\, dz_u. \tag{3.20}$$

According to Theorem 3.2, 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

$$\mathrm{E}_t[x_{t'}] = e^{-\kappa(t'-t)} x_t + \theta\left(1 - e^{-\kappa(t'-t)}\right), \tag{3.21}$$

$$\mathrm{Var}_t[x_{t'}] = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa(t'-t)}\right). \tag{3.22}$$

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 sub-figure 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 e.g. Figure 3.6. As a matter of fact, it can be shown that an Ornstein-Uhlenbeck process with probability one will sooner or later become negative.

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

$$dx_t = [\varphi(t) - \kappa x_t]\, dt + \beta(t)\, dz_t = \kappa\, [\theta(t) - x_t]\, dt + \beta(t)\, dz_t. \tag{3.23}$$

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

$$\mathrm{E}_t[x_{t'}] = e^{-\kappa(t'-t)}x_t + \int_t^{t'} \varphi(u)e^{-\kappa(t'-u)}\, du, \tag{3.24}$$

$$\mathrm{Var}_t[x_{t'}] = \int_t^{t'} \beta(u)^2 e^{-2\kappa(t'-u)}\, du. \tag{3.25}$$

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

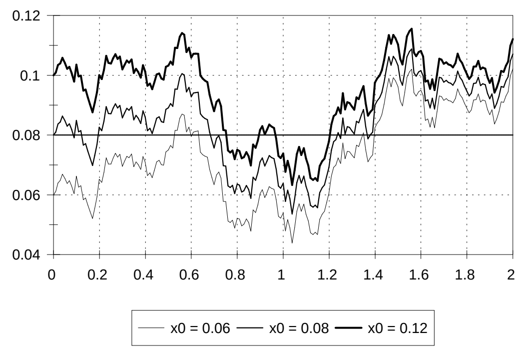
One of the earliest (but still frequently applied) dynamic models of the term structure of interest rates is based on the assumption that the short-term interest rate follows an Ornstein-Uhlenbeck process, cf. Section 7.4. In an extension of that model, the short-term interest rate is assumed to follow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

3.8.3 Square root processes

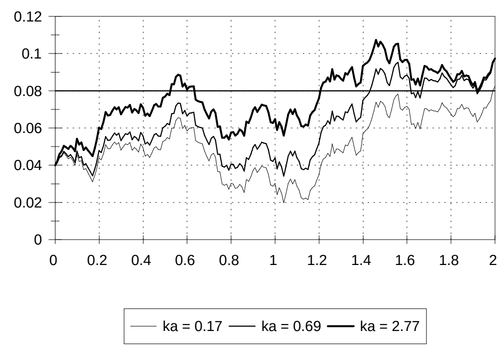
Another stochastic process frequently applied in term structure models is the so-called square root process. A one-dimensional stochastic process $x = (x_t)_{t \geq 0}$ is said to be a **square root process**, if its dynamics is of the form

$$dx_t = [\varphi - \kappa x_t]\, dt + \beta\sqrt{x_t}\, dz_t = \kappa\, [\theta - x_t]\, dt + \beta\sqrt{x_t}\, dz_t, \tag{3.26}$$

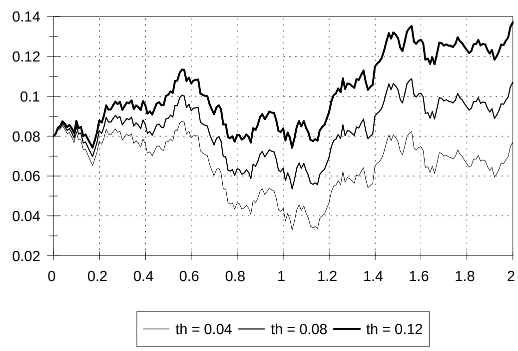
where $\varphi = \kappa\theta$. Here, φ, θ, β, and κ are positive constants. We assume that the initial value of the process x_0 is positive, so that the square root function can be applied. The only difference to the dynamics of an Ornstein-Uhlenbeck process is the term $\sqrt{x_t}$ in the volatility.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



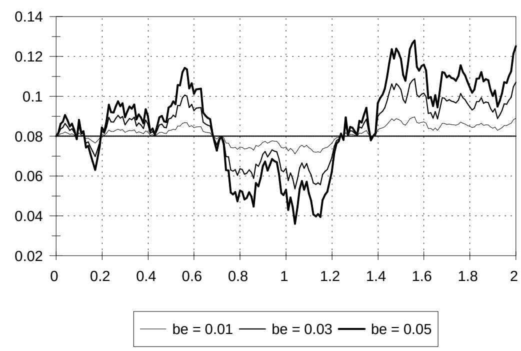
(a) Different initial values x_0



(b) Different κ-values; $x_0 = 0.04$



(c) Different θ-values



(d) Different β-values

Figure 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 $\beta = 0.03$.

is now $\beta^2 x_t$ which is proportional to the level of the process. A square root process also exhibits mean reversion.

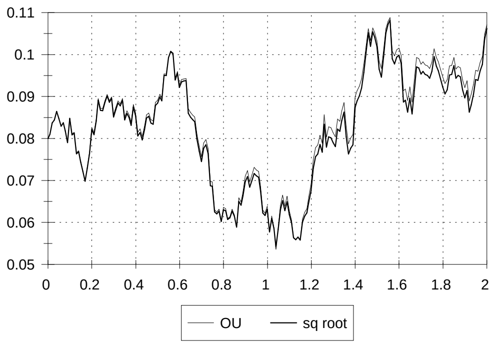
A square root process can only take on non-negative values. To see this, note that if the value should become zero, then the drift is positive and the volatility zero, and therefore the value of the process will with certainty become positive immediately after (zero is a so-called reflecting barrier). 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.[6] Hence, the value space for a square root process is either $\mathcal{S} = [0, \infty)$ or $\mathcal{S} = (0, \infty)$.

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

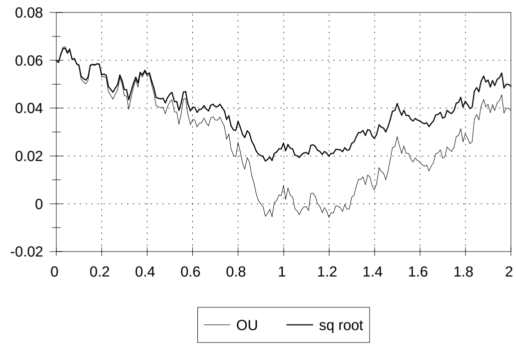
$$x_{t_i} = x_{t_{i-1}} + \kappa[\theta - x_{t_{i-1}}](t_i - t_{i-1}) + \beta\sqrt{x_{t_{i-1}}}\varepsilon_i\sqrt{t_i - t_{i-1}}.$$

Variations in the different parameters will have similar effects as for the Ornstein-Uhlenbeck process, which is illustrated in Figure 3.5. Instead, let us compare the paths for a square root process

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



(a) Initial value $x_0 = 0.08$, same random numbers as in Figure 3.5



(b) Initial value $x_0 = 0.06$, different random numbers

Figure 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.

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.

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

$$x_{t'} = x_t e^{-\kappa(t'-t)} + \frac{\varphi}{\kappa}\left(1 - e^{-\kappa(t'-t)}\right) + \beta \int_t^{t'} e^{-\kappa(t'-u)} \sqrt{x_u}\, dz_u. \tag{3.27}$$

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, cf. Theorem 3.1, we easily get

$$\mathrm{E}_t[x_{t'}] = e^{-\kappa(t'-t)} x_t + \theta\left(1 - e^{-\kappa(t'-t)}\right) = \frac{\varphi}{\kappa} + (x_t - \varphi)\, e^{-\kappa(t'-t)}. \tag{3.28}$$

To compute the variance we apply the second equation of Theorem 3.1:

$$
\begin{aligned}
\operatorname{Var}_t[x_{t'}] &= \operatorname{Var}_t\left[\beta \int_t^{t'} e^{-\kappa(t'-u)} \sqrt{x_u}\, dz_u\right] \\
&= \beta^2 \int_t^{t'} e^{-2\kappa(t'-u)}\, \mathrm{E}_t[x_u]\, du \\
&= \beta^2 \int_t^{t'} e^{-2\kappa(t'-u)} \left(\frac{\varphi}{\kappa} + (x_t - \varphi)\, e^{-\kappa(u-t)}\right) du \\
&= \frac{\beta^2 \varphi}{\kappa} \int_t^{t'} e^{-2\kappa(t'-u)}\, du + \frac{\beta^2}{\kappa}\left(x_t - \frac{\varphi}{\kappa}\right) e^{-2\kappa t' + \kappa t} \int_t^{t'} e^{\kappa u}\, du \\
&= \frac{\beta^2 \varphi}{2\kappa^2}\left(1 - e^{-2\kappa(t'-t)}\right) + \frac{\beta^2}{\kappa}\left(x_t - \frac{\varphi}{\kappa}\right)\left(e^{-\kappa(t'-t)} - e^{-2\kappa(t'-t)}\right) \\
&= \frac{\beta^2 x_t}{\kappa}\left(e^{-\kappa(t'-t)} - e^{-2\kappa(t'-t)}\right) + \frac{\beta^2 \theta}{2\kappa}\left(1 - e^{-\kappa(t'-t)}\right)^2. \qquad (3.29)
\end{aligned}
$$

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, given the value x_t, the value $x_{t'}$ with $t' > t$ is non-centrally χ^2-distributed. More precisely, the probability density function for $x_{t'}$ is

$$
f_{x_{t'}|x_t}(x) = f_{\chi^2_{a,b}}(2cx),
$$

where

$$
\begin{aligned}
c &= \frac{2\kappa}{\beta^2\left(1 - e^{-\kappa(t'-t)}\right)}, \\
b &= c x_t e^{-\kappa(t'-t)}, \\
a &= \frac{4\varphi}{\beta^2},
\end{aligned}
$$

and where $f_{\chi^2_{a,b}}(\cdot)$ denotes the probability density function for a non-centrally χ^2-distributed random variable with a degrees of freedom and non-centrality parameter b.

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, cf. 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 analyze than models based on Ornstein-Uhlenbeck processes.

3.9 Multi-dimensional processes

So far we have only considered one-dimensional processes, i.e. processes with a value space which is $\mathbb{R}$ or a subset of $\mathbb{R}$. Some models in the following chapters will involve multi-dimensional processes, which have values in (a subset of) $\mathbb{R}^K$ for some integer $K > 1$. A multi-dimensional process can also be considered as a vector of one-dimensional processes. In this section we will briefly introduce some multi-dimensional processes and a multi-dimensional version of Itô's Lemma. A note on the notation: vectors are printed in boldface. Matrices are indicated by a double line under the symbol. We treat all vectors as column vectors. The symbol $^\top$ denotes transposition, so that e.g. $(a,b)^\top$ represents the (column) vector $\begin{pmatrix} a \\ b \end{pmatrix}$. If $\boldsymbol{a}$ is a (column) vector, then $\boldsymbol{a}^\top$ is the corresponding row vector.

A **K-dimensional (standard) Brownian motion** $\boldsymbol{z} = (z_1, \dots, z_K)^\top$ is a stochastic process where the individual components z_i are mutually independent one-dimensional (standard) Brownian motions. If we let $\boldsymbol{0} = (0, \dots, 0)^\top$ 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) $\boldsymbol{z}_0 = \boldsymbol{0}$,

(ii) for all $t, t' \geq 0$ with $t < t'$: $\boldsymbol{z}_{t'} - \boldsymbol{z}_t \sim \boldsymbol{N}(\boldsymbol{0}, (t'-t)\underline{\underline{I}})$ [normally distributed increments],

(iii) for all $0 \leq t_0 < t_1 < \cdots < t_n$, 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 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 $\underline{\underline{b}}$. As for standard Brownian motions, we can also define multi-dimensional generalized Brownian motions, which simply are vectors of independent one-dimensional generalized Brownian motions.

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

$$d\boldsymbol{x}_t = \boldsymbol{\mu}(\boldsymbol{x}_t, t)\, dt + \underline{\underline{\sigma}}(\boldsymbol{x}_t, t)\, d\boldsymbol{z}_t, \tag{3.30}$$

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

$$\begin{aligned} dx_{it} &= \mu_i(\boldsymbol{x}_t, t)\, dt + \boldsymbol{\sigma}_i(x_t, t)^\top\, d\boldsymbol{z}_t \\ &= \mu_i(\boldsymbol{x}_t, t)\, dt + \sum_{k=1}^{K} \sigma_{ik}(\boldsymbol{x}_t, t)\, dz_{kt}, \quad i = 1, \dots, K, \end{aligned} \tag{3.31}$$

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 (i.e. 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

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

The volatility of the i'th component is given by $\|\boldsymbol{\sigma}_i(\boldsymbol{x}_t,t)\|$. It is clear from these computations that the elements of the matrix $\underline{\underline{\sigma}}(\boldsymbol{x}_t,t)$ determine all variances and covariances over infinitesimal periods. To be precise, the variance-covariance matrix is $\underline{\underline{\Sigma}}(\boldsymbol{x}_t,t)\,dt = \underline{\underline{\sigma}}(\boldsymbol{x}_t,t)\underline{\underline{\sigma}}(\boldsymbol{x}_t,t)\,dt$. Note that the individual component processes are generally not mutually independent since the drift and volatility of one component will generally depend on the values of the other components. Also note that means, variances, and covariances over non-infinitesimal intervals such as $[t,t']$ will generally depend on the entire path of the process over this interval.

Finally, 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 3.6 *Let $\boldsymbol{x} = (\boldsymbol{x}_t)_{t\geq 0}$ be an Itô process in $\mathbb{R}^K$ with dynamics*

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

which componentwise is equivalent to

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

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

$$\begin{aligned}dy_t = &\left(\frac{\partial g}{\partial t}(\boldsymbol{x}_t,t) + \sum_{i=1}^K \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t,t)\mu_{it} + \frac{1}{2}\sum_{i=1}^K\sum_{j=1}^K \frac{\partial^2 g}{\partial x_i \partial x_j}(\boldsymbol{x}_t,t)\gamma_{ijt}\right)dt\\ &+ \sum_{i=1}^K \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t,t)\sigma_{i1t}\,dz_{1t} + \dots + \sum_{i=1}^K \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t,t)\sigma_{iKt}\,dz_{Kt},\end{aligned} \tag{3.32}$$

where we have introduced the notation

$$\gamma_{ij} = \sigma_{i1}\sigma_{j1} + \cdots + \sigma_{iK}\sigma_{jK},$$

which is exactly the covariance between the processes x_i and x_j.

The result can also be written as

$$dy_t = \frac{\partial g}{\partial t}(\boldsymbol{x}_t, t)\, dt + \sum_{i=1}^{K} \frac{\partial g}{\partial x_i}(\boldsymbol{x}_t, t)\, dx_{it} + \frac{1}{2}\sum_{i=1}^{K}\sum_{j=1}^{K} \frac{\partial^2 g}{\partial x_i \partial x_j}(\boldsymbol{x}_t, t)(dx_{it})(dx_{jt}), \tag{3.33}$$

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.

Furthermore, the result can be expressed using vector and matrix notation:

$$dy_t = \left(\frac{\partial g}{\partial t}(\boldsymbol{x}_t, t) + \left(\frac{\partial g}{\partial x}(\boldsymbol{x}_t, t)\right)^{\top} \boldsymbol{\mu}_t + \frac{1}{2}\operatorname{tr}\left(\underline{\underline{\sigma}}_t \underline{\underline{\sigma}}_t^{\top} \left[\frac{\partial^2 g}{\partial x^2}(\boldsymbol{x}_t, t)\right]\right)\right) dt + \left(\frac{\partial g}{\partial x}(\boldsymbol{x}_t, t)\right)^{\top} \underline{\underline{\sigma}}_t\, d\boldsymbol{z}_t, \tag{3.34}$$

where

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

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

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 $\underline{\underline{\hat{\sigma}}}$, i.e. a matrix with $\hat{\sigma}_{ik} = 0$ for $k > i$. This is the so-called Cholesky decomposition. Hence, we may write the dynamics as

$$\begin{aligned} dx_{1t} &= \mu_1(\boldsymbol{x}_t, t)\, dt + \sigma_{11}(\boldsymbol{x}_t, t)\, dz_{1t} \\ dx_{2t} &= \mu_2(\boldsymbol{x}_t, t)\, dt + \sigma_{21}(\boldsymbol{x}_t, t)\, dz_{1t} + \sigma_{22}(\boldsymbol{x}_t, t)\, dz_{2t} \\ &\vdots \\ dx_{Kt} &= \mu_K(\boldsymbol{x}_t, t)\, dt + \sigma_{K1}(\boldsymbol{x}_t, t)\, dz_{1t} + \sigma_{K2}(\boldsymbol{x}_t, t)\, dz_{2t} + \cdots + \sigma_{KK}(\boldsymbol{x}_t, t)\, dz_{Kt} \end{aligned} \tag{3.35}$$

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 it's coefficient σ_{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 σ_{21} is fixed to match the covariance between changes to x_1 and x_2 and then σ_{22} can be chosen so that $\sqrt{\sigma_{21}^2 + \sigma_{22}^2}$ 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

$$\begin{aligned}
dx_{1t} &= \mu_1(\boldsymbol{x}_t, t)\, dt + V_1(\boldsymbol{x}_t, t)\, d\hat{z}_{1t} \\
dx_{2t} &= \mu_2(\boldsymbol{x}_t, t)\, dt + V_2(\boldsymbol{x}_t, t)\, d\hat{z}_{2t} \\
&\vdots \\
dx_{Kt} &= \mu_K(\boldsymbol{x}_t, t)\, dt + V_K(\boldsymbol{x}_t, t)\, d\hat{z}_{Kt}
\end{aligned} \tag{3.36}$$

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 (3.36) and (3.35) are meant to represent the same dynamics, we must have

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

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

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 (3.26) 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 that the world was different from the world we live in, e.g. if investors were risk-neutral instead of risk-averse. A different world is represented mathematically by a different probability measure. Hence, we need to be able to analyze stochastic variables and processes under different probability measures. In this section we will briefly discuss how we can change the probability measure.

If the state space Ω has only finitely many elements, we can write it as $\Omega = \{\omega_1, \dots, \omega_n\}$. As before, the set of events, i.e. 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 ξ 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, i.e. $x = (x_t)_{t \in [0,T]}$. 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

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

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

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

Here the first and the third equalities follow from the definition of ξ. The second equality follows from the *law of iterated expectations*, which says that the expectation today of what we expect tomorrow for a given random variable realized later is equal to today's expectation of that random variable. This is a very intuitive result. For a more formal statement and proof, see Øksendal (1998). The following result turns out to be very useful in our dynamic models of he economy. Let $x = (x_t)_{t\in[0,T]}$ be any stochastic process. Then we have

$$\mathrm{E}_t^{\mathbb{Q}}\left[x_{t'}\right] = \mathrm{E}_t^{\mathbb{P}}\left[\frac{\xi_{t'}}{\xi_t} x_{t'}\right]. \tag{3.37}$$

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.[7]. 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

$$\xi_t^\lambda = \exp\left\{-\int_0^t \lambda_s\, dz_s - \frac{1}{2}\int_0^t \lambda_s^2\, ds\right\}. \tag{3.38}$$

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 3.5) 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

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

From (3.37), we get that

$$\mathrm{E}_t^{\mathbb{Q}^\lambda}\left[x_{t'}\right] = \mathrm{E}_t^{\mathbb{P}}\left[\frac{\xi_{t'}^\lambda}{\xi_t^\lambda} x_{t'}\right] = \mathrm{E}_t^{\mathbb{P}}\left[x_{t'} \exp\left\{-\int_t^{t'} \lambda_s\, dz_s - \frac{1}{2}\int_t^{t'} \lambda_s^2\, ds\right\}\right] \tag{3.39}$$

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

Theorem 3.7 (Girsanov) *The process $z^\lambda = (z_t^\lambda)_{t\in[0,T]}$ defined by*

$$z_t^\lambda = z_t + \int_0^t \lambda_s\, ds, \quad 0 \le t \le T, \tag{3.40}$$

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

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

[7] Basically, λ must be 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, i.e. the expectation $\mathrm{E}^{\mathbb{P}}\left[\exp\left\{\frac{1}{2}\int_0^T \lambda_t^2\, dt\right\}\right]$ is finite.

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

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

where $\hat{\varphi} = \varphi - \sigma\lambda$. Consequently, the future values of x are normally distributed both under $\mathbb{P}$ and $\mathbb{Q}^\lambda$. From (3.21) and (3.22), 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 $\theta = \varphi/\kappa$):

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

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 (Adapted from Björk (1998).) 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 3.3 (Adapted from Björk (1998).) 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 3.4 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 is 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, i.e. $\mu_{it} = \mu_i x_{it}$, $\sigma_{it} = \sigma_i x_{it}$, and $\rho_t = \rho$ with μ_i, σ_i, and ρ being constants.

EXERCISE 3.5 Find the dynamics of the process ξ^λ defined in (3.38).

Chapter 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 through time and states. Hence, the pricing of fixed income securities follow 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 rate 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 (i.e. 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 riskless profit, i.e. 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 that 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.

In Section 4.4, 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 prices of risk). We introduce in Section 4.5 the concept of a representative agent and show that in a complete market, we may assume that the economy is inhabited by a single 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 4.6 we show how to generalize the results to the more realistic case with dividends at other points in time.

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, 1970), Debreu (1953, 1954, 1959), Negishi (1960), and Ross (1978). Textbook presentations of discrete-time asset pricing theory can be found in, e.g., Ingersoll (1987), Huang and Litzenberger (1988), Cochrane (2001), LeRoy and Werner (2001), and Duffie (2001, Chs. 1–4). As already discussed in Section 3.2.1 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. The presentation below is structured very similarly to that of Duffie (2001, Ch. 6), but many technical details are left out here. These details are not important for a basic understanding of continuous-time asset pricing models. Other textbook presentations with more details and proofs include Dothan (1990) and Karatzas and Shreve (1998). 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).

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 care 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, $\boldsymbol{z} = (\boldsymbol{z}_t)_{t\in[0,T]}$. Think of $d\boldsymbol{z}_t$ as d exogenous shocks to the economy at time t. All the uncertainty that affects the investors stems from these exogenous shocks. This includes *financial uncertainty*, i.e. uncertainty about the evolution of prices and interest rates, future expected returns, volatilities, and correlations, but also *non-financial uncertainty*, e.g. uncertainty about prices of consumption goods and uncertainty about future labor income of the agents. The state space Ω is in this case the set of all paths of the Brownian motion $\boldsymbol{z}$. Note that since a Brownian motion has infinitely many possible paths, we have an infinite state space.

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 one instantaneously riskless and N risky assets. Let us first describe the instantaneously riskless asset. Let r_t denote the continuously compounded, instantaneously riskless interest rate at time t, i.e. the rate of return over an infinitesimal interval $[t, t+dt]$ is $r_t\,dt$. The instantaneously riskless asset is a continuous roll-over of such instantaneously riskless 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* riskless 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 by the intercept of a yield curve estimate, which can be obtained by the methods discussed in Section 1.6 on page 13. 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 riskless (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. Money market rates are therefore expected to be higher than similar bond market rates.

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

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

Here $\mu_i = (\mu_{it})$ denotes the (relative) drift, and $\sigma_{ij} = (\sigma_{ijt})$ reflects the sensitivity of the relative 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. If we let $\boldsymbol{\sigma}_{it}$ be the vector $(\sigma_{i1t}, \dots, \sigma_{idt})^\top$, the price dynamics of asset i can be rewritten as

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

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 in 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", e.g. 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 returns 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, i.e. $\|\boldsymbol{\sigma}_{it}\| = \left(\sum_{j=1}^d \sigma_{ijt}^2\right)^{1/2}$.

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 riskless 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 the trading strategy over an 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 terms of mathematics, 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,

$$dV_t^{\alpha,\boldsymbol{\theta}} = \alpha_t r_t e^{\int_0^t r_u\, du}\, dt + \boldsymbol{\theta}_t^\top d\boldsymbol{P}_t. \tag{4.1}$$

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 so that there are k assets left. 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:

(i) $V_0^{\alpha,\boldsymbol{\theta}} < 0$ and $V_T^{\alpha,\boldsymbol{\theta}} \geq 0$ with probability one,

(ii) $V_0^{\alpha,\boldsymbol{\theta}} \leq 0$, $V_T^{\alpha,\boldsymbol{\theta}} \geq 0$ with probability one, and $V_T^{\alpha,\boldsymbol{\theta}} > 0$ with strictly positive probability.

A trading strategy $(\alpha, \boldsymbol{\theta})$ satisfying (i) 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 (ii) 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 replicating trading strategy. If the

[1] Two vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are called linearly independent if $k_1\boldsymbol{a} + k_2\boldsymbol{b} = \boldsymbol{0}$ implies $k_1 = k_2 = 0$, i.e. $\boldsymbol{a}$ 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*.

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, cf. Black and Scholes (1973) and Merton (1973).

Although the definition of arbitrage focuses on payoffs at time T, it does cover shorter term riskless gains. Suppose for example that we can construct a trading strategy with a non-positive initial value (i.e. 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 $n = 1, 2, \ldots$. 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 looses $\alpha 2^{n-1}$ 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, i.e. 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

In stead 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 mechanisms 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 markets price 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 strictly positive process $\zeta = (\zeta_t)$ with $\zeta_0 = 1$ and the property that the product of the state-price deflator and the price of an asset is a martingale, i.e. $(\zeta_t P_{it})$ is a martingale for any $i = 1, \ldots, N$ and $(\zeta_t \exp\{\int_0^t r_u \, du\})$ is a martingale. In particular, for all $t < t' \leq T$, we have

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

or

$$P_{it} = \mathrm{E}_t \left[\frac{\zeta_{t'}}{\zeta_t} P_{it'} \right]. \tag{4.2}$$

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, if a zero-coupon bond maturing at time T is traded, its time t price must be

$$B_t^T = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} \right]. \tag{4.3}$$

Due to the linearity of the value of a trading strategy, the pricing relation (4.2) also holds for any self-financing trading strategy:

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

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

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

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 $m_t = -r_t$, i.e. the relative drift of a state-price deflator is equal to the negative of the short-term interest rate. We will later give a characterization of the sensitivity vector $\boldsymbol{v}$.

Given a state-price deflator we can price any asset. But can we be sure that a state-price deflator exist? It turns out that the existence of a state-price deflator is basically equivalent to 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 super-martingales 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, i.e. 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

$$\mathrm{E}\left[\int_0^{\Delta t} e^{-\delta s}\left(u\left(c_s - \frac{\varepsilon P_{i0}}{\Delta t}\right) - u(c_s)\right) ds + \int_t^{t+\Delta t} e^{-\delta s}\left(u\left(c_s + \frac{\varepsilon P_{it}}{\Delta t}\right) - u(c_s)\right) ds\right] \leq 0.$$

Dividing by ε and letting $\varepsilon \to 0$, we obtain

$$\mathrm{E}\left[-\frac{P_{i0}}{\Delta t}\int_0^{\Delta t} e^{-\delta s} u'(c_s)\, ds + \frac{P_{it}}{\Delta t}\int_t^{t+\Delta t} e^{-\delta s} u'(c_s)\, ds\right] \leq 0.$$

Letting $\Delta t \to 0$, we arrive at

$$\mathrm{E}\left[-P_{i0} u'(c_0) + P_{it} e^{-\delta t} u'(c_t)\right] \leq 0,$$

or, equivalently,

$$P_{i0} u'(c_0) \geq \mathrm{E}\left[e^{-\delta t} P_{it} u'(c_t)\right].$$

The reverse inequality can be shown similarly by considering the "opposite" perturbation, i.e. 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

$$P_{it} = \mathrm{E}_t\left[e^{-\delta(t'-t)} \frac{u'(c_{t'})}{u'(c_t)} P_{it'}\right], \quad t \leq t' \leq T. \tag{4.4}$$

With intermediate dividends this relation is slightly different, cf. Section 4.6.

Comparing (4.2) and (4.4), we see that $\zeta_t = e^{-\delta t} u'(c_t)/u'(c_0)$ 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 lead to 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, cf. the discussion in Section 4.4.

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:

(i) $\mathbb{Q}$ is equivalent to $\mathbb{P}$,

(ii) 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,

(iii) the Radon-Nikodym derivative $d\mathbb{Q}/d\mathbb{P}$ has finite variance.

In particular, if $\mathbb{Q}$ is a risk-neutral probability measure, then

$$P_{it} = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t'} r_s \, ds} P_{it'} \right] \tag{4.5}$$

for any $t < t' \leq T$. Under some technical conditions on $\boldsymbol{\theta}$, see Duffie (2001, p. 109), the same relation holds for any self-financing trading strategy $(\alpha, \boldsymbol{\theta})$, i.e.

$$V_t^{\alpha,\boldsymbol{\theta}} = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^{t'} r_s \, ds} V_{t'}^{\alpha,\boldsymbol{\theta}} \right]. \tag{4.6}$$

These relations show that the risk-neutral probability measure (together with the short-term interest rate process) captures the market-wide pricing information. The price of a particular asset follows from the risk-neutral probability measure and the asset-specific payoff.

The existence of a risk-neutral probability measure is closely related to absence of arbitrage:

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 (4.6) 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 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 basically 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* $t \le T$, *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 (3.37). 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 (3.37), 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. □

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, following 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 dynamics of the discounted prices is

$$d\bar{\boldsymbol{P}}_t = \operatorname{diag}(\bar{\boldsymbol{P}}_t) \left[(\boldsymbol{\mu}_t - r_t \mathbf{1}) \, dt + \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t \right]. \tag{4.7}$$

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 $\xi_t = \mathrm{E}_t[d\mathbb{Q}/d\mathbb{P}]$. This is martingale since, for any $t < t'$, we have $\mathrm{E}_t[\xi_{t'}] = \mathrm{E}_t[\mathrm{E}_{t'}[d\mathbb{Q}/d\mathbb{P}]] = \mathrm{E}_t[d\mathbb{Q}/d\mathbb{P}] = \xi_t$ due to the law of iterated expectations (see the discussion in Section 3.10). Then it follows from the *Martingale Representation Theorem*, see Theorem 3.3, that a d-dimensional process $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$ exists such that

$$d\xi_t = -\xi_t \boldsymbol{\lambda}_t^{\top} \, d\boldsymbol{z}_t,$$

or, equivalently (using $\xi_0 = \mathrm{E}[d\mathbb{Q}/d\mathbb{P}] = 1$),

$$\xi_t = \exp\left\{ -\frac{1}{2} \int_0^t \|\boldsymbol{\lambda}_s\|^2 \, ds - \int_0^t \boldsymbol{\lambda}_s^{\top} \, d\boldsymbol{z}_s \right\}. \tag{4.8}$$

According to *Girsanov's Theorem*, i.e. Theorem 3.7, the process $\boldsymbol{z}^{\mathbb{Q}} = (\boldsymbol{z}_t^{\mathbb{Q}})$ defined by

$$d\boldsymbol{z}_t^{\mathbb{Q}} = d\boldsymbol{z}_t + \boldsymbol{\lambda}_t \, dt, \quad \boldsymbol{z}_0^{\mathbb{Q}} = 0, \tag{4.9}$$

is then 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 (4.7), we obtain

$$d\bar{\boldsymbol{P}}_t = \operatorname{diag}(\bar{\boldsymbol{P}}_t) \left[\left(\boldsymbol{\mu}_t - r_t \mathbf{1} - \underline{\underline{\sigma}}_t \boldsymbol{\lambda}_t \right) dt + \underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t^{\mathbb{Q}} \right]. \tag{4.10}$$

If discounted prices are to be $\mathbb{Q}$-martingales, the drift must be zero, so we must have that

$$\underline{\underline{\sigma}}_t \boldsymbol{\lambda}_t = \boldsymbol{\mu}_t - r_t \mathbf{1}. \tag{4.11}$$

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, $\lambda_1, \dots, \lambda_d$ (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 (4.8). The technical 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.) We summarize these findings as follows:

Theorem 4.5 *If a risk-neutral probability measure exists, there must be a solution to* (4.11) *for all* t*. If a solution* $\boldsymbol{\lambda}_t$ *exists for all* t *and the process* $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$ *satisfies technical conditions, then a risk-neutral probability measure exists.*

Any process $\boldsymbol{\lambda} = (\boldsymbol{\lambda}_t)$ solving (4.11) is called a **market price of risk process**. To understand this terminology, note that the i'th equation in the system (4.11) 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.

According to the theorem above, we basically have a one-to-one relation between risk-neutral probability measures and market prices of risk. Combining this 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. Then

$$d\bar{\boldsymbol{P}}_t = \operatorname{diag}(\bar{\boldsymbol{P}}_t)\underline{\underline{\sigma}}_t \, d\boldsymbol{z}_t^{\mathbb{Q}} \tag{4.12}$$

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 (4.8) 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

$$d\zeta_t = -\zeta_t\left[r_t\, dt + \boldsymbol{\lambda}_t^{\top}\, d\boldsymbol{z}_t\right]. \tag{4.13}$$

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. 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 (4.11), 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 (4.11). We can write one solution as

$$\boldsymbol{\lambda}_t^* = \underline{\underline{\hat{\sigma}}}_t^\top \left(\underline{\underline{\hat{\sigma}}}_t \underline{\underline{\hat{\sigma}}}_t^\top \right)^{-1} (\hat{\boldsymbol{\mu}}_t - r_t \mathbf{1}), \tag{4.14}$$

where $\underline{\underline{\hat{\sigma}}}_t$ and $\hat{\boldsymbol{\mu}}_t$ were defined in Section 4.2.3. In the special case where $k = d$, we have the unique solution

$$\boldsymbol{\lambda}_t^* = \underline{\underline{\hat{\sigma}}}_t^{-1} (\hat{\boldsymbol{\mu}}_t - r_t \mathbf{1}).$$

4.4 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 $\boldsymbol{z}$ 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, i.e.

$$\mathcal{M} = \left\{ V_T^{\alpha,\boldsymbol{\theta}} \mid (\alpha, \boldsymbol{\theta}) \text{ self-financing with } V_t^{\alpha,\boldsymbol{\theta}} \text{ bounded from below for all } t \in [0, T] \right\}.$$

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, i.e. 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).*

Clearly, a necessary (but not sufficient) condition for the market to be complete is that at least d risky asset are traded — if $N < d$, the matrix $\underline{\underline{\sigma}}_t$ cannot have rank d. If $\underline{\underline{\sigma}}_t$ has rank d, then there is exactly one solution to the system of equations (4.11) 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 (4.11) 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.

Real financial markets are probably not complete in a broad sense, since most investors face restrictions on the trading strategies they can invest in, e.g. short-selling and portfolio mix restrictions, and are exposed to risks that cannot be fully hedged by any financial investments, e.g. labor 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 $\boldsymbol{z}$ into $(\boldsymbol{Z}, \hat{\boldsymbol{Z}})$, where $\boldsymbol{Z}$ is k-dimensional and $\hat{\boldsymbol{Z}}$ 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 $\hat{\boldsymbol{Z}}$, representing non-hedgeable risk in interest rates, expected returns, and volatilities and correlations, respectively. Or other variables important for the investor, e.g. his labor income, may be sensitive to $\hat{\boldsymbol{Z}}$. 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

$$\Lambda_t = \left(\underline{\underline{\sigma}}_t\right)^{-1} (\boldsymbol{\mu}_t - r_t \mathbf{1}),$$

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.

4.5 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

(i) given the asset prices, each agent has chosen an optimal trading strategy according to her preferences and endowments,

(ii) markets clear, i.e. total demand equal 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

$$\zeta^l_t = e^{-\delta^l t} u_l'(c^l_t)/u_l'(c^l_0),$$

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 good 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). 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.

4.6 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, where as 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

$$B_t^{t'} = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'} r_u\, du}\right]. \tag{4.15}$$

In terms of a state-price deflator ζ, we have

$$B_t^{t'} = \mathrm{E}_t\left[\frac{\zeta_{t'}}{\zeta_t}\right]. \tag{4.16}$$

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

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{t'} r_u\,du} P_{t'} + \int_t^{t'} e^{-\int_t^s r_u\,du} D_s\,ds\right] \tag{4.17}$$

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

$$P_t = \mathrm{E}_t\left[\frac{\zeta_{t'}}{\zeta_t} P_{t'} + \int_t^{t'} \frac{\zeta_s}{\zeta_t} D_s\,ds\right].$$

The link between state-price deflators and the marginal rate of substitution of an agent is still the same: $\zeta_t = e^{-\delta t} u'(c_t)/u'(c_0)$ is valid state-price deflator.

Pricing expressions for assets that have both continuous and lump-sum dividends can be obtained by combining the expressions above appropriately.

4.7 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 in 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.
- 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.

4.8 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 (4.7).

Chapter 5

The economics of the term structure of interest rates

5.1 Introduction

A bond is nothing but 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 presence 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 concerned with the number of units of goods they consume and not the dollar value of these goods, the relations found in the first sections of this chapter apply to real interest rates. However, most traded bonds are nominal, i.e. 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 loose hypotheses. The most well-known of these is the expectation hypothesis, which postulates a close relation between current interest rates or bond returns and expected future interest rates or bond returns. Many economists still seem to rely on the validity of this hypothesis, and a lot of man power has been spend 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 hypothesis 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 maximizing an expected time-additive utility function, $\mathrm{E}[\int_0^T e^{-\delta t} u(C_t)\, dt]$. As discussed in Section 4.5, 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

$$\zeta_t = e^{-\delta t} \frac{u'(C_t)}{u'(C_0)}.$$

Assume that the aggregate consumption process $C = (C_t)$ has dynamics of the form

$$dC_t = C_t \left[\mu_{Ct}\, dt + \boldsymbol{\sigma}_{Ct}^{\top}\, d\boldsymbol{z}_t \right],$$

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 applied to the function $g(C,t) = e^{-\delta t} u'(C)/u'(C_0)$. Since the relevant derivatives are

$$\frac{\partial g}{\partial t} = -\delta g(C,t), \quad \frac{\partial g}{\partial C} = e^{-\delta t} \frac{u''(C)}{u'(C_0)} = \frac{u''(C)}{u'(C)} g(C,t), \quad \frac{\partial^2 g}{\partial C^2} = e^{-\delta t} \frac{u'''(C)}{u'(C_0)} = \frac{u'''(C)}{u'(C)} g(C,t),$$

the dynamics of $\zeta = (\zeta_t)$ is

$$d\zeta_t = \zeta_t \left[\left(-\delta - \left(\frac{-C_t u''(C_t)}{u'(C_t)} \right) \mu_{Ct} + \frac{1}{2} C_t^2 \frac{u'''(C_t)}{u'(C_t)} \|\boldsymbol{\sigma}_{Ct}\|^2 \right) dt - \left(\frac{-C_t u''(C_t)}{u'(C_t)} \right) \boldsymbol{\sigma}_{Ct}^{\top}\, d\boldsymbol{z}_t \right]. \quad (5.1)$$

Recalling from Section 4.3.1 that the equilibrium short-term interest rate equals minus the relative drift of the state-price deflator, we can write the short rate as

$$r_t = \delta + \frac{-C_t u''(C_t)}{u'(C_t)} \mu_{Ct} - \frac{1}{2} C_t^2 \frac{u'''(C_t)}{u'(C_t)} \|\boldsymbol{\sigma}_{Ct}\|^2. \quad (5.2)$$

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 behind this is that when the agents of the economy are impatient and has 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 (5.2) 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 will the agents appreciate the sure payments from the riskless asset and hence the lower a return is 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 (5.2) simplifies to

$$r_t = \delta + \gamma\mu_{Ct} - \frac{1}{2}\gamma(1+\gamma)\|\boldsymbol{\sigma}_{Ct}\|^2. \tag{5.3}$$

In particular, we see that if the drift and variance rates of aggregate consumption are constant, i.e. 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).

We can also characterize the equilibrium term structure of interest rates in terms of the expectations and uncertainty about future aggregate consumption.[1] The equilibrium time t price of a zero-coupon bond paying one consumption unit at time $T \geq t$ is given by

$$B_t^T = \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\right] = e^{-\delta(T-t)}\frac{\mathrm{E}_t\left[u'(C_T)\right]}{u'(C_t)}, \tag{5.4}$$

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 (5.4) becomes

$$\begin{aligned} e^{-\delta(T-t)}\frac{\mathrm{E}_t\left[u'(C_T)\right]}{u'(C_t)} &\approx e^{-\delta(T-t)}\left(1 + \frac{u''(C_t)}{u'(C_t)}\mathrm{E}_t[C_T - C_t] + \frac{1}{2}\frac{u'''(C_t)}{u'(C_t)}\mathrm{Var}_t[C_T - C_t]\right) \\ &\approx 1 - \delta(T-t) + e^{-\delta(T-t)}\frac{C_t u''(C_t)}{u'(C_t)}\mathrm{E}_t\left[\frac{C_T}{C_t} - 1\right] \\ &\quad + \frac{1}{2}e^{-\delta(T-t)}C_t^2\frac{u'''(C_t)}{u'(C_t)}\mathrm{Var}_t\left[\frac{C_T}{C_t}\right], \end{aligned}$$

[1] The presentation is adapted from Breeden (1986).

where $\mathrm{Var}_t[\,\cdot\,]$ denotes the variance conditional on the information available at time t, and we have used the approximation $e^{-\delta(T-t)} \approx 1 - \delta(T-t)$. Substituting the approximations of both sides into (5.4) and rearranging, we find the following approximate expression for the zero-coupon yield:

$$y_t^T \approx \delta + e^{-\delta(T-t)} \left(\frac{-C_t u''(C_t)}{u'(C_t)} \right) \frac{\mathrm{E}_t\left[C_T/C_t - 1\right]}{T-t} - \frac{1}{2} e^{-\delta(T-t)} C_t^2 \frac{u'''(C_t)}{u'(C_t)} \frac{\mathrm{Var}_t\left[C_T/C_t\right]}{T-t}. \quad (5.5)$$

Again assuming $u' > 0$, $u'' < 0$, and $u''' > 0$, we can state the following conclusions. The equilibrium 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 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.

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, Ingersoll, and Ross (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

$$\frac{d\eta_t}{\eta_t} = g(x_t)\, dt + \xi(x_t)\, dz_{1t}, \quad (5.6)$$

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 be 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 passes 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

$$dx_t = m(x_t)\, dt + v_1(x_t)\, dz_{1t} + v_2(x_t)\, dz_{2t}, \quad (5.7)$$

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

$$\frac{dP_t}{P_t} = \mu_t\, dt + \sigma_{1t}\, dz_{1t} + \sigma_{2t}\, dz_{2t}. \tag{5.8}$$

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 riskless 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 modeled 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 riskless asset, his wealth W_t will evolve as

$$\begin{aligned} dW_t = &\left\{ r_t W_t + W_t \alpha_t \left(g(x_t) - r_t \right) + W_t \pi_t \left(\mu_t - r_t \right) - c_t \right\} dt \\ &+ W_t \alpha_t \xi(x_t)\, dz_{1t} + W_t \pi_t \sigma_{1t}\, dz_{1t} + W_t \pi_t \sigma_{2t}\, dz_{2t}. \end{aligned} \tag{5.9}$$

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

$$J(W, x, t) = \sup_{(\alpha_s, \pi_s, c_s)_{s \in [t,T]}} \mathrm{E}_t \left[\int_t^T e^{-\delta(s-t)} u(c_s)\, ds \right],$$

i.e. 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

$$\alpha^* = \frac{-J_W}{W J_{WW}} \left[(g - r) \frac{\sigma_1^2 + \sigma_2^2}{\xi^2 \sigma_2^2} - (\mu - r) \frac{\sigma_1}{\xi \sigma_2^2} \right] + \frac{-J_{Wx}}{W J_{WW}} \frac{\sigma_2 v_1 - \sigma_1 v_2}{\xi \sigma_2}, \tag{5.10}$$

$$\pi^* = \frac{-J_W}{W J_{WW}} \left[-\frac{\sigma_1}{\xi \sigma_2^2} (g - r) + \frac{1}{\sigma_2^2} (\mu - r) \right] + \frac{-J_{Wx}}{W J_{WW}} \frac{v_2}{\sigma_2}. \tag{5.11}$$

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 lend 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, i.e. 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 choice.

In an equilibrium, we have $\pi^* = 0$ for the representative agent, and hence (5.11) implies that

$$\mu - r = \frac{\sigma_1}{\xi}(g - r) - \frac{\left(\frac{-J_{Wx}}{WJ_{WW}}\right)}{\left(\frac{-J_W}{WJ_{WW}}\right)}\sigma_2 v_2. \tag{5.12}$$

Substituting this into the expression for α^* and using the fact that $\alpha^* = 1$ in equilibrium, we get that

$$\begin{aligned} 1 &= \left(\frac{-J_W}{WJ_{WW}}\right)\left[(g-r)\frac{\sigma_1^2+\sigma_2^2}{\xi^2\sigma_2^2} - \frac{\sigma_1}{\xi}\frac{\sigma_1}{\xi\sigma_2^2}(g-r) + \frac{\left(\frac{-J_{Wx}}{WJ_{WW}}\right)}{\left(\frac{-J_W}{WJ_{WW}}\right)}\sigma_2 v_2 \frac{\sigma_1}{\xi\sigma_2^2}\right] \\ &\quad + \left(\frac{-J_{Wx}}{WJ_{WW}}\right)\frac{\sigma_2 v_1 - \sigma_1 v_2}{\xi\sigma_2} \\ &= \left(\frac{-J_W}{WJ_{WW}}\right)\frac{g-r}{\xi^2} + \left(\frac{-J_{Wx}}{WJ_{WW}}\right)\frac{v_1}{\xi}. \end{aligned}$$

Consequently, the equilibrium short-term interest rate can be written as

$$r = g - \left(\frac{-WJ_{WW}}{J_W}\right)\xi^2 + \frac{J_{Wx}}{J_W}\xi v_1. \tag{5.13}$$

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 riskless 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 riskless investment is relatively more attractive, so to establish market clearing the interest rate has to decrease.

- The last term in (5.13) 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 ξv_1. Suppose that high values of the state variable represents good states of the economy, where the wealth of the agent is high. Then the marginal utility J_W will be decreasing in x, i.e. $J_{Wx} < 0$. If the state variable and the growth rate of the economy are positively correlated, we see from (5.10) 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 (5.13).

We see from (5.12) that the market prices of risk are given by

$$\lambda_1 = \frac{g - r}{\xi}, \quad \lambda_2 = -\frac{\left(\frac{-J_{Wx}}{WJ_{WW}}\right)}{\left(\frac{-J_W}{WJ_{WW}}\right)}v_2 = -\frac{J_{Wx}}{J_W}v_2. \tag{5.14}$$

Applying the relation

$$g - r = \left(\frac{-W J_{WW}}{J_W} \right) \xi^2 - \frac{J_{Wx}}{J_W} \xi v_1,$$

we can rewrite λ_1 as

$$\lambda_1 \equiv \left(\frac{-W J_{WW}}{J_W} \right) \xi - \frac{J_{Wx}}{J_W} v_1. \tag{5.15}$$

5.4 Equilibrium term structure models

5.4.1 Production-based models

As a special case of their general equilibrium model with production, Cox, Ingersoll, and Ross (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 (5.13) 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, i.e.

$$g(x) = k_1 x, \qquad \xi(x)^2 = k_2 x,$$

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

$$\begin{aligned} dx_t &= \kappa \left(\bar{x} - x_t \right) dt + \rho \sigma_x \sqrt{x_t}\, dz_{1t} + \sqrt{1-\rho^2} \sigma_x \sqrt{x_t}\, dz_{2t} \\ &= \kappa \left(\bar{x} - x_t \right) dt + \sigma_x \sqrt{x_t}\, d\bar{z}_t, \end{aligned}$$

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

$$dr_t = \kappa \left(\bar{r} - r_t \right) dt + \sigma_r \sqrt{r_t}\, d\bar{z}_t, \tag{5.16}$$

where $\bar{r} = k\bar{x}$ and $\sigma_r = \sqrt{k} \sigma_x$. The market prices of risk given in (5.14) and (5.15) simplify to

$$\lambda_1 = \xi(x) = \sqrt{k_2 x} = \sqrt{k_2 / k}\, \sqrt{r}, \qquad \lambda_2 = 0.$$

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 are 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

$$dC_t = C_t \left[\mu_{Ct}\, dt + \sigma_C\, dz_t \right],$$

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

$$d\mu_{Ct} = \kappa \left(\bar{\mu}_C - \mu_{Ct} \right) dt + \theta\, dz_t.$$

The representative agent is assumed to have a constant relative risk aversion of γ. It follows from (5.3) 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 $dr_t = \gamma d\mu_{Ct}$, i.e.

$$dr_t = \kappa \left(\bar{r} - r_t \right) dt + \sigma_r\, dz_t, \tag{5.17}$$

where $\sigma_r = \gamma\theta$ and $\bar{r} = \gamma\bar{\mu}_C + \delta - \frac{1}{2}\gamma(1+\gamma)\sigma_C^2$. 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 affine term structure model in this way. Assume that the expected growth rate and the variance rate of aggregate consumption are affine in some state variables, i.e.

$$\mu_{Ct} = a_0 + \sum_{i=1}^{n} a_i x_{it}, \qquad \|\boldsymbol{\sigma}_{Ct}\|^2 = b_0 + \sum_{i=1}^{n} b_i x_{it},$$

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 e.g. 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 is 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]

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

[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.

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 applies. 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 money growth. 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 $S = (S_t)$ of an asset follows the stochastic process

$$dS_t = S_t \left[\mu_{St} \, dt + \boldsymbol{\sigma}_{St}^\top \, d\boldsymbol{z}_t \right],$$

then we know that

$$\mu_{St} - r_t = \boldsymbol{\sigma}_{St}^\top \boldsymbol{\lambda}_t \tag{5.18}$$

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

$$\xi_t \equiv \mathrm{E}_t \left[\frac{d\mathbb{Q}}{d\mathbb{P}} \right] = \exp\left\{ -\frac{1}{2} \int_t^T \|\boldsymbol{\lambda}_s\|^2 \, ds - \int_t^T \boldsymbol{\lambda}_s^\top \, d\boldsymbol{z}_s \right\}.$$

The real price of an asset paying no dividends in the time interval $[t, T]$ can then be written as

$$P_t = \mathrm{E}_t \left[\frac{\zeta_T}{\zeta_t} P_T \right] = \mathrm{E}_t^{\mathbb{Q}} \left[e^{-\int_t^T r_s \, ds} P_T \right].$$

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

$$dI_t = I_t \left[i_t \, dt + \boldsymbol{\sigma}_{It}^\top \, d\boldsymbol{z}_t \right]. \tag{5.19}$$

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 volatility vector of the inflation rate.

[4] Such results are reported by, e.g., Fama (1981), Fama and Gibbons (1982), Chen, Roll, and Ross (1986), and Marshall (1992).

Consider now a nominal bank account which over the next instant promises a riskless 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 has the same value as the account. An application of Itô's Lemma implies a real price dynamics of

$$dN_t = N_t \left[\left(\tilde{r}_t - i_t + \|\boldsymbol{\sigma}_{It}\|^2 \right) dt - \boldsymbol{\sigma}_{It}^\top \, d\boldsymbol{z}_t \right]. \tag{5.20}$$

Note that the real return on this instantaneously nominally riskless asset, dN_t/N_t, is risky. Since the percentage volatility vector is given by $-\boldsymbol{\sigma}_{It}$, 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

$$\tilde{r}_t = r_t + i_t - \|\boldsymbol{\sigma}_{It}\|^2 - \boldsymbol{\sigma}_{It}^\top \boldsymbol{\lambda}_t, \tag{5.21}$$

i.e. 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 riskless.

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

$$dP_t = P_t \left[\left(\tilde{\mu}_{Pt} - i_t - \tilde{\boldsymbol{\sigma}}_{Pt}^\top \boldsymbol{\sigma}_{It} + \|\boldsymbol{\sigma}_{It}\|^2 \right) dt + \left(\tilde{\boldsymbol{\sigma}}_{Pt} - \boldsymbol{\sigma}_{It} \right)^\top \, d\boldsymbol{z}_t \right].$$

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 (5.21). 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 (5.18), 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

$$\tilde{\mu}_{Pt} - \tilde{r}_t = \tilde{\boldsymbol{\sigma}}_{Pt}^\top \tilde{\boldsymbol{\lambda}}_t, \tag{5.22}$$

where $\tilde{\boldsymbol{\lambda}}_t$ is the nominal market price of risk vector defined by

$$\tilde{\boldsymbol{\lambda}}_t = \boldsymbol{\sigma}_{It} + \boldsymbol{\lambda}_t. \tag{5.23}$$

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 is given by

$$d\tilde{\zeta}_t = -\tilde{\zeta}_t\left[\tilde{r}_t\,dt + \tilde{\boldsymbol{\lambda}}_t^{\top}\,d\boldsymbol{z}_t\right] \tag{5.24}$$

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

$$\tilde{\xi}_t \equiv \mathrm{E}_t\left[\frac{d\tilde{\mathbb{Q}}}{d\mathbb{P}}\right] = \exp\left\{-\frac{1}{2}\int_t^T \|\tilde{\boldsymbol{\lambda}}_s\|^2\,ds - \int_t^T \tilde{\boldsymbol{\lambda}}_s^{\top}\,d\boldsymbol{z}_s\right\}.$$

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, cf. Section 5.4, typically apply to real interest rates and real market prices of risk. Due to the relations (5.21) and (5.23), 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 $\operatorname{Cov}(x, y) = \mathrm{E}(xy) - \mathrm{E}(x)\,\mathrm{E}(y)$, we can write

$$\begin{aligned}
\tilde{B}_t^T &= \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\frac{I_t}{I_T}\right] \\
&= \mathrm{E}_t\left[\frac{\zeta_T}{\zeta_t}\right]\mathrm{E}_t\left[\frac{I_t}{I_T}\right] + \operatorname{Cov}_t\left(\frac{\zeta_T}{\zeta_t}, \frac{I_t}{I_T}\right) \\
&= B_t^T\,\mathrm{E}_t\left[\frac{I_t}{I_T}\right] + \operatorname{Cov}_t\left(\frac{\zeta_T}{\zeta_t}, \frac{I_t}{I_T}\right).
\end{aligned} \tag{5.25}$$

From the dynamics of the state-price deflator and the price index, we get

$$\frac{\zeta_T}{\zeta_t} = \exp\left\{-\int_t^T \left(r_s + \frac{1}{2}\|\boldsymbol{\lambda}_s\|^2\right) ds - \int_t^T \boldsymbol{\lambda}_s^\top\, d\boldsymbol{z}_s\right\},$$

$$\frac{I_t}{I_T} = \exp\left\{-\int_t^T \left(i_s - \frac{1}{2}\|\boldsymbol{\sigma}_{Is}\|^2\right) ds - \int_t^T \boldsymbol{\sigma}_{Is}^\top\, d\boldsymbol{z}_s\right\},$$

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

$$r_t = \delta - \frac{C_t u''(C_t)}{u'(C_t)}\mu_{Ct} - \frac{1}{2}C_t^2\frac{u'''(C_t)}{u'(C_t)}\|\boldsymbol{\sigma}_{Ct}\|^2 \tag{5.26}$$

and the market price of risk vector is given by

$$\boldsymbol{\lambda}_t = \left(-\frac{C_t u''(C_t)}{u'(C_t)}\right)\boldsymbol{\sigma}_{Ct}. \tag{5.27}$$

By substituting the expression (5.27) for λ_t into (5.21), 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, i.e. $u(C) = C^{1-\gamma}/(1-\gamma)$, and both the aggregate consumption and the price index follow geometric Brownian motions, we get constant rates

$$r = \delta + \gamma\mu_C - \frac{1}{2}\gamma(1+\gamma)\|\boldsymbol{\sigma}_C\|^2, \tag{5.28}$$

$$\tilde{r} = r + i - \|\boldsymbol{\sigma}_I\|^2 - \gamma\boldsymbol{\sigma}_I^\top\boldsymbol{\sigma}_C. \tag{5.29}$$

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 (probably unrealistic) case of 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 set-up of Bakshi and Chen (1996) closely.

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, i.e. $\tilde{M}_t = M_t/I_t$. At time t the agent has the opportunity to invest in a nominally riskless 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:

$$\sup_{(C_t,M_t)} \mathrm{E}\left[\int_0^\infty e^{-\delta t} u\left(C_t, M_t/I_t\right)\,dt\right]$$
$$\text{s.t. } \mathrm{E}\left[\int_0^\infty \zeta_t\left(C_t + \frac{M_t}{I_t}\tilde{r}_t\right)\,dt\right] \le W_0.$$

The first order conditions are

$$e^{-\delta t} u_C(C_t, M_t/I_t) = \psi\zeta_t, \tag{5.30}$$
$$e^{-\delta t} u_M(C_t, M_t/I_t) = \psi\zeta_t\tilde{r}_t, \tag{5.31}$$

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

$$\zeta_t = e^{-\delta t} \frac{u_C(C_t, \tilde{M}_t)}{u_C(C_0, \tilde{M}_0)}. \tag{5.32}$$

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,

$$\frac{1}{I_t} = \mathrm{E}_t\left[\int_t^\infty e^{-\delta(s-t)}\frac{u_C(C_s, \tilde{M}_s)}{u_C(C_t, \tilde{M}_t)}\frac{\tilde{r}_s}{I_s}\,ds\right]. \tag{5.33}$$

Substituting the first optimality condition (5.30) into the second (5.31), we see that the nominal short rate is given by

$$\tilde{r}_t = \frac{u_M(C_t, M_t/I_t)}{u_C(C_t, M_t/I_t)}. \tag{5.34}$$

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 riskless 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 (5.33) and (5.34), we get that the price index must satisfy the recursive relation

$$\frac{1}{I_t} = \mathrm{E}_t\left[\int_t^\infty e^{-\delta(s-t)}\frac{u_M(C_s, \tilde{M}_s)}{u_C(C_t, \tilde{M}_t)}\frac{1}{I_s}\,ds\right]. \tag{5.35}$$

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 (5.32), 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

$$\begin{aligned} r_t = \delta &+ \left(\frac{-C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) \mu_{Ct} + \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) \tilde{\mu}_{Mt} \\ &- \frac{1}{2} \frac{C_t^2 u_{CCC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \|\boldsymbol{\sigma}_{Ct}\|^2 - \frac{1}{2} \frac{\tilde{M}_t^2 u_{CMM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \|\tilde{\boldsymbol{\sigma}}_{Mt}\|^2 - \frac{C_t \tilde{M}_t u_{CCM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \boldsymbol{\sigma}_{Ct}^\top \tilde{\boldsymbol{\sigma}}_{Mt}, \end{aligned} \tag{5.36}$$

while the market price of risk vector is

$$\begin{aligned} \lambda_t &= \left(-\frac{C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) \boldsymbol{\sigma}_{Ct} + \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) \tilde{\boldsymbol{\sigma}}_{Mt} \\ &= \left(-\frac{C_t u_{CC}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) \boldsymbol{\sigma}_{Ct} + \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) (\boldsymbol{\sigma}_{Mt} - \boldsymbol{\sigma}_{It}). \end{aligned} \tag{5.37}$$

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 (5.21) is also valid in the present setting. Substituting the expression (5.37) for the market price of risk into (5.21), we obtain

$$\tilde{r}_t - r_t - i_t + \|\boldsymbol{\sigma}_{It}\|^2 = -\left(-\frac{C_t u''(C_t)}{u'(C_t)} \right) \boldsymbol{\sigma}_{It}^\top \boldsymbol{\sigma}_{Ct} - \left(\frac{-\tilde{M}_t u_{CM}(C_t, \tilde{M}_t)}{u_C(C_t, \tilde{M}_t)} \right) \boldsymbol{\sigma}_{It}^\top \tilde{\boldsymbol{\sigma}}_{Mt}. \tag{5.38}$$

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 (5.36), we see that the real short rate becomes

$$\begin{aligned} r_t = \delta &+ [1 - \varphi(1-\gamma)]\mu_{Ct} - (1-\varphi)(1-\gamma)\tilde{\mu}_{Mt} - \frac{1}{2}[1-\varphi(1-\gamma)][2-\varphi(1-\gamma)]\|\boldsymbol{\sigma}_{Ct}\|^2 \\ &+ \frac{1}{2}(1-\varphi)(1-\gamma)[1-(1-\varphi)(1-\gamma)]\|\tilde{\boldsymbol{\sigma}}_{Mt}\|^2 + (1-\varphi)(1-\gamma)[1-\varphi(1-\gamma)]\boldsymbol{\sigma}_{Ct}^\top \tilde{\boldsymbol{\sigma}}_{Mt}, \end{aligned} \tag{5.39}$$

which for $\gamma = 1$ simplifies to

$$r_t = \delta + \mu_{Ct} - \|\boldsymbol{\sigma}_{Ct}\|^2. \tag{5.40}$$

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 (5.34), the nominal short rate is

$$\tilde{r}_t = \frac{1-\varphi}{\varphi} \frac{C_t}{\tilde{M}_t}. \tag{5.41}$$

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 (5.35). 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 (5.35), we get

$$\frac{1}{k} = \frac{1-\varphi}{\varphi} \int_t^\infty e^{-\delta(s-t)} \operatorname{E}_t \left[\left(\frac{C_s}{C_t} \right)^{1-\gamma} \left(\frac{M_s}{M_t} \right)^{-1} \right] ds.$$

Inserting the relations

$$\begin{aligned} \frac{C_s}{C_t} &= \exp\left\{ \left(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2 \right)(s-t) + \boldsymbol{\sigma}_C^\top (\boldsymbol{z}_s - \boldsymbol{z}_t) \right\}, \\ \frac{M_s}{M_t} &= \exp\left\{ \left(\mu_M - \frac{1}{2}\|\boldsymbol{\sigma}_M\|^2 \right)(s-t) + \boldsymbol{\sigma}_M^\top (\boldsymbol{z}_s - \boldsymbol{z}_t) \right\}, \end{aligned}$$

and applying a standard rule for expectations of lognormal variables, we get

$$\begin{aligned} \frac{1}{k} = \frac{1-\varphi}{\varphi} \int_t^\infty \exp\Big\{ \Big(-\delta + (1-\gamma)(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2) - \mu_M + \|\boldsymbol{\sigma}_M\|^2 \\ + \frac{1}{2}(1-\gamma)^2\|\boldsymbol{\sigma}_C\|^2 - (1-\gamma)\boldsymbol{\sigma}_C^\top \boldsymbol{\sigma}_M \Big)(s-t) \Big\}\, ds, \end{aligned}$$

which implies that the conjecture is true with

$$k = \frac{\varphi}{1-\varphi} \left(\delta - (1-\gamma)(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2) + \mu_M - \|\boldsymbol{\sigma}_M\|^2 - \frac{1}{2}(1-\gamma)^2\|\boldsymbol{\sigma}_C\|^2 + (1-\gamma)\boldsymbol{\sigma}_C^\top \boldsymbol{\sigma}_M \right).$$

From an application of Itô's Lemma, it follows that the price index also follows a geometric Brownian motion

$$dI_t = I_t \left[i\, dt + \boldsymbol{\sigma}_I^\top\, d\boldsymbol{z}_t \right], \tag{5.42}$$

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

$$\tilde{r}_t = \frac{1-\varphi}{\varphi} k = \delta - (1-\gamma)(\mu_C - \frac{1}{2}\|\boldsymbol{\sigma}_C\|^2) + \mu_M - \|\boldsymbol{\sigma}_M\|^2 - \frac{1}{2}(1-\gamma)^2\|\boldsymbol{\sigma}_C\|^2 + (1-\gamma)\boldsymbol{\sigma}_C^\top \boldsymbol{\sigma}_M, \tag{5.43}$$

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 (5.39). 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

$$r_t = \delta + \gamma\mu_C - \gamma\|\boldsymbol{\sigma}_C\|^2 \left[\frac{1}{2}(1+\gamma) + \varphi(1-\gamma) \right], \tag{5.44}$$

which is also a constant. In comparison with (5.28) for the case where money has no real effects, the last term in the equation above is new.

Another example

Bakshi and Chen (1996) also study another model specification in which both nominal and real short rates are time-varying, but evolve independently of each other. To obtain stochastic interest rates we have to specify more general processes for aggregate consumption and money supply than the geometric Brownian motions used above. They 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

$$r_t = \delta + \alpha_C + (\kappa_C - \sigma_C^2) x_t, \tag{5.45}$$

while the nominal short rate is

$$\tilde{r}_t = \frac{(\delta + \mu_M^*)(\delta + \mu_M^* + \kappa_g \theta_g)}{\delta + \mu_M^* + (\kappa_g + \sigma_g^2) g_t}. \tag{5.46}$$

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

$$B_t^T = e^{-a(T-t) - b(T-t)x},$$

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, i.e. $\rho_{CM} = 0$, 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. Such relations were discussed already by Fisher (1896) and 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, Ingersoll, and Ross (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 behavior 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. 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 agent 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 agent is risk-neutral, his marginal utility is constant, which implies that

$$\mathrm{E}_t\left[\frac{B_{t'}^T}{B_t^T} \right] = e^{\delta(t'-t)}, \tag{5.47}$$

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 5.2, it cannot hold when interest rates are uncertain.

A slight variation of the above is to align all expected continuously compounded returns, i.e. $\frac{1}{t'-t}\,\mathrm{E}_t[\ln\left(B_{t'}^T / B_t^T\right)]$ 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 riskless return over $[t, t+dt]$ is $r_t\, dt$, so for any zero-coupon bond, the hypothesis claims that

$$\mathrm{E}_t \left[\frac{dB_t^T}{B_t^T} \right] = r_t\, dt, \quad \text{for all } T > t, \tag{5.48}$$

or, equivalently,[5] that

$$B_t^T = \mathrm{E}_t \left[e^{-\int_t^T r_s\, ds} \right], \quad \text{for all } T > t.$$

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, i.e.

$$\frac{1}{B_t^T} = \mathrm{E}_t \left[e^{\int_t^T r_s\, ds} \right], \quad \text{for all } T > t \tag{5.49}$$

or, equivalently,

$$B_t^T = \left(\mathrm{E}_t \left[e^{\int_t^T r_s\, ds} \right] \right)^{-1}, \quad \text{for all } T > t.$$

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

[5] Here and later we use that, under suitable regularity conditions, the relative drift rate of an Itô process $x = (x_t)$ is given by the process $\mu = (\mu_t)$ if and only if $x_t = \mathrm{E}_t[x_T \exp\{\int_t^T \mu_s\, ds\}]$. Suppose first that the relative drift rate is given by μ so that $dx_t = x_t[\mu_t\, dt + \boldsymbol{\sigma}_t^\top\, d\boldsymbol{z}_t]$. Then an application of Itô's Lemma reveals that the process $x_t \exp\{\int_0^t \mu_s\, ds\}$ is a martingale so that $x_t \exp\{\int_0^t \mu_s\, ds\} = \mathrm{E}_t[x_T \exp\{\int_0^T \mu_s\, ds\}]$ and hence $x_t = \mathrm{E}_t[x_T \exp\{\int_t^T \mu_s\, ds\}]$.

The absolute drift of x is the limit of $\frac{1}{\Delta t} \mathrm{E}_t[x_{t+\Delta t} - x_t]$ as $\Delta t \to 0$. If $x_t = \mathrm{E}_t[x_T \exp\{\int_t^T \mu_s\, ds\}]$ for all t, then

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

i.e. the relative drift rate equals $-\mu_t$.

$e^{\int_t^T r_s\,ds}$ 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

$$y_t^T = -\frac{1}{T-t}\ln B_t^T = \mathrm{E}_t\left[\frac{1}{T-t}\int_t^T r_s\,ds\right], \tag{5.50}$$

or, equivalently,

$$B_t^T = e^{-\mathrm{E}_t\left[\int_t^T r_s\,ds\right]}, \quad \text{for all } T > t.$$

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, i.e. that forward rates are unbiased estimates of future spot rates. In symbols,

$$f_t^T = \mathrm{E}_t[r_T], \quad \text{for all } T > t.$$

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 three 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, i.e. $f'' > 0$, 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 $\boldsymbol{z}$. Assuming absence of arbitrage there exists a market price of risk process λ so that for any maturity T, the zero-coupon bond price dynamics is of the form

$$dB_t^T = B_t^T\left[\left(r_t + \left(\boldsymbol{\sigma}_t^T\right)^\top \boldsymbol{\lambda}_t\right)dt + \left(\boldsymbol{\sigma}_t^T\right)^\top d\boldsymbol{z}_t\right], \tag{5.51}$$

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 (5.48), we see that the local expectation hypothesis will hold if and only if $\left(\boldsymbol{\sigma}_t^T\right)^\top \boldsymbol{\lambda}_t = 0$ 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 (5.51) show that

$$d\left(\frac{1}{B_t^T}\right) = \frac{1}{B_t^T}\left[\left(-r_t - \left(\boldsymbol{\sigma}_t^T\right)^\top \boldsymbol{\lambda}_t + \|\boldsymbol{\sigma}_t^T\|^2\right) dt - \left(\boldsymbol{\sigma}_t^T\right)^\top d\boldsymbol{z}_t\right].$$

On the other hand, according to the hypothesis (5.49) the relative drift of $1/B_t^T$ equals $-r_t$; cf. a previous footnote. To match the two expressions for the drift, we must have

$$\left(\boldsymbol{\sigma}_t^T\right)^\top \boldsymbol{\lambda}_t = \|\boldsymbol{\sigma}_t^T\|^2, \quad \text{for all } T. \tag{5.52}$$

Is this possible? Cox, Ingersoll, and Ross (1981a) conclude that it is impossible. If the exogenous shock $\boldsymbol{z}$ 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 cannot hold when interest rates are stochastic. However, as pointed out by McCulloch (1993) and Fisher and Gilles (1998), in multi-dimensional cases the key condition (5.52) may indeed hold, at least in very special cases. Let φ be a d-dimensional function with the property that $\|\varphi(\tau)\|^2$ is independent of τ. Define $\boldsymbol{\lambda}_t = 2\varphi(0)$ and $\boldsymbol{\sigma}_t^T = \varphi(0) - \varphi(T-t)$. Then (5.52) is indeed satisfied. However, all such functions φ seem to generate very strange bond price dynamics. The examples given in the two papers mentioned above are

$$\varphi(\tau) = k\begin{pmatrix}\sqrt{2e^{-\tau} - e^{-2\tau}} \\ 1 - e^{-\tau}\end{pmatrix}, \qquad \varphi(\tau) = k_1\begin{pmatrix}\cos(k_2\tau) \\ \sin(k_2\tau)\end{pmatrix},$$

where k, k_1, k_2 are constants.

As discussed above, the rate or return version implies that the absolute drift rate of the log-bond price equals the short rate. We can see from (5.50) that the same is true for the yield-to-maturity version and hence the unbiased version.[6] On the other hand Itô's Lemma and (5.51) imply that

$$d\left(\ln B_t^T\right) = \left(r_t + (\boldsymbol{\sigma}_t^T)^\top \boldsymbol{\lambda}_t - \frac{1}{2}\|\boldsymbol{\sigma}_t^T\|^2\right) dt + \left(\boldsymbol{\sigma}_t^T\right)^\top d\boldsymbol{z}_t. \tag{5.53}$$

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.

[6] This is analogous to the previous footnote. According to the hypothesis,

$$\frac{1}{\Delta t}\,\mathrm{E}_t\left[\ln B_{t+\Delta t}^T - \ln B_t^T\right] = \frac{1}{\Delta t}\,\mathrm{E}_t\left[-\,\mathrm{E}_{t+\Delta t}\left[\int_{t+\Delta t}^T r_s\,ds\right] + \mathrm{E}_t\left[\int_t^T r_s\,ds\right]\right] = \frac{1}{\Delta t}\,\mathrm{E}_t\left[\int_t^{t+\Delta t} r_s\,ds\right],$$

which approaches r_t as $\Delta t \to 0$.

5.6.3 The weak expectation hypothesis

Above we looked at versions of the *pure* expectation hypothesis that all aligns 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

$$\mathrm{E}_t\left[\frac{dB_t^T}{B_t^T}\right] = (r_t + g(T-t))\ dt$$

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 (5.51), 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.

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 $h(0)=0$, i.e. 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 $r_t + h(T-t)$.[7] Comparing with (5.53), we see that also 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 requires 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 has 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 bonds should exceed the expected returns on short bonds to compensate for the higher price fluctuations of long 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 used in the usual sense of the word. Short

[7] 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,

$$\begin{aligned}\frac{1}{\Delta t}\mathrm{E}_t\left[\ln B_{t+\Delta t}^T - \ln B_t^T\right] &= \frac{1}{\Delta t}\mathrm{E}_t\left[-\int_{t+\Delta t}^T (\mathrm{E}_{t+\Delta t}[r_s] + h(s-(t+\Delta t)))\,ds + \int_t^T (\mathrm{E}_t[r_s] + h(s-t))\,ds\right] \\ &= \frac{1}{\Delta t}\mathrm{E}_t\left[\int_t^{t+\Delta t} r_s\,ds\right] - \frac{1}{\Delta t}\left(\int_{t+\Delta t}^T h(s-(t+\Delta t))\,ds - \int_t^T h(s-t)\,ds\right).\end{aligned}$$

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.

bonds are not necessarily more liquid than long 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.[8] 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 favorite 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, Ingersoll, and Ross (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 than 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, e.g. 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, e.g., Duffie and Epstein (1992), Wang (1996), Riedel (1999, 2000), Wachter (2002). The effects of central banks on the term structure are discussed and modeled by, e.g., Babbs and Webber (1994), Balduzzi, Bertola, and Foresi (1997), and Piazzesi (2001).

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, i.e. $f_t^{t',T} - \mathrm{E}_t[y_{t'}^T]$. This exercise will give a link between the term premium and a state-price deflator $\zeta = (\zeta_t)$.

[8] In a sense the liquidity preference hypothesis simply says that all investors prefer short bonds.

(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 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

$$\mathrm{E}_{t_0}\left[\frac{1}{B_{t_1}^{t_2}}\right] = \frac{1}{\mathrm{E}_{t_0}\left[B_{t_1}^{t_2}\right]}. \qquad (*)$$

(d) Show that (*) can only hold under full certainty. *Hint: Use Jensen's inequality.*

EXERCISE 5.3 Go through the derivations in Section 5.5.3.

EXERCISE 5.4 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 constrains 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, i.e. $\tilde{B}_t^T$. 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 the term premium is positive for some maturities and negative for others?

Chapter 6

Introduction to dynamic term structure models

6.1 Introduction

Chapter 4 reviewed general asset pricing theory. In this chapter we will explore some of the consequences the general theory has for term structure modeling and the pricing of some standard classes of derivatives. These results will be derived without specifying a concrete term structure model so that they apply in more or less all the models we shall consider in the following chapters.

One of the key observations of Chapter 4 is that the general pricing mechanisms in the financial market can be captured by a risk-neutral probability measure. Section 6.2 shows that even other probability measures will be helpful in the computation of the prices of derivative assets such as options on bonds, caps, floors, and swaptions.

In Section 6.3 we discuss the implications of the general asset pricing theory for the pricing of forwards and futures. We give a general characterization of both the forward price and the futures price and discuss the difference between these prices. In particular, we obtain general pricing formulas for bond futures and Eurodollar-futures.

In Section 6.4 we will briefly describe and discuss the special features of American-style derivatives, i.e. derivatives that entitle the owner to exercise the contract at all points in time before the final expiration date, or a least at several pre-specified points in time. Closed-form expressions for the prices of American-style derivatives are only present in the very trivial cases where premature exercise can be ruled out. Hence, in general, one must resort to numerical methods to compute both prices and hedge rations for American-style derivatives.

The asset pricing relations of Chapter 4 were developed under quite general assumptions on the dynamics of prices and interest rates. In Section 6.5 we will specialize this framework by assuming that the values of interest rates, expected returns, and volatilities are given as functions of some common state variable and that this state variable follows a diffusion process. The state variable may be one- or multi-dimensional. We will refer to such models as diffusion models. The main advantage of this added structure is that then, for most assets, the price can be written as some function of the state variable and time, and this function can be computed by solving a partial differential equation. Sometimes it is easier to solve partial differential equations than to compute the necessary risk-neutral expectations involving the future asset dividends and short-term interest rates. All the classic and also many modern dynamic term structure models are diffusion models so this section also serves as an introduction to many concrete term structure models.

The best known model for the pricing of derivative securities is the Black-Scholes-Merton model developed by Black and Scholes (1973) and Merton (1973). The model was originally created for the pricing of stock options, but a variant of the model (the so-called Black model introduced by Black (1976)) has later been applied to the pricing of many other derivative securities, including some of the interest rate dependent securities discussed in Chapter 2. In Section 6.6 we will briefly go through the Black-Scholes-Merton model and Black's variant and state Black's formula for the pricing of selected fixed income securities. However, we will also point out that the assumptions underlying Black's model are completely unacceptable when it comes to interest rate dependent securities.

Section 6.7 gives an overview of the concrete continuous-time term structure models we will describe in detail in Chapters 7–11. The distinction between absolute pricing models and relative pricing models is discussed. We will also outline some criteria for the choice of which of the numerous term structure models to use.

6.2 Additional probability measures convenient for pricing

In Chapter 4 we discussed the idea of computing prices using a risk-neutral probability measure. For some applications it turns out to be convenient to use different probability measures. First we discuss the general idea and then we introduce some particularly useful probability measures.

6.2.1 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 (4.5) 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, i.e. 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, cf. Theorem 3.6,

$$d\left(\frac{P_t}{S_t}\right) = \frac{P_t}{S_t}\left[\left(\mu_{Pt} - \mu_{St} + \|\boldsymbol{\sigma}_{St}\|^2 - \boldsymbol{\sigma}_{St}^\top \boldsymbol{\sigma}_{Pt}\right) dt + \left(\boldsymbol{\sigma}_{Pt} - \boldsymbol{\sigma}_{St}\right)^\top d\boldsymbol{z}_t\right]. \tag{6.1}$$

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

$$\left(\boldsymbol{\sigma}_{Pt} - \boldsymbol{\sigma}_{St}\right)^\top \boldsymbol{\lambda}_t^S = \mu_{Pt} - \mu_{St} + \|\boldsymbol{\sigma}_{St}\|^2 - \boldsymbol{\sigma}_{St}^\top \boldsymbol{\sigma}_{Pt}. \tag{6.2}$$

Then we can define a probability measure $\mathbb{Q}^S$ by the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}^S}{d\mathbb{P}} = \exp\left\{-\frac{1}{2}\int_0^T \|\boldsymbol{\lambda}_t^S\|^2\, ds - \int_0^T \left(\boldsymbol{\lambda}_t^S\right)^\top dz_t\right\}.$$

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 (6.1) 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 (6.2)? 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 (6.2) 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.

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

$$P_t = S_t\, \mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{P_T}{S_T}\right]. \tag{6.3}$$

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 in terms of S_t. In the following subsections we look at some important examples.

6.2.2 A zero-coupon bond as the numeraire – 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 $t \le T$ 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, i.e. 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 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

$$P_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[P_T \right]. \tag{6.4}$$

Under the probability measure $\mathbb{Q}^T$, the process $\boldsymbol{z}^T$ defined by

$$d\boldsymbol{z}_t^T = d\boldsymbol{z}_t + \left(\boldsymbol{\lambda}_t - \boldsymbol{\sigma}_t^T \right) \, dt, \qquad \boldsymbol{z}_0^T = 0, \tag{6.5}$$

is a standard Brownian motion. To compute the price from (6.4) 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 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

$$d\boldsymbol{z}_t^T = d\boldsymbol{z}_t^{\mathbb{Q}} - \boldsymbol{\sigma}_t^T \, dt, \tag{6.6}$$

which follows from (4.9) and (6.5). 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.

6.2.3 An annuity as the numeraire – swap martingale measures

As discussed above it is often computationally advantageous for the pricing of European options to use as a numeraire the zero-coupon bond maturing at the same time as the option. In this subsection we show that for the pricing of European swaptions it is computationally convenient to use another numeraire and hence another probability measure. A European payer swaption with expiration time T_0 and an exercise rate of K gives the right to enter into a payer swap with some face value H at time T_0. If the right is exercised, the holder must at given future points in time, $T_i = T_0 + i\delta$ where $i = 1, \dots, n$, pay $H\delta K$, but will receive payments $H\delta l_{T_i - \delta}^{T_i}$ that are still unknown. From (2.33) on page 39 we have that the time T_0 value of the payoff of such a swaption can be expressed as

$$\mathcal{P}_{T_0} = \left(\sum_{i=1}^{n} B_{T_0}^{T_i} \right) H\delta \max \left(\tilde{l}_{T_0}^{\delta} - K, 0 \right),$$

where $\tilde{l}^{\delta}_{T_0}$ is the (equilibrium) swap rate prevailing at time T_0.

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 $t \le T_0$ equals $G_t = \sum_{i=1}^n B_t^{T_i}$. In particular, the payoff of the swaption can be restated as

$$\mathcal{P}_{T_0} = G_{T_0} H \delta \max\left(\tilde{l}^{\delta}_{T_0} - K, 0\right),$$

and the payoff expressed in units of the annuity bond is simply $H\delta \max\left(\tilde{l}^{\delta}_{T_0} - K, 0\right)$. 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. This type of martingale measure was introduced by Jamshidian (1997). The price of the European payer swaption can now be written as

$$\mathcal{P}_t = G_t \, \mathrm{E}_t^{\mathbb{Q}^G}\left[\frac{\mathcal{P}_{T_0}}{G_{T_0}}\right] = G_t H \delta \, \mathrm{E}_t^{\mathbb{Q}^G}\left[\max\left(\tilde{l}^{\delta}_{T_0} - K, 0\right)\right], \tag{6.7}$$

so we only need to know the distribution of the swap rate $\tilde{l}^{\delta}_{T_0}$ under the swap martingale measure. In Chapter 11 we will look at a model that is based on the assumption that $\tilde{l}^{\delta}_{T_0}$ is lognormally distributed under the swap martingale measure. This results in a Black-Scholes-Merton-type formula for the price of the European swaption very similar to the pricing formula applied by many practitioners, cf. Equation (6.60) on page 136.

6.2.4 A general pricing formula 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. Of course, a similar result is valid for European put options. 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

$$C_T = \max(h_T - K, 0).$$

For an option on a traded asset, h_T is the price of the underlying asset at the expiry date. For an option on a given interest rate, h_T denotes the value of this interest rate at the expiry date. We can rewrite the payoff as

$$C_T = (h_T - K)\, \mathbf{1}_{\{h_T > K\}},$$

where $\mathbf{1}_{\{h_T > K\}}$ is the indicator for the event $h_T > K$. This indicator is a random variable whose value will be 1 if the realized value of h_T turns out to be larger than K and the value is 0 otherwise.

According to (6.4) the time t price of the option is[1]

$$\begin{aligned} C_t &= B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max(h_T) - K, 0)\right] \\ &= B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[(h_T - K)\mathbf{1}_{\{h_T > K\}}\right] \\ &= B_t^T \left(\mathrm{E}_t^{\mathbb{Q}^T} \left[h_T \mathbf{1}_{\{h_T > K\}}\right] - K \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\mathbf{1}_{\{h_T > K\}}\right] \right) \\ &= B_t^T \left(\mathrm{E}_t^{\mathbb{Q}^T} \left[h_T \mathbf{1}_{\{h_T > K\}}\right] - K \mathbb{Q}_t^T (h_T > K) \right) \\ &= B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[h_T \mathbf{1}_{\{h_T > K\}}\right] - K B_t^T \mathbb{Q}_t^T (h_T > K). \end{aligned} \tag{6.8}$$

Here $\mathbb{Q}_t^T(h_T > K)$ denotes the probability (using the probability measure $\mathbb{Q}^T$) of $h_T > K$ 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 h_t is then a valid numeraire with a corresponding probability measure $\mathbb{Q}^h$. Applying (6.3) for both the numeraires B_t^T and h_t, we get

$$\begin{aligned} B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[h_T \mathbf{1}_{\{h_T > K\}}\right] &= h_t \, \mathrm{E}_t^{\mathbb{Q}^h} \left[\mathbf{1}_{\{h_T > K\}}\right] \\ &= h_t \mathbb{Q}_t^h (h_T > K). \end{aligned}$$

The call price is therefore

$$C_t = h_t \mathbb{Q}_t^h (h_T > K) - K B_t^T \mathbb{Q}_t^T (h_T > K). \tag{6.9}$$

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 particular, for a call option on a zero-coupon bond maturing at time $S > T$ we get that

$$C_t = B_t^S \mathbb{Q}_t^S (B_T^S > K) - K B_t^T \mathbb{Q}_t^T (B_T^S > K), \tag{6.10}$$

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 this equation in later chapters to derive closed-form option pricing formulas in specific models of the term structure of interest rates.

6.3 Forward prices and futures prices

As we noted in Section 2.5, it is relatively easy to show that the forward price and the futures price for the same settlement date and the same underlying asset are identical if there is no

[1] In the computation we use the fact that the expected value of the indicator of an event is equal to the probability of that event. This follows from the general definition of an expected value, $\mathrm{E}[g(\omega)] = \int_{\omega \in \Omega} g(\omega) f(\omega) \, d\omega$, where $f(\omega)$ is the probability density function of the state ω and the integration is over all possible states. The set of possible states can be divided into two sets, namely the set of states ω for which $h_T > K$ and the set of ω for which $h_T \leq K$. Consequently,

$$\begin{aligned} \mathrm{E}[\mathbf{1}_{\{h_T > K\}}] &= \int_{\omega \in \Omega} \mathbf{1}_{\{h_T > K\}} f(\omega) \, d\omega \\ &= \int_{\omega : h_T > K} 1^\top f(\omega) \, d\omega + \int_{\omega : h_T \leq K} 0^\top f(\omega) \, d\omega \\ &= \int_{\omega : h_T > K} f(\omega) \, d\omega, \end{aligned}$$

which is exactly the probability of the event $h_T > K$.

uncertainty about future interest rates. The original proof of this result was given by Cox, Ingersoll, and Ross (1981b). Armed with our general asset pricing theory, we can now characterize forward and futures prices under less restrictive assumptions. In particular, we can allow for stochastic interest rates.

6.3.1 Forward prices

A forward with maturity date T and delivery price K provides a payoff of $S_T - K$ at time T, where S is the underlying variable, typically the price of an asset or a specific interest rate. For forwards contracted upon at time t, 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. At the settlement date the forward price equals the value of the underlying variable, i.e. $F_T^T = S_T$. It follows from the results on the pricing under the risk-neutral measure that the forward price F_t^T is fixed so that

$$\begin{aligned} 0 &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(S_T - F_t^T\right)\right] \\ &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} S_T\right] - F_t^T\,\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right] \\ &= \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} S_T\right] - F_t^T B_t^T, \end{aligned}$$

where the last equality is due to (4.15). Hence, the forward price is given by

$$F_t^T = \frac{\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} S_T\right]}{B_t^T}. \tag{6.11}$$

If the underlying variable is the price of a traded asset with no payments in the period $[t, T]$, we have

$$\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} S_T\right] = S_t,$$

so that the forward price can be written as

$$F_t^T = \frac{S_t}{B_t^T}.$$

This expression is consistent with the results for bond forwards given in Section 2.3 on page 22.

Applying a well-known property of covariances, we have that

$$\begin{aligned} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du} S_T\right] &= \mathrm{Cov}_t^{\mathbb{Q}}\left(e^{-\int_t^T r_u\,du}, S_T\right) + \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\right]\mathrm{E}_t^{\mathbb{Q}}[S_T] \\ &= \mathrm{Cov}_t^{\mathbb{Q}}\left(e^{-\int_t^T r_u\,du}, S_T\right) + B_t^T\,\mathrm{E}_t^{\mathbb{Q}}[S_T]. \end{aligned}$$

Upon substitution of this into (6.11) we get the following expression for the forward price:

$$F_t^T = \mathrm{E}_t^{Q}[S_T] + \frac{\mathrm{Cov}_t^{\mathbb{Q}}\left(e^{-\int_t^T r_u\,du}, S_T\right)}{B_t^T}. \tag{6.12}$$

We can also characterize the forward price in terms of the T-forward martingale measure. 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

$$\frac{0}{B_t^T} = \mathrm{E}_t^{\mathbb{Q}^T}\left[\frac{S_T - F_t^T}{B_T^T}\right]$$

and hence

$$F_t^T = \mathrm{E}_t^{\mathbb{Q}^T}[S_T] = \mathrm{E}_t^{Q^T}[F_T^T],$$

which implies that the forward price F_t^T is a $\mathbb{Q}^T$-martingale.

6.3.2 Futures prices

Consider a futures with final settlement at time T. Let Φ_t^T be the futures price at time t. The futures price at the settlement time is $\Phi_T^T = S_T$. We assume 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 spot martingale measure $\mathbb{Q}$, so that in particular*

$$\Phi_t^T = \mathrm{E}_t^{\mathbb{Q}}[S_T]. \tag{6.13}$$

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 $t, t+\Delta t, t+2\Delta t, \dots, t+N\Delta t \equiv T$. This proof is originally due to Cox, Ingersoll, and Ross (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

$$e^{\left(r_{t+(N-1)\Delta t}+\dots+r_t\right)\Delta t}\Phi_{t+N\Delta t}^T = e^{\left(r_{t+(N-1)\Delta t}+\dots+r_t\right)\Delta t}\Phi_T^T = e^{\left(r_{t+(N-1)\Delta t}+\dots+r_t\right)\Delta t}S_T.$$

The continuous-time limit of this is $e^{\int_t^T r_u\,du}S_T$. 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,

$$\Phi_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r_u\,du}\left(e^{\int_t^T r_u\,du}S_T\right)\right] = \mathrm{E}_t^{\mathbb{Q}}[S_T],$$

as was to be shown. □

The theorem 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

$$\Phi_t^{T,\text{cpn}} = \mathrm{E}_t^{\mathbb{Q}}\left[\sum_{T_i > T} Y_i B_T^{T_i}\right] = \sum_{T_i > T} Y_i \,\mathrm{E}_t^{\mathbb{Q}}\left[B_T^{T_i}\right] = \sum_{T_i > T} Y_i \Phi_t^{T,T_i}, \tag{6.14}$$

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.

Finally, consider Eurodollar futures contracts. As discussed in Section 2.6 on page 25 the marking-to-market payments of a Eurodollar futures are based on the changes in the *actual* futures price

$$\mathcal{E}_t^T = 100 - 0.25\left(100 - \tilde{\mathcal{E}}_t^T\right) = 100 - 25\varphi_t^T,$$

where $\tilde{\mathcal{E}}_t^T$ is the *quoted* futures price for the final settlement time T, and φ_t^T is the corresponding LIBOR futures rate. At the final settlement time T the quoted futures price is fixed at the value

$$\tilde{\mathcal{E}}_T^T = 100\left(1 - l_T^{T+0.25}\right) = 100\left(1 - 4\left[(B_T^{T+0.25})^{-1} - 1\right]\right),$$

so that 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 the analysis above 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

$$\tilde{\mathcal{E}}_T^T = 4\mathcal{E}_t^T - 300 = 500 - 400\,\mathrm{E}_t^{\mathbb{Q}}\left[(B_T^{T+0.25})^{-1}\right]. \tag{6.15}$$

6.3.3 A comparison of forward prices and futures prices

From (6.12) and (6.13) we get that the difference between the forward price F_t^T and the futures price Φ_t^T is given by

$$F_t^T - \Phi_t^T = \frac{\mathrm{Cov}_t^{\mathbb{Q}}\left(e^{-\int_t^T r_u\,du}, S_T\right)}{B_t^T}. \tag{6.16}$$

The forward price and the futures price will only be identical if the two random variables S_T and $\exp\left(-\int_t^T r_u\,du\right)$ are uncorrelated under the risk-neutral probability measure. Of course, this is true if the short rate r_t is constant or deterministic, in which case we recover the standard result. The forward price is larger [smaller] than the futures price if the variables $\exp\left(-\int_t^T r_u\,du\right)$ and S_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 S_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.

6.4 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 H_τ denote the payoff if the derivative is exercised at time $\tau \leq T$. In general, H_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 tell 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

$$\tau(I) = \min\{s \in [t, T] \mid I(\omega, s) = 1\}.$$

This is called a stopping time in the literature on stochastic processes. By our earlier analysis, the value of getting the payoff $H_{\tau(I)}$ at time $\tau(I)$ is given by $\mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau(I)} r_u\, du} H_{\tau(I)}\right]$. 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

$$P_t = \sup_{I \in \mathcal{I}[t,T]} \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau(I)} r_u\, du} H_{\tau(I)}\right]. \tag{6.17}$$

An optimal exercise strategy I^* is such that

$$P_t = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^{\tau_{I^*}} r_u\, du} H_{\tau_{I^*}}\right].$$

Note that the optimal exercise strategy and the price of the derivative must be solved for simultaneously. We will return to the valuation of American-style derivatives in the next section.

6.5 Diffusion models and the fundamental partial differential equation

Many financial models assume the existence of one or several so-called **state variables**, i.e. 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. The state variables of term structure models should be informative for the term structure. 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, cf. Section 3.4 on page 50. 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, i.e. models with several state variables.

6.5.1 One-factor diffusion models

We assume that a single, one-dimensional, state variable contains all the relevant information, i.e. that 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

$$dx_t = \alpha(x_t, t)\, dt + \beta(x_t, t)\, dz_t, \tag{6.18}$$

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 (4.9) to write the dynamics of the state variable under the risk-neutral probability measure as

$$dx_t = [\alpha(x_t, t) - \beta(x_t, t)\lambda(x_t, t)]\; dt + \beta(x_t, t)\, dz_t^{\mathbb{Q}}. \tag{6.19}$$

We also assume that the short interest rate depends at most on x and t, i.e. $r_t = r(x_t, t)$.

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 (see Theorem 3.5 on page 55) that the dynamics of $P_t = P(x_t, t)$ is

$$dP_t = P_t\left[\mu(x_t, t)\, dt + \sigma(x_t, t)\, dz_t\right], \tag{6.20}$$

where the functions μ and σ are defined by

$$\mu(x, t)P(x, t) = \frac{\partial P}{\partial t}(x, t) + \frac{\partial P}{\partial x}(x, t)\alpha(x, t) + \frac{1}{2}\frac{\partial^2 P}{\partial x^2}(x, t)\beta(x, t)^2, \tag{6.21}$$

$$\sigma(x, t)P(x, t) = \frac{\partial P}{\partial x}(x, t)\beta(x, t). \tag{6.22}$$

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 6.2 *The function* P *defined by*

$$P(x,t) = \mathrm{E}^{\mathbb{Q}}_{x,t}\left[e^{-\int_t^T r(x_u,u)\,du} H(x_T,T)\right] \tag{6.23}$$

satisfies the partial differential equation

$$\frac{\partial P}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial P}{\partial x}(x,t) + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) - r(x,t)P(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T), \tag{6.24}$$

together with the terminal condition

$$P(x,T) = H(x,T), \quad x \in \mathcal{S}. \tag{6.25}$$

The relation between expectations and partial differential equations is generally known as the Feynman-Kac theorem, cf. Øksendal (1998, 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, cf. (6.19). 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.

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 (4.11) in particular must hold for this asset, we have that

$$\lambda(x,t) = \frac{\frac{\alpha(x,t)}{x} - r(x,t)}{\frac{\beta(x,t)}{x}} = \frac{\alpha(x,t) - r(x,t)x}{\beta(x,t)}$$

or

$$\alpha(x,t) - \beta(x,t)\lambda(x,t) = r(x,t)x. \tag{6.26}$$

By insertion of this expression, the PDE (6.24) reduces to

$$\frac{\partial P}{\partial t}(x,t) + r(x,t)\left(x\frac{\partial P}{\partial x}(x,t) - P(x,t)\right) + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T). \tag{6.27}$$

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. This approach has proven successful in the pricing of stock options with the Black-Scholes-Merton model as the prime example. However, it is inappropriate to price interest rate derivatives just by modeling 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. Just as the relevant market price of risk in the Black-Scholes-Merton setting can be extracted from the current market price of the underlying stock, the market prices of interest rate risk can be extracted from the entire current market yield curve. Given the current yield curve (i.e. current bond prices) and an assumption on the volatility of the curve, all yield curve derivatives (bond options and futures, caps, floors, swaptions, etc.) can be priced by the no-arbitrage principle, i.e. without precise knowledge of preferences, etc. In a sense, such models have infinitely many state variables, namely bond prices or yields of all maturities. We will consider such term structure models in Chapters 10 and 11. For the term structure models based on one or a few state variables we need some knowledge of the market price of risk, either from an equilibrium model or by a reasonable assumption.

In many models the PDE (6.24) with the appropriate terminal condition can be solved explicitly for a large number of interesting securities. The solution is often found by coming up with a qualified guess and verifying that the guess is correct by substitution into the equation. Alternatively, the PDE can in some cases be restated as another partial differential equation which has a solution that is already known. In other models explicit solutions for the important securities cannot be found, so that it is necessary to resort to numerical solution techniques as those introduced in Chapter 16.

Hedging

In a model with a single one-dimensional state variable a locally riskless 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

$$\begin{aligned}
d\Pi_t &= \alpha_t r(x_t,t)\,dt + \theta_t\,dP_2(x_t,t) \\
&= r(x_t,t)\left(1 - \frac{\sigma_1(x_t,t)}{\sigma_2(x_t,t)}\right) P_1(x_t,t)\,dt \\
&\quad + \frac{\sigma_1(x_t,t)P_1(x_t,t)}{\sigma_2(x_t,t)P_2(x_t,t)}\left(\mu_2(x_t,t)P_2(x_t,t)\,dt + \sigma_2(x_t,t)P_2(x_t,t)\,dz_t\right) \\
&= \left(r(x_t,t) + \frac{\sigma_1(x_t,t)}{\sigma_2(x_t,t)}\left(\mu_2(x_t,t) - r(x_t,t)\right)\right) P_1(x_t,t)\,dt + \sigma_1(x_t,t)P_1(x_t,t)\,dz_t \\
&= \left(r(x_t,t) + \sigma_1(x_t,t)\lambda(x_t,t)\right) P_1(x_t,t)\,dt + \sigma_1(x_t,t)P_1(x_t,t)\,dz_t \\
&= \mu_1(x_t,t)P_1(x_t,t)\,dt + \sigma_1(x_t,t)P_1(x_t,t)\,dz_t \\
&= dP_{1t},
\end{aligned}$$

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 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, e.g. coupon bonds, swaps, caps, and floors. Theorem 6.2 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. Then Theorem 6.2 can be applied in order to separately find the current value of each of these payments after which the value of the security follows from a simple summation. Suppose that the security provides payments $H_j(x_{T_j})$ at time T_j for $j = 1, 2, \dots, N$, where $T_1 < T_2 < \dots < T_N$. Then the price of the security at time $t < T_1$ is given by

$$P(x_t, t) = \sum_{j=1}^{N} P_j(x_t, t),$$

where $P_j(x, t)$ solves the PDE

$$\begin{aligned} \frac{\partial P_j}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial P_j}{\partial x}(x,t) \\ + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P_j}{\partial x^2}(x,t) - r(x,t)P_j(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0, T_j), \end{aligned}$$

with the terminal condition

$$P_j(x, T_j) = H_j(x), \quad x \in \mathcal{S}.$$

Here is an alternative pricing approach. 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, i.e. in the intervals (T_j, T_{j+1}), the price of the security will satisfy the PDE (6.24). If the price of the security is to be priced by numerical techniques, this alternative approach is simpler than that explained in the previous paragraphs, since only one PDE needs to be solved rather than one PDE for each payment date.

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 (4.17) 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 6.2 can be extended to show that the function P in this case will solve the PDE

$$\frac{\partial P}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial P}{\partial x}(x,t) + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) - r(x,t)P(x,t) + h(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T),$$

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, i.e. $h(x,t) = q(x,t)P(x,t)$, the PDE can be written as

$$\frac{\partial P}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial P}{\partial x}(x,t) + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 P}{\partial x^2}(x,t) - \big(r(x,t) - q(x,t)\big)P(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T). \tag{6.28}$$

In this case the price can be written as

$$P(x_t,t) = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T [r(x_u,u) - q(x_u,u)]\,du} H(x_T,T)\right] \tag{6.29}$$

Futures prices

In Section 6.3.2 we showed that the futures price of an asset is generally given by the risk-neutral expectation of the price of the underlying asset at the final settlement date. In our diffusion setting, the futures price is $\Phi_t^T = \Phi^T(x_t,t)$ and we can write the pricing relation as

$$\Phi^T(x_t,t) = \mathrm{E}_t^{\mathbb{Q}}\left[S(x_T,T)\right],$$

where S denotes the price of the underlying asset. Comparing this with (6.29) we see that we can think of the futures price as being the price of an asset that has a terminal payment of $S(x_T,T)$ and a continuous dividend rate equal to $r(x_t,t)S(x_t,t)$. From (6.28) we get that the futures price function $\Phi^T(x,t)$ must satisfy the PDE

$$\frac{\partial \Phi^T}{\partial t}(x,t) + (\alpha(x,t) - \beta(x,t)\lambda(x,t))\frac{\partial \Phi^T}{\partial x}(x,t) + \frac{1}{2}\beta(x,t)^2\frac{\partial^2 \Phi^T}{\partial x^2}(x,t) = 0, \quad (x,t) \in \mathcal{S} \times [0,T). \tag{6.30}$$

American-style derivatives

In a diffusion model with a 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 PDE (6.24) in the continuation region corresponding to the optimal exercise strategy, i.e. for $(x,t) \in \mathcal{C}^*$. But since the continuation region is not known, but is part of the solution, it is much harder to solve the PDE for American-style derivatives than for European-style derivatives. Explicit solutions have only been obtained in trivial cases where premature exercise is known to be inoptimal, and the American-style derivative is therefore effectively a European-style security. This is for example the case for an American call option on a zero-coupon bond. However, numerical solution techniques for PDEs can, with some modifications, also be applied to the case of American-style derivatives; see Chapter 16.

6.5.2 Multi-factor diffusion models

Assume now that the short-term interest rate and the securities 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

$$d\boldsymbol{x}_t = \boldsymbol{\alpha}(\boldsymbol{x}_t, t)\, dt + \underline{\underline{\beta}}(\boldsymbol{x}_t, t)\, d\boldsymbol{z}_t, \tag{6.31}$$

where $\boldsymbol{z}$ is a vector of n independent standard Brownian motions. Here,

$$\boldsymbol{\alpha}(\boldsymbol{x}_t, t) = (\alpha_1(\boldsymbol{x}_t, t), \dots, \alpha_n(\boldsymbol{x}_t, t))^\top$$

is the vector of expected changes in each of the n state variables, and $\underline{\underline{\beta}}(\boldsymbol{x}_t, t)$ is an $n \times n$-dimensional matrix, which contains information about the variances and covariances of changes in the different state variables. To be more precise, the instantaneous variance-covariance matrix of changes in the state variables is given by $\underline{\underline{\beta}}(\boldsymbol{x}_t, t)\underline{\underline{\beta}}(\boldsymbol{x}_t, t)^\top\, dt$. We can write (6.31) componentwise as

$$dx_{it} = \alpha_i(\boldsymbol{x}_t, t)\, dt + \boldsymbol{\beta}_i(\boldsymbol{x}_t, t)^\top\, d\boldsymbol{z}_t = \alpha_i(\boldsymbol{x}_t, t)\, dt + \sum_{j=1}^n \beta_{ij}(\boldsymbol{x}_t, t)\, dz_{jt}.$$

As discussed in Section 3.4, the covariance between changes in the i'th and the j'th state variable over the next infinitesimal time period is given by

$$\text{Cov}_t(dx_{it}, dx_{jt}) = \sum_{k=1}^n \beta_{ik}(\boldsymbol{x}_t, t)\beta_{jk}(\boldsymbol{x}_t, t)\, dt,$$

while the variance of the change in the i'th state variable is

$$\text{Var}_t(dx_{it}) = \sum_{k=1}^n \beta_{ik}(\boldsymbol{x}_t, t)^2\, dt.$$

In particular, the volatility of the i'th state variable is the standard deviation

$$\|\boldsymbol{\beta}_i(\boldsymbol{x}_t, t)\| = \sqrt{\sum_{k=1}^n \beta_{ik}(\boldsymbol{x}_t, t)^2},$$

where $\|\cdot\|$ denotes the length of the vector. The instantaneous correlation between changes in the i'th and the j'th state variable is

$$\rho_{ij}(\boldsymbol{x}_t, t) = \frac{\text{Cov}_t(dx_{it}, dx_{jt})}{\sqrt{\text{Var}_t(dx_{it})}\sqrt{\text{Var}_t(dx_{jt})}} = \frac{\sum_{k=1}^n \beta_{ik}(\boldsymbol{x}_t, t)\beta_{jk}(\boldsymbol{x}_t, t)}{\|\boldsymbol{\beta}_i(\boldsymbol{x}_t, t)\|\, \|\boldsymbol{\beta}_j(\boldsymbol{x}_t, t)\|}.$$

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 (see Theorem 3.6 on page 65) that the dynamics of P_t is

$$\frac{dP_t}{P_t} = \mu(\boldsymbol{x}_t, t)\,dt + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\,dz_{jt}, \tag{6.32}$$

where the functions μ and σ_j are defined as

$$\begin{aligned}\mu(\boldsymbol{x}, t)P(\boldsymbol{x}, t) = {} & \frac{\partial P}{\partial t}(\boldsymbol{x}, t) + \sum_{j=1}^{n} \frac{\partial P}{\partial x_j}(\boldsymbol{x}, t)\alpha_j(\boldsymbol{x}, t) \\ & + \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n} \frac{\partial^2 P}{\partial x_j \partial x_k}(\boldsymbol{x}, t)\rho_{jk}(\boldsymbol{x}, t)\,\|\boldsymbol{\beta}_j(\boldsymbol{x}, t)\|\,\|\boldsymbol{\beta}_k(\boldsymbol{x}, t)\|,\end{aligned} \tag{6.33}$$

$$\sigma_j(\boldsymbol{x}, t)P(\boldsymbol{x}, t) = \sum_{k=1}^{n} \frac{\partial P}{\partial x_k}(\boldsymbol{x}, t)\beta_{kj}(\boldsymbol{x}, t). \tag{6.34}$$

We also know that for a market price of risk $\boldsymbol{\lambda}(\boldsymbol{x}_t, t)$, we have

$$\mu(\boldsymbol{x}_t, t) = r(\boldsymbol{x}_t, t) + \sigma(\boldsymbol{x}_t, t)\boldsymbol{\lambda}(\boldsymbol{x}_t, t) = r(\boldsymbol{x}_t, t) + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\lambda_j(\boldsymbol{x}_t, t). \tag{6.35}$$

Substituting in μ and σ, we arrive at a PDE as summarized in the following multi-dimensional version of Theorem 6.2:

Theorem 6.3 *The function P defined by*

$$P(\boldsymbol{x}, t) = \mathrm{E}^{\mathbb{Q}}_{\boldsymbol{x},t}\left[e^{-\int_t^T r(\boldsymbol{x}_u, u)\,du} H(\boldsymbol{x}_T, T)\right] \tag{6.36}$$

satisfies the partial differential equation

$$\begin{aligned}&\frac{\partial P}{\partial t}(\boldsymbol{x}, t) + \sum_{j=1}^{n}\left(\alpha_j(\boldsymbol{x}, t) - \sum_{k=1}^{n}\beta_{jk}(\boldsymbol{x}, t)\lambda_k(\boldsymbol{x}, t)\right)\frac{\partial P}{\partial x_j}(\boldsymbol{x}, t) \\ &+ \frac{1}{2}\sum_{j=1}^{n}\sum_{k=1}^{n}\rho_{jk}(\boldsymbol{x}, t)\,\|\boldsymbol{\beta}_j(\boldsymbol{x}, t)\|\,\|\boldsymbol{\beta}_k(\boldsymbol{x}, t)\|\frac{\partial^2 P}{\partial x_j \partial x_k}(\boldsymbol{x}, t) - r(\boldsymbol{x}, t)P(\boldsymbol{x}, t) = 0, \quad (\boldsymbol{x}, t) \in \mathcal{S} \times [0, T),\end{aligned} \tag{6.37}$$

and the terminal condition

$$P(\boldsymbol{x}, T) = H(\boldsymbol{x}, T), \quad \boldsymbol{x} \in \mathcal{S}.$$

Using matrix notation the PDE can be written more compactly as

$$\begin{aligned}&\frac{\partial P}{\partial t}(\boldsymbol{x}, t) + \left(\boldsymbol{\alpha}(\boldsymbol{x}, t) - \underline{\underline{\beta}}(\boldsymbol{x}, t)\boldsymbol{\lambda}(\boldsymbol{x}, t)\right)^{\top}\frac{\partial P}{\partial \boldsymbol{x}}(\boldsymbol{x}, t) \\ &\quad + \frac{1}{2}\operatorname{tr}\left(\underline{\underline{\beta}}(\boldsymbol{x}, t)\underline{\underline{\beta}}(\boldsymbol{x}, t)^{\top}\frac{\partial^2 P}{\partial \boldsymbol{x}^2}(\boldsymbol{x}, t)\right) - r(\boldsymbol{x}, t)P(\boldsymbol{x}, t) = 0, \quad (\boldsymbol{x}, t) \in \mathcal{S} \times [0, T),\end{aligned} \tag{6.38}$$

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.

In Chapter 8 we will study specific multi-factor diffusion models of the term structure of interest rates. In some of these models explicit solutions to the PDE can be found for some important securities.

6.6 The Black-Scholes-Merton model and Black's variant

The best known model for derivative pricing is the Black-Scholes-Merton model developed by Black and Scholes (1973) and Merton (1973) for the pricing of European options on stocks. This model also serves as an example of a diffusion model. Practitioners often apply slightly modified versions of the Black-Scholes-Merton model and option pricing formula to price other derivatives 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, as we shall discuss at the end of the section, the use of the Black-76 approach to fixed-income securities is not theoretically justified.

6.6.1 The Black-Scholes-Merton model

The critical assumptions underlying the Black-Scholes-Merton option pricing model are that the riskless 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, i.e.

$$dS_t = \mu(S_t, t)\, dt + \sigma S_t\, dz_t, \tag{6.39}$$

where z is a standard Brownian motion, σ is a constant, and μ is a "nice" function. 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 $\mathcal{S} = \mathbb{R}_+$. Furthermore, we assume that the underlying asset has no payments in the life of the derivative security.

We seek to determine the price of a derivative security that at time T pays off $H(S_T, T)$, which depends on the price of the underlying asset S_T and on no other uncertain variables. The time t price P_t of the derivative asset is then given by $P_t = P(S_t, t)$ where

$$P(S,t) = \mathrm{E}^{\mathbb{Q}}_{S,t}\left[e^{-\int_t^T r\, du} H(S_T, T)\right] = e^{-r[T-t]}\, \mathrm{E}^{\mathbb{Q}}_t \left[H(S_T, T)\right]$$

and the risk-neutral dynamics of the underlying asset price is

$$dS_t = rS_t\, dt + \sigma S_t\, dz^{\mathbb{Q}}_t,$$

i.e. a geometric Brownian motion so that S_T is lognormally distributed. The function $P(S, t)$ solves the PDE

$$\frac{\partial P}{\partial t}(S,t) + rS\frac{\partial P}{\partial S}(S,t) + \frac{1}{2}\sigma^2 S^2 \frac{\partial^2 P}{\partial S^2}(S,t) = rP(S,t), \quad (S,t) \in \mathcal{S} \times [0, T). \tag{6.40}$$

In addition, the price function must satisfy the terminal condition of the form

$$P(S,T) = H(S,T), \quad \text{for all } S \in \mathcal{S}, \tag{6.41}$$

where $\mathcal{S}$ is the set of all possible terminal prices S_T of the underlying asset.

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 (6.40) with the relevant terminal condition or by calculating the discounted risk-neutral expected payoff, i.e.

$$C(S_t,t) = e^{-r[T-t]}\,\mathrm{E}^{\mathbb{Q}}_{S_t,t}\left[\max(S_T-K,0)\right].$$

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:[2]

$$C(S_t,t) = S_t N\left(d_1(S_t,t)\right) - Ke^{-r[T-t]} N\left(d_2(S_t,t)\right), \tag{6.42}$$

where

$$d_1(S_t,t) = \frac{\ln(S_t/K) + r[T-t]}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}, \tag{6.43}$$

$$d_2(S_t,t) = \frac{\ln(S_t/K) + r[T-t]}{\sigma\sqrt{T-t}} - \frac{1}{2}\sigma\sqrt{T-t} = d_1(S_t,t) - \sigma\sqrt{T-t}. \tag{6.44}$$

It can be verified that the function $C(S,t)$ defined in (6.42) solves the PDE (6.40) with the relevant terminal condition. Applying the put-call parity for European stock options, we get the price formula for a European put option:

$$\pi(S_t,t) = Ke^{-r[T-t]} N\left(-d_2(S_t,t)\right) - S_t N\left(-d_1(S_t,t)\right). \tag{6.45}$$

Strictly speaking, to derive the option pricing formulas above, the only assumption needed on the price of the underlying asset is that S_T (given S_t) is lognormally distributed in the risk-neutral world. Letting $\sigma\sqrt{T-t}$ denote the standard deviation of $\ln S_T$, the formulas above will hold. However, if we want to use the pricing formulas for options on the same underlying asset, but with different maturities, we must have that the uncertain prices of the underlying asset at different future points in time must all be lognormally distributed in the risk-neutral world. This is ensured by the assumption (6.39).

For the case where the underlying asset has payments in the life of the derivative asset, the same procedure applies with some minor corrections, which can be found in Hull (2003).

[2] 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.

6.6.2 Black's model

Black (1976) introduced a variant of the Black-Scholes-Merton model, which he applied for the pricing of European options on commodity futures. Let us consider a European call option that expires at T, has an exercise price K, and is written on a commodity futures expiring at T^*, where $T^* \geq T$. The futures price at time t is denoted by $\Phi_t^{T^*}$. The payoff of the option at time T is $\max(\Phi_T^{T^*} - K, 0)$.

To price this option, Black assumed a constant riskless interest rate and that the futures price $\Phi_T^{T^*}$ at expiry of the option is lognormally distributed in the risk-neutral world. The standard deviation of $\ln \Phi_T^{T^*}$ (with the information available at time t) is denoted by $\sigma\sqrt{T-t}$. It is further assumed that the expectation of the time T futures price equals the current futures price, i.e. $\mathrm{E}_t^{\mathbb{Q}}[\Phi_T^{T^*}] = \Phi_t^{T^*}$.

We can see that this is correct when the riskless interest rate is constant and the asset underlying the futures contract has a price S_t that follows a process of the form $dS_t = S_t[r\,dt + \sigma\,dz_t^{\mathbb{Q}}]$ in a risk-neutral world. As discussed in Section 2.5, the futures price is then identical to the forward price, which is given by $F_t^{T^*} = S_t e^{r[T^*-t]}$. From Itô's Lemma (Theorem 3.5 on page 55) we get that

$$dF_t^{T^*} = \sigma F_t^{T^*}\,dz_t^{\mathbb{Q}},$$

from which it follows that the expected change in the forward price (and hence the futures price) is equal to zero, so that $\mathrm{E}_t^{\mathbb{Q}}[\Phi_T^{T^*}] = \Phi_t^{T^*}$, and $\Phi_T^{T^*} = F_T^{T^*}$ is lognormally distributed with

$$\ln \Phi_T^{T^*} \sim N\left(\ln \Phi_t^{T^*} - \frac{1}{2}\sigma^2[T-t], \sigma^2[T-t]\right). \tag{6.46}$$

Furthermore, under these assumptions the volatility of the futures price equals the volatility of the price of the underlying asset.

According to the risk-neutral pricing principle, the option price can be written as

$$C_t = e^{-r[T-t]}\,\mathrm{E}_t^{\mathbb{Q}}\left[\max\left(\Phi_T^{T^*} - K, 0\right)\right].$$

Applying (6.46) and Theorem A.4 of Appendix A, we can compute the price as

$$C_t = e^{-r[T-t]}\left[\Phi_t^{T^*} N\left(\hat{d}_1(\Phi_t^{T^*}, t)\right) - KN\left(\hat{d}_2(\Phi_t^{T^*}, t)\right)\right], \tag{6.47}$$

where

$$\hat{d}_1(\Phi_t^{T^*}, t) = \frac{\ln(\Phi_t^{T^*}/K)}{\sigma\sqrt{T-t}} + \frac{1}{2}\sigma\sqrt{T-t}, \tag{6.48}$$

$$\hat{d}_2(\Phi_t^{T^*}, t) = \frac{\ln(\Phi_t^{T^*}/K)}{\sigma\sqrt{T-t}} - \frac{1}{2}\sigma\sqrt{T-t} = \hat{d}_1(\Phi_t^{T^*}, t) - \sigma\sqrt{T-t}. \tag{6.49}$$

The expression (6.47) is called **Black's formula**. For European put options we similarly have that

$$\pi_t = e^{-r[T-t]}\left[KN\left(-\hat{d}_2(\Phi_t^{T^*}, t)\right) - \Phi_t^{T^*} N\left(-\hat{d}_1(\Phi_t^{T^*}, t)\right)\right].$$

Analogously to the Black-Scholes-Merton model it is not strictly necessary to assume that the futures price $\Phi_t^{T^*}$ follows a geometric Brownian motion. It suffices that the future futures price $\Phi_T^{T^*}$ is lognormally distributed in a hypothetical risk-neutral world and that the expected change in the futures price equals zero. The parameter σ is then just a measure for the standard deviation over the life of the option, but it is still often referred to as the volatility of the futures price.

Since the original development, Black's model has been adapted by market participants for the pricing of various fixed income securities, such as bond options, caps/floors, and swaptions. Since the payoff of these securities depends on future interest rates, the original Black assumption of constant interest rates is of course inappropriate. The pricing formulas are derived by first computing the expected payoff in a risk-neutral world and then discounting by the current riskless discount factor. The uncertainty of future interest rates is taken into account when the expected payoff is computed, but not in the discounting. Let us look at some examples:

Bond options As in Section 2.7, we consider a European call option with expiration time T and exercise price K, written on a bond with price B_t. Black's model applied to bond options involves the forward price $F_t^{T,\text{cpn}}$ of the underlying bond with delivery at expiration of the option. According to Theorem 2.2 on page 24, the forward price is

$$F_t^{T,\text{cpn}} = \frac{\sum_{T_i > T} Y_i B_t^{T_i}}{B_t^T} = \frac{B_t - \sum_{t < T_i < T} Y_i B_t^{T_i}}{B_t^T}, \tag{6.50}$$

where $T_1 < T_2 < \cdots < T_n$ are the payment dates and Y_i denotes the payment of the bond at time T_i.

To apply Black's model, the following assumptions must be made:

(a) the futures price equals the forward price;

(b) the forward price at the expiration time of the option, $F_T^{T,\text{cpn}} = B_T$, is lognormally distributed in the risk-neutral world with the standard deviation of $\ln F_T^{T,\text{cpn}}$ given by $\sigma\sqrt{T-t}$;

(c) the expected change in the forward price of the bond between time t and T equals zero in the risk-neutral world;

(d) the option price can be computed as the risk-neutral expected payoff multiplied by the current riskless discount factor.

The distributional assumption is satisfied if the forward price $F_t^{T,\text{cpn}}$ follows a stochastic process with a constant relative volatility σ and a drift of zero. The expected payoff in a risk-neutral world is then

$$\mathrm{E}_t^{\mathbb{Q}}\left[\max\left(B_T - K, 0\right)\right] = F_t^{T,\text{cpn}} N\left(\hat{d}_1(F_t^{T,\text{cpn}}, t)\right) - KN\left(\hat{d}_2(F_t^{T,\text{cpn}}, t)\right),$$

where the functions $\hat{d}_1$ and $\hat{d}_2$ are as in (6.48)–(6.49). By multiplying the expected payoff with the riskless discount factor, i.e. the zero-coupon bond price B_t^T, we arrive at *Black's formula for a European call option on a bond*:

$$\begin{aligned} C_t^{K,T,\text{cpn}} &= B_t^T \left[F_t^{T,\text{cpn}} N\left(\hat{d}_1(F_t^{T,\text{cpn}}, t)\right) - KN\left(\hat{d}_2(F_t^{T,\text{cpn}}, t)\right)\right], \\ &= \left(B_t - \sum_{t<T_i<T} Y_i B_t^{T_i}\right) N\left(\hat{d}_1(F_t^{T,\text{cpn}}, t)\right) - KB_t^T N\left(\hat{d}_2(F_t^{T,\text{cpn}}, t)\right). \end{aligned} \tag{6.51}$$

Similarly, the price of a European put option on a bond is

$$\pi_t^{K,T,\text{cpn}} = KB_t^T N\left(-\hat{d}_2(F_t^{T,\text{cpn}}, t)\right) - \left(B_t - \sum_{t<T_i<T} Y_i B_t^{T_i}\right) N\left(-\hat{d}_1(F_t^{T,\text{cpn}}, t)\right). \tag{6.52}$$

Caps and floors As discussed in Section 2.8, a cap can be seen as a portfolio of caplets. The i'th caplet gives a payoff at time T_i of

$$\mathcal{C}^i_{T_i} = H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right), \tag{6.53}$$

cf. (2.11). Under the assumptions that

(a) seen from time t, the risk-neutral distribution of $l^{T_i}_{T_i-\delta} = L^{T_i-\delta,T_i}_{T_i-\delta}$ is lognormal with the standard deviation of $\ln l^{T_i}_{T_i-\delta}$ given by $\sigma_i\sqrt{T_i-\delta-t}$;

(b) the expected change in the forward rate $L^{T_i-\delta,T_i}_t$ between time t and $T_i-\delta$ is equal to zero in a risk-neutral world;

(c) we can discount the risk-neutral expected payoff with the current discount factor;

we obtain *Black's formula for the caplet price*

$$\mathcal{C}^i_t = H\delta B^{T_i}_t\left[L^{T_i-\delta,T_i}_t N\left(\hat{d}^i_1(L^{T_i-\delta,T_i}_t, t)\right) - KN\left(\hat{d}^i_2(L^{T_i-\delta,T_i}_t, t)\right)\right], \quad t < T_i - \delta, \tag{6.54}$$

where the functions $\hat{d}^i_1$ and $\hat{d}^i_2$ are given by

$$\hat{d}^i_1(L^{T_i-\delta,T_i}_t, t) = \frac{\ln(L^{T_i-\delta,T_i}_t/K)}{\sigma_i\sqrt{T_i-\delta-t}} + \frac{1}{2}\sigma_i\sqrt{T_i-\delta-t}, \tag{6.55}$$

$$\hat{d}^i_2(L^{T_i-\delta,T_i}_t, t) = \hat{d}^i_1(L^{T_i-\delta,T_i}_t, t) - \sigma_i\sqrt{T_i-\delta-t}. \tag{6.56}$$

The assumptions (a)-(b) are satisfied if the forward rate $L^{T_i-\delta,T_i}_t$ in the risk-neutral world follows the process

$$dL^{T_i-\delta,T_i}_t = \sigma_i L^{T_i-\delta,T_i}_t\, dz^{\mathbb{Q}}_t \tag{6.57}$$

with a constant volatility σ_i. The price for the entire cap is obtained by summation:

$$\mathcal{C}_t = H\delta\sum_{i=1}^n B^{T_i}_t\left[L^{T_i-\delta,T_i}_t N\left(\hat{d}^i_1(L^{T_i-\delta,T_i}_t, t)\right) - KN\left(\hat{d}^i_2(L^{T_i-\delta,T_i}_t, t)\right)\right], \quad t \le T_0. \tag{6.58}$$

As discussed in Section 2.8 the price must be adjusted slightly if the first payment is already known. For a floor the corresponding formula is

$$\mathcal{F}_t = H\delta\sum_{i=1}^n B^{T_i}_t\left[KN\left(-\hat{d}^i_2(L^{T_i-\delta,T_i}_t, t)\right) - L^{T_i-\delta,T_i}_t N\left(-\hat{d}^i_1(L^{T_i-\delta,T_i}_t, t)\right)\right], \quad t \le T_0. \tag{6.59}$$

Swaptions Let us look at a European payer swaption which was introduced in Section 2.9.2. From (2.33) on page 39 we have that the payoff of a payer swaption at the expiration time T_0 can be expressed as

$$\mathcal{P}_{T_0} = \left(\sum_{i=1}^n B^{T_i}_{T_0}\right) H\delta \max\left(\tilde{l}^\delta_{T_0} - K, 0\right),$$

where $\tilde{l}^\delta_{T_0}$ is the (equilibrium) swap rate, and K is the exercise rate. *Black's formula for the price of a European payer swaption* is

$$\mathcal{P}_t = H\delta\left(\sum_{i=1}^n B^{T_i}_t\right)\left[\tilde{L}^{\delta,T_0}_t N\left(\hat{d}_1(\tilde{L}^{\delta,T_0}_t, t)\right) - KN\left(\hat{d}_2(\tilde{L}^{\delta,T_0}_t, t)\right)\right], \quad t < T_0, \tag{6.60}$$

where the functions $\hat{d}_1$ and $\hat{d}_2$ are as in (6.48) and (6.49) with $T = T_0$. As in Section 2.9, $\tilde{L}_t^{\delta,T_0}$ is the forward swap rate. By analogy, the following expression for the price of a European receiver swaption is obtained:

$$\mathcal{R}_t = H\delta \left(\sum_{i=1}^{n} B_t^{T_i} \right) \left[KN\left(-\hat{d}_2(\tilde{L}_t^{\delta,T_0}, t)\right) - \tilde{L}_t^{\delta,T_0} N\left(-\hat{d}_1(\tilde{L}_t^{\delta,T_0}, t)\right)\right], \quad t < T_0, \tag{6.61}$$

The assumptions underlying the formula are that the swap rate $\tilde{l}_{T_0}^{\delta} = \tilde{L}_{T_0}^{\delta,T_0}$ at the expiration date of the swaption is lognormally distributed, or more precisely that $\ln \tilde{l}_{T_0}^{\delta} = \ln \tilde{L}_{T_0}^{\delta,T_0}$ is risk-neutrally normally distributed with variance $\sigma^2[T_0 - t]$, and that the risk-neutral expectation of the change in the forward swap rate is zero. These assumptions are satisfied if the forward swap rate $\tilde{L}_t^{\delta,T_0}$ in the risk-neutral world follows the stochastic process

$$d\tilde{L}_t^{\delta,T_0} = \sigma \tilde{L}_t^{\delta,T_0} \, dz_t^{\mathbb{Q}}. \tag{6.62}$$

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 that value of σ, which by substitution into the Black-Scholes-Merton formula (6.42), 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 are expressed in terms of implicit interest rate volatilities computed with reference to the Black pricing formula.

According to (6.58) 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 one-year and a two-year cap on the one-year LIBOR rate are traded, the unique value of σ_1 that makes Black's price equal to the market price of the one-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 two-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, i.e. 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.[3] 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 volatilities. The picture is the same whether implicit or historical forward rate volatilities are used.

If we consider formula (2.27) 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

$$\mathrm{Var}_t[\tilde{l}_{T_0}^{\delta}] = \mathrm{Var}_t \left[\sum_{i=1}^{n} w_i L_{T_0}^{T_i - \delta, T_i} \right] = \sum_{i=1}^{n} \sum_{j=1}^{n} w_i w_j \sigma_i \sigma_j \rho_{ij},$$

[3] See for example the discussion in Hull (2003, Ch. 22).

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.[4] 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 Black's formula for swaptions, cf. formula (6.60).

6.6.3 Problems in applying Black's model to fixed income securities

As already hinted upon above, the assumptions underlying the application of Black's formula on interest rate dependent securities are highly problematic. Let us take a closer look at the critical points.

Firstly, the lognormality assumption for bond prices and interest rates is doubtful. For several reasons the price of a bond cannot follow a geometric Brownian motion throughout its life. We know that the price converges to the terminal payment of the bond as the maturity date approaches. Furthermore, the bond price is limited from above by the sum of the future bond payments under the appropriate assumption that all forward rates are non-negative. When the bond price approaches its upper limit or the maturity date approaches, the volatility of the bond price has to go to zero. The volatility of the bond price will therefore depend on both the level of the price and the time to maturity. A lognormality assumption can at best be an approximation to the true distribution. In addition, the forward price and the futures price on a bond are not necessarily equal when the interest rate uncertainty is taken into account. It is less clear whether it is reasonable to assume that future interest rates are lognormally distributed, and that the expected changes in the forward rates and the forward swap rates are zero in a risk-neutral world. We will discuss this further in later chapters.

Secondly, the multiplication of the current discount factor and the risk-neutral expectation of the payoff does not lead to the correct price. In fact, as we have seen in Section 6.2.2, this is true if we take the expectation under the appropriate forward martingale measure instead of the risk-neutral measure.

Thirdly, simultaneous applications of Black's formula to different derivative securities are inconsistent. If for example we apply Black's formula for the pricing of a European option on zero-coupon bond, we must assume that the price of the zero-coupon bond is lognormally distributed. If we also apply Black's formula for the pricing of a European option on a coupon bond, we must assume that the price of the coupon bond is lognormally distributed. Since the price of the coupon bond is a weighted average of the prices of zero-coupon bonds, cf. (1.2) on page 6, and a sum of lognormally distributed random variables is not lognormally distributed, the assumptions are inconsistent.[5] Similarly, the swap rate is a linear combination of forward rates according to (2.27) on page 37. When Black's formula is applied for the pricing of caplets, it is implicitly assumed that the relevant

[4] These considerations are taken from Rebonato (1996, Sec. 1.4).

[5] A similar problem is present when the Black-Scholes-Merton formula is used both for the pricing of options on a stock index and options on the individual stocks. If the prices of the individual stocks are lognormally distributed, the value of the index will not be lognormally distributed. However, it can be shown that the distribution of a sum of "many" lognormally distributed random variables is very accurately approximated by a lognormal distribution with carefully selected parameters, cf. Turnbull and Wakeman (1991).

forward rates are lognormally distributed. Then the swap rate will not be lognormally distributed, so that it is inconsistent to use Black's formula for swaptions also. Furthermore, lognormality assumptions for both interest rates and bond prices are inconsistent.

Several research papers suggest other models for bond option pricing that are also based on specific assumptions on the evolution of the price of the underlying bond. The most prominent examples are Ball and Torous (1983) and Schaefer and Schwartz (1987). A critical analysis of such models can be seen in Rady and Sandmann (1994). A problem in applying these models is that the assumptions on the price dynamics for different bonds may be inconsistent, and hence the option pricing formula obtained in the model will only be valid for options on one particular bond.

To ensure consistent pricing of different fixed income securities we must model the evolution of the entire term structure of interest rates. In many of the consistent term structure models we shall discuss in the following chapters, we will obtain relatively simple and internally consistent pricing formulas for many of the popular fixed income securities. As we shall see in Chapter 11, it is in fact possible to construct consistent term structure models in which Black's formula is the correct pricing formula for some securities, but, even in those models, applications of Black's formula for different classes of securities are inconsistent.

6.7 An overview of continuous-time 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 has 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:[6]

(a) **flexible:** the model should be able to handle most situations of practical interest, i.e. 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 (e.g. 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;

[6] The presentation is in part based on Rogers (1995).

(f) **theoretically sound:** the model should be consistent with the broadly accepted principles for the behavior 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.

6.7.1 Overall categorization

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, i.e. 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, i.e. 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, e.g. 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, etc., 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 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 that this state variable follows a Markov diffusion model. 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 can only be applied in diffusion models. All well-known absolute pricing models of the term structure are diffusion models. In contrast, relative pricing models are typically formulated as non-diffusion models at the outset, which is natural since the evolution of the entire term structure must be modeled and the term structure consists of, in principle, infinitely many values. In order to enjoy the benefits of diffusion models, non-diffusion models are sometimes successfully reformulated as diffusion models. We will consider examples of this idea in Chapter 10.

6.7.2 Some frequently applied models

Apparently, the first dynamic model of the term structure of interest rates was introduced by Merton (1970). His model is a diffusion model with a single state variable, which is the short-term interest r_t itself. It makes good sense to use an interest rate as the state variable in models of bond prices and other interest rate derivatives. There is an obvious practical advantage in using the short rate as the state variable. As we have seen above it is necessary to specify how the short rate depends on the state variable chosen, which is evident when the short rate itself is the state variable. Furthermore, if another state variable x is used and r_t is a monotonic function of x_t, we can express all relevant functions in terms of r instead of x. Therefore we might as well use r as the state variable from the beginning. Merton's assumptions imply that the short rate follows a generalized Brownian motion under the risk-neutral probability measure, i.e.

$$dr_t = \hat{\varphi}\, dt + \beta\, dz_t^{\mathbb{Q}}, \tag{6.63}$$

where $\hat{\varphi}$ and β are constants.

Following Merton's idea, many other one-factor diffusion models with r as the single state variable have been suggested in the literature. They all take a certain risk-neutral dynamics of the short rate of the form

$$dr_t = \hat{\alpha}(r_t)\, dt + \beta(r_t)\, dz_t^{\mathbb{Q}},$$

where $\hat{\alpha}$ and β are well-behaved functions. Either the functional form of $\hat{\alpha}$ is assumed directly or it is derived from assumptions on the real-world drift α and the market price of risk λ. If the model should be used for more than just computing prices at a given date, it is necessary to know both α and λ. The pricing differences between the models stem from differences in the specification of the functions $\hat{\alpha}$ and β. It turns out that models in which $\hat{\alpha}$ and β are affine functions of the current value of r are particularly tractable and allow many closed-form pricing equations to be derived.[7] Such models are called **affine models**.

The two most famous term structure models are the one-factor affine diffusion models introduced by Vasicek (1977) and Cox, Ingersoll, and Ross (1985b). In the Vasicek model the basic assumption is that the short rate follows an Ornstein-Uhlenbeck process, cf. Section 3.8.2, and that the market price of risk is constant, i.e.

$$dr_t = \kappa[\theta - r_t]\, dt + \beta\, dz_t, \qquad \lambda_t = \lambda \text{ constant.} \tag{6.64}$$

The risk-neutral drift of the short rate is then $\kappa[\theta - r_t] - \beta\lambda = (\kappa\theta - \beta\lambda) - \kappa r_t$, which is affine in r_t. The variance rate is simply β^2, which is constant and hence a (degenerate) affine function of r_t. In the Cox-Ingersoll-Ross (or CIR) model the assumption is that the short rate follows a square-root process, cf. Section 3.8.3, and that the market price of risk is proportional to the square-root of the short rate, i.e.

$$dr_t = \kappa[\theta - r_t]\, dt + \beta\sqrt{r_t}\, dz_t, \qquad \lambda_t = \lambda\sqrt{r_t}. \tag{6.65}$$

[7] A function of the form $f(x) = a_0 + a_1 x$ is affine in x. The function is only linear in a strict mathematical terminology if $a_0 = 0$.

The risk-neutral drift of the short rate is then $\kappa[\theta - r_t] - \beta\lambda r_t = \kappa\theta - (\kappa + \beta\lambda)r_t$, which is affine in r_t. The variance rate is $\beta^2 r_t$, which is also affine in r_t. Chapter 7 provides a detailed analysis of these most important one-factor diffusion models and mentions a number of other, non-affine one-factor models.

A common problem for the one-factor models is that because all bond prices are assumed to be affected by a single exogenous variable, the price changes in any two bonds over an infinitesimal time period will be perfectly correlated, which conflicts with empirical evidence. Empirical studies by for example Litterman and Scheinkman (1991) and Stambaugh (1988) conclude that at least two, and perhaps three or four, state variables are necessary for the model to give a reasonable description of actual yield curve movements. This motivates the study of multi-factor diffusion models. In these models, the short rate is assumed to be a function of several state variables and each of these state variables are assumed to follow some diffusion process. Again it is common to distinguish between affine models and non-affine models, where an affine model is a model in which the risk-neutral drift rates and the variance and covariance rates of the state variables are affine functions of the current value of the state variables. Beaglehole and Tenney (1991) and Longstaff and Schwartz (1992a) have suggested two-factor affine models that extend the Vasicek model and the CIR model, respectively. We will review these and other multi-factor models in Chapter 8.

The one- and multi-factor diffusion models are absolute pricing models. They involve a small number of state variables and constant parameters. The derived prices and interest rates will also be functions of the state variables and these 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 (i.e. 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, e.g. 3 months, 6 months, and 12 months, with a similar compounding period. The market models are all based on as assumption 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 Concluding remarks

This chapter has further developed the consequences of asset pricing theory for the term structure modeling and derivatives pricing. The results obtained will be applied to specific models of the term structure of interest rates in the following chapters.

Chapter 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

$$dr_t = \alpha(r_t, t)\, dt + \beta(r_t, t)\, dz_t, \tag{7.1}$$

where $z = (z_t)_{t \geq 0}$ is a standard Brownian motion under the real-world probability measure $\mathbb{P}$. The market price of risk at time t is of the form $\lambda(r_t, t)$. The short rate dynamics under the risk-neutral probability measure $\mathbb{Q}$ (i.e. the spot martingale measure) is therefore

$$dr_t = \hat{\alpha}(r_t, t)\, dt + \beta(r_t, t)\, dz_t^{\mathbb{Q}}, \tag{7.2}$$

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, i.e. the set of values which the short rate can have with strictly positive probability.[1]

A model of the type (7.2) 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 $T - t$, cf. Theorem 7.1 below. In time inhomogeneous models, these considerations are not valid, which renders the analysis of such models more complicated. Furthermore, time homogeneity seems to be a realistic property: why should the drift and the volatility of the short rate depend on the calendar date? Surely, the drift and the volatility change

[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.

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 2, 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 and 6: 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 classical models of Merton (1970), Vasicek (1977), and Cox, Ingersoll, and Ross (1985b). Some non-affine models are outlined and discussed in Section 7.6. Section 7.7 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 7.8 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 6.2 on page 126 is then

$$\frac{\partial P}{\partial t}(r,t) + \hat{\alpha}(r)\frac{\partial P}{\partial r}(r,t) + \frac{1}{2}\beta(r)^2\frac{\partial^2 P}{\partial r^2}(r,t) - rP(r,t) = 0, \quad (r,t) \in \mathcal{S} \times [0,T), \tag{7.3}$$

with the terminal condition

$$P(r,T) = H(r), \quad r \in \mathcal{S}, \tag{7.4}$$

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-adjusted drift rate $\hat{\alpha}(r)$ and the variance rate $\beta(r)^2$ are affine functions of the short rate, i.e. of the form

$$\hat{\alpha}(r) = \hat{\varphi} - \hat{\kappa} r, \qquad \beta(r)^2 = \delta_1 + \delta_2 r, \tag{7.5}$$

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, i.e. 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

$$dr_t = (\hat{\varphi} - \hat{\kappa} r_t)\; dt + \sqrt{\delta_1 + \delta_2 r_t}\, dz_t^{\mathbb{Q}}. \tag{7.6}$$

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 (7.6), $B^T(r,t)$ 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 (7.3) with the terminal condition $B^T(r,T) = 1$.

Theorem 7.1 *In the model* (7.6) *the time t price of a zero-coupon bond maturing at time T is given as*

$$B^T(r,t) = e^{-a(T-t)-b(T-t)r}, \tag{7.7}$$

where the functions $a(\tau)$ and $b(\tau)$ satisfy the following system of ordinary differential equations:

$$\frac{1}{2}\delta_2 b(\tau)^2 + \hat{\kappa} b(\tau) + b'(\tau) - 1 = 0, \quad \tau > 0, \tag{7.8}$$

$$a'(\tau) - \hat{\varphi} b(\tau) + \frac{1}{2}\delta_1 b(\tau)^2 = 0, \quad \tau > 0, \tag{7.9}$$

together with the conditions $a(0) = b(0) = 0$.

Proof: We will show that the price $B^T(r,t)$ in (7.7) is a solution to the partial differential equation (7.3). 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

$$\begin{aligned} \frac{\partial B^T}{\partial t}(r,t) &= B^T(r,t)\left(a'(T-t) + b'(T-t)r\right), \\ \frac{\partial B^T}{\partial r}(r,t) &= -B^T(r,t)b(T-t), \\ \frac{\partial^2 B^T}{\partial r^2}(r,t) &= B^T(r,t)b(T-t)^2. \end{aligned} \tag{7.10}$$

After substituting these derivatives into (7.3) and dividing through by $B^T(r,t)$, we get

$$a'(T-t) + b'(T-t)r - b(T-t)\hat{\alpha}(r) + \frac{1}{2}b(T-t)^2\beta(r)^2 - r = 0, \quad (r,t) \in \mathcal{S} \times [0,T). \tag{7.11}$$

Substituting (7.5) into (7.11) and gathering terms involving r, we find that the functions a and b must satisfy the equation

$$\begin{aligned} &\left(a'(T-t) - \hat{\varphi} b(T-t) + \frac{1}{2}\delta_1 b(T-t)^2\right) \\ &\qquad + \left(\frac{1}{2}\delta_2 b(T-t)^2 + \hat{\kappa} b(T-t) + b'(T-t) - 1\right) r = 0, \quad (r,t) \in \mathcal{S} \times [0,T). \end{aligned}$$

This can only be true if (7.8) and (7.9) hold.[2] □

Conversely, it can be shown that the zero-coupon bond prices $B^T(r,t)$ are only of the exponential-affine form (7.7), if the drift rate and the variance rate are affine functions of the short rate as

[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.

in (7.5).[3]

The differential equations (7.8)–(7.9) are called **Ricatti equations**. The functions a and b are determined by first solving (7.8) with the condition $b(0) = 0$ to obtain the b-function. The solution to (7.9) with the condition $a(0) = 0$ can be written in terms of the b-function as

$$a(\tau) = \hat{\varphi} \int_0^\tau b(u)\, du - \frac{1}{2}\delta_1 \int_0^\tau b(u)^2\, du, \tag{7.12}$$

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 (7.6). 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

$$B^T(r,t) = e^{-y^T(r,t)(T-t)},$$

cf. (1.10) on page 9. It follows from (7.7) that

$$y^T(r,t) = -\frac{\ln B^T(r,t)}{T-t} = \frac{a(T-t)}{T-t} + \frac{b(T-t)}{T-t} r, \tag{7.13}$$

i.e. any zero-coupon rate is an affine function of the short rate. If b is positive, all zero-coupon rates 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 rates of short maturities more than zero-coupon rates of long maturities, which seems to be a reasonable property. Note that the zero-coupon rate for a fixed time to maturity of τ can be written as

$$y^{t+\tau}(r,t) = \frac{a(\tau)}{\tau} + \frac{b(\tau)}{\tau} r, \tag{7.14}$$

which is independent of t, which again stems from the time homogeneity of the model.

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

$$f^T(r,t) = -\frac{\frac{\partial B^T}{\partial T}(r,t)}{B^T(r,t)},$$

cf. (1.14) on page 9. From (7.7) we get that

$$f^T(r,t) = a'(T-t) + b'(T-t) r. \tag{7.15}$$

Hence, the forward rates are also affine in the short rate r. For a fixed time to maturity τ, the forward rate is

$$f^{t+\tau}(r,t) = a'(\tau) + b'(\tau) r.$$

[3] For details see Duffie (2001, Sec. 7E).

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 martingale measures are 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

$$dB_t^T = B_t^T \left[\left(r_t + \sigma^T(r_t,t)\lambda(r_t,t) \right) dt + \sigma^T(r_t,t)\, dz_t \right], \tag{7.16}$$

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 (7.10) that the sensitivity term is

$$\sigma^T(r,t) = -b(T-t)\beta(r). \tag{7.17}$$

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 $\sigma^T(r,t)$, i.e. $b(T-t)\beta(r)$. In equilibrium, risky assets will *normally* have an expected rate of return that exceeds the locally riskfree 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 (7.13) imply that

$$d\bar{y}_t^\tau = \frac{b(\tau)}{\tau}\alpha(r_t)\, dt + \frac{b(\tau)}{\tau}\beta(r_t)\, dz_t \tag{7.18}$$

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 $\alpha(r_t) = \hat{\alpha}(r_t) + \lambda(r_t)\beta(r_t)$ is 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 (7.15) 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

Theorem 2.1 on page 23 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 (7.7), the forward price becomes

$$F^{T,S}(r,t) = \exp\left\{ -\left[a(S-t) - a(T-t) \right] - \left[b(S-t) - b(T-t) \right] r \right\}, \tag{7.19}$$

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 6.3.2 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 can be found by solving the partial differential equation (6.30), which in a time homogeneous one-factor model is

$$\frac{\partial \Phi^{T,S}}{\partial t}(r,t) + \hat{\alpha}(r)\frac{\partial \Phi^{T,S}}{\partial r}(r,t) + \frac{1}{2}\beta(r)^2\frac{\partial^2 \Phi^{T,S}}{\partial r^2}(r,t) = 0, \quad \forall (r,t) \in \mathcal{S} \times [0,T), \tag{7.20}$$

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 (7.6). 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*

$$\Phi^{T,S}(r,t) = e^{-\tilde{a}(T-t)-\tilde{b}(T-t)r}, \tag{7.21}$$

where the functions $\tilde{a}(\tau)$ and $\tilde{b}(\tau)$ satisfy the following system of ordinary differential equations:

$$\frac{1}{2}\delta_2\tilde{b}(\tau)^2 + \hat{\kappa}\tilde{b}(\tau) + \tilde{b}'(\tau) = 0, \quad \tau \in (0,T), \tag{7.22}$$

$$\tilde{a}'(\tau) - \hat{\varphi}\tilde{b}(\tau) + \frac{1}{2}\delta_1\tilde{b}(\tau)^2 = 0, \quad \tau \in (0,T), \tag{7.23}$$

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 (7.23) with $\tilde{a}(0) = a(S-T)$ can generally be written as

$$\tilde{a}(\tau) = a(S-T) + \hat{\varphi}\int_0^\tau \tilde{b}(u)\,du - \frac{1}{2}\delta_1\int_0^\tau \tilde{b}(u)^2\,du. \tag{7.24}$$

The proof of this theorem is analogous to the proof of Theorem 7.1, since the PDE (7.20) is almost identical to the PDE (7.3) satisfied by the zero-coupon bond price. The last claim in the theorem above is left for the reader as Exercise 7.3. The claim implies that, for $\delta_2 = 0$, the futures price becomes

$$\Phi^{T,S}(r,t) = e^{-\tilde{a}(T-t)-[b(S-t)-b(T-t)]r}. \tag{7.25}$$

Comparing with the forward price expression (7.19), we see that, for $\delta_2 = 0$, we have

$$\frac{\frac{\partial F^{T,S}}{\partial r}(r,t)}{F^{T,S}(r,t)} = \frac{\frac{\partial \Phi^{T,S}}{\partial r}(r,t)}{\Phi^{T,S}(r,t)},$$

i.e. 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 Theorem 2.2 on page 24 that the forward price at time t for delivery at time T is given by

$$F^{T,\mathrm{cpn}}(r,t) = \sum_{T_i > T} Y_i F^{T,T_i}(r,t), \tag{7.26}$$

into which we can insert (7.19) on the right-hand side. From (6.14) we get that the same relation holds for futures prices:

$$\Phi^{T,\mathrm{cpn}}(r,t) = \sum_{T_i > T} Y_i \Phi^{T,T_i}(r,t), \tag{7.27}$$

into which we can insert (7.21) on the right-hand side.

For Eurodollar futures we have from (6.15) on page 123 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 (7.22)–(7.23) 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

$$\hat{a}(\tau) = -a(0.25) + \hat{\varphi}\int_0^\tau \hat{b}(u)\,du - \frac{1}{2}\delta_1\int_0^\tau \hat{b}(u)^2\,du. \tag{7.28}$$

The quoted Eurodollar futures price is therefore

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400e^{-\hat{a}(T-t)-\hat{b}(T-t)r}. \tag{7.29}$$

If $\delta_2 = 0$, we have $\hat{b}(\tau) = b(\tau) - b(\tau + 0.25)$.

7.2.3 European options on coupon bonds: Jamshidian's trick

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**.

Let us first recall the notation of Section 2.7. Since prices now depend on the short rate, we let $C^{K,T,S}(r,t)$ denote the price at time t, given the prevailing short rate $r_t = r$, of a European call option expiring at time T with an exercise price of K written on a zero-coupon bond paying one unit of account at time S. Here, $t \le T < S$. We also consider an option on a coupon bond with payments Y_i at time T_i ($i = 1, 2, \dots, n$), where $T_1 < T_2 < \dots < T_n$. The price of this 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. $C^{K,T,\mathrm{cpn}}(r,t)$ denotes the price at time t of a European call with expiry date T, an exercise price of K, and with the coupon bond as its underlying asset. Jamshidian's result can then be formulated as follows:

Theorem 7.3 *In an affine one-factor model, where the zero-coupon bond prices are given by* (7.7) *with* $b(\tau) > 0$ *for all* τ*, the price of a European call on a coupon bond is*

$$C^{K,T,cpn}(r,t) = \sum_{T_i > T} Y_i C^{K_i,T,T_i}(r,t), \tag{7.30}$$

where $K_i = B^{T_i}(r^*, T)$*, and* r^* *is defined as the solution to the equation*

$$B(r^*, T) = K. \tag{7.31}$$

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*, i.e.

$$B(r^*,T) = \sum_{T_i>T} Y_i B^{T_i}(r^*,T) = K. \tag{7.32}$$

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

$$\begin{aligned}\max\left(\sum_{T_i>T} Y_i B^{T_i}(r_T,T) - K, 0\right) &= \sum_{T_i>T} Y_i B^{T_i}(r_T,T) - K \\ &= \sum_{T_i>T} Y_i \left(B^{T_i}(r_T,T) - K_i\right) \\ &= \sum_{T_i>T} Y_i \max\left(B^{T_i}(r_T,T) - K_i, 0\right).\end{aligned}$$

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, 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:

$$\begin{aligned}
C^{K,T,\text{cpn}}(r,t) &= \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-\int_t^T r_u\,du}\max\left(B(r_T,T)-K,0\right)\right] \\
&= \mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-\int_t^T r_u\,du}\sum_{T_i>T} Y_i \max\left(B^{T_i}(r_T,T)-K_i,0\right)\right] \\
&= \sum_{T_i>T} Y_i\,\mathrm{E}^{\mathbb{Q}}_{r,t}\left[e^{-\int_t^T r_u\,du}\max\left(B^{T_i}(r_T,T)-K_i,0\right)\right] \\
&= \sum_{T_i>T} Y_i C^{K_i,T,T_i}(r,t),
\end{aligned}$$

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]

From (6.10) we get that 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

$$B^{T_i}(r_T,T) > K_i \;\Leftrightarrow\; r_T < r^*$$

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

$$\begin{aligned}
C^{K,T,\text{cpn}}(r,t) &= \sum_{T_i>T} Y_i\left\{B^{T_i}(r,t)\,\mathbb{Q}^{T_i}_{r,t}(r_T<r^*) - K_i B^T(r,t)\,\mathbb{Q}^{T}_{r,t}(r_T<r^*)\right\} \\
&= \sum_{T_i>T} Y_i B^{T_i}(r,t)\,\mathbb{Q}^{T_i}_{r,t}(r_T<r^*) - K B^T(r,t)\,\mathbb{Q}^{T}_{r,t}(r_T<r^*).
\end{aligned} \tag{7.33}$$

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.

[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 discuss the approximation more closely in Chapter 12.

7.3 Merton's model

7.3.1 The short rate process

Apparently, the first dynamic, 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, i.e.

$$dr_t = \hat{\varphi}\,dt + \beta\,dz_t^{\mathbb{Q}}, \tag{7.34}$$

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

$$r_T = r_t + \hat{\varphi}[T-t] + \beta[z_T^{\mathbb{Q}} - z_t^{\mathbb{Q}}], \quad t < T. \tag{7.35}$$

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

$$\mathrm{E}_{r,t}^{\mathbb{Q}}[r_T] = r + \hat{\varphi}[T-t]$$

and variance

$$\mathrm{Var}_{r,t}^{\mathbb{Q}}[r_T] = \beta^2[T-t].$$

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 them into your mattress until the loan is due. The difference between the proceeds and the repayment is a riskless 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 they earn a zero interest rate. Hence, both nominal and real interest rates should stay non-negative.[6]

---

[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.

7.3.2 Bond pricing

Merton's model is of the affine form (7.6) 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,

$$B^T(r,t) = e^{-a(T-t)-b(T-t)r}. \tag{7.36}$$

According to (7.8), the function $b(\tau)$ solves the simple ordinary differential equation $b'(\tau) = 1$ with $b(0) = 0$, which implies that

$$b(\tau) = \tau. \tag{7.37}$$

The function $a(\tau)$ can then be determined from (7.12):

$$a(\tau) = \hat{\varphi} \int_0^\tau u\, du - \frac{1}{2}\beta^2 \int_0^\tau u^2\, du = \frac{1}{2}\hat{\varphi}\tau^2 - \frac{1}{6}\beta^2\tau^3. \tag{7.38}$$

Note that 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. Equation (7.14) implies 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 (7.18) 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, all zero-coupon rates will change by the same. In other words, the yield curve will only change by parallel shifts. (See also Exercise 7.1.) 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 relatively simple setting.

7.3.4 Forwards and futures

By substituting the expressions (7.37) and (7.38) into (7.19), we get that the forward price on a zero-coupon bond under Merton's assumptions is

$$F^{T,S}(r,t) = \exp\left\{-\frac{1}{2}\left[(S-t)^2 - (T-t)^2\right] + \frac{1}{6}\beta^2\left[(S-t)^3 - (T-t)^3\right] - (S-T)r\right\}.$$

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 (7.24), the futures price can be written as

$$\Phi^{T,S}(r,t) = \exp\left\{\frac{1}{2}\hat{\varphi}(S-T)(S+T-2t) - \frac{1}{6}\beta^2(S-T)^2(2T+S-3t) - (S-T)r\right\}.$$

Forward and futures prices on coupon bonds can be found by inserting the expressions above into (7.26) and (7.27).

In Eq. (7.29), we get $\hat{b}(\tau) = b(\tau) - b(\tau + 0.25) = -0.25$ and from (7.28) we conclude that

$$\hat{a}(\tau) = -a(0.25) - 0.25\hat{\varphi}\tau - \frac{1}{2}(0.25)^2\beta^2\tau = -\frac{1}{2}(0.25)^2\hat{\varphi} + \frac{1}{6}(0.25)^3\beta^2 - 0.25\hat{\varphi}\tau - \frac{1}{2}(0.25)^2\beta^2\tau.$$

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

To price European options on zero-coupon bonds in Merton's setting, we shall apply the T-forward martingale measure technique introduced in Section 6.2.2 on page 117. The price on a European call option with expiry date T and exercise price K written on a zero-coupon bond maturing at time S is generally given by

$$C^{K,T,S}(r,t) = B^T(r,t)\,\mathrm{E}_{r,t}^{\mathbb{Q}^T}\left[\max\left(B^S(r_T,T) - K, 0\right)\right]. \tag{7.39}$$

To compute the expectation on the right-hand side we must know the distribution of $B^S(r_T,T)$ under the T-forward martingale measure given that $r_t = r$. Recall that the forward price for delivery at time T of a zero-coupon bond expiring at time S according to (2.3) on page 23 is

$$F_t^{T,S} = \frac{B_t^S}{B_t^T}.$$

In particular, $F_T^{T,S} = B_T^S$ is determining the payoff of the option.

We will find the distribution of $B_T^S = B^S(r_T,T)$ by deriving the dynamics of the forward price $F_t^{T,S}$. From Section 6.2.2 we have that the forward price $F_t^{T,S}$ is a martingale under the T-forward martingale measure, so the forward price has a drift rate of zero under this probability measure. The forward price is a function of the prices of two zero-coupon bonds. We can therefore determine the volatility of the forward price by an application of Itô's Lemma for functions of multiple stochastic processes (see Theorem 3.6). First note that the volatility of the forward price will depend only on the volatilities of the two zero-coupon bond prices, so that we need not worry about the drift rates of these prices. Also note that volatilities are invariant to changes of probability measure. In Merton's model, the relative volatility of the zero-coupon bond maturing at time S is $\sigma^S(r_t,t) = -\beta[S-t]$, cf. (7.17) and (7.37), so that

$$dB_t^S = \dots\, dt - \beta[S-t]B_t^S\, dz_t^T, \qquad dB_t^T = \dots\, dt - \beta[T-t]B_t^T\, dz_t^T.$$

It now follows from Itô's Lemma (check it!) that

$$\begin{aligned} dF_t^{T,S} &= \left(\sigma^S(r_t,t) - \sigma^T(r_t,t)\right)F_t^{T,S}\, dz_t^T \\ &= -\beta[S-T]F_t^{T,S}\, dz_t^T. \end{aligned} \tag{7.40}$$

Hence, the forward price follows a geometric Brownian motion with a drift rate of zero (under the T-forward martingale measure). In particular, $\ln B_T^S = \ln F_T^{T,S}$ is normally distributed with variance

$$v(t,T,S)^2 \equiv \operatorname{Var}_{r,t}^{Q^T}\left[\ln F_T^{T,S}\right] = \beta^2(S-T)^2(T-t) \tag{7.41}$$

and mean

$$m(t,T,S) \equiv \mathrm{E}_{r,t}^{Q^T}\left[\ln F_T^{T,S}\right] = \ln F_t^{T,S} - \frac{1}{2}v(t,T,S)^2, \tag{7.42}$$

cf. the analysis of the geometric Brownian motion in Section 3.8.1 on page 56.

We can now compute the price of the call option in (7.39) by an application of Theorem A.4 in Appendix A:

$$\begin{aligned} C^{K,T,S}(r,t) &= B^T(r,t)\,\mathrm{E}_{r,t}^{\mathbb{Q}^T}\left[\max\left(F^{T,S}(r_T,T)-K,0\right)\right] \\ &= B^T(r,t)\left\{\mathrm{E}_{r,t}^{\mathbb{Q}^T}\left[F^{T,S}(r_T,T)\right]N(d_1) - KN(d_2)\right\} \\ &= B^S(r,t)N(d_1) - KB^T(r,t)N(d_2), \end{aligned} \tag{7.43}$$

where we have used the fact that

$$\mathrm{E}_{r,t}^{\mathbb{Q}^T}\left[F^{T,S}(r_T,T)\right] = F^{T,S}(r,t) = \frac{B^S(r,t)}{B^T(r,t)}.$$

Here, d_1 and d_2 are given by

$$d_1 = \frac{m(t,T,S)-\ln K}{v(t,T,S)} + v(t,T,S) = \frac{1}{v(t,T,S)}\ln\left(\frac{B^S(r,t)}{KB^T(r,t)}\right) + \frac{1}{2}v(t,T,S), \tag{7.44}$$

$$d_2 = d_1 - v(t,T,S) = \frac{1}{v(t,T,S)}\ln\left(\frac{B^S(r,t)}{KB^T(r,t)}\right) - \frac{1}{2}v(t,T,S), \tag{7.45}$$

and $v(t,T,S) = \beta[S-T]\sqrt{T-t}$.

Note that the pricing formula (7.43) has the same structure as the Black-Scholes-Merton formula: the price on the underlying asset multiplied by one probability minus the present value of the exercise price multiplied by another probability. In addition, the relevant "uncertainty index" is here $v(t,T,S)$, where

$$v(t,T,S)^2 = \operatorname{Var}_{r,t}^{\mathbb{Q}^T}\left[\ln B^S(r_T,T)\right] = \operatorname{Var}_{r,t}\left[\ln B^S(r_T,T)\right],$$

which corresponds to the term $\sigma^2[T-t] = \operatorname{Var}_{S,t}[\ln S_T]$ in the Black-Scholes-Merton formula.

The price of a European call option on a coupon bond can be found by combining the pricing formula (7.43) and Jamshidian's trick of Theorem 7.3:

$$\begin{aligned} C^{K,T,\mathrm{cpn}}(r,t) &= \sum_{T_i>T} Y_i\left\{B^{T_i}(r,t)N(d_1^i) - K_iB^T(r,t)N(d_2^i)\right\} \\ &= \sum_{T_i>T} Y_iB^{T_i}(r,t)N(d_1^i) - B^T(r,t)\sum_{T_i>T} Y_iK_iN(d_2^i) \\ &= \sum_{T_i>T} Y_iB^{T_i}(r,t)N(d_1^i) - KB^T(r,t)N(d_2^i), \end{aligned} \tag{7.46}$$

where

$$\begin{aligned} d_1^i &= \frac{1}{v(t,T,T_i)}\ln\left(\frac{B^{T_i}(r,t)}{K_iB^T(r,t)}\right) + \frac{1}{2}v(t,T,T_i), \\ d_2^i &= d_1^i - v(t,T,T_i), \\ v(t,T,T_i) &= \beta[T_i-T]\sqrt{T-t}, \end{aligned}$$

and we have used the fact that the d_2^i's are identical, cf. 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 future 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:

$$dr_t = \kappa[\theta - r_t]\,dt + \beta\,dz_t, \tag{7.47}$$

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 on page 59, 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. 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

$$\mathrm{E}_{r,t}\left[r_T\right] = \theta + (r - \theta)e^{-\kappa[T-t]}, \tag{7.48}$$

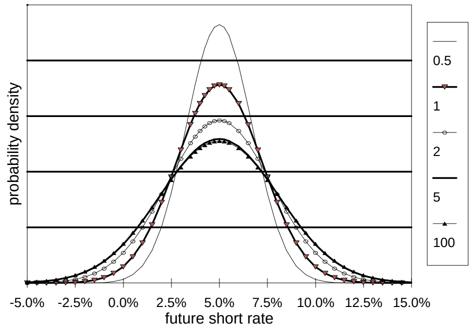
$$\mathrm{Var}_{r,t}\left[r_T\right] = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa[T-t]}\right) \tag{7.49}$$

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 $T - t = (\ln 2)/\kappa$.

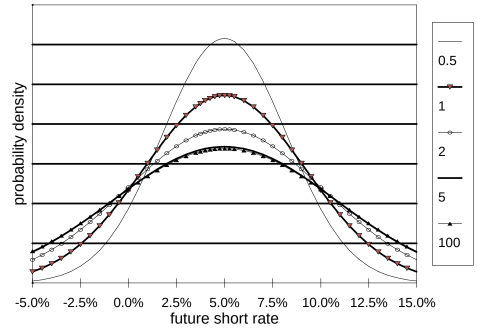
Like other Gaussian models, Vasicek's model assigns a positive probability to negative values of the future short rate (and all other future rates), despite the inappropriateness of this property. Figure 7.1 illustrates the distribution of the short rate $\frac{1}{2}, 1, 2, 5$, and 100 years into the 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

$$\mathbb{P}_{r,t}(r_T < 0) = \mathbb{P}_{r,t}\left(\frac{r_T - \mathrm{E}_{r,t}[r_T]}{\sqrt{\mathrm{Var}_{r,t}[r_T]}} < -\frac{\mathrm{E}_{r,t}[r_T]}{\sqrt{\mathrm{Var}_{r,t}[r_T]}}\right) = N\left(-\frac{\mathrm{E}_{r,t}[r_T]}{\sqrt{\mathrm{Var}_{r,t}[r_T]}}\right),$$

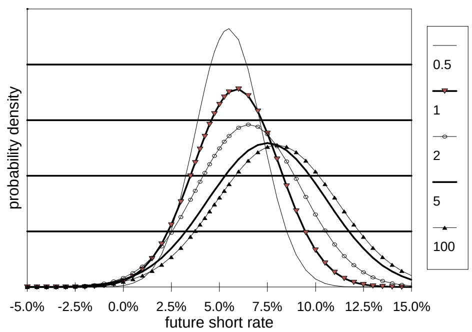
into which we can insert (7.48) and (7.49). Clearly, this probability is increasing in the interest rate volatility β and decreasing in the speed of adjustment κ, in the long-term level θ, and in the current level of the short rate r.



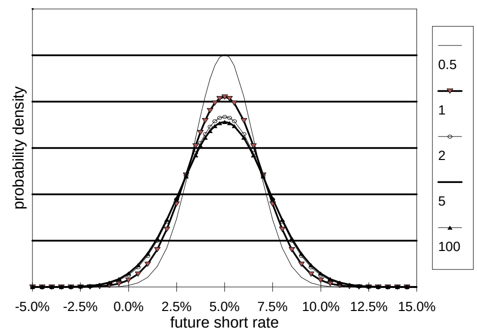
(a) $\kappa = 0.36,\ \theta = 0.05,\ \beta = 0.0265$



(b) $\kappa = 0.36,\ \theta = 0.05,\ \beta = 0.05$



(c) $\kappa = 0.36,\ \theta = 0.08,\ \beta = 0.0265$



(d) $\kappa = 0.7,\ \theta = 0.05,\ \beta = 0.0265$

Figure 7.1: The distribution of r_T for $T - t = 0.5, 1, 2, 5, 100$ years given a current short rate of $r_t = 0.05$.

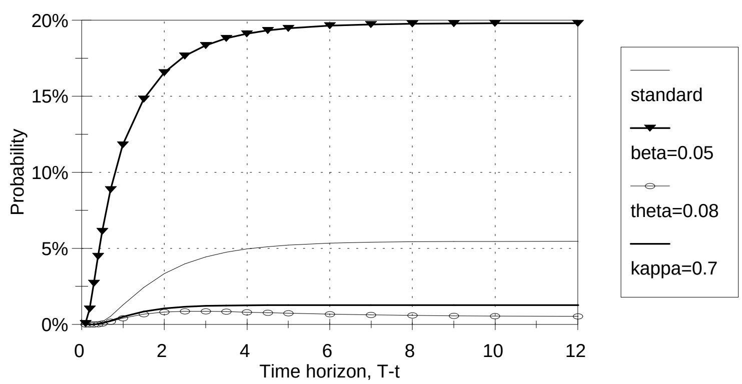


Figure 7.2: The probability that r_T is negative given $r_t = 0.05$ as a function of the horizon $T - t$. The benchmark parameter values are $\kappa = 0.36$, $\theta = 0.05$, and $\beta = 0.0265$.

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

$$\begin{aligned} dr_t &= \kappa[\theta - r_t]\,dt + \beta\left(dz_t^{\mathbb{Q}} - \lambda\,dt\right) \\ &= \kappa[\hat{\theta} - r_t]\,dt + \beta\,dz_t^{\mathbb{Q}}, \end{aligned} \tag{7.50}$$

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 (7.50) is of the form (7.6) 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

$$B^T(r,t) = e^{-a(T-t)-b(T-t)r}, \tag{7.51}$$

where $b(\tau)$ satisfies the ordinary differential equation

$$\kappa b(\tau) + b'(\tau) - 1 = 0, \quad b(0) = 0,$$

which has the solution

$$b(\tau) = \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right), \tag{7.52}$$

and from (7.12) we get

$$a(\tau) = \kappa\hat{\theta}\int_0^\tau b(u)\,du - \frac{1}{2}\beta^2\int_0^\tau b(u)^2\,du = y_\infty[\tau - b(\tau)] + \frac{\beta^2}{4\kappa}b(\tau)^2. \tag{7.53}$$

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", i.e. 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

$$\frac{\partial B^T}{\partial r}(r,t) = -b(T-t)B^T(r,t), \qquad \frac{\partial^2 B^T}{\partial r^2}(r,t) = b(T-t)^2 B^T(r,t).$$

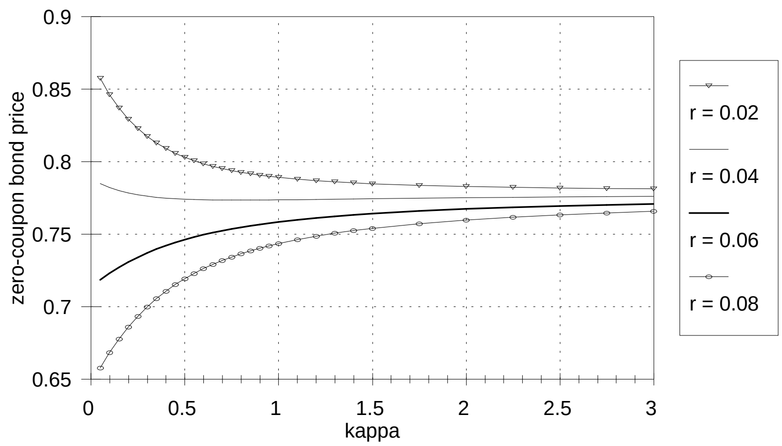


Figure 7.3: The price of a 5 year zero-coupon bond as a function of the speed of adjustment parameter κ for different values of the current short rate r. The other parameter values are $\theta = 0.05$, $\beta = 0.03$, and $\lambda = -0.15$.

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 Figure 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 Figure 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 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 (7.16) on page 148 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.

[7] Note that $\hat{\theta}$ goes to θ for $\kappa \to \infty$.

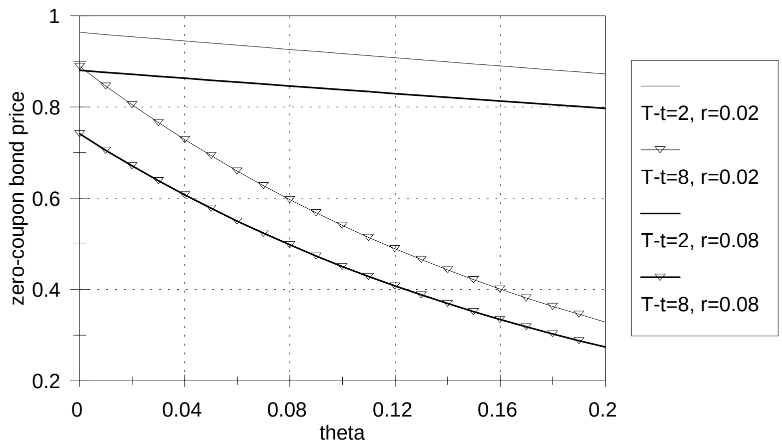


Figure 7.4: The price of a zero-coupon bond $B^T(r,t)$ as a function of the long-term level θ for different combinations of the time to maturity and the current short rate. The other parameter values are $\kappa = 0.3$, $\beta = 0.03$, and $\lambda = -0.15$.

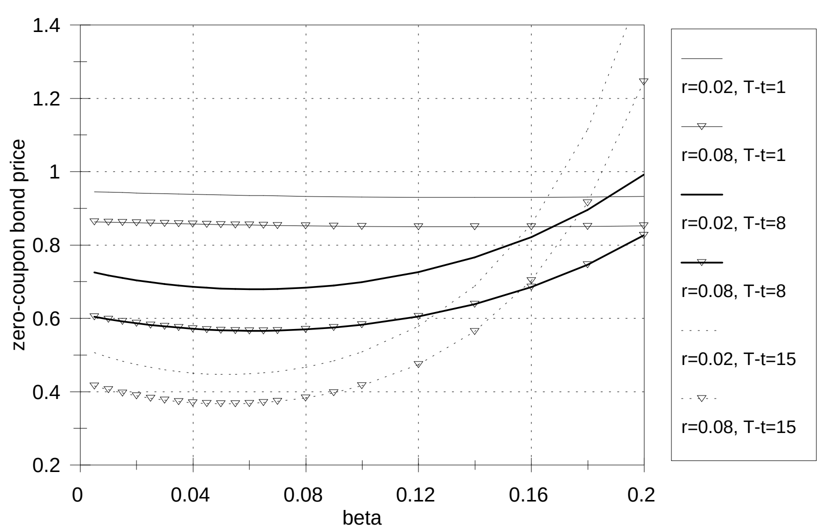


Figure 7.5: The price of a zero-coupon bond $B^T(r,t)$ as a function of the volatility parameter β for different combinations of the time to maturity $T - t$ and the current short rate r. The values of the fixed parameters are $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = -0.15$.

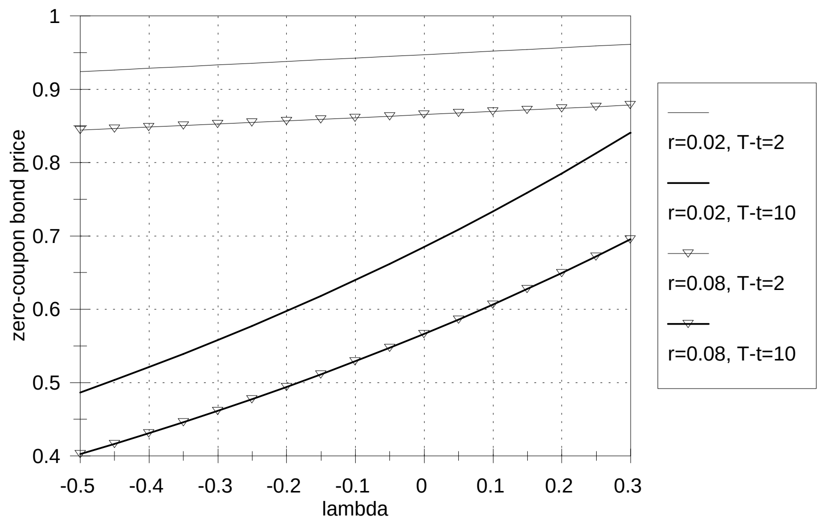


Figure 7.6: The price of zero-coupon bonds $B^T(r,t)$ as a function of λ for different combinations of the time to maturity $T-t$ and the current short rate r. The values of the fixed parameters are $\kappa = 0.3$, $\theta = 0.05$, and $\beta = 0.03$.

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, Figure 7.7 depicts the discount function, i.e. 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 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 (7.13) on page 147 the zero-coupon rate $y^T(r,t)$ 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

$$a'(\tau) = y_\infty[1 - b'(\tau)] + \frac{\beta^2}{2\kappa} b(\tau)b'(\tau), \tag{7.54}$$

$$b'(\tau) = e^{-\kappa\tau}, \tag{7.55}$$

so that an application of l'Hôspital's rule implies that

$$\lim_{\tau\to 0} \frac{b(\tau)}{\tau} = 1 \qquad \text{and} \qquad \lim_{\tau\to 0} \frac{a(\tau)}{\tau} = 0,$$

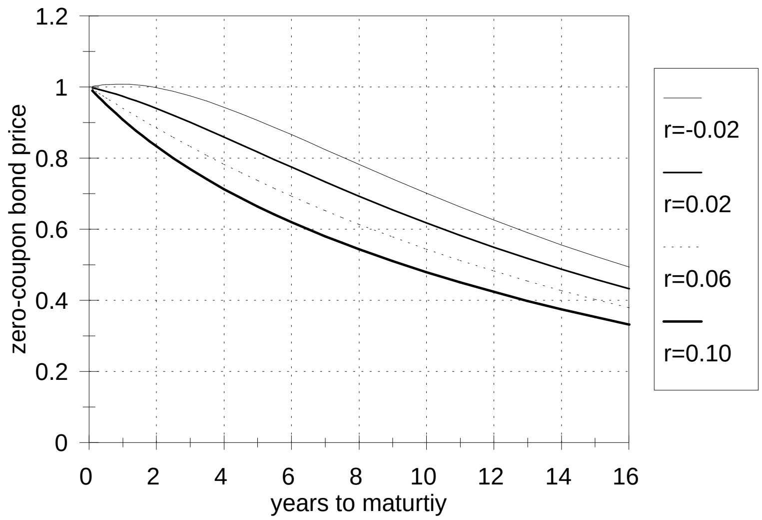


Figure 7.7: The price of zero-coupon bonds $B^T(r,t)$ as a function of the time to maturity $T-t$. The parameter values are $\kappa = 0.3$, $\theta = 0.05$, $\beta = 0.03$, and $\lambda = -0.15$.

and thus

$$\lim_{T\to t} y^T(r,t) = r,$$

i.e. 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

$$\lim_{T\to\infty} y^T(r,t) = y_\infty.$$

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, i.e. increasing in T up to some maturity T^* and then decreasing for longer maturities.

Proof: The zero-coupon rate $y^T(r,t)$ 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 (7.53). We are interested in the relation between the zero-coupon rate and the time to maturity $T - t$, i.e. the function $Y(\tau) = y^{t+\tau}(r,t)$. Defining $h(\tau) = b(\tau)/\tau$, we have that

$$Y(\tau) = y_\infty + h(\tau)\left(\frac{\beta^2}{4\kappa}b(\tau) + r - y_\infty\right).$$

A straightforward computation gives 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

$$Y'(\tau) = h'(\tau)\left(r - y_\infty + \frac{\beta^2}{4\kappa}g(\tau)\right). \tag{7.56}$$

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 (7.56) 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, i.e. $Y'(\tau) < 0$ 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 (7.56) 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 shows 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 (7.15) on page 147, the instantaneous forward rate $f^T(r,t)$ prevailing at time t is given by

$$f^T(r,t) = a'(T-t) + b'(T-t)r.$$

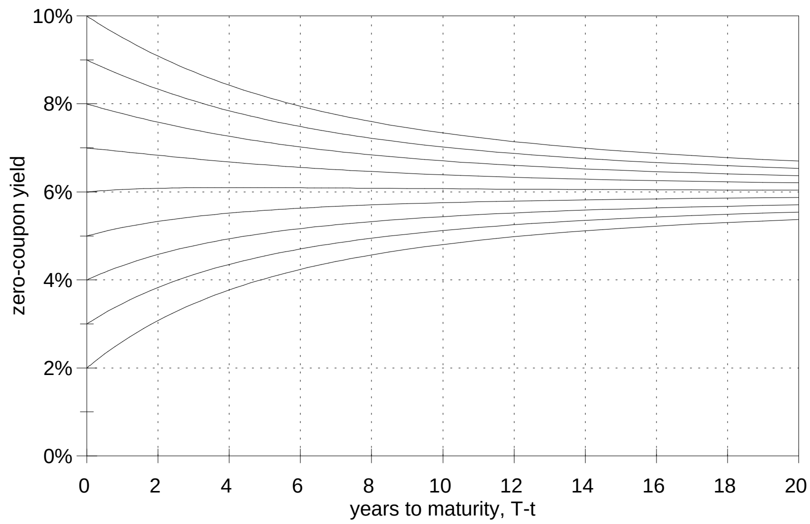


Figure 7.8: The yield curve for different values of the short rate. The parameter values are $\kappa = 0.3$, $\theta = 0.05$, $\beta = 0.03$, and $\lambda = -0.15$. The long rate is then $y_\infty = 6\%$. The yield curve is increasing for $r < 5.75\%$, decreasing for $r > 6.5\%$, and *humped* for intermediate 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 (7.54) and (7.55) this expression can be rewritten as

$$\begin{aligned} f^T(r,t) &= -\left(1 - e^{-\kappa[T-t]}\right)\left(\frac{\beta^2}{2\kappa^2}\left(1 - e^{-\kappa[T-t]}\right) - \hat{\theta}\right) + e^{-\kappa[T-t]} r \\ &= \left(1 - e^{-\kappa[T-t]}\right)\left(y_\infty + \frac{\beta^2}{2\kappa^2} e^{-\kappa[T-t]}\right) + e^{-\kappa[T-t]} r. \end{aligned} \tag{7.57}$$

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 (7.52) and (7.53) into the general expression

$$F^{T,S}(r,t) = \exp\left\{-\left[a(S-t) - a(T-t)\right] - \left[b(S-t) - b(T-t)\right] r\right\},$$

cf. (7.19).

In Vasicek's model the δ_2 parameter in the general dynamics (7.6) 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 (7.24) 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 (7.26) and (7.27).

For Eurodollar futures, (7.29) 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 (7.28) 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

To find the price of a European call option on a zero-coupon bond in Vasicek's model we follow the same procedure as in Merton's model. The price can generally be written as

$$C^{K,T,S}(r,t) = B^T(r,t)\, \mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[\max\left(F^{T,S}(r_T,T) - K, 0\right)\right],$$

where $F^{T,S}(r_T,T) = B^S(r_T,T)$. We seek the distribution of the forward price $F^{T,S}_T = F^{T,S}(r_T,T)$ under the T-forward martingale measure $\mathbb{Q}^T$. We have already seen that $\sigma^S(r_t,t) = -\beta b(S-t)$ in Vasicek's model, so that the dynamics of the forward price becomes

$$\begin{aligned} dF^{T,S}_t &= \left(\sigma^S(r_t,t) - \sigma^T(r_t,t)\right) F^{T,S}_t\, dz^T_t \\ &= -\beta[b(S-t) - b(T-t)] F^{T,S}_t\, dz^T_t. \end{aligned} \tag{7.58}$$

Hence, the forward price follows a geometric Brownian motion with a drift rate of zero and a time-dependent volatility (under the T-forward martingale measure). In particular, $\ln B^S_T = \ln F^{T,S}_T$ is normally distributed with variance

$$\begin{aligned} v(t,T,S)^2 \equiv \mathrm{Var}^{Q^T}_{r,t}\left[\ln F^{T,S}_T\right] &= \beta^2 \int_t^T \left[b(S-u) - b(T-u)\right]^2\, du \\ &= \frac{\beta^2}{2\kappa^3}\left(1 - e^{-\kappa[S-T]}\right)^2 \left(1 - e^{-2\kappa[T-t]}\right) \end{aligned} \tag{7.59}$$

and mean

$$m(t,T,S) \equiv \mathrm{E}^{Q^T}_{r,t}\left[\ln F^{T,S}_T\right] = \ln F^{T,S}_t - \frac{1}{2} v(t,T,S)^2, \tag{7.60}$$

cf. Section 3.8.1 on page 56. Theorem A.4 in Appendix A now implies that

$$\begin{aligned} C^{K,T,S}(r,t) &= B^T(r,t)\, \mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[\max\left(F^{T,S}(r_T,T) - K, 0\right)\right] \\ &= B^T(r,t)\left\{ \mathrm{E}^{\mathbb{Q}^T}_{r,t}\left[F^{T,S}(r_T,T)\right] N(d_1) - K N(d_2) \right\} \\ &= B^S(r,t) N(d_1) - K B^T(r,t) N(d_2), \end{aligned} \tag{7.61}$$

where

$$d_1 = \frac{1}{v(t,T,S)} \ln\left(\frac{B^S(r,t)}{K B^T(r,t)}\right) + \frac{1}{2} v(t,T,S), \tag{7.62}$$

$$d_2 = d_1 - v(t,T,S). \tag{7.63}$$

This result was first derived by Jamshidian (1989).

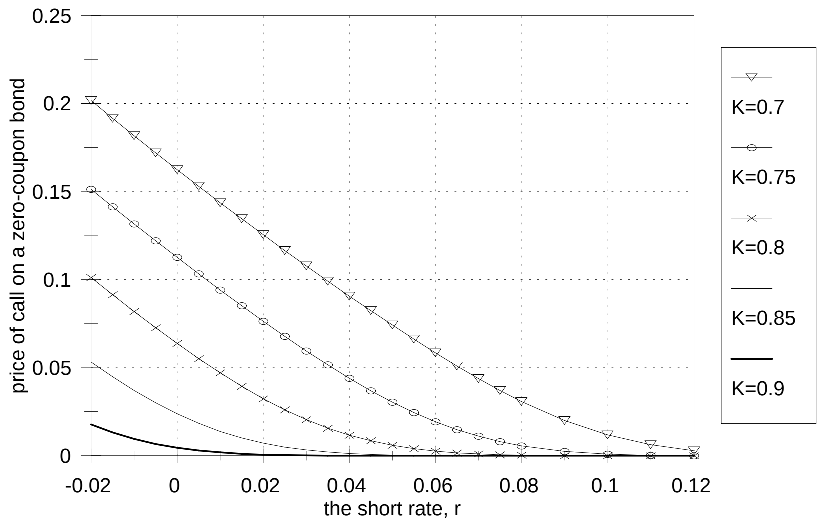


Figure 7.9: The price of a European call option on a zero-coupon bond as a function of the current short rate r. The option expires in $T - t = 0.5$ years, while the bond matures in $S - t = 5$ years. The prices are computed using Vasicek's model with parameter values $\beta = 0.03$, $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = -0.15$.

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 7.2 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 Figure 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 Figure 7.5, the bond price will decrease with β for low values of β, and this effect can be so strong that the option 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:

$$\begin{aligned} C^{K,T,\mathrm{cpn}}(r,t) &= \sum_{T_i > T} Y_i \left\{ B^{T_i}(r,t) N(d_1^i) - K_i B^T(r,t) N(d_2^i) \right\} \\ &= \sum_{T_i > T} Y_i B^{T_i}(r,t) N(d_1^i) - K B^T(r,t) N(d_2^i), \end{aligned} \tag{7.64}$$

where K_i is defined as $K_i = B^{T_i}(r^*, T)$, r^* is given as the solution to the equation $B(r^*, T) = K$,

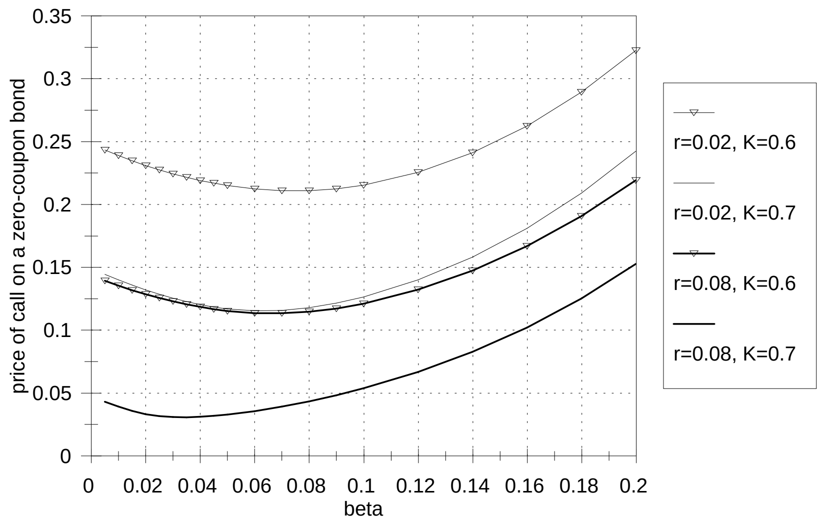


Figure 7.10: The price of a European call option on a zero-coupon bond as a function of the interest rate volatility β. The option expires in $T - t = 0.5$ years, while the bond matures in $\tau - t = 5$ years. The prices are computed using Vasicek's model with the parameter values $\kappa = 0.3$, $\theta = 0.05$, and $\lambda = -0.15$.

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 used that we know that all the d_2^i's are identical.

7.5 The Cox-Ingersoll-Ross model

7.5.1 The short rate process

Probably the most popular one-factor model, both among academics and practitioners, was suggested by Cox, Ingersoll, and Ross (1985b). They assume that the short rate follows a square root process

$$dr_t = \kappa \left[\theta - r_t\right] dt + \beta \sqrt{r_t}\, dz_t, \tag{7.65}$$

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. The only difference relative to Vasicek's short rate process is the specification of the volatility, which 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 behavior – whether the relation between volatility and short rate is of the form $\beta\sqrt{r}$

is not so clear, cf. the discussion in Section 7.7. The short rate in the CIR model cannot become negative, which is a major advantage relative to Vasicek's model. 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; see Section 3.8.3 for details.

As discussed in Section 5.4, the CIR model is a special case of a comprehensive general equilibrium model of the financial markets developed by the same authors in another article, Cox, Ingersoll, and Ross (1985a). The short rate process (7.65) *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

$$dr_t = (\hat{\varphi} - \hat{\kappa} r_t)\, dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}}. \tag{7.66}$$

Since this is of the form (7.6) 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

$$dr_t = \hat{\kappa}\left[\hat{\theta} - r_t\right]\, dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}},$$

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 changed 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. To be more precise, the probability density function (under the real-world probability measure) for r_T given r_t is

$$f_{r_T|r_t}(r) = f_{\chi^2_{2q+2,2u}}(2cr),$$

where

$$c = \frac{2\kappa}{\beta^2\left(1 - e^{-\kappa[T-t]}\right)}, \qquad u = cr_t e^{-\kappa[T-t]}, \qquad q = \frac{2\kappa\theta}{\beta^2} - 1,$$

and where $f_{\chi^2_{a,b}}(\cdot)$ denotes the probability density function for a non-centrally χ^2-distributed random variable with a degrees of freedom and non-centrality parameter b. The mean and variance of r_T are

$$\begin{aligned}
\mathrm{E}_{r,t}[r_T] &= \theta + (r - \theta)e^{-\kappa[T-t]}, \\
\mathrm{Var}_{r,t}[r_T] &= \frac{\beta^2 r}{\kappa}\left(e^{-\kappa[T-t]} - e^{-2\kappa[T-t]}\right) + \frac{\beta^2\theta}{2\kappa}\left(1 - e^{-\kappa[T-t]}\right)^2.
\end{aligned}$$

[8] In their general model r is in fact the *real* short-term interest rate and not the nominal short-term interest rate that we can observe. However, in practice the model is used for the nominal rates.

Note that the mean is just as in Vasicek's model, cf. (7.48), while the expression for the variance is slightly more complicated than in the Vasicek model, cf. (7.49). 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

Since the CIR model is affine, Theorem 7.1 implies that the price of a zero-coupon bond maturing at time T is

$$B^T(r,t) = e^{-a(T-t)-b(T-t)r}, \tag{7.67}$$

where the functions $a(\tau)$ and $b(\tau)$ solve the ordinary differential equations (7.8) and (7.9), which for the CIR model become

$$\frac{1}{2}\beta^2 b(\tau)^2 + \hat{\kappa} b(\tau) + b'(\tau) - 1 = 0, \tag{7.68}$$

$$a'(\tau) - \kappa\theta b(\tau) = 0 \tag{7.69}$$

with the conditions $a(0) = b(0) = 0$. The solution to these equations is

$$b(\tau) = \frac{2(e^{\gamma\tau}-1)}{(\gamma+\hat{\kappa})(e^{\gamma\tau}-1)+2\gamma}, \tag{7.70}$$

$$a(\tau) = -\frac{2\hat{\kappa}\hat{\theta}}{\beta^2}\left(\ln(2\gamma) + \frac{1}{2}(\hat{\kappa}+\gamma)\tau - \ln\left[(\gamma+\hat{\kappa})(e^{\gamma\tau}-1)+2\gamma\right]\right), \tag{7.71}$$

where $\gamma = \sqrt{\hat{\kappa}^2 + 2\beta^2}$, cf. Exercise 7.4.

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 (7.16) 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 (similar to the Vasicek model) not on θ.

7.5.3 The yield curve

Next we study the zero-coupon yield curve $T \mapsto y^T(r,t)$. From (7.13) 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, i.e. 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 7.7.)

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 using (7.68) and (7.69) can be rewritten as

$$f^T(r,t) = r + \hat{\kappa}\left[\hat{\theta} - r\right] b(T-t) - \frac{1}{2}\beta^2 r b(T-t)^2. \tag{7.72}$$

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 (7.70) and (7.71) 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 (7.19). The forward price on a coupon bond follows from (7.26).

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 (7.22), 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 (7.24), which in the present case is

$$\tilde{a}(\tau) = a(S-T) + \kappa\theta\int_0^\tau \tilde{b}(u)\,du.$$

The solution is

$$\tilde{b}(\tau) = \frac{2\hat{\kappa}b(S-T)}{\beta^2 b(S-T)\left(e^{\hat{\kappa}\tau}-1\right) + 2\hat{\kappa}e^{\hat{\kappa}\tau}}, \tag{7.73}$$

$$\tilde{a}(\tau) = a(S-T) - \frac{2\kappa\theta}{\beta^2}\ln\left(\frac{\tilde{b}(\tau)e^{\hat{\kappa}\tau}}{b(S-T)}\right). \tag{7.74}$$

The futures price on a coupon bond follows from (7.27).

According to (7.29) the quoted Eurodollar futures price is

$$\tilde{\mathcal{E}}^T(r,t) = 500 - 400 e^{-\hat{a}(T-t) - \hat{b}(T-t)r},$$

where the functions $\hat{a}$ and $\hat{b}$ in the CIR model can be computed to be

$$\hat{b}(\tau) = \frac{-2\hat{\kappa} b(0.25)}{-\beta^2 b(0.25)\left(e^{\hat{\kappa}\tau} - 1\right) + 2\hat{\kappa} e^{\hat{\kappa}\tau}}, \tag{7.75}$$

$$\hat{a}(\tau) = -a(0.25) - \frac{2\kappa\theta}{\beta^2} \ln\left(-\frac{\hat{b}(\tau) e^{\hat{\kappa}\tau}}{b(0.25)}\right). \tag{7.76}$$

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 $r_t = r$). However, this is much more complicated in the CIR model than in the models of Merton and Vasicek. Another approach is based on (6.10), which states that

$$C^{K,T,S}(r,t) = B^S(r,t)\, \mathbb{Q}^S_t\left(B^S(r_T,T) > K\right) - K B^T(r,t)\, \mathbb{Q}^T_t\left(B^S(r_T,T) > K\right), \tag{7.77}$$

where $\mathbb{Q}^S$ denotes the S-forward martingale measure and $\mathbb{Q}^T$, as before, is the T-forward martingale measure. We have $B^S(r_T,T) = \exp\{-a(S-T) - b(S-T)r_T\}$ so that

$$B^S(r_T,T) > K \quad \Leftrightarrow \quad r_T < \frac{-\ln K - a(S-T)}{b(S-T)}$$

and the option price can be rewritten as

$$C^{K,T,S}(r,t) = B^S(r,t)\, \mathbb{Q}^S_t\left(r_T < \frac{-\ln K - a(S-T)}{b(S-T)}\right) - K B^T(r,t)\, \mathbb{Q}^T_t\left(r_T < \frac{-\ln K - a(S-T)}{b(S-T)}\right). \tag{7.78}$$

Let us derive the dynamics of r_t under the τ-forward martingale measure $\mathbb{Q}^\tau$, where τ is some general future date. Then we can afterwards use this with $\tau = S$ and $\tau = T$ to obtain the two probabilities in the option pricing formula above. The sensitivity term of the price of the zero-coupon bond maturing at time τ is $\sigma^\tau_t = -\beta b(\tau - t)\sqrt{r_t}$. From (6.6) we have that

$$dz^{\mathbb{Q}}_t = dz^\tau_t + \sigma^\tau_t\, dt = dz^\tau_t - \beta b(\tau - t)\sqrt{r_t}\, dt.$$

Combining this with (7.66), we see that we can write the dynamics of the short rate under the τ-forward martingale measure as

$$\begin{aligned} dr_t &= (\hat{\varphi} - \hat{\kappa} r_t)\, dt + \beta\sqrt{r_t}\left(dz^\tau_t - \beta b(\tau - t)\sqrt{r_t}\, dt\right) \\ &= \left(\hat{\varphi} - \left[\hat{\kappa} + \beta^2 b(\tau - t)\right] r_t\right) dt + \beta\sqrt{r_t}\, dz^\tau_t. \end{aligned} \tag{7.79}$$

As always the volatility term is not altered when changing the measure. 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 (7.78) 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, Ingersoll, and Ross:

$$C^{K,T,S}(r,t) = B^S(r,t)\chi^2(h_1; f, g_1) - KB^T(r,t)\chi^2(h_2; f, g_2), \tag{7.80}$$

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

$$f = \frac{4\kappa\theta}{\beta^2}, \qquad h_1 = 2\bar{r}\left(\xi + \psi + b(S-T)\right), \qquad h_2 = 2\bar{r}(\xi + \psi),$$

$$g_1 = \frac{2\xi^2 r e^{\gamma[T-t]}}{\xi + \psi + b(S-T)}, \qquad g_2 = \frac{2\xi^2 r e^{\gamma[T-t]}}{\xi + \psi},$$

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, i.e. 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, Ingersoll, and Ross (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:

$$C^{K,T,\text{cpn}}(r,t) = \sum_{T_i > T} Y_i B^{T_i}(r,t)\chi^2(h_{1i}; f, g_{1i}) - K B^T(r,t)\chi^2(h_2; f, g_2), \tag{7.81}$$

where

$$h_{1i} = 2r^*\left(\xi + \psi + b(T_i - T)\right), \qquad h_2 = 2r^*(\xi+\psi), \qquad g_{1i} = \frac{2\xi^2 r e^{\gamma[T-t]}}{\xi + \psi + b(T_i - T)},$$

and f, g_2, ξ, and ψ are defined just below (7.80). This result was first derived by Longstaff (1993).

7.6 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, Derman, and Toy (1990), the so-called BDT model. The general time homogeneous version of the Black-Karasinski model is

$$d(\ln r_t) = \kappa[\theta - \ln r_t]\, dt + \beta\, dz_t^{\mathbb{Q}}, \tag{7.82}$$

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, i.e. r_T is lognormally distributed. A pleasant consequence of this 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 gives that

$$dr_t = \left(\left[\kappa\theta + \frac{1}{2}\beta^2\right] r_t - \kappa r_t \ln r_t\right) dt + \beta r_t\, dz_t^{\mathbb{Q}}.$$

There are no closed-form pricing expressions neither for bonds nor forwards, futures, and 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 (7.82) the future short rate is lognormally distributed. Another model with this property is the one where the short rate follows a geometric Brownian motion

$$dr_t = r_t\left[\alpha\, dt + \beta\, dz_t^{\mathbb{Q}}\right], \tag{7.83}$$

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.[10]

In addition to the lack of nice pricing formulas, the lognormal models (7.82) and (7.83) have another very inappropriate property. As shown by Hogan and Weintraub (1993) these models imply that, for all t, T, S with $t < T < S$,

$$\mathrm{E}^{\mathbb{Q}}_{r,t}\left[(B^S(r_T, T))^{-1}\right] = \infty. \tag{7.84}$$

As noted by Sandmann and Sondermann (1997) this result has two inexpedient consequences, which we state in the following theorem.

Theorem 7.6 *In the lognormal one-factor models* (7.82) *and* (7.83) *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, i.e. for $t \le T < S$

$$\mathrm{E}^{\mathbb{Q}}_{r,t}\left[\exp\left\{\int_T^S r_u\, du\right\}\right] = \infty.$$

(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 gives that[11]

$$\begin{aligned} B^S(r,T) &= \mathrm{E}^{\mathbb{Q}}_{r,T}\left[\exp\left\{-\int_T^S r_u\, du\right\}\right] \\ &= \mathrm{E}^{\mathbb{Q}}_{r,T}\left[\left(\exp\left\{\int_T^S r_u\, du\right\}\right)^{-1}\right] > \left(\mathrm{E}^{\mathbb{Q}}_{r,T}\left[\exp\left\{\int_T^S r_u\, du\right\}\right]\right)^{-1} \end{aligned}$$

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[12]

$$\mathrm{E}^{\mathbb{Q}}_{r,t}\left[\exp\left\{\int_T^S r_u\, du\right\}\right] > \mathrm{E}^{\mathbb{Q}}_{r,t}\left[(B^S(r_T,T))^{-1}\right] = \infty,$$

where the equality comes from (7.84).

From (6.15) 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 (7.84) with $S = T + 0.25$ into the expression above we get 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.

[10] 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 (7.83) is described by Hansen and Jørgensen (2000).

[11] Jensen's inequality says that if X is a random variable and $f(x)$ is a convex function, then $\mathrm{E}[f(X)] > f(\mathrm{E}[X])$.

[12] Here we apply the law of iterated expectations: $\mathrm{E}^{\mathbb{Q}}_{r,t}\left[\mathrm{E}^{\mathbb{Q}}_{r_T,T}[Y]\right] = \mathrm{E}^{\mathbb{Q}}_{r,t}[Y]$ for any random variable Y.

It can be shown that the problematic relation (7.84) 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.

In summary, the lognormal models (7.82) and (7.83) 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, cf. the discussion in the next section.

Finally, let us mention some models that are neither affine nor lognormal. The model

$$dr_t = \kappa \left[\theta - r_t\right] \, dt + \beta r_t \, dz_t^{\mathbb{Q}} \tag{7.85}$$

was suggested by Brennan and Schwartz (1980) and Courtadon (1982). Despite the relatively simple dynamics, no explicit pricing formulas have been derived neither 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.

7.7 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, Lo, and MacKinlay (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, e.g. maximum likelihood [as in Marsh and Rosenfeld (1983) and Ogden (1987)] or various moment matching methods [as in Andersen and Lund (1997), Chan, Karolyi, Longstaff, and Sanders (1992), and Dell' Aquila, Ronchetti, and Trojani (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 e.g. 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, e.g. one or three months, as an approximation to the short rate. However, Knez, Litterman, and Scheinkman (1994) and Duffee (1996) argue that special features of the trading in one-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, Long Jr., and Pearson (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 enters the real-world dynamics of the short rate in (7.47) 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, i.e. 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.[13] The model's results concerning the dynamics of interest rates

[13] For example, Brown and Dybvig (1986) find that the parameter estimates of the CIR model fluctuate considerably over time.

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, Karolyi, Longstaff, and Sanders (1992) [henceforth abbreviated CKLS], who consider time homogeneous models of the type

$$dr_t = (\theta - \kappa r_t)\, dt + \beta r_t^{\gamma}\, dz_t. \tag{7.86}$$

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 (7.83) ($\gamma = 1$, $\theta = 0$). 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 1964–1989. They estimate eight different restricted models and 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 (7.83) 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 (7.13) 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 (7.8) and (7.9) numerically. For non-affine models the relation can be found by numerically solving the PDE (7.3) 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, i.e. 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 (1964–1989). 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 (7.85).

Finally, applying a different estimation method and a different data set (weekly observations of three month U.S. government bond yields over the period 1954-1995), Andersen and Lund (1997) estimate γ to 0.676, which is much lower than the estimate of CKLS. Christensen, Poulsen, and Sørensen (2001) discuss some general problems in estimating a process like (7.86), and using a maximum likelihood estimation procedure and 1982-1995 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.

7.8 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, Ingersoll, and Ross 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 perform 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 unproblematic, however. 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 parameterization as outlined in Chapter 1. If one instead applies the parameterization 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.

7.9 Exercises

EXERCISE 7.1 (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, i.e. 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 (7.3) on page 145 to show that

$$\frac{1}{2}\beta(r)^2(T-t)^2 - \hat{\alpha}(r)(T-t) + h'(T-t)(T-t) + h(T-t) = 0 \qquad (*)$$

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 7.2 (Call on zero-coupon bonds in Vasicek's model) Figure 7.9 on page 167 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, cf. (7.61). 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 7.3 (Futures on bonds) Show the last claim in Theorem 7.2.

EXERCISE 7.4 (CIR zero-coupon bond price) Show that the functions b and a given by (7.70) and (7.71) solve the ordinary differential equations (7.68) and (7.69).

EXERCISE 7.5 (Comparison of prices in the models of Vasicek and CIR) Compare the prices according to Vasicek's model (7.47) and the CIR-model (7.65) 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 7.6 (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.

Chapter 8

Multi-factor diffusion models

8.1 What is wrong with one-factor models?

The preceding chapter gave an overview over one-factor diffusion models of the term structure of interest rates. All the 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, cf. Chapter 2. 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 accepted selected one-factor models. Furthermore, particularly the CIR model is based on a dynamics of the short rate that has 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, e.g. the yield curve is sometimes decreasing for short maturities and then increasing for longer maturities.

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

$$\bar{y}_t^\tau = \frac{a(\tau)}{\tau} + \frac{b(\tau)}{\tau} r_t,$$

cf. (7.13). 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, i.e. 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

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

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).

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, e.g. Rebonato (1996, Ch. 2) who uses data from the U.K. bond market.

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, cf. the discussion in Section 8.3.1.

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 three 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 1986–1991 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 with 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 1989–1992.

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 (approx. 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, i.e. 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, Harjes, and Kroner (1996), Andersen and Lund (1997), Vetzal (1997), Balduzzi, Das, and Foresi (1998), Boudoukh, Richardson, Stanton, and Whitelaw (1999), and Dai and Singleton (2000) 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.

8.2 Multi-factor diffusion models of the term structure

In this section we review the notation and the general results in multi-factor diffusion models, which were first discussed in Chapter 6. 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, i.e. the set of possible states. In the expression (8.1) above, $\boldsymbol{\alpha}$ 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, i.e. $\underline{\underline{\beta}}(\boldsymbol{x}_t, t)$ 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, cf. Øksendal (1998). 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

$$\mu(\boldsymbol{x}_t, t) = r(\boldsymbol{x}_t, t) + \sum_{j=1}^{n} \sigma_j(\boldsymbol{x}_t, t)\lambda_j(\boldsymbol{x}_t, t),$$

where μ denotes the expected rate of return on the asset, and $\sigma_1, \dots, \sigma_n$ are the volatility terms, i.e. 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 (6.35) on page 131.

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

$$\hat{\boldsymbol{\alpha}}(\boldsymbol{x}, t) = \boldsymbol{\alpha}(\boldsymbol{x}, t) - \underline{\underline{\beta}}(\boldsymbol{x}, t)\boldsymbol{\lambda}(\boldsymbol{x}, t),$$

i.e.

$$\hat{\alpha}_i(\boldsymbol{x}, t) = \alpha_i(\boldsymbol{x}, t) - \sum_{j=1}^{n} \beta_{ij}(\boldsymbol{x}, t)\lambda_j(\boldsymbol{x}, t),$$

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 Chapter 6, 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

$$\begin{aligned} \frac{\partial P}{\partial t}(\boldsymbol{x}, t) &+ \sum_{i=1}^{n} \hat{\alpha}_i(\boldsymbol{x}, t)\frac{\partial P}{\partial x_i}(\boldsymbol{x}, t) \\ &+ \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \gamma_{ij}(\boldsymbol{x}, t)\frac{\partial^2 P}{\partial x_i \partial x_j}(\boldsymbol{x}, t) - r(\boldsymbol{x}, t)P(\boldsymbol{x}, t) = 0, \quad (\boldsymbol{x}, t) \in \mathcal{S} \times [0, T), \end{aligned} \tag{8.2}$$

with the appropriate terminal condition

$$P(\boldsymbol{x}, T) = H(\boldsymbol{x}), \quad \boldsymbol{x} \in \mathcal{S}.$$

Here, $\gamma_{ij} = \sum_{k=1}^{n} \beta_{ik}\beta_{jk}$ is the (i, j)'th element of the variance-covariance matrix $\underline{\underline{\beta}}\,\underline{\underline{\beta}}^\top$. 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 as 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 as an expectation under the T-forward martingale measure

$$P(\boldsymbol{x}, t) = B^T(\boldsymbol{x}, t)\, \mathrm{E}^{\mathbb{Q}^T}_{\boldsymbol{x},t}\left[H(\boldsymbol{x}_T)\right], \tag{8.3}$$

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,

$$d\boldsymbol{x}_t = \hat{\boldsymbol{\alpha}}(\boldsymbol{x}_t)\, dt + \underline{\underline{\beta}}(\boldsymbol{x}_t)\, d\boldsymbol{z}^{\mathbb{Q}}_t. \tag{8.4}$$

8.3 Multi-factor affine diffusion models

We focus first on so-called affine models, which in a multi-factor setting were introduced by Duffie and Kan (1996). In the affine multi-factor models the dynamics of the vector of state variables is of the form

$$d\boldsymbol{x}_t = \left(\hat{\boldsymbol{\varphi}} - \underline{\underline{\hat{\kappa}}}\boldsymbol{x}_t\right)\, dt + \underline{\underline{\Gamma}} \begin{pmatrix} \sqrt{\nu_1(\boldsymbol{x}_t)} & 0 & \dots & 0 \\ 0 & \sqrt{\nu_2(\boldsymbol{x}_t)} & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & \dots & \sqrt{\nu_n(\boldsymbol{x}_t)} \end{pmatrix} d\boldsymbol{z}^{\mathbb{Q}}_t, \tag{8.5}$$

where

$$\nu_j(\boldsymbol{x}) = \delta_{j0} + \boldsymbol{\delta}_j^{\top}\boldsymbol{x} = \delta_{j0} + \sum_{k=1}^{n} \delta_{jk} x_k.$$

Here, $\hat{\boldsymbol{\varphi}} = (\hat{\varphi}_1, \dots, \hat{\varphi}_n)^{\top}$ and $\boldsymbol{\delta}_j = (\delta_{j1}, \dots, \delta_{jn})^{\top}$ for $j = 1, \dots, n$ are all constant vectors, $\delta_{10}, \dots, \delta_{n0}$ are constant scalars, and $\underline{\underline{\hat{\kappa}}}$ and $\underline{\underline{\Gamma}}$ are constant $n \times n$ matrices.

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 (8.5) *and* the short rate $r_t = r(\boldsymbol{x}_t)$ is an affine function of $\boldsymbol{x}$, i.e. a constant scalar ξ_0 and a constant n-dimensional vector $\boldsymbol{\xi} = (\xi_1, \dots, \xi_n)^{\top}$ exist such that

$$r(\boldsymbol{x}) = \xi_0 + \boldsymbol{\xi}^{\top}\boldsymbol{x} = \xi_0 + \sum_{i=1}^{n} \xi_i x_i. \tag{8.6}$$

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. Note that if the vector $\boldsymbol{\lambda}(\boldsymbol{x})$ of market prices of risk is also an affine function of $\boldsymbol{x}$, the drift of the state variables will also be affine under the real-world probability measure.

We will first consider two-factor affine models, both because the notation and the statement of results are simpler and because most of the multi-factor models studied in the literature have two factors. Subsequently, we will briefly extend the analysis to the general n-factor models.

8.3.1 Two-factor affine diffusion models

In a two-factor affine model the dynamics of the vector of state variables is of the form

$$d\boldsymbol{x}_t = \left(\hat{\boldsymbol{\varphi}} - \underline{\underline{\hat{\kappa}}}\boldsymbol{x}_t\right)\, dt + \underline{\underline{\Gamma}} \begin{pmatrix} \sqrt{\nu_1(\boldsymbol{x}_t)} & 0 \\ 0 & \sqrt{\nu_2(\boldsymbol{x}_t)} \end{pmatrix} d\boldsymbol{z}^{\mathbb{Q}}_t, \tag{8.7}$$

which can be written componentwise as

$$
\begin{aligned}
dx_{1t} &= (\hat{\varphi}_1 - \hat{\kappa}_{11}x_{1t} - \hat{\kappa}_{12}x_{2t})\, dt \\
&\qquad + \Gamma_{11}\sqrt{\delta_{10} + \delta_{11}x_{1t} + \delta_{12}x_{2t}}\, dz^{\mathbb{Q}}_{1t} + \Gamma_{12}\sqrt{\delta_{20} + \delta_{21}x_{1t} + \delta_{22}x_{2t}}\, dz^{\mathbb{Q}}_{2t}
\end{aligned}
\tag{8.8}
$$

$$
\begin{aligned}
dx_{2t} &= (\hat{\varphi}_2 - \hat{\kappa}_{21}x_{1t} - \hat{\kappa}_{22}x_{2t})\, dt \\
&\qquad + \Gamma_{21}\sqrt{\delta_{10} + \delta_{11}x_{1t} + \delta_{12}x_{2t}}\, dz^{\mathbb{Q}}_{1t} + \Gamma_{22}\sqrt{\delta_{20} + \delta_{21}x_{1t} + \delta_{22}x_{2t}}\, dz^{\mathbb{Q}}_{2t}.
\end{aligned}
\tag{8.9}
$$

The short rate is given by

$$r_t = \xi_0 + \xi_1 x_{1t} + \xi_2 x_{2t}. \tag{8.10}$$

To ensure that the square roots are well-defined, we require that

$$\nu_j(\boldsymbol{x}) = \delta_{j0} + \delta_{j1}x_1 + \delta_{j2}x_2 \geq 0, \quad j = 1, 2,$$

for all the values that $\boldsymbol{x} = (x_1, x_2)^\top$ can have, i.e. for all $\boldsymbol{x} \in \mathcal{S}$. Duffie and Kan state two conditions on the parameters that imply both that the process $\boldsymbol{x} = (\boldsymbol{x}_t)$ is well-defined by the stochastic differential equation above and that $\nu_j(\boldsymbol{x}_t) > 0$ for all t (with probability 1). These conditions are satisfied in the specific models we consider in the rest of the chapter.

Under the assumption (8.5) the drift $\hat{\boldsymbol{\alpha}}(\boldsymbol{x}) = \hat{\boldsymbol{\varphi}} - \underline{\underline{\hat{\kappa}}}\boldsymbol{x}$ is clearly an affine function of $\boldsymbol{x}$. Furthermore, the variance-covariance matrix $\underline{\underline{\beta}}(\boldsymbol{x})\underline{\underline{\beta}}(\boldsymbol{x})^\top$ is also affine in $\boldsymbol{x}$, in the sense that any element of the matrix is an affine function of $\boldsymbol{x}$. To see this, we use (8.8) and (8.9) to derive that the variance of the instantaneous change in each of the state variables is equal to $\operatorname{Var}^{\mathbb{Q}}_t(dx_{it}) = \gamma_i(x_{1t}, x_{2t})^2\, dt$, where

$$
\begin{aligned}
\gamma_i(x_1, x_2)^2 &= \Gamma^2_{i1}\left[\delta_{10} + \delta_{11}x_1 + \delta_{12}x_2\right] + \Gamma^2_{i2}\left[\delta_{20} + \delta_{21}x_1 + \delta_{22}x_2\right] \\
&= \left(\delta_{10}\Gamma^2_{i1} + \delta_{20}\Gamma^2_{i2}\right) + \left(\delta_{11}\Gamma^2_{i1} + \delta_{21}\Gamma^2_{i2}\right) x_1 + \left(\delta_{12}\Gamma^2_{i1} + \delta_{22}\Gamma^2_{i2}\right) x_2,
\end{aligned}
$$

which is an affine function of the state variables. The covariance between instantaneous changes in the two state variables is $\operatorname{Cov}^{\mathbb{Q}}_t(dx_{1t}, dx_{2t}) = \gamma_{12}(x_{1t}, x_{2t})\, dt$, where

$$
\begin{aligned}
\gamma_{12}(x_1, x_2) &= \Gamma_{11}\Gamma_{21}\left[\delta_{10} + \delta_{11}x_1 + \delta_{12}x_2\right] + \Gamma_{12}\Gamma_{22}\left[\delta_{20} + \delta_{21}x_1 + \delta_{22}x_2\right] \\
&= (\Gamma_{11}\Gamma_{21}\delta_{10} + \Gamma_{12}\Gamma_{21}\delta_{20}) + (\Gamma_{11}\Gamma_{21}\delta_{11} + \Gamma_{12}\Gamma_{21}\delta_{21})\, x_1 \\
&\qquad + (\Gamma_{11}\Gamma_{21}\delta_{12} + \Gamma_{12}\Gamma_{21}\delta_{22})\, x_2,
\end{aligned}
$$

which is also an affine function of the state variables. Hence, all the elements of the variance-covariance matrix are affine in the state variables. In this way, the class of affine multi-factor models is a natural generalization of the class of affine one-factor models.

Analogously to the one-factor analysis, we get that the price B^T_t of a zero-coupon bond maturing at time T in a multi-factor affine diffusion model can be written as an exponential-affine function of the vector of state variables. The following theorem states the precise result:

Theorem 8.1 *In an affine two-factor diffusion model where the short rate is of the form* (8.10) *and the state variables follow the process* (8.7)*, the zero-coupon bond price* $B^T_t = B^T(x_{1t}, x_{2t}, t)$ *is given by the function*

$$B^T(x_1, x_2, t) = \exp\left\{-a(T-t) - b_1(T-t)x_1 - b_2(T-t)x_2\right\}, \tag{8.11}$$

where the functions b_1, b_2, and a solve the following system of ordinary differential equations

$$\begin{aligned} b_1'(\tau) = & -\hat{\kappa}_{11} b_1(\tau) - \hat{\kappa}_{21} b_2(\tau) - \frac{1}{2}\delta_{11}\left(\Gamma_{11} b_1(\tau) + \Gamma_{21} b_2(\tau)\right)^2 \\ & - \frac{1}{2}\delta_{21}\left(\Gamma_{12} b_1(\tau) + \Gamma_{22} b_2(\tau)\right)^2 + \xi_1, \end{aligned} \tag{8.12}$$

$$\begin{aligned} b_2'(\tau) = & -\hat{\kappa}_{12} b_1(\tau) - \hat{\kappa}_{22} b_2(\tau) - \frac{1}{2}\delta_{12}\left(\Gamma_{11} b_1(\tau) + \Gamma_{21} b_2(\tau)\right)^2 \\ & - \frac{1}{2}\delta_{22}\left(\Gamma_{12} b_1(\tau) + \Gamma_{22} b_2(\tau)\right)^2 + \xi_2, \end{aligned} \tag{8.13}$$

$$\begin{aligned} a'(\tau) = & \hat{\varphi}_1 b_1(\tau) + \hat{\varphi}_2 b_2(\tau) - \frac{1}{2}\delta_{10}\left(\Gamma_{11} b_1(\tau) + \Gamma_{21} b_2(\tau)\right)^2 \\ & - \frac{1}{2}\delta_{20}\left(\Gamma_{12} b_1(\tau) + \Gamma_{22} b_2(\tau)\right)^2 + \xi_0, \end{aligned} \tag{8.14}$$

with the initial conditions $a(0) = b_1(0) = b_2(0) = 0$.

The result can be demonstrated by verifying that the function $B^T(x_1, x_2, t)$ given by (8.11) is a solution to the partial differential equation (8.2) in the affine model if the functions $a(\tau)$, $b_1(\tau)$, and $b_2(\tau)$ solve the ordinary differential equations (8.12)-(8.14). The terminal condition on the price, $B^T(x_1, x_2, T) = 1$ for all $(x_1, x_2) \in \mathcal{S}$, implies that $a(0)$, $b_1(0)$, and $b_2(0)$ all have to be zero. Except for the increased notational complexity, the proof is identical to the one-factor case and is therefore omitted. Note that to determine a, b_1, and b_2, one must first solve the two differential equations (8.12) and (8.13) simultaneously, and then a follows from the b_i's by an integration.

In the following sections we will look at specific affine models in which the Ricatti equations (8.12)-(8.14) have explicit solutions. This is the case in Gaussian models and so-called multi-factor CIR models. For other specifications of the affine model, the Ricatti equations must be solved numerically, see e.g. Duffie and Kan (1996). The Ricatti equations can be solved faster numerically than the partial differential equation.

In an affine two-factor model the zero-coupon yields $\bar{y}_t^\tau = y_t^{t+\tau} = -(\ln B_t^{t+\tau})/\tau$ take the form

$$\bar{y}^\tau(x_1, x_2) = \frac{a(\tau)}{\tau} + \frac{b_1(\tau)}{\tau} x_1 + \frac{b_2(\tau)}{\tau} x_2,$$

and the forward rates $\bar{f}_t^\tau = f_t^{t+\tau} = -\partial(\ln B_t^{t+\tau})/\partial T$ are of the form

$$\bar{f}^\tau(x_1, x_2) = a'(\tau) + b_1'(\tau) x_1 + b_2'(\tau) x_2.$$

Let us also look at the volatility of the price of a zero-coupon bond with a fixed maturity date T. Since $\frac{\partial B^T}{\partial x_i}(x_1, x_2, t) = -b_i(T-t) B^T(x_1, x_2, t)$, Itô's Lemma for functions of several variables (see Theorem 3.6 on page 65) implies that

$$\frac{dB_t^T}{B_t^T} = r(x_{1t}, x_{2t})\, dt + \sigma_1^T(x_{1t}, x_{2t}, t)\, dz_{1t}^{\mathbb{Q}} + \sigma_2^T(x_{1t}, x_{2t}, t)\, dz_{2t}^{\mathbb{Q}},$$

where

$$\begin{aligned} \sigma_1^T(x_1, x_2, t) &= -\left(b_1(T-t)\Gamma_{11} + b_2(T-t)\Gamma_{21}\right)\sqrt{\delta_{10} + \delta_{11} x_1 + \delta_{12} x_2}, \\ \sigma_2^T(x_1, x_2, t) &= -\left(b_1(T-t)\Gamma_{12} + b_2(T-t)\Gamma_{22}\right)\sqrt{\delta_{20} + \delta_{21} x_1 + \delta_{22} x_2}. \end{aligned}$$

The volatility of the zero-coupon bond price is therefore $\|\sigma^T(x_1, x_2, t)\|$, where

$$\begin{aligned} \|\sigma^T(x_1, x_2, t)\|^2 &= \sigma_1^T(x_1, x_2, t)^2 + \sigma_2^T(x_1, x_2, t)^2 \\ &= \left(b_1(T-t)\Gamma_{11} + b_2(T-t)\Gamma_{21}\right)^2\left(\delta_{10} + \delta_{11} x_1 + \delta_{12} x_2\right) \\ &\quad + \left(b_1(T-t)\Gamma_{12} + b_2(T-t)\Gamma_{22}\right)^2\left(\delta_{20} + \delta_{21} x_1 + \delta_{22} x_2\right). \end{aligned}$$

Similarly, the dynamics of the zero-coupon yield $\bar{y}_t^\tau$ will be of the form

$$d\bar{y}_t^\tau = \dots\, dt + \sigma_{y1}(x_{1t}, x_{2t}, \tau)\, dz_{1t}^{\mathbb{Q}} + \sigma_{y2}(x_{1t}, x_{2t}, \tau)\, dz_{2t}^{\mathbb{Q}},$$

where

$$\sigma_{y1}(x_1, x_2, \tau) = \left(\frac{b_1(\tau)}{\tau}\Gamma_{11} + \frac{b_2(\tau)}{\tau}\Gamma_{21} \right) \sqrt{\delta_{10} + \delta_{11}x_1 + \delta_{12}x_2},$$
$$\sigma_{y2}(x_1, x_2, \tau) = \left(\frac{b_1(\tau)}{\tau}\Gamma_{12} + \frac{b_2(\tau)}{\tau}\Gamma_{22} \right) \sqrt{\delta_{20} + \delta_{21}x_1 + \delta_{22}x_2},$$

and we have omitted the drift rate for clarity. We can see that, in a two-factor model, zero-coupon yields with different maturities are not perfectly correlated since the covariance is

$$\mathrm{Cov}_t^{\mathbb{Q}}\left(d\bar{y}_t^{\tau_1}, d\bar{y}_t^{\tau_2}\right) = \left(\sigma_{y1}(x_{1t}, x_{2t}, \tau_1)\sigma_{y1}(x_{1t}, x_{2t}, \tau_2) + \sigma_{y2}(x_{1t}, x_{2t}, \tau_1)\sigma_{y2}(x_{1t}, x_{2t}, \tau_2)\right) dt,$$

and, hence, the correlation is

$$\mathrm{Corr}_t^{\mathbb{Q}}\left(d\bar{y}_t^{\tau_1}, d\bar{y}_t^{\tau_2}\right) = \frac{\sigma_{y1}(\boldsymbol{x}_t, \tau_1)\sigma_{y1}(\boldsymbol{x}_t, \tau_2) + \sigma_{y2}(\boldsymbol{x}_t, \tau_1)\sigma_{y2}(\boldsymbol{x}_t, \tau_2)}{\sqrt{\sigma_{y1}(\boldsymbol{x}_t, \tau_1)^2 + \sigma_{y2}(\boldsymbol{x}_t, \tau_1)^2}\sqrt{\sigma_{y1}(\boldsymbol{x}_t, \tau_2)^2 + \sigma_{y2}(\boldsymbol{x}_t, \tau_2)^2}},$$

which in general will be less than 1.

The above analysis was based on two unspecified state variables x_1 and x_2. Due to the fact that all prices, interest rates, volatilities, etc., are functions of x_1 and x_2, it is typically possible to shift to another pair of state variables $\tilde{x}_1$ and $\tilde{x}_2$ and express prices, rates, etc., in terms of the new variables. This is convenient if the new variables $\tilde{x}_1$ and $\tilde{x}_2$ are easier to observe than the original variables x_1 and x_2 since the price expressions are then simpler to apply in practice, and it will be easier to estimate the model parameters. Of course, we could have stated the model in terms of $\tilde{x}_1$ and $\tilde{x}_2$ from the beginning, but it may be easier to develop the pricing formulas using x_1 and x_2. See Section 8.5.2 for an important example.

8.3.2 n-factor affine diffusion models

In the general n-factor affine model where the vector of state variables follows the process (8.5), and the short rate is given as in (8.6), the zero-coupon bond price $B_t^T = B^T(\boldsymbol{x}_t, t)$ is given by the function

$$B^T(\boldsymbol{x}, t) = \exp\left\{-a(T-t) - \boldsymbol{b}(T-t)^\top \boldsymbol{x}\right\} = \exp\left\{ -a(T-t) - \sum_{j=1}^n b_j(T-t)x_j \right\}, \qquad (8.15)$$

where the functions $a(\tau), b_1(\tau), \dots, b_n(\tau)$ solve the following system of ordinary differential equations:

$$b_i'(\tau) = -\sum_{j=1}^n \hat{\kappa}_{ji} b_j(\tau) - \frac{1}{2}\sum_{k=1}^n \delta_{ki} \left(\sum_{j=1}^n \Gamma_{jk} b_j(\tau) \right)^2 + \xi_i, \quad i = 1, \dots, n, \qquad (8.16)$$

$$a'(\tau) = \sum_{j=1}^n \hat{\varphi}_j b_j(\tau) - \frac{1}{2}\sum_{k=1}^n \delta_{k0} \left(\sum_{j=1}^n \Gamma_{jk} b_j(\tau) \right)^2 + \xi_0, \qquad (8.17)$$

with the initial conditions $a(0) = b_1(0) = \dots = b_n(0) = 0$. Conversely, under certain regularity conditions, the zero-coupon bond prices are only of the exponential-affine form if $\hat{\boldsymbol{\alpha}}$, $\underline{\underline{\beta}}\,\underline{\underline{\beta}}^\top$, and r are affine functions of $\boldsymbol{x}$, cf. 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

$$\bar{y}^\tau(\boldsymbol{x}) = \frac{a(\tau)}{\tau} + \sum_{j=1}^n \frac{b_j(\tau)}{\tau} x_j, \tag{8.18}$$

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

$$\sigma_j^T(\boldsymbol{x}, t) = -\sqrt{\delta_{j0} + \delta_{j1}x_1 + \dots \delta_{jn}x_n} \sum_{k=1}^n \Gamma_{kj} b_k(T-t). \tag{8.19}$$

8.3.3 European options on coupon bonds

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, cf. (7.30) on page 151. 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, cf. Section 2.9.2 on page 38) 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 to 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 (2002) 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.

8.4 Multi-factor Gaussian diffusion models

8.4.1 General analysis

The simplest affine multi-factor term structure models are the Gaussian models, which were first studied by Langetieg (1980). A Gaussian model is an affine model of the form (8.5) where the volatility functions $\nu_j(\boldsymbol{x})$ are constant so that the vectors $\boldsymbol{\delta}_j$ are equal to zero. Since the diagonal matrix is multiplied by the constant matrix $\underline{\underline{\Gamma}}$, we can and will assume that all the $\nu_j(\boldsymbol{x})$'s are equal to 1, i.e. $\delta_{j0} = 1$. In the Gaussian models the dynamics of the state variables is therefore

$$d\boldsymbol{x}_t = \left(\hat{\boldsymbol{\varphi}} - \underline{\underline{\hat{\kappa}}}\boldsymbol{x}_t\right) dt + \underline{\underline{\Gamma}}\, d\boldsymbol{z}_t^{\mathbb{Q}}. \tag{8.20}$$

Analogously to the analysis for one-dimensional Ornstein-Uhlenbeck processes in Section 3.8.2 on page 59, we get that future values of the vector of state variables are n-dimensionally normally distributed. In particular, the individual state variables are normally distributed. Since the short rate is a linear combination of these state variables, it follows that future values of the short rate are normally distributed in these 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] In a Gaussian model the sensitivities of the zero-coupon bond prices σ_j^T defined in (8.19) depend only on the time to maturity of the bond,

$$\sigma_j^T(t) = -\sum_{k=1}^n \Gamma_{kj} b_k(T-t).$$

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 (8.15). According to (8.3), the price of a European call option on a zero-coupon bond is

$$C^{K,T,S}(\boldsymbol{x},t) = B^T(\boldsymbol{x},t)\,\mathrm{E}^{\mathbb{Q}^T}_{\boldsymbol{x},t}\left[\max\left(B^S(\boldsymbol{x}_T,T)-K,0\right)\right].$$

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. As in the Gaussian one-factor models studied in Chapter 7, we conclude that

$$C^{K,T,S}(\boldsymbol{x},t) = B^S(\boldsymbol{x},t)N\left(d_1\right) - KB^T(\boldsymbol{x},t)N\left(d_2\right),$$

where

$$d_1 = \frac{1}{v(t,T,S)}\ln\left(\frac{B^S(\boldsymbol{x},t)}{KB^T(\boldsymbol{x},t)}\right) + \frac{1}{2}v(t,T,S), \tag{8.21}$$

$$d_2 = d_1 - v(t,T,S), \tag{8.22}$$

and

$$\begin{aligned} v(t,T,S)^2 &= \mathrm{Var}_t^{\mathbb{Q}^T}\left[\ln F_T^{T,S}\right] \\ &= \sum_{j=1}^n \int_t^T \left(\sigma_j^S(u) - \sigma_j^T(u)\right)^2 du \\ &= \sum_{j=1}^n \int_t^T \left(\sum_{k=1}^n \Gamma_{kj}\left[b_k(S-u) - b_k(T-u)\right]\right)^2 du. \end{aligned} \tag{8.23}$$

Let us focus on the case of two factors. In a Gaussian two-factor diffusion model the dynamics of the state variables is of the form

$$dx_{1t} = \left(\hat{\varphi}_1 - \hat{\kappa}_{11}x_{1t} - \hat{\kappa}_{12}x_{2t}\right)dt + \Gamma_{11}\,dz_{1t}^{\mathbb{Q}} + \Gamma_{12}\,dz_{2t}^{\mathbb{Q}}, \tag{8.24}$$

$$dx_{2t} = \left(\hat{\varphi}_2 - \hat{\kappa}_{21}x_{1t} - \hat{\kappa}_{22}x_{2t}\right)dt + \Gamma_{21}\,dz_{1t}^{\mathbb{Q}} + \Gamma_{22}\,dz_{2t}^{\mathbb{Q}}. \tag{8.25}$$

The ordinary differential equations (8.12) and (8.13) reduce to

$$b_1'(\tau) = -\hat{\kappa}_{11}b_1(\tau) - \hat{\kappa}_{21}b_2(\tau) + \xi_1, \quad b_1(0) = 0, \tag{8.26}$$

$$b_2'(\tau) = -\hat{\kappa}_{12}b_1(\tau) - \hat{\kappa}_{22}b_2(\tau) + \xi_2, \quad b_2(0) = 0, \tag{8.27}$$

[1] Generally the moments depend on the eigenvalues and the eigenvectors of the matrix $\underline{\underline{\hat{\kappa}}}$, cf. the discussion in Langetieg (1980).

while the equation for the function a becomes

$$\begin{aligned} a'(\tau) &= \hat{\varphi}_1 b_1(\tau) + \hat{\varphi}_2 b_2(\tau) - \frac{1}{2}\left(\Gamma_{11} b_1(\tau) + \Gamma_{21} b_2(\tau)\right)^2 \\ &\quad - \frac{1}{2}\left(\Gamma_{12} b_1(\tau) + \Gamma_{22} b_2(\tau)\right)^2 + \xi_0, \quad a(0) = 0. \end{aligned} \tag{8.28}$$

According to (8.23), the variance term in the option price is given by

$$\begin{aligned} v(t,T,S)^2 &= \left(\Gamma_{11}^2 + \Gamma_{12}^2\right) \int_t^T \left[b_1(S-u) - b_1(T-u)\right]^2 du \\ &\quad + \left(\Gamma_{21}^2 + \Gamma_{22}^2\right) \int_t^T \left[b_2(S-u) - b_2(T-u)\right]^2 du \\ &\quad + 2\left(\Gamma_{11}\Gamma_{21} + \Gamma_{12}\Gamma_{22}\right) \int_t^T \left[b_1(S-u) - b_1(T-u)\right]\left[b_2(S-u) - b_2(T-u)\right] du. \end{aligned} \tag{8.29}$$

8.4.2 A specific example: 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 in which

$$dr_t = \kappa\left[\hat{\theta} - r_t\right] dt + \beta\, dz_t^{\mathbb{Q}} = \left(\hat{\varphi} - \kappa r_t\right) dt + \beta\, dz_t^{\mathbb{Q}},$$

cf. Section 7.4 on page 157. 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:

$$dr_t = \left(\hat{\varphi} + \varepsilon_t - \kappa_r r_t\right) dt + \beta_r\, dz_{1t}^{\mathbb{Q}}, \tag{8.30}$$

$$d\varepsilon_t = -\kappa_\varepsilon \varepsilon_t\, dt + \beta_\varepsilon \rho\, dz_{1t}^{\mathbb{Q}} + \beta_\varepsilon \sqrt{1-\rho^2}\, dz_{2t}^{\mathbb{Q}}. \tag{8.31}$$

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. 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 two-factor model, 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

$$b_1'(\tau) = -\kappa_r b_1(\tau) + 1, \qquad b_1(0) = 0, \tag{8.32}$$

$$b_2'(\tau) = b_1(\tau) - \kappa_\varepsilon b_2(\tau), \qquad b_2(0) = 0. \tag{8.33}$$

The first of these is identical to the differential equation solved in the original one-factor Vasicek model, cf. Section 7.4.2 on page 159, so we know that the solution is

$$b_1(\tau) = \frac{1}{\kappa_r}\left(1 - e^{-\kappa_r \tau}\right).$$

Next, it can be verified that the solution to the equation for b_2 is given by

$$b_2(\tau) = \frac{1}{\kappa_r[\kappa_r - \kappa_\varepsilon]} e^{-\kappa_r \tau} - \frac{1}{\kappa_\varepsilon[\kappa_r - \kappa_\varepsilon]} e^{-\kappa_\varepsilon \tau} + \frac{1}{\kappa_r \kappa_\varepsilon}.$$

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

$$\begin{aligned}
a(\tau) &= a(\tau) - a(0) = \int_0^\tau a'(u)\, du \\
&= \hat{\varphi}\int_0^\tau b_1(u)\, du - \frac{1}{2}\beta_r^2 \int_0^\tau b_1(u)^2\, du - \frac{1}{2}\beta_\varepsilon^2 \int_0^\tau b_2(u)^2\, du - \rho\beta_r\beta_\varepsilon \int_0^\tau b_1(u) b_2(u)\, du \\
&= \frac{\hat{\varphi}}{\kappa_r}(\tau - b_1(\tau)) - \frac{1}{\kappa_r^2}\left(\frac{1}{2}\beta_r^2 - \frac{\rho\beta_r\beta_\varepsilon}{\kappa_r - \kappa_\varepsilon} + \frac{\beta_\varepsilon^2}{2(\kappa_r - \kappa_\varepsilon)^2}\right)\left(\tau - b_1(\tau) - \frac{1}{2}\kappa_r b_1(\tau)^2\right) \\
&\quad - \frac{\beta_\varepsilon^2}{4\kappa_\varepsilon^3(\kappa_r - \kappa_\varepsilon)}\left(2\tau\kappa_\varepsilon - 3 + 4e^{-\kappa_\varepsilon \tau} - e^{-2\kappa_\varepsilon \tau}\right) \\
&\quad - \frac{\rho\beta_r\beta_\varepsilon}{\kappa_r\kappa_\varepsilon(\kappa_r - \kappa_\varepsilon)}\left(\tau - b_1(\tau) - \frac{1 - e^{-\kappa_\varepsilon \tau}}{\kappa_\varepsilon} + \frac{1 - e^{-(\kappa_r + \kappa_\varepsilon)\tau}}{\kappa_r + \kappa_\varepsilon}\right).
\end{aligned}$$

The relevant variance $v(t,T,S)^2$ entering the price of an option on a zero-coupon bond follows from (8.29):

$$\begin{aligned}
v(t,T,S)^2 &= \beta_r^2 \int_t^T \left[b_1(S-u) - b_1(T-u)\right]^2 du + \beta_\varepsilon^2 \int_t^T \left[b_2(S-u) - b_2(T-u)\right]^2 du \\
&\quad + 2\rho\beta_r\beta_\varepsilon \int_t^T \left[b_1(S-u) - b_1(T-u)\right]\left[b_2(S-u) - b_2(T-u)\right] du,
\end{aligned}$$

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.

Other Gaussian multi-factor models have been studied by Langetieg (1980) and Beaglehole and Tenney (1991).

8.5 Multi-factor CIR models

8.5.1 General analysis

A term structure model is said to be an n-factor CIR model if the short rate equals the sum of the n state variables, $r_t = \sum_{j=1}^n x_{jt}$, and the risk-neutral dynamics of the n state variables is of the form

$$dx_{jt} = (\hat{\varphi}_j - \hat{\kappa}_j x_{jt})\, dt + \beta_j \sqrt{x_{jt}}\, dz_{jt}^{\mathbb{Q}}, \quad j = 1, \dots, n,$$

where $\hat{\varphi}_j$, $\hat{\kappa}_j$, and β_j are constants. Note that the state variables are mutually independent and that each state variable follows a square root process, just as the short rate process in the one-factor CIR model, cf. Section 7.5 on page 168. In particular, the processes will have non-negative values at all points in time.[2] Also note that a multi-factor CIR model is an affine model of the form (8.5), where (i) the matrix $\underline{\underline{\hat{\kappa}}}$ is diagonal with $\hat{\kappa}_1, \dots, \hat{\kappa}_n$ in the diagonal and zeros in all other entries, (ii) the matrix $\underline{\underline{\Gamma}}$ is the identity matrix, i.e. the matrix with ones in the diagonal and zeros in all other entries, and (iii) $\nu_j(\boldsymbol{x}_t) = \beta_j^2 x_{jt}$.

In the multi-factor CIR models the ordinary differential equations (8.16) and (8.17) can be solved explicitly. The solutions can be computed from the known expressions for $a(\tau)$ and $b(\tau)$ in the one-factor CIR model. To see this, first recall from Chapter 6 that the zero-coupon bond price can be written as

$$B^T(\boldsymbol{x}, t) = \mathrm{E}^{\mathbb{Q}}\left[\exp\left\{-\int_t^T r_u \, du\right\} \middle| \boldsymbol{x}_t = \boldsymbol{x}\right].$$

Using the relation $r_u = x_{1u} + \dots + x_{nu}$ and the independence of the state variables, we can rewrite the zero-coupon bond price as

$$\begin{aligned} B^T(\boldsymbol{x}, t) &= \mathrm{E}^{\mathbb{Q}}\left[\exp\left\{-\sum_{j=1}^n \int_t^T x_{ju} \, du\right\} \middle| \boldsymbol{x}_t = \boldsymbol{x}\right] \\ &= \mathrm{E}^{\mathbb{Q}}\left[\prod_{j=1}^n \exp\left\{-\int_t^T x_{ju} \, du\right\} \middle| \boldsymbol{x}_t = \boldsymbol{x}\right] \\ &= \prod_{j=1}^n \mathrm{E}^{\mathbb{Q}}\left[\exp\left\{-\int_t^T x_{ju} \, du\right\} \middle| x_{jt} = x_j\right]. \end{aligned}$$

Because each state variable x_j follows a process of the same type as the short rate in the one-factor CIR model, we get

$$\mathrm{E}^{\mathbb{Q}}\left[\exp\left\{-\int_t^T x_{ju} \, du\right\} \middle| x_{jt} = x_j\right] = \exp\left\{-a_j(T-t) - b_j(T-t)x_j\right\},$$

where

$$\begin{aligned} b_j(\tau) &= \frac{2(e^{\gamma_j \tau} - 1)}{(\gamma_j + \hat{\kappa}_j)(e^{\gamma_j \tau} - 1) + 2\gamma_j}, \\ a_j(\tau) &= -\frac{2\hat{\varphi}_j}{\beta_j^2}\left(\ln(2\gamma_j) + \frac{1}{2}(\hat{\kappa}_j + \gamma_j)\tau - \ln\left[(\gamma_j + \hat{\kappa}_j)(e^{\gamma_j \tau} - 1) + 2\gamma_j\right]\right), \end{aligned}$$

and $\gamma_j = \sqrt{\hat{\kappa}_j^2 + 2\beta_j^2}$, cf. (7.67), (7.70), and (7.71). Consequently, the zero-coupon bond price is

$$B^T(\boldsymbol{x}, t) = \prod_{j=1}^n \exp\left\{-a_j(T-t) - b_j(T-t)x_j\right\} = \exp\left\{-a(T-t) - \sum_{j=1}^n b_j(T-t)x_j\right\},$$

where $a(\tau) = \sum_{j=1}^n a_j(\tau)$. The risk-neutral dynamics of the zero-coupon bond price is

$$\frac{dB_t^T}{B_t^T} = r(\boldsymbol{x}_t) \, dt + \sum_{j=1}^n \sigma_j^T(\boldsymbol{x}_t, t) \, dz_{jt}^{\mathbb{Q}},$$

where the sensitivities σ_j^T are given by

$$\sigma_j^T(\boldsymbol{x}, t) = -\beta_j \sqrt{x_j} b_j(T-t). \tag{8.34}$$

[2] As in the one-factor CIR model, the process x_j will be strictly positive if $2\hat{\varphi}_j \geq \beta_j^2$.

8.5.2 A specific example: the Longstaff-Schwartz model

Model description

The prevalent 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, Ingersoll, and Ross (1985a). With some empirical support Longstaff and Schwartz assume that the economy has one state variable, x_1, that affects only expected returns on productive investments and another state variable, x_2, that affects 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 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, etc., of the agents in the economy (and an appropriate scaling of the two state variables), the equilibrium short rate is exactly the sum of the two state variables,

$$r_t = x_{1t} + x_{2t}. \tag{8.35}$$

Furthermore, the market price of risk associated with x_1, i.e. $\lambda_1(\boldsymbol{x}, t)$, 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 motion under the risk-neutral probability measure $\mathbb{Q}$ is given by

$$dz_{1t}^{\mathbb{Q}} = dz_{1t}, \qquad dz_{2t}^{\mathbb{Q}} = dz_{2t} + \frac{\lambda}{\beta_2}\sqrt{x_{2t}}\, dt. \tag{8.36}$$

The dynamics of the state variables under the risk-neutral measure becomes

$$dx_{1t} = (\hat{\varphi}_1 - \hat{\kappa}_1 x_{1t})\, dt + \beta_1 \sqrt{x_{1t}}\, dz_{1t}^{\mathbb{Q}},$$
$$dx_{2t} = (\hat{\varphi}_2 - \hat{\kappa}_2 x_{2t})\, dt + \beta_2 \sqrt{x_{2t}}\, dz_{2t}^{\mathbb{Q}},$$

where $\hat{\varphi}_1 = \varphi_1$, $\hat{\kappa}_1 = \kappa_1$, $\hat{\varphi}_2 = \varphi_2$, and $\hat{\kappa}_2 = \kappa_2 + \lambda$.

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

$$B^T(x_1, x_2, t) = \exp\left\{-a(T-t) - b_1(T-t)x_1 - b_2(T-t)x_2\right\}, \tag{8.37}$$

where $a(\tau) = a_1(\tau) + a_2(\tau)$,

$$b_j(\tau) = \frac{2(e^{\gamma_j \tau} - 1)}{(\gamma_j + \hat{\kappa}_j)(e^{\gamma_j \tau} - 1) + 2\gamma_j}, \quad j = 1, 2,$$
$$a_j(\tau) = -\frac{2\hat{\varphi}_j}{\beta_j^2}\left(\ln(2\gamma_j) + \frac{1}{2}(\hat{\kappa}_j + \gamma_j)\tau - \ln\left[(\gamma_j + \hat{\kappa}_j)(e^{\gamma_j \tau} - 1) + 2\gamma_j\right]\right), \quad j = 1, 2,$$

and $\gamma_j = \sqrt{\hat{\kappa}_j^2 + 2\beta_j^2}$.

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 (8.35), we get $dr_t = dx_{1t} + dx_{2t}$, i.e.

$$dr_t = (\varphi_1 + \varphi_2 - \kappa_1 x_{1t} - \kappa_2 x_{2t})\, dt + \beta_1 \sqrt{x_{1t}}\, dz_{1t} + \beta_2 \sqrt{x_{2t}}\, dz_{2t}.$$

The instantaneous variance is $\text{Var}_t(dr_t) = v_t\, dt$, where

$$v_t = \beta_1^2 x_{1t} + \beta_2^2 x_{2t}, \tag{8.38}$$

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 (8.35) and (8.38) imply that

$$x_{1t} = \frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}, \qquad x_{2t} = \frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}. \tag{8.39}$$

The dynamics of r and v can then be rewritten as

$$\begin{aligned} dr_t = &\left(\varphi_1 + \varphi_2 - \frac{\kappa_1 \beta_2^2 - \kappa_2 \beta_1^2}{\beta_2^2 - \beta_1^2} r_t - \frac{\kappa_2 - \kappa_1}{\beta_2^2 - \beta_1^2} v_t\right) dt \\ &+ \beta_1 \sqrt{\frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}}\, dz_{1t} + \beta_2 \sqrt{\frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}}\, dz_{2t}, \end{aligned} \tag{8.40}$$

$$\begin{aligned} dv_t = &\left(\beta_1^2 \varphi_1 + \beta_2^2 \varphi_2 - \beta_1^2 \beta_2^2 \frac{\kappa_1 - \kappa_2}{\beta_2^2 - \beta_1^2} r_t - \frac{\beta_2^2 \kappa_2 - \beta_1^2 \kappa_1}{\beta_2^2 - \beta_1^2} v_t\right) dt \\ &+ \beta_1^3 \sqrt{\frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}}\, dz_{1t} + \beta_2^3 \sqrt{\frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}}\, dz_{2t}. \end{aligned} \tag{8.41}$$

Since both x_1 and x_2 stay non-negative, it follows from (8.39) 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 (8.40) and (8.41) 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 (8.39) into (8.37), we can write the zero-coupon bond price as a function of r and v:

$$B^T(r, v, t) = \exp\left\{-a(T-t) - \tilde{b}_1(T-t) r - \tilde{b}_2(T-t) v\right\}, \tag{8.42}$$

where

$$\begin{aligned} \tilde{b}_1(\tau) &= \frac{\beta_2^2 b_1(\tau) - \beta_1^2 b_2(\tau)}{\beta_2^2 - \beta_1^2}, \\ \tilde{b}_2(\tau) &= \frac{b_2(\tau) - b_1(\tau)}{\beta_2^2 - \beta_1^2}. \end{aligned}$$

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 for $\tau \to 0$ 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, $B^T(0,0,t) < 1$. Finally, $B^T(r,v,t) \to 0$ for $r \to \infty$ (in which case $v \to \infty$).

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 (i.e. first increasing, then decreasing), it can have a trough (i.e. first decreasing, then increasing) or both a hump and a trough. We can see most of these shapes in Figure 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.

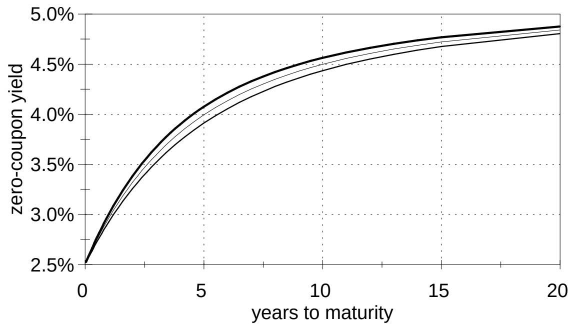
The dynamics of the zero-coupon bond price is of the form

$$\begin{aligned}
\frac{dB_t^T}{B_t^T} &= r_t\, dt - \beta_1 \sqrt{x_{1t}} b_1(T-t)\, dz_{1t}^{\mathbb{Q}} - \beta_2 \sqrt{x_{2t}} b_2(T-t)\, dz_{2t}^{\mathbb{Q}} \\
&= (r_t - \lambda x_{2t} b_2(T-t))\, dt - \beta_1 \sqrt{x_{1t}} b_1(T-t)\, dz_{1t} - \beta_2 \sqrt{x_{2t}} b_2(T-t)\, dz_{2t} \\
&= \left(r_t + \frac{\lambda}{\beta_2^2 - \beta_1^2} b_2(T-t)(\beta_1^2 r_t - v_t) \right) dt \\
&\quad - \beta_1 \sqrt{\frac{\beta_2^2 r_t - v_t}{\beta_2^2 - \beta_1^2}} b_1(T-t)\, dz_{1t} - \beta_2 \sqrt{\frac{v_t - \beta_1^2 r_t}{\beta_2^2 - \beta_1^2}} b_2(T-t)\, dz_{2t},
\end{aligned}$$

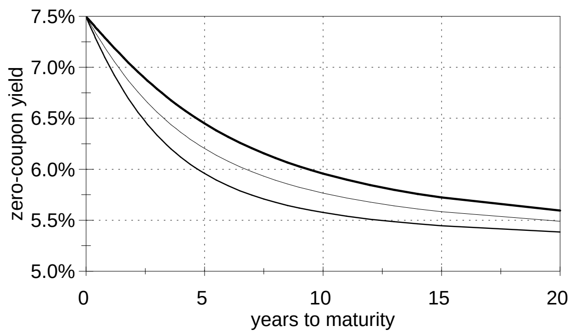
where we have applied (8.34), (8.36), and (8.39). The so-called term premium, i.e. the expected rate of return on a zero-coupon bond in excess of the short rate, is $\lambda b_2(T-t)(\beta_1^2 r_t - v_t)/(\beta_2^2 - \beta_1^2)$, which is positive if $\lambda < 0$. It is consistent with empirical studies 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 given by

$$\|\sigma^T(r,v,t)\|^2 = \frac{\beta_1^2 \beta_2^2}{\beta_2^2 - \beta_1^2} \left(b_1(T-t)^2 - b_2(T-t)^2 \right) r_t + \frac{\beta_2^2 b_2(T-t)^2 - \beta_1^2 b_1(T-t)^2}{\beta_2^2 - \beta_1^2} v_t.$$

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 $T-t$.



(a) Low current short rate



(b) High current short rate

(c) Medium current short rate

Figure 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 $(0.25\beta_2^2 + 0.75\beta_1^2)r$.

Options and other derivatives

Longstaff and Schwartz state a pricing formula for a European call option on a zero-coupon bond, which in our notation looks as follows:

$$\begin{aligned} C^{K,T,S}(r,v,t) &= B^S(r,v,t)\chi^2\left(\theta_1,\theta_2;4\varphi_1/\beta_1^2,4\varphi_2/\beta_2^2,\omega_1[\beta_2^2 r-v],\omega_2[v-\beta_1^2 r]\right) \\ &\quad - KB^T(r,v,t)\chi^2\left(\hat\theta_1,\hat\theta_2;4\varphi_1/\beta_1^2,4\varphi_2/\beta_2^2,\hat\omega_1[\beta_2^2 r-v],\hat\omega_2[v-\beta_1^2 r]\right), \end{aligned} \tag{8.43}$$

where

$$\theta_i = \frac{-4\gamma_i^2[a(S-T)+\ln K]}{\beta_i^2(e^{\gamma_i[T-t]}-1)^2\hat b_i(S-t)}, \quad i=1,2,$$

$$\hat\theta_i = \frac{-4\gamma_i^2[a(S-T)+\ln K]}{\beta_i^2(e^{\gamma_i[T-t]}-1)^2\hat b_i(T-t)\hat b_i(S-T)}, \quad i=1,2,$$

$$\omega_i = \frac{4\gamma_i e^{\gamma_i[T-t]}\hat b_i(S-t)}{\beta_i^2(\beta_2^2-\beta_1^2)(e^{\gamma_i[T-t]}-1)\hat b_i(S-T)}, \quad i=1,2,$$

$$\hat\omega_i = \frac{4\gamma_i e^{\gamma_i[T-t]}\hat b_i(T-t)}{\beta_i^2(\beta_2^2-\beta_1^2)(e^{\gamma_i[T-t]}-1)}, \quad i=1,2,$$

$$\hat b_i(\tau) = \frac{\gamma_i b_i(\tau)}{e^{\gamma_i\tau}-1}, \quad i=1,2.$$

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, see page 173, 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 (8.43) 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 numerically solve the fundamental partial differential equation with the appropriate terminal conditions, cf. 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

$$\begin{aligned}\frac{\partial P}{\partial t}(x_1,x_2,t) + (\hat{\varphi}_1 - \hat{\kappa}_1 x_1)\frac{\partial P}{\partial x_1}(x_1,x_2,t) + (\hat{\varphi}_2 - \hat{\kappa}_2 x_2)\frac{\partial P}{\partial x_2}(x_1,x_2,t)\\ + \frac{1}{2}\beta_1^2 x_1 \frac{\partial^2 P}{\partial x_1^2}(x_1,x_2,t) + \frac{1}{2}\beta_2^2 x_2 \frac{\partial^2 P}{\partial x_2^2}(x_1,x_2,t)\\ - (x_1+x_2)P(x_1,x_2,t) = 0, \quad (x_1,x_2,t) \in \mathbb{R}_+ \times \mathbb{R}_+ \times [0,T).\end{aligned}$$

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.

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, cf. 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, while 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.

8.6 Other multi-factor diffusion models

8.6.1 Models with stochastic consumer prices

Cox, Ingersoll, and Ross (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 (e.g. dollars) of nominal securities, i.e. 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

$$B^T(I,r,t) = I^{-1} \exp\left\{ - \left[a(T-t) + \left(\pi - \frac{1}{2}\beta_I^2\right)(T-t)\right] - b(T-t)r \right\},$$

where the functions $a(\tau)$ and $b(\tau)$ are exactly as in the one-factor CIR model, cf. (7.70) and (7.71) on page 170.

In their second extension 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, Ingersoll, and Ross (1985b).[3] Other affine models of this type have been studied by Chen and Scott (1993).

8.6.2 Models with stochastic long-term level and volatility

The Longstaff-Schwartz model described in Section 8.5.2 is not the only model that includes stochastic interest rate volatility. Vetzal (1997) considers the model

$$\begin{aligned} dr_t &= (\varphi_1 - \kappa_1 r_t) \, dt + \beta_t r_t^\gamma \, dz_{1t}, \\ d(\ln \beta_t^2) &= \left(\varphi_2 - \kappa_2 \ln \beta_t^2\right) dt + \rho \xi \, dz_{1t} + \sqrt{1-\rho^2} \xi \, dz_{2t}, \end{aligned}$$

while Andersen and Lund (1997) study the special case where the two processes are independent, i.e. $\rho = 0$. Both articles focus on the estimation and testing of the models. None of the models are affine or able to produce closed-form expressions for prices on bonds or options. Also Boudoukh, Richardson, Stanton, and Whitelaw (1999) consider a model with stochastic volatility of the short rate, but they use a non-parametric estimation technique in order to estimate the drift of the short rate and both the drift and the volatility of the short rate volatility.

Balduzzi, Das, Foresi, and Sundaram (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

[3] In addition, Cox, Ingersoll, and Ross (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.

form

$$\begin{aligned}
dr_t &= \kappa_r[\theta_t - r_t]\,dt + \sqrt{v_t}\,dz_{1t},\\
d\theta_t &= \kappa_\theta[\bar\theta - \theta_t]\,dt + \beta_\theta\,dz_{2t},\\
dv_t &= \kappa_v[\bar v - v_t]\,dt + \rho\beta_v\sqrt{v_t}\,dz_{1t} + \sqrt{1-\rho^2}\beta_v\sqrt{v_t}\,dz_{3t}.
\end{aligned}$$

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

$$\begin{aligned}
dr_t &= \left(\kappa_r[\theta_t - r_t] - \lambda_r v_t\right)\,dt + \sqrt{v_t}\,dz^{\mathbb{Q}}_{1t},\\
d\theta_t &= \left(\kappa_\theta[\bar\theta - \theta_t] - \lambda_\theta\beta_\theta\right)\,dt + \beta_\theta\,dz^{\mathbb{Q}}_{2t},\\
dv_t &= \left(\kappa_v[\bar v - v_t] - \lambda_v v_t\right)\,dt + \rho\beta_v\sqrt{v_t}\,dz^{\mathbb{Q}}_{1t} + \sqrt{1-\rho^2}\beta_v\sqrt{v_t}\,dz^{\mathbb{Q}}_{3t},
\end{aligned}$$

where λ_r, λ_θ, and λ_v are constants. The model is affine so that the zero-coupon bond prices are

$$B^T(r,\theta,v,t) = \exp\left\{-a(T-t) - b_1(T-t)r - b_2(T-t)\theta - b_3(T-t)v\right\}.$$

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, cf. (8.16) and (8.17). The model can produce a wide variety of interesting yield curve shapes. The estimation of the model is also discussed.

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

$$\begin{aligned}
dr_t &= \kappa_r[\theta_t - r_t]\,dt + \sqrt{v_t}\,dz_{1t},\\
d\theta_t &= \kappa_\theta[\bar\theta - \theta_t]\,dt + \beta_\theta\sqrt{\theta_t}\,dz_{2t},\\
dv_t &= \kappa_v[\bar v - v_t]\,dt + \beta_v\sqrt{v_t}\,dz_{3t},
\end{aligned}$$

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 $\bar v$. Since the model is affine, the zero-coupon bond prices are of the form

$$B^T(r,\theta,v,t) = \exp\left\{-a(T-t) - b_1(T-t)r - b_2(T-t)\theta - b_3(T-t)v\right\}.$$

Although the model is not a multi-factor CIR model, 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) divide the class of affine models into subclasses, and for each subclass they find the most general model that satisfies the conditions $\nu_j(\boldsymbol{x}) \geq 0$ $(j = 1, \dots, n)$ which ensure that the process for the state variables is well-defined. They focus on the three-factor models and show how the three-factor models suggested in the literature (e.g. the models of Chen and of Balduzzi, Das, Foresi, and Sundaram) fit into this classification. They find that the suggested models can be extended and, based on an empirical test using the historical evolution of zero-coupon yields of three different maturities, they conclude that the extensions are important in order for the models to be realistic. However, explicit expressions for the functions a, b_1, b_2, and b_3 have not been found in the extended models.

8.6.3 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 yields 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, Ma, and Yong (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.

8.6.4 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, e.g. 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 that gives a direct relation between security prices and representative yields for the different maturity segments.

As shown in (8.18), 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 (8.16) and (8.17) 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, i.e.

$$\bar{y}^{\tau_j}(\boldsymbol{x}) = x_j,$$

we must have that

$$b_j(\tau_j) = \tau_j, \qquad a(\tau_j) = b_i(\tau_j) = 0, \quad i \neq j. \tag{8.44}$$

These conditions impose very complicated restrictions on the parameters in the drift and the volatility terms in the dynamics of the state variables, i.e. 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 (8.44) have to be solved numerically.

An alternative procedure is to start with an affine model with other state variables so that the conditions (8.44) 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 (i.e. the key rates) is also an affine model.

8.6.5 Quadratic models

In Section 7.6 we gave a short introduction to quadratic one-factor models, i.e. 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, i.e.

$$r_t = \xi + \boldsymbol{\psi}^\top \boldsymbol{x}_t + \boldsymbol{x}_t^\top \underline{\underline{\Theta}}\,\boldsymbol{x}_t = \xi + \sum_{i=1}^n \psi_i x_{it} + \sum_{i=1}^n \sum_{j=1}^n \Theta_{ij} x_{it} x_{jt}.$$

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.

8.7 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, Hendershott, and Sanders (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.

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 will come to the defense of the 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 parameterization of the discount function as e.g. a cubic spline or a Nelson-Siegel parameterization described in Chapter 1. Probably, an equally good fit to the observed market prices and an economically more appropriate yield curve estimate can be obtained by applying the parameterization 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.

Chapter 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 (i.e. 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 by a carefully chosen time-dependent function. The model is said to be calibrated to the market term structure. The resulting model is time inhomogeneous. The calibrated model is consistent with the observed term structure and is therefore a relative pricing model (or pure no-arbitrage model), cf. 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 modeled 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 (7.6) on page 145 by deterministic functions, we get the short rate dynamics

$$dr_t = (\hat{\varphi}(t) - \hat{\kappa}(t) r_t)\, dt + \sqrt{\delta_1(t) + \delta_2(t) r_t}\, dz_t^{\mathbb{Q}} \tag{9.1}$$

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*

$$B^T(r,t) = e^{-a(t,T) - b(t,T) r} \tag{9.2}$$

and the functions $a(t,T)$ and $b(t,T)$ satisfy the following system of differential equations:

$$\frac{1}{2}\delta_2(t) b(t,T)^2 + \hat{\kappa}(t) b(t,T) - \frac{\partial b}{\partial t}(t,T) - 1 = 0, \tag{9.3}$$

$$\frac{\partial a}{\partial t}(t,T) + \hat{\varphi}(t) b(t,T) - \frac{1}{2}\delta_1(t) b(t,T)^2 = 0 \tag{9.4}$$

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 (9.3) and (9.4) by first solving (9.3) for $b(t,T)$ and then substituting that solution into (9.4), 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

$$y^T(r,t) = \frac{a(t,T)}{T-t} + \frac{b(t,T)}{T-t} r \tag{9.5}$$

and

$$f^T(r,t) = \frac{\partial a}{\partial T}(t,T) + \frac{\partial b}{\partial T}(t,T) r. \tag{9.6}$$

Both expressions are affine in r.

Next, let us look at the term structures of volatilities in these models, i.e. 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

$$dy_t^{t+\tau} = \dots\, dt + \frac{b(t,t+\tau)}{\tau}\beta(r_t,t)\, dz_t \tag{9.7}$$

and

$$df_t^{t+\tau} = \dots\, dt + \left.\frac{\partial b(t,T)}{\partial T}\right|_{T=t+\tau}\beta(r_t,t)\, dz_t, \tag{9.8}$$

respectively. Focusing on the volatilities, we have omitted the rather complicated drift terms.

From (9.3) 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, i.e. the ordinary differential equation (7.8) on page 146. 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 e.g. 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, i.e. 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 $a(T,T) = 0$, Eq. (9.4) 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,

$$a(0,T) = \int_0^T \hat{\varphi}(t) b(T-t)\, dt - \frac{\delta_1}{2}\int_0^T b(T-t)^2\, dt. \tag{9.9}$$

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, $\bar{B}(T)$, i.e.

$$a(0,T) = -b(T) r_0 - \ln \bar{B}(T) \tag{9.10}$$

for any time-to-maturity T. We can then determine $\hat{\varphi}(t)$ by comparing (9.9) and (9.10). In the extensions of the models of Merton and Vasicek we are able to find an explicit expression for $\hat{\varphi}(t)$, while 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

$$dr_t = \hat{\varphi}(t)\,dt + \beta\,dz_t^{\mathbb{Q}}, \tag{9.11}$$

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

$$a(t,T) = \int_t^T \hat{\varphi}(u)(T-u)\,du - \frac{1}{2}\beta^2 \int_t^T (T-u)^2\,du, \tag{9.12}$$

cf. 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* (9.11) *with*

$$\hat{\varphi}(t) = \bar{f}'(t) + \beta^2 t \tag{9.13}$$

for all t will be identical to the current term structure. In this case we have

$$a(t,T) = -\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right) - (T-t)\bar{f}(t) + \frac{1}{2}\beta^2 t(T-t)^2, \tag{9.14}$$

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 (9.9), 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,[1] 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

$$\frac{\partial^2 a}{\partial T^2}(0,T) = \hat{\varphi}(T) - \beta^2 T. \tag{9.15}$$

[1]If $h(t,T)$ is a deterministic function, which is differentiable in T, then

$$\frac{\partial}{\partial T}\left(\int_0^T h(t,T)\,dt\right) = h(T,T) + \int_0^T \frac{\partial h}{\partial T}(t,T)\,dt.$$

We wish to satisfy the relation (9.10), i.e.

$$a(0,T) = -Tr_0 - \ln \bar{B}(T).$$

Recall from (1.14) on page 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

$$\frac{\partial^2 a}{\partial T^2}(0,T) = \bar{f}'(T), \tag{9.16}$$

where we have assumed that the term structure of forward rates is differentiable. Comparing (9.15) and (9.16), we obtain the stated result.

Substituting (9.13) into (9.12), 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

$$\int_t^T \bar{f}(u)\,du = \int_0^T \bar{f}(u)\,du - \int_0^t \bar{f}(u)\,du = -\ln \bar{B}(T) + \ln \bar{B}(t) = -\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right). \tag{9.17}$$

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, i.e. 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 (7.43) on page 156 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 (7.50) by a time-dependent function $\hat{\theta}(t)$, we get the following short rate dynamics under the spot martingale measure:

$$dr_t = \kappa\left[\hat{\theta}(t) - r_t\right]dt + \beta\,dz_t^{\mathbb{Q}}. \tag{9.18}$$

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 (9.18) 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 on page 59 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,

$$\operatorname{Var}_{r,t}[r_T] = \operatorname{Var}^{\mathbb{Q}}_{r,t}[r_T] = \beta^2\int_t^T e^{-2\kappa[T-u]}\,du = \frac{\beta^2}{2\kappa}\left(1 - e^{-2\kappa[T-t]}\right),$$

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 spot martingale 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

$$B^T(r,t) = e^{-a(t,T) - b(T-t)r}, \tag{9.19}$$

where

$$b(\tau) = \frac{1}{\kappa}\left(1 - e^{-\kappa\tau}\right), \tag{9.20}$$

$$a(t,T) = \kappa\int_t^T \hat{\theta}(u)b(T-u)\,du + \frac{\beta^2}{4\kappa}b(T-t)^2 + \frac{\beta^2}{2\kappa^2}\left(b(T-t) - (T-t)\right). \tag{9.21}$$

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* (9.18) *with*

$$\hat{\theta}(t) = \bar{f}(t) + \frac{1}{\kappa}\bar{f}'(t) + \frac{\beta^2}{2\kappa^2}\left(1 - e^{-2\kappa t}\right) \tag{9.22}$$

will be identical to the current term structure of interest rates. In this case we have

$$a(t,T) = -\ln\left(\frac{\bar{B}(T)}{\bar{B}(t)}\right) - b(T-t)\bar{f}(t) + \frac{\beta^2}{4\kappa}b(T-t)^2\left(1 - e^{-2\kappa t}\right). \tag{9.23}$$

Proof: From Eq. (9.21) it follows that

$$a(0,t) = \kappa \int_0^t \hat{\theta}(u) b(t-u)\, du + \frac{\beta^2}{4\kappa} b(t)^2 + \frac{\beta^2}{2\kappa^2} \left(b(t) - t\right). \tag{9.24}$$

Repeated differentiations yield

$$\frac{\partial a}{\partial t}(0,t) = \kappa \int_0^t \hat{\theta}(u) e^{-\kappa[t-u]}\, du - \frac{1}{2}\beta^2 b(t)^2 \tag{9.25}$$

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 footnote 1, page 209). Consequently,

$$\hat{\theta}(t) = \frac{\beta^2}{2\kappa^2}\left(1 - e^{-2\kappa t}\right) + \frac{1}{\kappa}\frac{\partial^2 a}{\partial t^2}(0,t) + \frac{\partial a}{\partial t}(0,t). \tag{9.26}$$

Differentiation of the expression (9.10), which we want to be satisfied, yields

$$\frac{\partial a}{\partial t}(0,t) = -\frac{\bar{B}'(t)}{\bar{B}(t)} - r_0 e^{-\kappa t} = \bar{f}(t) - r_0 e^{-\kappa t} \tag{9.27}$$

and

$$\frac{\partial^2 a}{\partial t^2}(0,t) = \bar{f}'(t) + \kappa r_0 e^{-\kappa t}.$$

Substituting these expressions into (9.26), we obtain (9.22).

Substituting (9.26) into (9.21), we get

$$\begin{aligned} a(t,T) = &\int_t^T \bar{f}(u)\left(1 - e^{-\kappa[T-u]}\right) du + \frac{1}{\kappa}\int_t^T \bar{f}'(u)\left(1 - e^{-\kappa[T-u]}\right) du \\ &+ \frac{\beta^2}{2\kappa^2}\int_t^T \left(1 - e^{-2\kappa u}\right)\left(1 - e^{-\kappa[T-u]}\right) du + \frac{\beta^2}{4\kappa} b(T-t)^2 + \frac{\beta^2}{2\kappa^2}\left(b(T-t) - (T-t)\right).\end{aligned}$$

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 (9.17) 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 (9.23). □

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, i.e.

$$C^{K,T,S}(r,t) = B^S(r,t)N(d_1) - KB^T(r,t)N(d_2), \tag{9.28}$$

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 analyzed in Section 7.5 in the same way as we extended the models of Merton and Vasicek, the short rate dynamics becomes[2]

$$dr_t = (\kappa\theta(t) - \hat{\kappa}r_t)\, dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}}. \tag{9.29}$$

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, cf. (7.70) on page 170, 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

$$\bar{f}(T) = b'(T)r_0 + \kappa \int_0^T \theta(u)b'(T-u)\, du, \quad T > 0.$$

[2] This extension was suggested already in the original article by Cox, Ingersoll, and Ross (1985b).

According to Heath, Jarrow, and Morton (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 conditions that guarantee a well-defined short rate process. Clearly, a necessary condition for $\theta(t)$ to be non-negative for all t is that

$$\bar{f}(T) \geq r_0 b'(T), \quad T > 0. \tag{9.30}$$

Not all forward rate curves satisfy this condition, cf. 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, e.g. using the techniques outlined in Chapter 16.

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, e.g. caps, floors, and swaptions, and then apply the model for the pricing of less liquid securities. 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 (i.e. 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 Eqs. (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 (9.7) 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

$$dr_t = \kappa(t)[\hat{\theta}(t) - r_t]\, dt + \beta(t)\, dz_t^{\mathbb{Q}}. \tag{9.31}$$

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 (9.28) still applies. However, the relevant volatility is now $v(t,T,S)$, where

$$v(t,T,S)^2 = \int_t^T \beta(u)^2 \left[b(u,S) - b(u,T)\right]^2\, du = [b(0,S) - b(0,T)]^2 \int_t^T \beta(u)^2 \left(\frac{\partial b}{\partial u}(0,u)\right)^{-2}\, du,$$

cf. Hull and White (1990a). Jamshidian's result (7.30) 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, i.e. dependent on the calendar time, cf. 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 to "stretch 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 Section 1.6, 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 parameterization

$$\bar{f}(t) = c_1 + c_2 e^{-kt} + c_3 t e^{-kt}, \tag{9.32}$$

cf. (1.30), ensures a nice and smooth forward rate curve and will presumably be more suitable in the calibration procedure.

No matter which of these parameterizations 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 9.2.

Recall that the cubic spline and the Nelson-Siegel parameterizations 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 parameterization of the discount function, e.g. (7.67) and the associated expressions for a and b in the Cox-Ingersoll-Ross model. Why not use such a parameterization instead of the cubic spline or the Nelson-Siegel parameterization? And if the parameterization 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 parameterization of the current yield curve and the model dynamics of interest rates. For example, if we use a Nelson-Siegel parameterization (9.32) of the current yield curve and let this yield curve evolve according to a dynamic model, e.g. the Hull-White model, will the future yield curves also be of the form (9.32)? Intuitively, it seems reasonable to use a parameterization which is consistent with the model dynamics, in the sense that the possible future yield curves can be written on the same parameterized from, although possibly with other parameter values.

Which parameterizations 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 parameterization $\bar{f}(t) = c_1 + c_2 t$ is consistent with the Ho-Lee model (9.11), i.e. 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 parameterization of the forward rate curve, which is consistent with the Hull-White model (9.18), is
$$\bar{f}(t) = c_1 e^{-kt} + c_2 e^{-2kt}.$$

- The Nelson-Siegel parameterization (9.32) is consistent neither with the Ho-Lee model nor the Hull-White model. However, the extended Nelson-Siegel parameterization
$$\bar{f}(t) = c_1 + c_2 e^{-kt} + c_3 t e^{-kt} + c_4 e^{-2kt}$$
is consistent with the Hull-White model.

Furthermore, it can be shown that the Nelson-Siegel parameterization is not consistent with any non-trivial one-factor diffusion model, cf. Filipović (1999).

9.8 Calibrated non-affine models

In Section 7.6 on page 174 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 (7.82) can thus be extended to

$$d(\ln r_t) = \kappa(t)(\theta(t) - \ln r_t)\, dt + \beta(t)\, dz_t^{\mathbb{Q}}, \tag{9.33}$$

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 Karasinksi 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, Derman, and Toy (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.[3] 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 (7.83) and cannot be calibrated to the observed yield curve.

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 8.7, 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

[3] It is not clear from Black, Derman, and Toy (1990) how a calibrated tree can be constructed, but Jamshidian (1991) fills this gap in their presentation.

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 on page 182, 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, cf. 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 8.6. 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

$$dr_t = (\hat{\varphi}(t) - \hat{\kappa}(t) r_t)\, dt + \beta \sqrt{r_t}\, dz_t^{\mathbb{Q}},$$

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, i.e. 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 (7.70) on page 170. 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 (9.30) 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

$$\bar{B}(t) = e^{-a(t)-b(t)r_0}, \tag{*}$$

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, cf. (7.51)–(7.53) on page 159.

(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 (9.22) 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 (*), i.e. as in the Vasicek model.

Chapter 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, Jarrow, and Morton (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 too deep 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

$$df_t^T = \alpha\left(t, T, (f_t^s)_{s \geq t}\right) dt + \sum_{i=1}^{n} \beta_i\left(t, T, (f_t^s)_{s \geq t}\right) dz_{it}, \quad 0 \leq t \leq T, \tag{10.1}$$

[1] The binomial model of Ho and Lee (1986) can be seen as a forerunner of the more complete and thorough HJM-analysis.

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. As discussed in Chapter 6 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

$$B_t^T = e^{-\int_t^T f_t^s \, ds} = e^{-y_t^T (T-t)}, \tag{10.2}$$

$$f_t^T = -\frac{\partial \ln B_t^T}{\partial T} = y_t^T + (T-t) \frac{\partial y_t^T}{\partial T}, \tag{10.3}$$

$$y_t^T = \frac{1}{T-t} \int_t^T f_t^s \, ds = -\frac{1}{T-t} \ln B_t^T. \tag{10.4}$$

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 modeling 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 (i.e. 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

$$r_t = f_t^t = \lim_{T \downarrow t} y_t^T = -\lim_{T \downarrow t} \frac{\partial \ln B_t^T}{\partial T}.$$

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 modeling 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.

[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.

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*

$$dB_t^T = B_t^T \left[\mu^T(t, (f_t^s)_{s \geq t})\, dt + \sum_{i=1}^n \sigma_i^T(t, (f_t^s)_{s \geq t})\, dz_{it} \right], \tag{10.5}$$

where

$$\mu^T(t, (f_t^s)_{s \geq t}) = r_t - \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du + \frac{1}{2} \sum_{i=1}^n \left(\int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du \right)^2, \tag{10.6}$$

$$\sigma_i^T(t, (f_t^s)_{s \geq t}) = - \int_t^T \beta_i(t, u, (f_t^s)_{s \geq t})\, du, \tag{10.7}$$

Proof: For simplicity we only proof 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 3.4 on page 54, 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 3.5 on page 55) implies that the dynamics of the zero coupon bond prices is

$$\begin{aligned} dB_t^T &= \left\{ -e^{-Y_t} \left(-r_t + \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du \right) + \frac{1}{2} e^{-Y_t} \left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du \right)^2 \right\} dt \\ &\quad - e^{-Y_t} \left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du \right) dz_t \\ &= B_t^T \Bigg[r_t - \int_t^T \alpha(t, u, (f_t^s)_{s \geq t})\, du + \frac{1}{2} \left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du \right)^2 dt \\ &\quad - \left(\int_t^T \beta(t, u, (f_t^s)_{s \geq t})\, du \right) dz_t \Bigg], \end{aligned}$$

which gives the one-factor version of (10.5). □

Now we turn to the behavior 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 $\boldsymbol{z}^{\mathbb{Q}} = (z_1^{\mathbb{Q}}, \dots, z_n^{\mathbb{Q}})^\top$ 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*

$$\hat{\alpha}(t, T, (f_t^s)_{s\geq t}) = \sum_{i=1}^n \beta_i(t, T, (f_t^s)_{s\geq t}) \int_t^T \beta_i(t, u, (f_t^s)_{s\geq t})\, du. \tag{10.8}$$

The relation (10.8) is called **the HJM drift restriction**. The drift restriction has important consequences: Firstly, the forward rate behavior 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 to separately specify the market prices of risk.

10.4 Three well-known special cases

Since the general HJM framework is quite abstract, we will in this section 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 $\beta(t, T, (f_t^s)_{s \geq t}) = \beta > 0$, i.e. the forward rate volatilities are identical for all maturities (independent of T) and constant over time (independent of t). From the HJM drift restriction (10.8), 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,

$$dr_t = \hat{\varphi}(t) \, dt + \beta \, dz_t^{\mathbb{Q}}, \tag{10.9}$$

where $\hat{\varphi}(t) = \partial f_0^t / \partial t + \beta^2 t$. From (10.9), 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 on page 209. 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

$$C_t^{K,T,S} = B_t^S N(d_1) - K B_t^T N(d_2), \tag{10.10}$$

where

$$d_1 = \frac{1}{v(t, T, S)} \ln\left(\frac{B_t^S}{K B_t^T}\right) + \frac{1}{2} v(t, T, S), \tag{10.11}$$

$$d_2 = d_1 - v(t, T, S), \tag{10.12}$$

$$v(t, T, S) = \beta [S - T] \sqrt{T - t}. \tag{10.13}$$

In addition, Jamshidian's trick for the pricing of European options on coupon bonds (see Theorem 7.3 on page 151) 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

$$\beta(t,T,(f_t^s)_{s\geq t}) = \beta e^{-\kappa[T-t]} \tag{10.14}$$

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

$$\begin{aligned} g(t) &= f_0^t + \int_0^t \frac{\beta^2}{\kappa} e^{-\kappa[t-u]} \left(1 - e^{-\kappa[t-u]}\right) du \\ &= f_0^t + \frac{\beta^2}{2\kappa^2} \left(1 - e^{-\kappa t}\right)^2. \end{aligned}$$

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 on page 210 reveals that the HJM one-factor model with forward rate volatilities given by (10.14) 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 (10.10), but where

$$v(t,T,S) = \frac{\beta}{\sqrt{2\kappa^3}}\left(1 - e^{-\kappa[S-T]}\right)\left(1 - e^{-2\kappa[T-t]}\right)^{1/2}. \tag{10.15}$$

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

$$dr_t = (\kappa\theta(t) - \hat{\kappa} r_t)\, dt + \beta\sqrt{r_t}\, dz_t^{\mathbb{Q}}$$

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 (10.7). 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

$$\beta(t,T,(f_t^s)_{s\geq t}) = \frac{4\gamma^2 e^{\gamma[T-t]}}{\left((\gamma + \hat{\kappa})(e^{\gamma[T-t]} - 1) + 2\gamma\right)^2}\beta\sqrt{r_t}. \tag{10.16}$$

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, cf. 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, i.e.

$$\beta_i(t, T, (f_t^s)_{s \geq t}) = \beta_i(t, T), \quad i = 1, 2, \dots, n.$$

To see this, first note that from the drift restriction (10.8) 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, cf. Theorem 3.2 on page 53. The future forward rates are therefore normally distributed under $\mathbb{Q}$. The short-term interest rate is $r_t = f_t^t$, i.e.

$$r_t = f_0^t + \int_0^t \hat{\alpha}(u, t)\, du + \sum_{i=1}^{n} \int_0^t \beta_i(u, t)\, dz_{iu}^{\mathbb{Q}}, \quad 0 \leq t, \tag{10.17}$$

which is also normally distributed under 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 $\beta_i(t, T, (f_t^s)_{s \geq t})$ 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*

$$C_t^{K,T,S} = B_t^S N(d_1) - K B_t^T N(d_2), \tag{10.18}$$

where

$$d_1 = \frac{1}{v(t, T, S)} \ln\left(\frac{B_t^S}{K B_t^T}\right) + \frac{1}{2} v(t, T, S), \tag{10.19}$$

$$d_2 = d_1 - v(t, T, S), \tag{10.20}$$

$$v(t, T, S) = \left(\sum_{i=1}^{n} \int_t^T \left[\int_T^S \beta_i(u, y)\, dy\right]^2 du\right)^{1/2}. \tag{10.21}$$

[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 $\mathbb{P}$.

Proof: We will apply the same procedure as we did in the diffusion models of Chapter 7, see e.g. the derivation of the option price in the Vasicek model in Section 7.4.5. The option price is given by

$$C_t^{K,T,S} = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max\left(B_T^S - K, 0\right)\right] = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max\left(F_T^{T,S} - K, 0\right)\right], \tag{10.22}$$

where $\mathbb{Q}^T$ denotes the T-forward martingale measure introduced in Section 6.2.2 on page 117. 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 (10.7) 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 3.2 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

$$\begin{aligned} v(t,T,S)^2 &= \int_t^T \left[\int_T^S \beta_1\,dy\right]^2 du + \int_t^T \left[\int_T^S \beta_2 e^{-\kappa[y-u]}\,dy\right]^2 du \\ &= \beta_1^2 [S-T]^2 [T-t] + \frac{\beta_2^2}{2\kappa^3}\left(1 - e^{-\kappa[S-T]}\right)^2 \left(1 - e^{-2\kappa[T-t]}\right), \end{aligned}$$

cf. (10.13) and (10.15).

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.

10.6 Diffusion representations of HJM models

As discussed immediately below the basic assumption (10.1) on page 221, 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

For the purpose of using numerical techniques for derivative pricing, it is crucial whether or not the relevant uncertainty can be described by some low-dimensional diffusion process. 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 n time steps has $n+1$ endnodes if it is recombining, but 2^n 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, cf. the analysis in Chapter 6. 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, etc., 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, see, e.g., Boyle, Broadie, and Glasserman (1997), Broadie and Glasserman (1997b), Carr and Yang (1997), Andersen (2000), and Longstaff and Schwartz (2001). 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*

$$dr_t = \left\{ \frac{\partial f_0^t}{\partial t} + \sum_{i=1}^n \int_0^t \frac{\partial \beta_i(u,t,(f_u^s)_{s\geq u})}{\partial t} \left[\int_u^t \beta_i(u,x,(f_u^s)_{s\geq u})\, dx \right] du + \sum_{i=1}^n \int_0^t \beta_i(u,t,(f_u^s)_{s\geq u})^2\, du \right.$$
$$\left. + \sum_{i=1}^n \int_0^t \frac{\partial \beta_i(u,t,(f_u^s)_{s\geq u})}{\partial t}\, dz_{iu}^{\mathbb{Q}} \right\} dt + \sum_{i=1}^n \beta_i(t,t,(f_t^s)_{s\geq t})\, dz_{it}^{\mathbb{Q}}. \quad (10.23)$$

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 (10.8). 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

$$\begin{aligned} r_t &= f_0^t + \int_0^t \hat{\alpha}(u,t,(f_u^s)_{s\geq u})\, du + \sum_{i=1}^n \int_0^t \beta_i(u,t,(f_u^s)_{s\geq u})\, dz_{iu}^{\mathbb{Q}} \\ &= f_0^t + \sum_{i=1}^n \int_0^t \beta_i(u,t,(f_u^s)_{s\geq u}) \left[\int_u^t \beta_i(u,x,(f_u^s)_{s\geq u})\, dx \right] du + \sum_{i=1}^n \int_0^t \beta_i(u,t,(f_u^s)_{s\geq u})\, dz_{iu}^{\mathbb{Q}}. \end{aligned} \quad (10.24)$$

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 3.4 on page 54). 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 (10.23). □

From (10.23) 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 (10.16) 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*

$$dr_t = \left[\frac{\partial f_0^t}{\partial t} + h(t)^2 \sum_{i=1}^n \int_0^t g_i(u)^2 \, du + \frac{h'(t)}{h(t)} (r_t - f_0^t) \right] dt + \sum_{i=1}^n g_i(t) h(t) \, dz_{it}^{\mathbb{Q}}, \tag{10.25}$$

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

$$\beta(t, T, (f_t^s)_{s \geq t}) = g(t) h(T), \tag{10.26}$$

where g and h are deterministic functions and h is continuously differentiable, non-zero, and never changing sign. First note that (10.24) and (10.26) imply that

$$r_t = f_0^t + h(t) \int_0^t g(u)^2 \left[\int_u^t h(x) \, dx \right] du + h(t) \int_0^t g(u) \, dz_u^{\mathbb{Q}}, \tag{10.27}$$

and, thus,

$$\int_0^t g(u) \, dz_u^{\mathbb{Q}} = \frac{1}{h(t)} (r_t - f_0^t) - \int_0^t g(u)^2 \left[\int_u^t h(x) \, dx \right] du. \tag{10.28}$$

The dynamics of r in Equation (10.23) 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 (10.28) can be written as the one-factor version of (10.25). □

Note that the Ho-Lee model and the Hull-White model studied in Section 10.4 both satisfy the condition (10.26).

[4] Carverhill claims that the h function can be different for each factor, i.e., $\beta_i(t, T, (f_t^s)_{s \geq t}) = g_i(t) h_i(T)$, but this is incorrect.

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 (10.18). It can be shown that with a volatility specification of the form (10.26), 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 on page 150 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, e.g., forward rates, zero-coupon bond yields, and zero-coupon bond prices do not depend on calendar time in itself, cf. 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 (10.7) 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 empirically seems to depend 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

$$\beta(t, T, (f_t^s)_{s \geq t}) = \beta(t, t, (f_t^s)_{s \geq t}) e^{-\int_t^T \kappa(x)\, dx} \tag{10.29}$$

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 (10.29) 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:

$$B_t^T = e^{-a(t,T) - b_1(t,T) r_t - b_2(t,T) \varphi_t},$$

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

$$dr_t = \left(\frac{\partial f_0^t}{\partial t} + \varphi_t - \kappa(t)[r_t - f_0^t]\right) dt + \beta(t,t,(f_t^s)_{s\geq t})\,dz_t^{\mathbb{Q}},$$
$$d\varphi_t = \left(\beta(t,t,(f_t^s)_{s\geq t})^2 - 2\kappa(t)\varphi_t\right) dt.$$

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, i.e. 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 allowable specification is $\beta_r(r,\varphi,t) = \beta r^\gamma$ for some non-negative constants β and γ, which, e.g., 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.

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, Chiarella, El-Hassan, and Zheng (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, i.e. of the form

$$\beta_i(t,T,(f_t^s)_{s\geq t}) = \beta_i(t,T,f_t^T). \tag{10.30}$$

[5] Li, Ritchken, and Sankarasubramanian (1995) show how to build a tree for this model, in which both European- and American-type term structure derivatives can be efficiently priced.

A model of this type, inspired by the Black-Scholes' stock option pricing model, is obtained by letting

$$\beta_i(t,T,f_t^T) = \gamma_i(t,T) f_t^T, \tag{10.31}$$

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 (10.31) 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, Jarrow, and Morton (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, i.e. 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 (10.30) to yield non-negative and non-exploding interest rates. One of the conditions is that the forward rate volatility is bounded from above. This is, obviously, not satisfied for proportional volatility models, i.e. models where (10.31) holds.

10.8 Concluding remarks

Empirical studies of various specifications of the HJM model framework have been performed on a variety of data sets by, e.g., Amin and Morton (1994), Flesaker (1993), Heath, Jarrow, and Morton (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 $\beta_i(t,T,(f_t^s)_{s\geq t})$ *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 parameterization 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 (7.57) on page 165 that the forward rates at time t are given by

$$\begin{aligned} f_t^T &= \left(1 - e^{-\kappa[T-t]}\right)\left(y_\infty + \frac{\beta^2}{2\kappa^2} e^{-\kappa[T-t]}\right) + e^{-\kappa[T-t]} r_t \\ &= y_\infty + \left(\frac{\beta^2}{2\kappa^2} + r_t - y_\infty\right) e^{-\kappa[T-t]} - \frac{\beta^2}{2\kappa^2} e^{-2\kappa[T-t]}, \end{aligned}$$

[6] This was shown by Morton (1988).

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 do generally not have this nice property, and neither do the HJM-models studied in this chapter. If we use a given parameterization of the initial forward curve, then we cannot be sure that the future forward curves can be described by the same parameterization 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

$$f_0^T = r_0 + \hat{\kappa}\left[\hat{\theta} - r\right] b(T) - \frac{1}{2}\beta^2 r b(T)^2,$$

cf. (7.72) on page 171, where the function $b(T)$ is given by (7.70). With such an initial forward rate curve, the one-factor HJM model with forward rate volatility function given by (10.16) is indistinguishable from the original time homogeneous one-factor CIR model.

Chapter 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, e.g. caps, floors, swaps, and swaptions, depend on periodically compounded interest rates such as spot LIBOR rates $l_t^{t+\delta}$, forward LIBOR rates $L_t^{T,T+\delta}$, spot swap rates $\tilde{l}_t^{\delta}$, and forward swap rates $\tilde{L}_t^{T,\delta}$. 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 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 $L_t^{T,T+\delta}$ 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.

In the recent literature 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 $L_t^{T,T+\delta}$ 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 7.6 on page 174, 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.

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 2.8, 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

$$\mathcal{C}^i_{T_i} = H\delta \max\left(l^{T_i}_{T_i-\delta} - K, 0\right) = H\delta \max\left(L^{T_i-\delta,T_i}_{T_i-\delta} - K, 0\right),$$

where H is the face value of the cap. A cap can be considered as a portfolio of caplets, namely one caplet for each payment date.

The definition of the forward martingale measures in Chapter 6 implies that 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, i.e.

$$\mathcal{C}^i_t = H\delta B^{T_i}_t \, \mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(L^{T_i-\delta,T_i}_{T_i-\delta} - K, 0\right)\right], \quad t < T_i - \delta. \tag{11.1}$$

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 riskless discount factor, while the remaining payoffs are valued as above. For more details see Section 2.8. The price of the corresponding floor is

$$\mathcal{F}_t = H\delta \sum_{i=1}^{n} B^{T_i}_t \, \mathrm{E}^{\mathbb{Q}^{T_i}}_t \left[\max\left(K - L^{T_i-\delta,T_i}_{T_i-\delta}, 0\right)\right], \quad t < T_0. \tag{11.3}$$

In order to compute the cap price from (11.2), we need knowledge of the distribution of $L^{T_i-\delta,T_i}_{T_i-\delta}$ 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 $L^{T_i-\delta,T_i}_t$ under $\mathbb{Q}^{T_i}$. The following argument shows that under the $\mathbb{Q}^{T_i}$ probability measure the drift rate of $L^{T_i-\delta,T_i}_t$ is zero, i.e. $L^{T_i-\delta,T_i}_t$ is a $\mathbb{Q}^{T_i}$-martingale. Remember from Eq. (1.9) on page 8 that

$$L^{T_i-\delta,T_i}_t = \frac{1}{\delta}\left(\frac{B^{T_i-\delta}_t}{B^{T_i}_t} - 1\right). \tag{11.4}$$

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 $B^{T_i}_t$ is a martingale. In particular, the ratio $B^{T_i-\delta}_t / B^{T_i}_t$ 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 that also 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 (i.e. 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

$$dL_t^{T_i-\delta,T_i} = \beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right) dz_t^{T_i}, \quad t < T_i - \delta, \quad i = 1,\dots,n, \tag{11.5}$$

where z^{T_i} is a one-dimensional standard Brownian motion under the T_i-forward martingale measure $\mathbb{Q}^{T_i}$. The symbol $(L_t^{T_j,T_j+\delta})_{T_j\geq t}$ 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

$$\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right) = \gamma(t, T_i-\delta, T_i) L_t^{T_i-\delta,T_i}$$

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 $B_t^{T_1}, B_t^{T_2}, \dots, B_t^{T_n}$. 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 modeled forward rates, i.e. $L_t^{T_0,T_1}, L_t^{T_1,T_2}, \dots, L_t^{T_{n-1},T_n}$. 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) for 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. 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, i.e. 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.

[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.

The terminal measure

We wish to describe the evolution in all the modeled 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 6.2.2 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

$$dz_t^{T_n} = dz_t^{T_i} + \sum_{j=i}^{n-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)}{1+\delta f_s(t, T_j, T_{j+1})}\, dt.$$

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

$$\begin{aligned}
dL_t^{T_i-\delta,T_i} &= \beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) \, dz_t^{T_i} \\
&= \beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) \left[dz_t^{T_n} - \sum_{j=i}^{n-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right)}{1 + \delta L_t^{T_j,T_{j+1}}} \, dt \right] \\
&= -\sum_{j=i}^{n-1} \frac{\delta\beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) \beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right)}{1 + \delta L_t^{T_j,T_{j+1}}} \, dt \\
&\quad + \beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) \, dz_t^{T_n}.
\end{aligned} \tag{11.9}$$

Note that the drift may involve some or all of the other modeled forward rates. Therefore, the vector of all the forward rates $(L_t^{T_0,T_1}, \ldots, L_t^{T_{n-1},T_n})$ 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, i.e. 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 modeled forward rates at and before time T. Let P_t denote the time t value of this asset (measured in monetary units, e.g. 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

$$P_t = B_t^{T_n} \, \mathrm{E}_t^{\mathbb{Q}^{T_n}} \left[\frac{H_{T_k}}{B_{T_k}^{T_n}} \right].$$

From (11.4) we have that

$$\begin{aligned}
\frac{1}{B_{T_k}^{T_n}} &= \frac{B_{T_k}^{T_k}}{B_{T_k}^{T_{k+1}}} \frac{B_{T_k}^{T_{k+1}}}{B_{T_k}^{T_{k+2}}} \cdots \frac{B_{T_k}^{T_{n-1}}}{B_{T_k}^{T_n}} \\
&= \left[1 + \delta L_{T_k}^{T_k,T_{k+1}}\right] \left[1 + \delta L_{T_k}^{T_{k+1},T_{k+2}}\right] \cdots \left[1 + \delta L_{T_k}^{T_{n-1},T_n}\right] \\
&= \prod_{j=k}^{n-1} \left[1 + \delta L_{T_k}^{T_j,T_{j+1}}\right]
\end{aligned}$$

[2] However, Andersen and Andreasen (2000) introduce a trick that may reduce the computational complexity considerably.

so that the price can be rewritten as

$$P_t = B_t^{T_n} \, \mathrm{E}_t^{\mathbb{Q}^{T_n}} \left[H_{T_k} \prod_{j=k}^{n-1} \left[1 + \delta L_{T_k}^{T_j,T_{j+1}} \right] \right]. \tag{11.10}$$

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:

$$P_t = B_t^{T_n} \, \mathrm{E}_t^{\mathbb{Q}^{T_n}} \left[H_{T_n} \right]. \tag{11.11}$$

In that case it suffices to simulate the evolution of the forward rates that determine the payoff of the security.

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, cf. the discussion in Section 6.2. 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 one 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, etc. Let us define

$$I(t) = \min \left\{ i \in \{1, 2, \dots, n\} : T_i \geq t \right\}$$

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

$$A_t^* = B_t^{T_{I(t)}} N_t = B_t^{T_{I(t)}} \prod_{j=0}^{I(t)-1} \frac{1}{B_{T_j}^{T_{j+1}}} = B_t^{T_{I(t)}} \prod_{j=0}^{I(t)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}} \right], \tag{11.12}$$

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 modeled 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

$$\frac{A^*_t}{A^*_T} = \frac{B_t^{T_{I(t)}} \prod_{j=0}^{I(t)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]}{B_T^{T_{I(T)}} \prod_{j=0}^{I(T)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]} = \frac{B_t^{T_{I(t)}}}{B_T^{T_{I(T)}}} \prod_{j=I(t)}^{I(T)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]^{-1},$$

we get that the price can be rewritten as

$$P_t = B_t^{T_{I(t)}} \,\mathrm{E}_t^{\mathbb{Q}^*} \left[\frac{H_T}{B_T^{T_{I(T)}}} \prod_{j=I(t)}^{I(T)-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]^{-1} \right]. \tag{11.13}$$

In particular, if T is one of the dates in the tenor structure, say $T = T_k$, we get

$$P_t = B_t^{T_{I(t)}} \,\mathrm{E}_t^{\mathbb{Q}^*} \left[H_{T_k} \prod_{j=I(t)}^{k-1} \left[1 + \delta L_{T_j}^{T_j,T_{j+1}}\right]^{-1} \right] \tag{11.14}$$

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 that the process z^* defined by

$$dz^*_t = dz_t^{T_i} - \left[\sigma_t^{T_{I(t)}} - \sigma_t^{T_i}\right] dt$$

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

$$dz^*_t = dz_t^{T_i} - \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right)}{1 + \delta L_t^{T_j,T_{j+1}}}\, dt. \tag{11.15}$$

Substituting this relation into (11.5), we can rewrite the dynamics of the forward rates under the spot LIBOR martingale measure as

$$\begin{aligned} dL_t^{T_i-\delta,T_i} &= \beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) dz_t^{T_i} \\ &= \beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) \left[dz^*_t + \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right)}{1 + \delta L_t^{T_j,T_{j+1}}}\, dt \right] \\ &= \sum_{j=I(t)}^{i-1} \frac{\delta\beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) \beta\left(t, T_j, T_{j+1}, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right)}{1 + \delta L_t^{T_j,T_{j+1}}}\, dt \\ &\quad + \beta\left(t, T_i - \delta, T_i, (L_t^{T_k,T_k+\delta})_{T_k \geq t}\right) dz^*_t. \end{aligned} \tag{11.16}$$

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, cf. (10.8) on page 224.

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 $\{T_1, T_2, \dots, T_n\}$, and where the size of the payment only depends on the modeled 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 (11.14). 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 (11.14) 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 Eq. (1.9) on page 8 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

$$\begin{aligned} L_t^{T_0,T_2} &= \frac{1}{T_2 - T_0}\left(\frac{B_t^{T_0}}{B_t^{T_2}} - 1\right) \\ &= \frac{1}{T_2 - T_0}\left(\frac{B_t^{T_0}}{B_t^{T_1}}\frac{B_t^{T_1}}{B_t^{T_2}} - 1\right) \\ &= \frac{1}{2\delta}\left(\left[1 + \delta L_t^{T_0,T_1}\right]\left[1 + \delta L_t^{T_1,T_2}\right] - 1\right), \end{aligned} \tag{11.17}$$

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 (11.18), determine the distribution of the forward rate $L_t^{T_0,T_2}$. A LIBOR market model based on three-month interest rates can hence also be used for the pricing of contracts that depend on six-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 caps is Black's formula, i.e. formula (6.58) on page 136. As discussed in Chapter 6, 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), while 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. Whereas Miltersen, Sandmann, and Sondermann derive the cap price formula using PDEs, we will follow Brace, Gatarek, and Musiela and use the forward martingale measure technique discussed in Chapter 6 since this simplifies the analysis considerably.

In the development of Black's cap price formula in Chapter 6, we assumed among other things that the forward rate $L_t^{T_i-\delta,T_i}$ was a martingale under the spot martingale measure $\mathbb{Q}$ and that the future value $L_{T_i-\delta}^{T_i-\delta,T_i}$ was lognormally distributed under $\mathbb{Q}$. However, as shown in Theorem 11.1 this forward rate is a martingale under the T_i-forward martingale measure and will therefore not be a martingale under the $\mathbb{Q}$-measure. (Remember: a change of measure corresponds to changing the drift rate.) 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**:

$$dL_t^{T_i-\delta,T_i} = L_t^{T_i-\delta,T_i}\gamma(t,T_i-\delta,T_i)\,dz_t^{T_i}, \quad i=1,2,\dots,n, \tag{11.18}$$

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

$$\ln L_{T_i-\delta}^{T_i-\delta,T_i} = \ln L_t^{T_i-\delta,T_i} - \frac{1}{2}\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du + \int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)\,dz_u^{T_i}.$$

Because γ is a deterministic function, it follows from Theorem 3.2 on page 53 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,

$$\ln L_{T_i-\delta}^{T_i-\delta,T_i} \sim N\left(\ln L_t^{T_i-\delta,T_i} - \frac{1}{2}\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du, \int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du\right)$$

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* (11.18) *the price of the caplet with payment date T_i at any time $t < T_i - \delta$ is given by*

$$\mathcal{C}_t^i = H\delta B_t^{T_i}\left[L_t^{T_i-\delta,T_i}N(d_{1i}) - KN(d_{2i})\right], \tag{11.19}$$

where

$$d_{1i} = \frac{\ln\left(L_t^{T_i-\delta,T_i}/K\right)}{v_L(t,T_i-\delta,T_i)} + \frac{1}{2}v_L(t,T_i-\delta,T_i), \tag{11.20}$$

$$d_{2i} = d_{1i} - v_L(t,T_i-\delta,T_i), \tag{11.21}$$

$$v_L(t,T_i-\delta,T_i) = \left(\int_t^{T_i-\delta}\gamma(u,T_i-\delta,T_i)^2\,du\right)^{1/2}. \tag{11.22}$$

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 (11.19) is identical to Black's formula (6.54) 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* (11.18) *the price of a cap at any time* $t < T_0$ *is given as*

$$\mathcal{C}_t = H\delta\sum_{i=1}^{n} B_t^{T_i}\left[L_t^{T_i-\delta,T_i}N(d_{1i})-KN(d_{2i})\right], \tag{11.23}$$

where d_{1i} *and* d_{2i} *are as in* (11.20) *and* (11.21).

For $t \geq T_0$ the first-coming payment of the cap is known and is therefore to be discounted with the riskless discount factor, while the remaining payments are to be valued as above. For details, see Section 2.8.

Analogously, the price of a floor under the assumption (11.18) is

$$\mathcal{F}_t = H\delta\sum_{i=1}^{n} B_t^{T_i}\left[KN\left(-d_{2i}\right)-L_t^{T_i-\delta,T_i}N\left(-d_{1i}\right)\right], \quad t < T_0. \tag{11.24}$$

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 (11.18) is replaced by

$$dL_t^{T_i-\delta,T_i} = L_t^{T_i-\delta,T_i}\sum_{j=1}^{d}\gamma_j(t,T_i-\delta,T_i)\,dz_{jt}^{T_i}, \tag{11.25}$$

where all $\gamma_j(t,T_i-\delta,T_i)$ are bounded and deterministic functions. Again, the cap price is given by (11.23) with the small change that $v_L(t,T_i-\delta,T_i)$ is to be computed as

$$v_L(t,T_i-\delta,T_i) = \left(\sum_{j=1}^{d}\int_t^{T_i-\delta}\gamma_j(u,T_i-\delta,T_i)^2\,du\right)^{1/2}. \tag{11.26}$$

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 (11.18). The reason is that when the forward LIBOR rates have volatilities proportional to their level, the volatility of the forward swap rate 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:

$$\mathcal{P}_t = H\delta \sum_{i=1}^{n} B_t^{T_i} \left[L_t^{T_i-\delta,T_i} N(d_{1i}^*) - KN(d_{2i}^*) \right], \quad t < T_0, \tag{11.27}$$

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 (11.18). This approximation delivers the price much faster than a Monte Carlo simulation. Brace, Gatarek, and Musiela provide numerical examples in which the price computed using the approximation (11.27) 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, cf. Exercise 11.1.

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, etc., 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 (11.17) we have that

$$L_t^{T_0,T_2} = \frac{1}{2\delta} \left(\left[1 + \delta L_t^{T_0,T_1} \right] \left[1 + \delta L_t^{T_1,T_2} \right] - 1 \right).$$

According to the assumption (11.18) 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 $L_t^{T_0,T_2}$. 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, e.g. 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 so far this issue has not been satisfactorily 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

$$\beta \left(t, T_i - \delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j \ge t} \right) = L_t^{T_i-\delta,T_i} \gamma(t, T_i - \delta, T_i),$$

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, alternative volatility specifications may be more realistic (see Section 11.6). Below we will consider a tractable and empirically relevant alternative LIBOR market model.

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 (6.54) 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 (i.e. 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 (11.18).[3]

Andersen and Andreasen consider a so-called CEV LIBOR market model where the forward rate volatility is given as

$$\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right) = \left(L_t^{T_i-\delta,T_i}\right)^{\alpha}\gamma(t, T_i-\delta, T_i), \quad i=1,\dots,n,$$

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

$$\mathcal{C}_t^i = H\delta B_t^{T_i}\left[L_t^{T_i-\delta,T_i}\left(1-\chi^2(a;b,c)\right) - K\chi^2(c;b',a)\right] \tag{11.28}$$

[3] Hull (2003, Ch. 15) has a detailed discussion of the similar phenomenon for stock and currency options.

[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

$$\begin{aligned}
&\frac{\partial\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)/\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)}{\partial L_t^{T_i-\delta,T_i}/L_t^{T_i-\delta,T_i}} \\
&\qquad = \frac{\partial\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)}{\partial L_t^{T_i-\delta,T_i}}\frac{L_t^{T_i-\delta,T_i}}{\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)} \\
&\qquad = \alpha\left(L_t^{T_i-\delta,T_i}\right)^{\alpha-1}\gamma(t, T_i-\delta, T_i)\frac{L_t^{T_i-\delta,T_i}}{\beta\left(t, T_i-\delta, T_i, (L_t^{T_j,T_j+\delta})_{T_j\geq t}\right)} = \alpha.
\end{aligned}$$

Cox and Ross (1976) study a similar variant of the Black-Scholes-Merton model for stock options.

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.

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, i.e. (6.60) on page 136, 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 (i.e. payment dates $T_{k+1}, \dots, T_n$) as a (k,n)-payer swap. Here we must have $1 \le k < n$. Let us by $\tilde{L}_t^{T_k,\delta}$ denote the forward swap rate prevailing at time $t \le T_k$ for a (k,n)-swap. Analogous to (2.30) on page 38, we have that

$$\tilde{L}_t^{T_k,\delta} = \frac{B_t^{T_k} - B_t^{T_n}}{\delta G_t^{k,n}}, \tag{11.29}$$

where we have introduced the notation

$$G_t^{k,n} = \sum_{i=k+1}^{n} B_t^{T_i}, \tag{11.30}$$

which is 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 (2.33) on page 39 we know that the value of this swaption at the expiration date T_k is given by

$$\mathcal{P}_{T_k}^{k,n} = G_{T_k}^{k,n} H \delta \max\left(\tilde{L}_{T_k}^{T_k,\delta} - K, 0\right). \tag{11.31}$$

As discussed in Section 6.2.3 on page 118, 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

$$P_t = G_t^{k,n}\, \mathrm{E}_t^{\mathbb{Q}^{k,n}}\left[\frac{H_{T_k}}{G_{T_k}^{k,n}}\right]. \tag{11.32}$$

The pricing formula (11.32) is particularly convenient for the (k,n)-swaption. Inserting the payoff from (11.31), we obtain a price of

$$\mathcal{P}_t^{k,n} = G_t^{k,n} H \delta\, \mathrm{E}_t^{\mathbb{Q}^{k,n}}\left[\max\left(\tilde{L}_{T_k}^{T_k,\delta} - K, 0\right)\right]. \tag{11.33}$$

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 (11.29), 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 modeled forward swap rates and (ii) the sensitivity functions that show 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 modeled forward swap rates.

Under the assumption that $\beta^{k,n}$ is proportional to the level of the forward swap rate, i.e.

$$d\tilde{L}_t^{T_k,\delta} = \tilde{L}_t^{T_k,\delta}\gamma^{k,n}(t)\, dz_t^{k,n} \tag{11.34}$$

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 formula (11.33) can be computed explicitly:

Theorem 11.6 *Under the assumption* (11.34) *the price of a European* (k,n)*-payer swaption is given by*

$$\mathcal{P}_t^{k,n} = \left(\sum_{i=k+1}^{n} B_t^{T_i}\right) H\delta\left[\tilde{L}_t^{T_k,\delta} N(d_1) - KN(d_2)\right], \quad t < T_k, \tag{11.35}$$

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 formula (6.60) with σ given by $\sigma = v_{k,n}(t)/\sqrt{T_k - t}$. Hence, the lognormal swap market model provides some theoretical support of 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 modeled 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 (11.34) 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 neither for caps, floors nor swaptions with other maturity dates. The correct prices of these securities must be computed using numerical methods, e.g. 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 (11.18) 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,

$$\gamma(t, T_i - \delta, T_i) = \gamma e^{-\kappa[T_i - \delta - t]}, \quad i = 1, \dots, n,$$

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, i.e. 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 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.

11.7 Exercises

EXERCISE 11.1 (Caplets and options on zero-coupon bonds) Assume that the lognormal LIBOR market model holds. Use the caplet formula (11.19) and the relations between caplets, floorlets, and European bond options known from Chapter 2 to show that the following pricing formulas for European options on zero-coupon bonds are valid:

$$\begin{aligned} C_t^{K,T_i-\delta,T_i} &= (1-K)B_t^{T_i}N(e_{1i}) - K[B_t^{T_i-\delta} - B_t^{T_i}]N(e_{2i}), \\ \pi_t^{K,T_i-\delta,T_i} &= K[B_t^{T_i-\delta} - B_t^{T_i}]N(-e_{2i}) - (1-K)B_t^{T_i}N(-e_{1i}), \end{aligned}$$

where

$$\begin{aligned} e_{1i} &= \frac{1}{v_L(t,T_i-\delta,T_i)} \ln\left(\frac{(1-K)B_t^{T_i}}{K[B_t^{T_i-\delta} - B_t^{T_i}]} \right) + \frac{1}{2} v_L(t,T_i-\delta,T_i), \\ e_{2i} &= e_{1i} - v_L(t,T_i-\delta,T_i), \end{aligned}$$

and $v_L(t,T_i-\delta,T_i)$ is given by (11.22) in the one-factor setting and by (11.26) in the multi-factor setting. Note that these pricing formulas only apply to options expiring at one of the time points $T_0, T_1, \dots, T_{n-1}$, 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.

Chapter 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 the portfolio management.

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. We will first describe the traditional, but still widely used, duration and convexity measures and discuss their relations to the dynamics of the term structure of interest rates. Then we will consider risk measures that are more directly linked to the dynamic term structure models we have analyzed in the previous chapters. Here we focus on diffusion models and emphasize models with a single state variable. We will compare the different risk measures and their use in the construction of so-called immunization strategies. Finally, we will show how the duration measure can be useful for the pricing of European options on bonds and hence the pricing of European swaptions.

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, i.e. 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 $T_1 < \dots < T_n$. 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 computed 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

$$D_t^{\text{Mac}} = -\frac{1 + \hat{y}_t^B}{B_t}\frac{dB_t}{d\hat{y}_t^B} = \frac{\sum_{T_i > t}(T_i - t) Y_i (1 + \hat{y}_t^B)^{-(T_i - t)}}{B_t} = \sum_{T_i > t} w^{\text{Mac}}(t, T_i)(T_i - t), \tag{12.2}$$

where the weights $w^{\text{Mac}}(t, T_i)$ 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

$$\frac{dB_t}{B_t} = -D_t^{\text{Mac}} \frac{d\hat{y}_t^B}{1 + \hat{y}_t^B}.$$

For bullet bonds, annuity bonds, and serial bonds an explicit expression for the Macaulay duration can be derived.[1] In many 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

$$\Delta B_t \approx \frac{dB_t}{dy_t^B} \Delta y_t^B,$$

[1] The formula for the Macaulay duration of a bullet bond can be found in many textbooks, e.g. Fabozzi (2000) and van Horne (2001).

and hence

$$\frac{\Delta B_t}{B_t} \approx -D_t^{\text{Mac}} \Delta y_t^B.$$

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

$$K_t^{\text{Mac}} = \frac{1}{2B_t} \frac{d^2 B_t}{d(y_t^B)^2} = \frac{1}{2} \sum_{T_i > t} w^{\text{Mac}}(t, T_i)(T_i - t)^2, \tag{12.3}$$

we can write the second-order approximation as

$$\frac{\Delta B_t}{B_t} \approx -D_t^{\text{Mac}} \Delta y_t^B + K_t^{\text{Mac}} \left(\Delta y_t^B\right)^2.$$

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, e.g. 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 e.g. Ingersoll, Skelton, and Weil (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) 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

$$D_t^{\text{FW}} = \sum_{T_i > t} w(t, T_i)(T_i - t), \tag{12.4}$$

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 set of weights and hence the two duration measures will be very close, see e.g. Table 12.1 on page 263.

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 $dy_t^{T_i} = d\hat{y}_t^{T_i}/(1 + \hat{y}_t^{T_i})$.

We can also define the **Fisher-Weil convexity** as

$$K_t^{\mathrm{FW}} = \frac{1}{2} \sum_{T_i > t} w(t, T_i)(T_i - t)^2. \tag{12.5}$$

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 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, i.e.

$$y_t^T = r_t + h(T - t),$$

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 follows a process of the type

$$dr_t = \alpha(r_t, t)\, dt + \beta(r_t, t)\, dz_t$$

in the real world and hence

$$dr_t = \hat{\alpha}(r_t, t)\, dt + \beta(r_t, t)\, dz_t^{\mathbb{Q}}$$

in a hypothetical risk-neutral world.

In such a model the price of any fixed income security will be given by a function solving the fundamental partial differential equation (7.3) on page 145. In particular, the price function of any zero-coupon bond $B^T(r,t)$ satisfies

$$\frac{\partial B^T}{\partial t}(r,t) + \hat{\alpha}(r,t)\frac{\partial B^T}{\partial r}(r,t) + \frac{1}{2}\beta(r,t)^2\frac{\partial^2 B^T}{\partial r^2}(r,t) - rB^T(r,t) = 0, \quad (r,t) \in \mathcal{S} \times [0,T),$$

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

$$h'(T-t)(T-t) + h(T-t) = \hat{\alpha}(r,t)(T-t) - \frac{1}{2}\beta(r,t)^2(T-t)^2, \quad (r,t) \in \mathcal{S} \times [0,T).$$

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.

We assume that the short rate r_t is the only state variable, and that it follows a process of the form

$$dr_t = \alpha(r_t, t)\, dt + \beta(r_t, t)\, dz_t.$$

For an asset with price $B_t = B(r_t, t)$, Itô's Lemma implies that

$$dB_t = \left(\frac{\partial B}{\partial t}(r_t, t) + \alpha(r_t, t) \frac{\partial B}{\partial r}(r_t, t) + \frac{1}{2}\beta(r_t, t)^2 \frac{\partial^2 B}{\partial r^2}(r_t, t) \right) dt + \frac{\partial B}{\partial r}(r_t, t)\beta(r_t, t)\, dz_t,$$

and hence

$$\begin{aligned}\frac{dB_t}{B_t} &= \left(\frac{1}{B(r_t, t)} \frac{\partial B}{\partial t}(r_t, t) + \alpha(r_t, t) \frac{1}{B(r_t, t)} \frac{\partial B}{\partial r}(r_t, t) + \frac{1}{2}\beta(r_t, t)^2 \frac{1}{B(r_t, t)} \frac{\partial^2 B}{\partial r^2}(r_t, t) \right) dt \\ &\quad + \frac{1}{B(r_t, t)} \frac{\partial B}{\partial r}(r_t, t)\beta(r_t, t)\, dz_t.\end{aligned}$$

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

$$D(r, t) = -\frac{1}{B(r, t)} \frac{\partial B}{\partial r}(r, t). \tag{12.6}$$

Note the similarity to the definition of the Macaulay duration. The unexpected 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

$$K(r, t) = \frac{1}{2B(r, t)} \frac{\partial^2 B}{\partial r^2}(r, t) \tag{12.7}$$

and the **time value** as

$$\Theta(r, t) = \frac{1}{B(r, t)} \frac{\partial B}{\partial t}(r, t). \tag{12.8}$$

Consequently, the rate of return on the asset over the next infinitesimal period of time can be written as

$$\frac{dB_t}{B_t} = \left(\Theta(r_t, t) - \alpha(r_t, t)D(r_t, t) + \beta(r_t, t)^2 K(r_t, t) \right) dt - D(r_t, t)\beta(r_t, t)\, dz_t. \tag{12.9}$$

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

$$\begin{aligned}D_\Pi(r, t) &= -\frac{1}{\Pi(r, t)} \frac{\partial \Pi}{\partial r}(r, t) \\ &= -\frac{1}{\Pi(r, t)} \left(N_1 \frac{\partial B_1}{\partial r}(r, t) + N_2 \frac{\partial B_2}{\partial r}(r, t) \right) \\ &= \frac{N_1 B_1(r, t)}{\Pi(r, t)} \left(-\frac{1}{B_1(r, t)} \frac{\partial B_1}{\partial r}(r, t) \right) + \frac{N_2 B_2(r, t)}{\Pi(r, t)} \left(-\frac{1}{B_2(r, t)} \frac{\partial B_2}{\partial r}(r, t) \right) \\ &= \eta_1(r, t)D_1(r, t) + \eta_2(r, t)D_2(r, t),\end{aligned} \tag{12.10}$$

[2] Recall that the volatility of an asset is defined as the standard deviation of the return on the asset over the next instant.

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 (12.9) 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

$$\Theta(r,t) - \hat{\alpha}(r,t)D(r,t) + \beta(r,t)^2 K(r,t) = r, \tag{12.11}$$

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 (12.11) 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]

Note that the rate of return on the security over the next instant can be rewritten as

$$\frac{dB_t}{B_t} = \left(r_t - \lambda(r_t,t)\beta(r_t,t)D(r_t,t)\right)dt - D(r_t,t)\beta(r_t,t)\,dz_t,$$

which only involves the duration, but neither 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, cf. Theorem 6.2 on page 126 and the subsequent discussion of hedging on page 127. 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. Straightforward differentiation implies that

$$\frac{\partial D}{\partial r}(r,t) = D(r,t)^2 - 2K(r,t) \tag{12.12}$$

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 will probably 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, i.e. 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

[3] In the Black-Scholes-Merton model the time value and the so-called Δ and Γ values are related in a similar way, cf. Hull (2003, Section 14.7). Apparently, Christensen and Sørensen (1994) were the 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.

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 (12.11) 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 (12.4), 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 (12.6) 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.

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, e.g. 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, e.g. 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

$$D(r,t) = -\frac{1}{B(r,t)}\frac{\partial B}{\partial r}(r,t) = \frac{1}{B(r,t)}\sum_{T_i>t} b(T_i-t)Y_i B^{T_i}(r,t) = \sum_{T_i>t} w(r,t,T_i)b(T_i-t),$$

[4]The duration $D(r,t)$ is well-defined by (12.6) 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.

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

$$K(r,t) = \frac{1}{2}\sum_{T_i>t} w(r,t,T_i) b(T_i - t)^2.$$

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

$$D^*(r,t) = b^{-1}\left(\sum_{T_i>t} w(r,t,T_i) b(T_i-t)\right). \tag{12.13}$$

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 (12.13) the time-denominated duration of a coupon bond is

$$\begin{aligned} D^*(r,t) &= -\frac{1}{\kappa}\ln\left(1 - \kappa \sum_{T_i>t} w(r,t,T_i) b(T_i-t)\right) \\ &= -\frac{1}{\kappa}\ln\left(1 - \sum_{T_i>t} w(r,t,T_i)(1 - e^{-\kappa[T_i-t]})\right) \\ &= -\frac{1}{\kappa}\ln\left(\sum_{T_i>t} w(r,t,T_i) e^{-\kappa[T_i-t]}\right). \end{aligned}$$

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

$$D(r,t) = \sum_{T_i>t} w(r,t,T_i) \frac{2(e^{\gamma[T_i-t]}-1)}{(\gamma+\hat{\kappa})(e^{\gamma[T_i-t]}-1)+2\gamma}.$$

Since

$$\begin{aligned} b^{-1}(y) &= \frac{1}{\gamma} \ln\left(1 + \frac{2\gamma y}{2-(\hat{\kappa}+\gamma)y}\right) \\ &= \frac{1}{\gamma} \ln\left(1 + 2\gamma\left[\frac{2}{y} - (\hat{\kappa}+\gamma)\right]^{-1}\right), \end{aligned}$$

the time-denominated duration of a coupon bond is

$$D^*(r,t) = \frac{1}{\gamma} \ln\left(1 + 2\gamma\left[\frac{2}{\sum_{T_i>t} w(r,t,T_i) b(T_i-t)} - (\hat{\kappa}+\gamma)\right]^{-1}\right). \tag{12.14}$$

□

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 Merton model the volatility of a zero-coupon bond is proportional to the time-to-maturity of the bond, cf. (7.17) and (7.37). 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 (7.70) on page 170. 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

time-to-maturity in 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

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.

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 (12.14).

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 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 Chapter 6 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 (12.10), 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

$$\eta(r,t) = \frac{\bar{D}(r,t) - D_2(r,t)}{D_1(r,t) - D_2(r,t)}. \tag{12.15}$$

If (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 (12.15) holds at any point in time, then the desired cash flow will be matched with certainty, i.e. 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 (12.12), 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 (12.11) 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

$$\eta_1(r,t) = \frac{(\bar{D}(r,t) - D_3(r,t))(K_2(r,t) - K_3(r,t)) - (D_2(r,t) - D_3(r,t))(\bar{K}(r,t) - K_3(r,t))}{(D_1(r,t) - D_3(r,t))(K_2(r,t) - K_3(r,t)) - (D_2(r,t) - D_3(r,t))(K_1(r,t) - K_3(r,t))}, \tag{12.16}$$

$$\eta_2(r,t) = \frac{(D_1(r,t) - D_3(r,t))(\bar{K}(r,t) - K_3(r,t)) - (\bar{D}(r,t) - D_3(r,t))(K_1(r,t) - K_3(r,t))}{(D_1(r,t) - D_3(r,t))(K_2(r,t) - K_3(r,t)) - (D_2(r,t) - D_3(r,t))(K_1(r,t) - K_3(r,t))}. \tag{12.17}$$

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, cf. (12.11). 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. The immunization strategy based on Macaulay durations is then defined by Eq. (12.15) 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 1000 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.

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 one 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 one year at the most, while the long bond matures five 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 one year before the liability is due, the portfolio is replaced by a position in the bond with only one 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 30000 simulations of the evolution of the yield curve in the CIR model over the 10-year period to the due date of the liability. 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 30000 simulated terminal portfolio values was 1000.01, which is very close to the desired value of 1000. The average absolute deviation was 0.124% of the desired portfolio value. In 29.1% of the 30000 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%, etc.[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

[5]Even with 30000 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.

2 portfolio adjustments per year						
	Macaulay criterion			short-long criterion		
	Macaulay	Fisher-Weil	CIR	Macaulay	Fisher-Weil	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		
	Macaulay	Fisher-Weil	CIR	Macaulay	Fisher-Weil	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		
	Macaulay	Fisher-Weil	CIR	Macaulay	Fisher-Weil	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%

Table 12.2: Results from the immunization of a 10-year liability based on 30000 simulations of the CIR model. The current short rate is 5%.

a portfolio of bonds with very similar maturities (a low convexity portfolio). The high convexity strategy deviates by more than five percent 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 percent 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 one percent 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 1000, 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.

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 one year, a bond maturing two years after the liability is due, and a bond maturing ten 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 more dramatic and in all of the 30000 simulated outcomes the deviation is less than 0.05% although the portfolio is only rebalanced once a month! The numbers under the column heading Hull-White are explained below.

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

12 portfolio adjustments per year						
	Identical durations			Identical durations and convexities		
	Macaulay	CIR	Hull-White	Macaulay	CIR	Hull-White
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%

Table 12.3: Results from the immunization of a 10-year liability of 1000 dollars based on 30000 simulations of the CIR model. The current short rate is 5%.

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, i.e. the yield curve of the CIR model. Table 12.3 shows the results of such an immunization strategy. 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. Let us consider a two-factor diffusion model where the state variables x_1 and x_2 are assumed to develop as

$$dx_{1t} = \alpha_1(x_{1t}, x_{2t})\,dt + \beta_{11}(x_{1t}, x_{2t})\,dz_{1t} + \beta_{12}(x_{1t}, x_{2t})\,dz_{2t}, \tag{12.18}$$

$$dx_{2t} = \alpha_2(x_{1t}, x_{2t})\,dt + \beta_{21}(x_{1t}, x_{2t})\,dz_{1t} + \beta_{22}(x_{1t}, x_{2t})\,dz_{2t}. \tag{12.19}$$

For a security with the price $B_t = B(x_{1t}, x_{2t}, t)$, Itô's Lemma implies that

$$\begin{aligned}\frac{dB_t}{B_t} &= \dots\, dt - D_1(x_{1t}, x_{2t}, t)\left[\beta_{11}(x_{1t}, x_{2t})\, dz_{1t} + \beta_{12}(x_{1t}, x_{2t})\, dz_{2t}\right] \\ &\quad - D_2(x_{1t}, x_{2t}, t)\left[\beta_{21}(x_{1t}, x_{2t})\, dz_{1t} + \beta_{22}(x_{1t}, x_{2t})\, dz_{2t}\right],\end{aligned} \tag{12.20}$$

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:

$$\begin{aligned}&\Theta(x_1, x_2, t) - \hat{\alpha}_1(x_1, x_2) D_1(x_1, x_2, t) - \hat{\alpha}_2(x_1, x_2) D_2(x_1, x_2, t) \\ &\quad + \gamma_1(x_1, x_2)^2 K_1(x_1, x_2, t) + \gamma_2(x_1, x_2)^2 K_2(x_1, x_2, t) + \gamma_{12}(x_1, x_2) K_{12}(x_1, x_2, t) = r(x_1, x_2).\end{aligned}$$

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

$$\begin{aligned}
K_j(x_1, x_2, t) &= \sum_{T_i > t} w(x_1, x_2, t, T_i) b_j(T_i - t)^2, \quad j = 1, 2, \\
K_{12}(x_1, x_2, t) &= \sum_{T_i > t} w(x_1, x_2, t, T_i) b_1(T_i - t) b_2(T_i - t), \\
\Theta(x_1, x_2, t) &= \sum_{T_i > t} w(x_1, x_2, t, T_i) \left(a'(T_i - t) + b_1'(T_i - t)x_1 + b_2'(T_i - t)x_2\right).
\end{aligned}$$

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 (12.18) and (12.19) the variance of the rate of return can be computed from (12.20):

$$\begin{aligned}
\mathrm{Var}_t\left(\frac{dB_t}{B_t}\right) &= \mathrm{Var}_t\left([D_1\beta_{11} + D_2\beta_{21}]\, dz_{1t} + [D_1\beta_{12} + D_2\beta_{22}]\, dz_{2t}\right) \\
&= \left([D_1\beta_{11} + D_2\beta_{21}]^2 + [D_1\beta_{12} + D_2\beta_{22}]^2\right) dt \\
&= \left(D_1^2\gamma_1^2 + D_2^2\gamma_2^2 + 2D_1D_2\gamma_{12}\right) dt,
\end{aligned}$$

where we for notational simplicity 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,

$$\sigma_B(x_1, x_2, t)^2 = \sigma^{t+D^*}(x_1, x_2, t)^2.$$

This equation can only be solved numerically. For an affine two-factor model the equation is of the form

$$\left(\sum_{T_i>t} w_i b_1(T_i - t)\right)^2 \gamma_1^2 + \left(\sum_{T_i>t} w_i b_2(T_i - t)\right)^2 \gamma_2^2 + 2\left(\sum_{T_i>t} w_i b_1(T_i - t)\right)\left(\sum_{T_i>t} w_i b_2(T_i - t)\right)\gamma_{12}$$
$$= b_1(D^*)^2\gamma_1^2 + b_2(D^*)^2\gamma_2^2 + 2b_1(D^*)b_2(D^*)\gamma_{12},$$

where we again have simplified the notation, e.g. w_i represents $w(x_1, x_2, t, T_i)$. 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 8.5.2 on page 195. 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 2.9.2, 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, $C_t^{K,T,\text{cpn}}$ 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 bonds after expiration of

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

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 $v = 0.002$.

the option, i.e. $B_t = \sum_{T_i > T} Y_i B_t^{T_i}$ where Y_i is the payment at time T_i. Wei's approximation is then given by the following relation:

$$C_t^{K,T,\text{cpn}} \approx \tilde{C}_t^{K,T,\text{cpn}} = \frac{B_t}{B_t^{t+D_t^*}} C_t^{K^*,T,t+D_t^*}, \tag{12.21}$$

where $K^* = K B_t^{t+D_t^*}/B_t$, and where D_t^* denotes the time-denominated duration of the cash flow of the underlying coupon bonds 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 on page 150) 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, e.g. 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 to 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 (2002) and Singleton and Umantsev (2002) introduce two approximations that may dominate (with respect to accuracy and computational speed) the duration-based approach discussed here, 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

$$C_t^{K,T,\mathrm{cpn}} \approx \frac{B_t}{B_t^S} C_t^{K_S,T,S}, \tag{12.22}$$

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 (12.21).

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 (12.22) is therefore equal to

$$\begin{aligned} C_t^{K,T,\mathrm{cpn}} - \frac{B_t}{B_t^S} C_t^{K_S,T,S} &= B_t^S \left(\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{\max(B_T - K, 0)}{B_T^S}\right] - \frac{B_t}{B_t^S} \, \mathrm{E}_t^{\mathbb{Q}^S}\left[\max\left(1 - \frac{KB_t^S}{B_t B_T^S}, 0\right)\right]\right) \\ &= B_t^S \, \mathrm{E}_t^{\mathbb{Q}^S}\left[\max\left(\frac{B_T}{B_T^S} - \frac{K}{B_T^S}, 0\right) - \max\left(\frac{B_t}{B_t^S} - \frac{K}{B_T^S}, 0\right)\right]. \end{aligned} \tag{12.23}$$

From the definition of the S-forward martingale measure it follows also that

$$\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{B_T}{B_T^S}\right] = \frac{B_t}{B_t^S} \tag{12.24}$$

and that

$$\mathrm{E}_t^{\mathbb{Q}^S}\left[\frac{K}{B_T^S}\right] = \frac{KB_t^T}{B_t^S}. \tag{12.25}$$

For *deep-in-the-money* call options both max-terms in (12.23) will with a high probability return the first argument, and it follows then from (12.24) 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 (12.24) and (12.25) 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 $\operatorname{Var}_t^{\mathbb{Q}^S}\left[B_T/B_t - B_T^S/B_t^S\right]$ 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 $S = T^{\text{mv}}$ which is the value of S that minimizes the variance of the difference in the relative price change over the next instant, i.e. $\operatorname{Var}_t^{\mathbb{Q}^S}\left[\frac{dB_t}{B_t} - \frac{dB_t^S}{B_t^S}\right]$. This idea also gives rise to an alternative time-denominated duration measure, $D_t^{\text{mv}} = T^{\text{mv}} - t$, 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, $D_t^* = D_t^{\text{mv}}$. 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 8.5.2. 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 (8.43), 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 (12.21) will be very valuable if it is sufficiently accurate.

two-month options					six-month options				
K	appr. price	abs. dev.	rel. dev.	std. dev.	K	appr. price	abs. dev.	rel. dev.	std. dev.
86	5.08407	$0.1 \cdot 10^{-5}$	0.000%	$1.8 \cdot 10^{-4}$	91	4.45364	$2.0 \cdot 10^{-5}$	0.000%	$4.8 \cdot 10^{-4}$
87	4.10368	$0.2 \cdot 10^{-5}$	0.000%	$1.7 \cdot 10^{-4}$	92	3.53250	$4.9 \cdot 10^{-5}$	0.001%	$4.0 \cdot 10^{-4}$
88	3.12553	$0.7 \cdot 10^{-5}$	0.000%	$1.5 \cdot 10^{-4}$	93	2.63434	$8.2 \cdot 10^{-5}$	0.003%	$3.4 \cdot 10^{-4}$
89	2.16242	$2.1 \cdot 10^{-5}$	0.001%	$1.2 \cdot 10^{-4}$	94	1.79287	$9.6 \cdot 10^{-5}$	0.005%	$3.2 \cdot 10^{-4}$
90	1.26608	$3.2 \cdot 10^{-5}$	0.003%	$1.0 \cdot 10^{-4}$	95	1.06678	$5.5 \cdot 10^{-5}$	0.005%	$3.1 \cdot 10^{-4}$
91	0.56030	$0.7 \cdot 10^{-5}$	0.001%	$0.9 \cdot 10^{-4}$	96	0.52036	$-3.3 \cdot 10^{-5}$	-0.006%	$2.4 \cdot 10^{-4}$
92	0.15992	$-2.7 \cdot 10^{-5}$	-0.017%	$0.7 \cdot 10^{-4}$	97	0.19074	$-9.6 \cdot 10^{-5}$	-0.050%	$2.0 \cdot 10^{-4}$
93	0.02442	$-2.0 \cdot 10^{-5}$	-0.083%	$0.7 \cdot 10^{-4}$	98	0.04576	$-7.7 \cdot 10^{-5}$	-0.168%	$2.0 \cdot 10^{-4}$
94	0.00163	$-0.4 \cdot 10^{-5}$	-0.253%	$0.4 \cdot 10^{-4}$	99	0.00576	$-2.7 \cdot 10^{-5}$	-0.474%	$1.5 \cdot 10^{-4}$
95	0.00001	$-0.0 \cdot 10^{-5}$	-1.545%	$0.1 \cdot 10^{-4}$	100	0.00021	$-0.2 \cdot 10^{-5}$	-1.051%	$0.5 \cdot 10^{-4}$

Table 12.5: Prices of two- and six-month European call options on a two-year bullet 8% bond in the Longstaff-Schwartz model. The underlying bond has a current price of 89.3400, a two-month forward price of 91.2042, a six-month forward price of 95.7687, and a time-denominated stochastic duration of 1.9086 years.

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.[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 two or six months written on an 8% bullet bond with a single annual payment date and a time-to-maturity of two or ten years. The parameters in the dynamics of the state variables, see (8.40) and (8.41), 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 two-year bond for various exercise prices around the forward-at-the-money value of K, i.e. B_t/B_t^T. The corresponding results for options on the ten-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.

All the approximate prices are correct to three decimals, and the percentage deviations are also

[6] The results shown are based on simulations of 10000 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.

two-month options					six-month options				
K	appr. price	abs. dev.	rel. dev.	std. dev.	K	appr. price	abs. dev.	rel. dev.	std. dev.
74	4.42874	$1.2{\cdot}10^{-4}$	0.003%	$1.9{\cdot}10^{-3}$	78	4.27344	$1.1{\cdot}10^{-3}$	0.027%	$4.4{\cdot}10^{-3}$
75	3.46569	$2.5{\cdot}10^{-4}$	0.007%	$1.6{\cdot}10^{-3}$	79	3.42836	$1.3{\cdot}10^{-3}$	0.037%	$4.3{\cdot}10^{-3}$
76	2.53643	$3.9{\cdot}10^{-4}$	0.015%	$1.4{\cdot}10^{-3}$	80	2.64289	$1.2{\cdot}10^{-3}$	0.045%	$4.3{\cdot}10^{-3}$
77	1.69005	$4.0{\cdot}10^{-4}$	0.024%	$1.4{\cdot}10^{-3}$	81	1.93654	$0.8{\cdot}10^{-3}$	0.042%	$4.2{\cdot}10^{-3}$
78	0.98799	$1.8{\cdot}10^{-4}$	0.018%	$1.3{\cdot}10^{-3}$	82	1.33393	$0.2{\cdot}10^{-3}$	0.015%	$3.8{\cdot}10^{-3}$
79	0.48542	$-1.7{\cdot}10^{-4}$	-0.036%	$1.0{\cdot}10^{-3}$	83	0.85064	$-0.5{\cdot}10^{-3}$	-0.063%	$3.1{\cdot}10^{-3}$
80	0.19080	$-3.9{\cdot}10^{-4}$	-0.202%	$0.9{\cdot}10^{-3}$	84	0.49430	$-1.1{\cdot}10^{-3}$	-0.220%	$2.8{\cdot}10^{-3}$
81	0.05666	$-3.2{\cdot}10^{-4}$	-0.570%	$0.9{\cdot}10^{-3}$	85	0.25641	$-1.3{\cdot}10^{-3}$	-0.508%	$2.6{\cdot}10^{-3}$
82	0.01267	$-1.6{\cdot}10^{-4}$	-1.263%	$0.8{\cdot}10^{-3}$	86	0.11491	$-1.2{\cdot}10^{-3}$	-1.001%	$2.7{\cdot}10^{-3}$
83	0.00185	$-0.5{\cdot}10^{-4}$	-2.424%	$0.5{\cdot}10^{-3}$	87	0.04372	$-0.8{\cdot}10^{-3}$	-1.786%	$2.6{\cdot}10^{-3}$

Table 12.6: Prices on two- and six-month European call options on a ten-year bullet 8% bond in the Longstaff-Schwartz model. The underlying bond has a current price of 76.9324, a two-month forward price of 78.5377, a six-month forward price of 82.4682, and a time-denominated stochastic duration of 4.8630 years.

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 two-month options on the two-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 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,

$$df_t^T = \alpha(t, T, (f_t^s)_{s \geq t})\, dt + \beta(t, T, (f_t^s)_{s \geq t})\, dz_t,$$

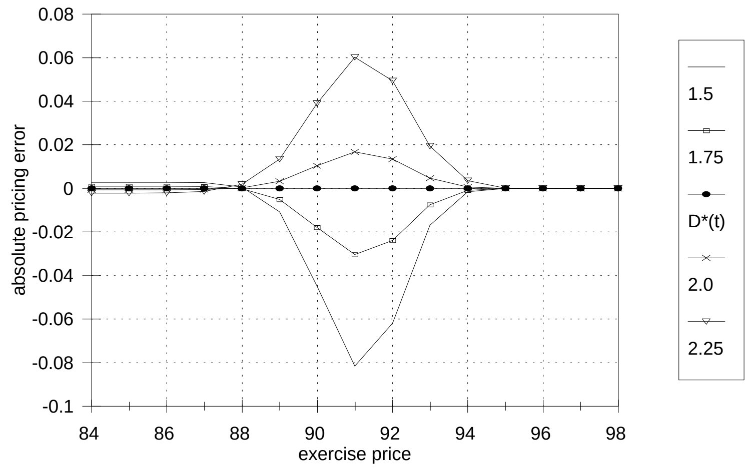


Figure 12.1: The absolute price errors for two-month options on two-year bullet bonds in the Longstaff-Schwartz-model for different maturities of the zero-coupon bond underlying the approximating option. The time-denominated stochastic duration of the underlying coupon bond is $D_t^* = 1.9086$ years, and the two-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, cf. Theorem 10.1 on page 223.[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 is to select a number of key interest rates, i.e. zero-coupon yields for certain representative maturities, e.g. 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 two-year zero-coupon yield is assumed to affect all zero-coupon yields of maturities between one year and five 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 two-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 two years and the preceding key rate maturity of one year. Similarly, the change in the two-year rate is assumed to cause a change of approximately

[7] If β is positive, the volatility of the zero-coupon bond is strictly speaking $-\sigma^T$.

0.0033 (33 basis points) in the four-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 that 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, e.g. an arbitrage-free dynamic model using these key rates as state variables, cf. the short discussion in Section 8.6.4 on page 203. 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 e.g. 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, cf. the discussion in Section 8.1. However, also these duration measures should 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.

Chapter 13

Mortgage-backed securities

13.1 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 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 issue 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 deviate from government bonds in several respects. Most importantly, the cash flow to the bond owners depend 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 behavior 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 13.2 gives an overview of some typical mortgages. Section 13.3 describes the standard class of mortgage-backed bonds, the so-called *pass-throughs*. Section 13.4 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 modeling of prepayment behavior and the effects on the valuation of mortgage-backed bonds. The option-based approach discussed in Section 13.5 focuses on determining the rational prepayment behavior 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 rationally prepay for other reasons. Hence, various modifications of the basic option-based approach are also considered. The empirical approach outlined in Section 13.6 is based on historical records of actual prepayment behavior 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 13.7. Section 13.8 offers a short introduction to other mortgage-backed securities than the standard pass-throughs.

13.2 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 lender according to a specified payment schedule. If the borrower fails to meet any of these payments, the lender has the right to take over 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. Legislation may impose restrictions on the mortgages that can be offered, e.g. require that the original face value of the mortgage can be at most 80% of the market value of the property.

Mortgage loans are typically long-term (e.g. 30-year) loans with a prespecified schedule of regular (e.g. monthly or quarterly) interest rate payments and repayments of the principal. Different types of mortgages are offered. Below we describe the most popular types. First, let us introduce some common 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 payments 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 nominal rate 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 repayment of the original face value of the loan is typically split over several dates into partial repayments. 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 $D(t_{n+1}) = D(t_n) - P(t_n)$, 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, e.g. the actual collection of payments from borrowers, preparing information on the fiscal implications of the mortgage, etc. 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.

13.2.1 Level-payment fixed-rate mortgages

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 date 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

$$D(t_n) = A\frac{1-(1+R)^{-(N-n)}}{R}. \tag{13.1}$$

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.

Note that the above equations give the scheduled cash flow and outstanding debt over the life of the mortgage, but as already mentioned the actual evolution of cash flow and outstanding debt can be different due to unscheduled prepayments.

13.2.2 Adjustable-rate mortgages

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, i.e. 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.

13.2.3 Other mortgage types

"Balloon mortgage": the contract rate is renegotiated at specific dates.

"Interest only mortgage", "endowment mortgage": the borrower pays only interest on the loan, at least for some initial period.

For more details, see Fabozzi (2000, Chap. 10).

13.2.4 Points

Above we have described various types of mortgages that borrowers may choose among. The borrowers may also choose between different maturities for a given type of a loan, e.g. 20 years or 30 years. Of course, the choice of maturity will typically affect the mortgage rate offered. In the U.S., the lending institutions offer additional flexibility. For a given loan type of a given maturity, the borrower may choose between different loans characterized by the contract rate and the so-called points. A mortgage with 0.5 points mean that the borrower has to pay 0.5% 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 which corresponds to 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 similar analysis.

13.3 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, but they are not necessarily completely identical.

Mortgage-backed bonds is 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 bond holders, 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 bond holders. Moreover, the costs of issuance of the bonds etc. 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 bond holders 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 bond holders with a delay of some weeks.

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 the United States, Germany, Denmark, and Sweden. The U.S. market initiated in the 1970s is by now far the largest of these markets. The mid-2002 total notional amount of U.S. mortgage-backed securities was more than 3.9 trillion U.S.-dollars, even higher than the 3.5 trillion U.S.-dollars notional amount of publicly traded U.S. government bonds (Longstaff 2002). 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 larger since in those countries a larger fraction of the mortgages are funded by the issuance of mortgage-backed bonds.

In the U.S., most mortgage-backed bonds are issued by three agencies: 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 issuing agency guarantees the payments to the bond holders even if borrowers default.[1] 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"). Ginnie Mae pass-throughs are even guaranteed by the U.S. government, but the bonds issued by the two other institutions are also considered virtually free of default risk. If a borrower defaults, the mortgage is prepaid by the agency. Some commercial banks and other financial institutions also issue mortgage-backed bonds. The credit quality of these bond issues are rated by the institutions that rate other bond issues such as corporate bonds, e.g. Standard & Poors and Moody's.

In Denmark, the institutions issuing the mortgage-backed bonds guarantee the payments to bond owners so the relevant default risk is that of the issuing institution, which currently seems to be negligible.

In the U.S., the pass-through bonds are issued at par. 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

[1] There are two types of guarantees. The owners of a *fully modified* pass-through are guaranteed a timely payment of both interest and principal. The owners of a *modified* pass-through are guaranteed a timely payment of interest, whereas the payment of principal takes place as it is collected from the borrowers, although with a maximum delay relative to schedule.

issued. The purpose of this practice is to form relatively large and liquid bond series in stead of many smaller bond series.

13.4 The prepayment option

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. In particular, the borrower may terminate the mortgage by repaying the total outstanding debt. In addition, 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, e.g. application fees, origination fees, credit evaluation charges, etc. 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.

In order to value a mortgage, we have to model the prepayment probability throughout the term of the mortgage. If 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 specifies 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

$$\Pi_{t_n} = 1 - e^{-\int_{t_{n-1}}^{t_n} \pi_t \, dt} \approx \int_{t_{n-1}}^{t_n} \pi_t \, dt \approx (t_n - t_{n-1}) \pi_{t_n} = \delta \pi_{t_n}. \tag{13.2}$$

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 behavior 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, e.g. 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 of the mortgage. This corresponds to the cash flow of a hypothetical non-callable bond – 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 in the money. In the present case means that 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 13.5.

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 (13.1).

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. Other borrowers that consider a prepayment may take an upward-sloping yield curve as a predictor of declining interest rates, which will make a prepayment more profitable in the future. Hence they will postpone the prepayment.

House sales. In the U.S., mortgages must be prepaid whenever the underlying property is being sold. In Denmark, 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. 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 be increasing in 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, e.g. 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. As emphasized by Longstaff (2002), 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. At least in Denmark, borrowers are primarily advised by the lending institutions. 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.

13.5 Rational prepayment models

13.5.1 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 modeling the prepayment behavior 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. 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, i.e. 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, i.e. 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 6.5 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)

$$\frac{\partial M}{\partial t}(r,t) + \hat{\alpha}(r)\frac{\partial M}{\partial r}(r,t) + \frac{1}{2}\beta(r)^2\frac{\partial^2 M}{\partial r^2}(r,t) - rM(r,t) = 0. \tag{13.3}$$

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,

$$M(r, t_n-) = M(r, t_n) + Y(t_n). \tag{13.4}$$

The standard approach to solving a PDE like (13.3) numerically is the finite difference approach. This is based on a discretization of time and state. For example, the valuation and possible exercise is only considered at time points $t \in \bar{T} \equiv \{0, \Delta t, 2\Delta t, \dots, \bar{N}\Delta t\}$, where $\bar{N}\Delta t = t_N$. The value space of the short rate is approximated by the finite space $r \in \bar{S} \equiv \{r_{\min}, r_{\min}+\Delta r, r_{\min}+2\Delta r, \dots, r_{\max}\}$. Hence we restrict ourselves to combinations of time points and short rates in the grid $\bar{S} \times \bar{T}$. For the mortgage considered here, it is helpful to have $t_n \in \bar{T}$ 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 that

$$M(r, t_N) = 0, \quad r \in \bar{S},$$

and therefore

$$M(r, t_N-) = Y(t_N), \quad r \in \bar{S}.$$

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}, i.e. we know $M(r, t_{n+1}-)$ for all $r \in \bar{S}$. 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 $r \in \bar{S}$, it is optimal to prepay at time t_n, if that leads to a lower mortgage value, i.e.

$$M^c(r, t_n) > D(t_n) + X(t_n).$$

The corresponding conditional prepayment probability $\Pi_{t_n} \equiv \Pi(r_{t_n}, t_n)$ is

$$\Pi(r, t_n) = \begin{cases} 1 & \text{if } M^c(r, t_n) > D(t_n) + X(t_n), \\ 0 & \text{if } M^c(r, t_n) \leq D(t_n) + X(t_n). \end{cases} \tag{13.5}$$

The mortgage value at time t_n is

$$\begin{aligned} M(r, t_n) &= \min\{M^c(r, t_n), D(t_n) + X(t_n)\} \\ &= (1 - \Pi(r, t_n))M^c(r, t_n) + \Pi(r, t_n)(D(t_n) + X(t_n)), \quad r \in \bar{S}. \end{aligned} \tag{13.6}$$

The value just before time t_n is

$$M(r, t_n-) = M(r, t_n) + Y(t_n), \quad r \in \bar{S}.$$

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 bond holder does not receive the prepayment costs. We assume that the entire scheduled payments are passed through to the bond holders, 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

$$B(r, t_N-) = Y(t_N), \quad r \in \bar{S}.$$

At any previous scheduled payment date t_n, we first compute the continuation values of the bond, i.e. $B^c(r, t_n)$, $r \in \bar{S}$, by the finite difference approximation. Then the bond value excluding the payment at t_n is

$$B(r,t_n) = (1 - \Pi(r,t_n))B^c(r,t_n) + \Pi(r,t_n)D(t_n) = \begin{cases} B^c(r,t_n) & \text{if } M^c(r,t_n) \leq D(t_n) + X(t_n), \\ D(t_n) & \text{if } M^c(r,t_n) > D(t_n) + X(t_n), \end{cases} \tag{13.7}$$

for any $r \in \bar{S}$. Then the scheduled payment can be added:

$$B(r, t_n-) = B(r, t_n) + Y(t_n), \quad r \in \bar{S}.$$

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 rates, see Hull (2003, Ch. 23).

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. 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 "inoptimal" 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, but 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 13.4. In the following subsections we discuss how these features can be incorporated into the option-based approach.

13.5.2 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 behavior. 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 behavior. 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 inoptimal 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:

$$B(r,t) = \sum_{i=1}^{M} w_{mt} B_m(r,t). \tag{13.8}$$

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 behavior. 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.

13.5.3 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 13.4, 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 an inoptimal 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

$$\Pi^e_{t_n} \equiv 1 - e^{-\int_{t_{n-1}}^{t_n} \lambda_t \, dt} \approx \int_{t_{n-1}}^{t_n} \lambda_t \, dt \approx (t_n - t_{n-1})\lambda_{t_n} = \delta \lambda_{t_n} \tag{13.9}$$

that the borrower will prepay the mortgage in the interval $(t_{n-1}, t_n]$ for "exogenous" reasons, i.e. 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, i.e. $\lambda_t = \lambda(r_t, t)$. 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 inoptimality. He assumes that each borrower 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

$$\begin{aligned}
\text{Prob(no prepayment)} &= \text{Prob}\left((\text{no optimal prepayment}) \text{ AND } (\text{no inoptimal prepayment})\right) \\
&= \text{Prob}\left((\text{no optimal prepayment})\right) \times \text{Prob}\left((\text{no inoptimal prepayment})\right) \\
&= e^{-\eta_t \Delta t} e^{-\lambda_t \Delta t} \\
&= e^{-(\eta_t + \lambda_t)\Delta t}.
\end{aligned}$$

Hence, the probability that a rational prepayment takes place in $(t_{n-1}, t_n]$ is

$$\Pi^r_{t_n} = 1 - e^{-\int_{t_{n-1}}^{t_n} (\eta_t + \lambda_t)\, dt} \approx \int_{t_{n-1}}^{t_n} (\eta_t + \lambda_t)\, dt \approx \delta(\eta_{t_n} + \lambda_{t_n}). \tag{13.10}$$

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 (13.9) and (13.10) – 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 inoptimal and Π^r when prepayment is optimal, i.e.

$$\Pi(r, t_n) = \begin{cases} \Pi^e(r, t_n) & \text{if } M^c(r, t_n) < D(t_n) + X(t_n), \\ \Pi^r(r, t_n) & \text{if } M^c(r, t_n) \geq D(t_n) + X(t_n). \end{cases} \tag{13.11}$$

In the backwards valuation iterative procedure, we replace Equation (13.6) by

$$\begin{aligned}
M(r, t_n) &= (1 - \Pi(r, t_n))M^c(r, t_n) + \Pi(r, t_n)(D(t_n) + X(t_n)) \\
&= \begin{cases} (1 - \Pi^e(r, t_n))M^c(r, t_n) + \Pi^e(r, t_n)(D(t_n) + X(t_n)), & \text{if } M^c(r, t_n) < D(t_n) + X(t_n), \\ (1 - \Pi^r(r, t_n))M^c(r, t_n) + \Pi^r(r, t_n)(D(t_n) + X(t_n)), & \text{if } M^c(r, t_n) \geq D(t_n) + X(t_n), \end{cases}
\end{aligned} \tag{13.12}$$

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 than in the pure option-based approach since the continuation value takes into account the possibility for making inoptimal prepayments or missing optimal prepayments in the future. Similarly, Equation (13.7) in the valuation of the bond has to be replaced by

$$\begin{aligned}
B(r, t_n) &= (1 - \Pi(r, t_n))B^c(r, t_n) + \Pi(r, t_n)D(t_n) \\
&= \begin{cases} (1 - \Pi^e(r, t_n))B^c(r, t_n) + \Pi^e(r, t_n)D(t_n) & \text{if } M^c(r, t_n) < D(t_n) + X(t_n), \\ (1 - \Pi^r(r, t_n))B^c(r, t_n) + \Pi^r(r, t_n)D(t_n) & \text{if } M^c(r, t_n) \geq D(t_n) + X(t_n). \end{cases}
\end{aligned} \tag{13.13}$$

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, inoptimal 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 consumption, etc.). The estimate for the hazard rate η implies that the average time between two successive checks for optimal prepayment (which is $1/\eta$) is eight 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 compare 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.

13.5.4 The option to default

The borrower also has an option to default on the mortgage. It may be optimal for the borrower to default if the market value of the mortgage exceeds the market value of the house. 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, Keenan, Muller, and Epperson (1992) and Deng, Quigley, and Van Order (2000) find that it is important to consider the prepayment option and the default option simultaneously. According to Deng, Quigley, and Van Order (2000), the inclusion of the default option is helpful in explaining empirical behavior.

13.5.5 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, e.g. 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, e.g., Stanton and Wallace (1998) and Longstaff (2002).

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, etc. 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, etc.) and durable goods (house, car, etc.). 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, e.g. 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, etc. Such models could also address the choice of mortgage, e.g. who should prefer fixed-rate mortgages and who should prefer adjustable-rate mortgages. However, such models are bound to be quite complex. The model of Campbell and Cocco (2003) is a good starting point.

13.6 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 behavior. The conditional prepayment rate for a mortgage pool is assumed to be a function of some explanatory state variables that have to be specified, i.e. $\Pi_t = \Pi(t, \boldsymbol{v}_t)$ where $\boldsymbol{v}$ is a vector of explanatory variables. Section 13.4 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, i.e. by constructing a tree or finite-difference based lattice or by performing Monte Carlo simulations. 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 $t \in \bar{T} \equiv \{0, \Delta t, 2\Delta t, \dots, \bar{N}\Delta t\}$, 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

$$\pi(t, \boldsymbol{v}_t) = \pi_0(t) e^{\boldsymbol{\theta}^\top \boldsymbol{v}_t}, \tag{13.14}$$

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 empirically 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 (13.2). 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 behavior *after* the estimation period.

Many recent empirical prepayment models are based on a periodic conditional prepayment rate of the form

$$\Pi(t, \boldsymbol{v}_t) = N(f(\boldsymbol{v}_t; \boldsymbol{\theta})), \tag{13.15}$$

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 to specify 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,

$$\Pi(t, \boldsymbol{v}_t) = \sum_{m=1}^{M} w_{mt} \Pi_m(t, \boldsymbol{v}_t), \tag{13.16}$$

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, e.g. 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 behavior in the future might be very different from the prepayment behavior 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.

For studies on the Danish mortgage-backed bond market, see Jakobsen (1992)...

13.7 Risk measures for mortgage-backed bonds

To be added in a later version...

13.8 Other mortgage-backed securities

Collateralized Mortgage Obligations (CMO's): McConnell and Singh (1994), Childs, Ott, and Riddiough (1996)

13.9 Concluding remarks

Difficult problem to solve!

Any model must be based on assumptions about interest rate dynamics and prepayment variables over a 30-year horizon.

Due to the difficulties in predicting prepayment behavior, bond investors may incorporate a "safety margin" in their valuation of mortgage-backed bonds. This will lead to lower bond prices and higher borrowing rates.

Other references: Boudoukh, Richardson, Stanton, and Whitelaw (1995), Boudoukh, Whitelaw, Richardson, and Stanton (1997)

Chapter 14

Credit risky securities

Bonds issued by corporations are credit risky. Two types of models are used for pricing corporate bonds.

Structural models that are based on assumptions about the specific issuing firm, e.g. an uncertain flow of earnings and a given or optimally derived capital structure: Merton (1974), Black and Cox (1976), Shimko, Tejima, and van Deventer (1993), Leland (1994), Longstaff and Schwartz (1995), Goldstein, Ju, and Leland (2001), Christensen, Flor, Lando, and Miltersen (2002).

Reduced form models: Jarrow, Lando, and Turnbull (1997), Lando (1998), Duffie and Singleton (1999).

Chapter 15

Stochastic interest rates and the pricing of stock and currency derivatives

15.1 Introduction

In the preceding chapters we have 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, e.g. 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 and the correlations between interest rates and the random variables that determine the payments of the security.

We will in this chapter first consider the pricing of stock options when we allow for the uncertain evolution of interest rates, in contrast to the classical 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 which practitioners often implement the Black-Scholes-Merton formula. In the Gaussian interest rate setting we will also derive similar pricing formulas for European options on forwards and futures on stocks. Subsequently, we consider 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 Chapter 6 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, i.e. 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, cf. 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

$$dS_t = S_t \left[r_t \, dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} \right)^\top \, d\boldsymbol{z}_t^{\mathbb{Q}} \right], \tag{15.2}$$

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, e.g. 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

$$\operatorname{Var}_t^{\mathbb{Q}}\left(dS_t/S_t\right) = \operatorname{Var}_t^{\mathbb{Q}}\left(\sum_{j=1}^n \sigma_{jt}^{\text{st}}\, dz_{jt}^{\mathbb{Q}}\right) = \sum_{j=1}^n \left(\sigma_{jt}^{\text{st}}\right)^2\, dt$$

so that the volatility of the stock is equal to the length of the vector $\boldsymbol{\sigma}_t^{\text{st}}$, i.e. $\|\boldsymbol{\sigma}_t^{\text{st}}\| = \sqrt{\sum_{j=1}^n \left(\sigma_{jt}^{\text{st}}\right)^2}$. 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 $\left(\boldsymbol{\sigma}_t^{\text{st}}\right)^\top \boldsymbol{\sigma}_t^T = \sum_{j=1}^n \sigma_{jt}^{\text{st}}\sigma_{jt}^T$. Consequently, the instantaneous correlation is $\left(\boldsymbol{\sigma}_t^{\text{st}}\right)^\top \boldsymbol{\sigma}_t^T / \left[\|\boldsymbol{\sigma}_t^{\text{st}}\| \cdot \|\boldsymbol{\sigma}_t^T\|\right]$.

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

$$dS_t = S_t\left[r_t\, dt + v_t^{\text{st}}\, dz_{1t}^{\mathbb{Q}}\right], \tag{15.4}$$

$$dB_t^T = B_t^T\left[r_t\, dt + \rho v_t^T\, dz_{1t}^{\mathbb{Q}} + \sqrt{1-\rho^2} v_t^T\, dz_{2t}^{\mathbb{Q}}\right], \tag{15.5}$$

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, cf. Theorem 3.6 on page 65. 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 martingale 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:

$$\begin{aligned} dF_t^T &= \frac{\partial g}{\partial S}(S_t, B_t^T)\, S_t \left(\boldsymbol{\sigma}_t^{\text{st}}\right)^\top\, d\boldsymbol{z}_t^T + \frac{\partial g}{\partial B}(S_t, B_t^T)\, B_t^T \left(\boldsymbol{\sigma}_t^T\right)^\top\, d\boldsymbol{z}_t^T \\ &= F_t^T \left(\boldsymbol{\sigma}_t^{\text{st}} - \boldsymbol{\sigma}_t^T\right)^\top\, d\boldsymbol{z}_t^T. \end{aligned}$$

A standard calculation yields

$$d(\ln F_t^T) = -\frac{1}{2}\|\boldsymbol{\sigma}_t^{\text{st}} - \boldsymbol{\sigma}_t^T\|^2\, dt + \left(\boldsymbol{\sigma}_t^{\text{st}} - \boldsymbol{\sigma}_t^T\right)^\top\, d\boldsymbol{z}_t^T,$$

and hence

$$\ln S_T = \ln F_T^T = \ln F_t^T - \frac{1}{2}\int_t^T \|\boldsymbol{\sigma}_u^{\text{st}} - \boldsymbol{\sigma}_u^T\|^2\, du + \int_t^T \left(\boldsymbol{\sigma}_u^{\text{st}} - \boldsymbol{\sigma}_u^T\right)^\top\, d\boldsymbol{z}_u^T. \tag{15.6}$$

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 (15.6) 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 (15.6) and Theorem 3.2 on page 53 that $\ln S_T = \ln F_T^T$ is normally distributed, i.e. $S_T = F_T^T$ 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

$$\begin{aligned} d_1 &= \frac{1}{v_F(t,T)} \ln \left(\frac{\mathrm{E}_t^{\mathbb{Q}^T}[F_T^T]}{K} \right) + \frac{1}{2} v_F(t,T), \\ d_2 &= d_1 - v_F(t,T), \\ v_F(t,T)^2 &\equiv \mathrm{Var}_t^{\mathbb{Q}^T}[\ln F_T^T]. \end{aligned}$$

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

$$\begin{aligned} v_F(t,T)^2 = \mathrm{Var}_t^{\mathbb{Q}^T}[\ln F_T^T] &= \mathrm{Var}_t^{\mathbb{Q}^T} \left[\int_t^T \left(\boldsymbol{\sigma}^{\text{st}}(u) - \boldsymbol{\sigma}^T(u) \right)^\top d\boldsymbol{z}_u^T \right] \\ &= \mathrm{Var}_t^{\mathbb{Q}^T} \left[\int_t^T \sum_{j=1}^n \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u) \right) dz_{ju}^T \right] \\ &= \sum_{j=1}^n \mathrm{Var}_t^{\mathbb{Q}^T} \left[\int_t^T \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u) \right) dz_{ju}^T \right] \\ &= \sum_{j=1}^n \int_t^T \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u) \right)^2 du, \\ &= \int_t^T \sum_{j=1}^n \left(\sigma_j^{\text{st}}(u) - \sigma_j^T(u) \right)^2 du, \\ &= \int_t^T \left\| \boldsymbol{\sigma}^{\text{st}}(u) - \boldsymbol{\sigma}^T(u) \right\|^2 du \\ &= \int_t^T \|\boldsymbol{\sigma}^{\text{st}}(u)\|^2 du + \int_t^T \|\boldsymbol{\sigma}^T(u)\|^2 du - 2 \int_t^T \boldsymbol{\sigma}^{\text{st}}(u)^\top \boldsymbol{\sigma}^T(u)\, du, \end{aligned}$$

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 3.2. 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

$$C_t = S_t N(d_1) - K B_t^T N(d_2), \tag{15.7}$$

where

$$\begin{aligned} d_1 &= \frac{1}{v_F(t,T)} \ln \left(\frac{S_t}{K B_t^T} \right) + \frac{1}{2} v_F(t,T), \\ d_2 &= d_1 - v_F(t,T). \end{aligned}$$

If the sensitivity vector of the stock is constant, we get

$$v_F(t,T)^2 = \|\boldsymbol{\sigma}^{\text{st}}\|^2 (T-t) + \int_t^T \|\boldsymbol{\sigma}^T(u)\|^2\, du - 2\int_t^T \left(\boldsymbol{\sigma}^{\text{st}}\right)^\top \boldsymbol{\sigma}^T(u)\, du. \tag{15.8}$$

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 (15.7) reduces to the famous Black-Scholes-Merton formula

$$C(t) = S_t N(d_1) - K e^{-r[T-t]} N(d_2), \tag{15.9}$$

where

$$\begin{aligned} d_1 &= \frac{1}{\|\boldsymbol{\sigma}^{\text{st}}\|\sqrt{T-t}} \ln\left(\frac{S_t}{K e^{-r[T-t]}}\right) + \frac{1}{2}\|\boldsymbol{\sigma}^{\text{st}}\|\sqrt{T-t}, \\ d_2 &= d_1 - \|\boldsymbol{\sigma}^{\text{st}}\|\sqrt{T-t}. \end{aligned}$$

The more general formula (15.7) was first shown by Merton (1973). It holds for all Gaussian term structure models, e.g. 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 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, e.g. (7.51) on page 159 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, also the relevant variance $v_F(t,T)^2$ in Merton's formula (15.9) differs from the Black-Scholes-Merton formula. The first of the three terms in (15.8) 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. For typical parameter values the two latter 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, e.g. based on (15.6). 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.

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, Cao, and Chen (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, e.g. 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 6.3, 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

$$dF_t^{\bar{T}} = F_t^{\bar{T}}\left[-\left(\boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top}\,d\boldsymbol{z}_t^{\mathbb{Q}}\right], \tag{15.10}$$

and, subsequently, that

$$d(\ln F_t^{\bar{T}}) = \left[-\frac{1}{2}\|\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)\right]dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)^{\top}\,d\boldsymbol{z}_t^{\mathbb{Q}}.$$

It follows that

$$\ln F_{\bar{T}}^{\bar{T}} = \ln F_t^{\bar{T}} + \int_t^{\bar{T}}\left[-\frac{1}{2}\|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)\right]du + \int_t^{\bar{T}}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\,d\boldsymbol{z}_u^{\mathbb{Q}}.$$

Equivalently,

$$S_{\bar{T}} = F_{\bar{T}}^{\bar{T}} = F_t^{\bar{T}}\exp\left\{\int_t^{\bar{T}}\left[-\frac{1}{2}\|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)\right]du + \int_t^{\bar{T}}\left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^{\top}\,d\boldsymbol{z}_u^{\mathbb{Q}}\right\}.$$

Therefore the futures price can be written as

$$\Phi_t^{\bar{T}} = F_t^{\bar{T}} \, \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left\{\int_t^{\bar{T}} \left[-\frac{1}{2}\|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 - \left(\boldsymbol{\sigma}_u^{\bar{T}}\right)^\top \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)\right] du + \int_t^{\bar{T}} \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^\top d\boldsymbol{z}_u^{\mathbb{Q}}\right\}\right].$$

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

$$\begin{aligned}
\Phi_t^{\bar{T}} &= F_t^{\bar{T}} \exp\left\{\int_t^{\bar{T}} \left[-\frac{1}{2}\|\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\|^2 - \boldsymbol{\sigma}^{\bar{T}}(u)^\top \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)\right] du\right\} \\
&\quad \times \mathrm{E}_t^{\mathbb{Q}}\left[\exp\left\{\int_t^{\bar{T}} \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)^\top d\boldsymbol{z}_u^{\mathbb{Q}}\right\}\right] \\
&= F_t^{\bar{T}} \exp\left\{-\int_t^{\bar{T}} \boldsymbol{\sigma}^{\bar{T}}(u)^\top \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du\right\},
\end{aligned} \tag{15.11}$$

where the last equality is a consequence of Theorem 3.2 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 Chapter 6 we know that the price of a European call option on a forward is

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T}\left[\max(F_T^{\bar{T}} - K, 0)\right], \tag{15.12}$$

where T is the expiry time and K is the exercise price. Let us find the dynamics in the forward price $F_t^{\bar{T}}$ under the T-forward martingale measure $\mathbb{Q}^T$. From (6.6) we can shift the probability measure from $\mathbb{Q}$ to $\mathbb{Q}^T$ by applying the relation

$$d\boldsymbol{z}_t^T = d\boldsymbol{z}_t^{\mathbb{Q}} - \boldsymbol{\sigma}_t^T \, dt. \tag{15.13}$$

Substituting this into (15.10), we can write the dynamics in $F_t^{\bar{T}}$ under $\mathbb{Q}^T$ as

$$dF_t^{\bar{T}} = F_t^{\bar{T}}\left[\left(\boldsymbol{\sigma}_t^T - \boldsymbol{\sigma}_t^{\bar{T}}\right)^\top \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right) dt + \left(\boldsymbol{\sigma}_t^{\mathrm{st}} - \boldsymbol{\sigma}_t^{\bar{T}}\right)^\top d\boldsymbol{z}_t^T\right].$$

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

$$\begin{aligned}
\ln F_T^{\bar{T}} = \ln F_t^{\bar{T}} &+ \int_t^T \left(\boldsymbol{\sigma}_u^T - \boldsymbol{\sigma}_u^{\bar{T}}\right)^\top \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right) du \\
&- \frac{1}{2}\int_t^T \|\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\|^2 \, du + \int_t^T \left(\boldsymbol{\sigma}_u^{\mathrm{st}} - \boldsymbol{\sigma}_u^{\bar{T}}\right)^\top d\boldsymbol{z}_u^T.
\end{aligned}$$

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

$$\begin{aligned}
\mu_F \equiv \mathrm{E}_t^{\mathbb{Q}^T}\left[\ln F_T^{\bar{T}}\right] = \ln F_t^{\bar{T}} &+ \int_t^T \left(\boldsymbol{\sigma}^T(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right)^\top \left(\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\right) du \\
&- \frac{1}{2}\int_t^T \|\boldsymbol{\sigma}^{\mathrm{st}}(u) - \boldsymbol{\sigma}^{\bar{T}}(u)\|^2 \, du
\end{aligned}$$

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 (15.12) 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

$$\begin{aligned} d_1 &= \frac{\mu_F - \ln K}{v_F} + v_F = \frac{\mu_F + \frac{1}{2} v_F^2 - \ln K}{v_F} + \frac{1}{2} v_F, \\ d_2 &= d_1 - v_F. \end{aligned}$$

Since

$$\mu_F + \frac{1}{2} v_F^2 = \ln F_t^{\bar{T}} + \int_t^T \left(\boldsymbol{\sigma}^T(u) - \boldsymbol{\sigma}^{\bar{T}}(u) \right)^\top \left(\boldsymbol{\sigma}^{\rm st}(u) - \boldsymbol{\sigma}^{\bar{T}}(u) \right) du,$$

we can replace $e^{\mu_F + \frac{1}{2} v_F^2}$ by $F_t^{\bar{T}} e^{\xi}$, where

$$\xi = \int_t^T \left(\boldsymbol{\sigma}^T(u) - \boldsymbol{\sigma}^{\bar{T}}(u) \right)^\top \left(\boldsymbol{\sigma}^{\rm st}(u) - \boldsymbol{\sigma}^{\bar{T}}(u) \right) du.$$

Hence, the option price can be rewritten as

$$C_t = B_t^T F_t^{\bar{T}} e^{\xi} N(d_1) - K B_t^T N(d_2), \tag{15.14}$$

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

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max(\Phi_T^{\bar{T}} - K, 0) \right]. \tag{15.15}$$

With deterministic volatilities we can apply (15.11) and insert $\Phi_T^{\bar{T}} = F_T^{\bar{T}} e^{-\psi(T,\bar{T},\bar{T})}$, where we have introduced the notation

$$\psi(T, U, \bar{T}) = \int_T^U \boldsymbol{\sigma}^{\bar{T}}(u)^\top \left(\boldsymbol{\sigma}^{\rm st}(u) - \boldsymbol{\sigma}^{\bar{T}}(u) \right) du.$$

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 $K e^{\psi(T,\bar{T},\bar{T})}$. From (15.14) it follows that the price of the futures option is

$$C_t = e^{-\psi(T,\bar{T},\bar{T})} \left[B_t^T F_t^{\bar{T}} e^{\xi} N(d_1) - K e^{\psi(T,\bar{T},\bar{T})} B_t^T N(d_2) \right],$$

which can be rewritten as

$$C_t = F_t^{\bar{T}} B_t^T e^{\xi - \psi(T,\bar{T},\bar{T})} N(d_1) - K B_t^T N(d_2), \tag{15.16}$$

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

$$C_t = \Phi_t^{\bar{T}} B_t^T e^{\xi - \psi(t,T,\bar{T})} N(d_1) - K B_t^T N(d_2) \tag{15.17}$$

with

$$d_1 = \frac{\ln(\Phi_t^{\bar{T}}/K) + \xi - \psi(t, T, \bar{T})}{v_F} + \frac{1}{2}v_F.$$

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, e.g. 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 $\breve{B}_t^T$ 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, $\breve{r}_t$ and $\breve{y}_t^T$ 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 $\breve{B}_t^T \varepsilon_t - B_t^T K$ 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, i.e.

$$F_t^T = \frac{\breve{B}_t^T}{B_t^T} \varepsilon_t. \tag{15.18}$$

This relation is consistent with the results on forward prices derived in Theorem 2.1 and in Section 6.3.1. 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 $\check{r}_t$, the expected percentage increase in the price of foreign currency, i.e. the exchange rate, must equal $r_t - \check{r}_t$. The dynamics of the spot exchange rate will therefore be of the form

$$d\varepsilon_t = \varepsilon_t \left[(r_t - \check{r}_t)\, dt + (\boldsymbol{\sigma}_t^{\varepsilon})^{\top} \, d\boldsymbol{z}_t^{\mathbb{Q}} \right], \tag{15.19}$$

where $\boldsymbol{z}^{\mathbb{Q}}$ 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, $\check{r}$, and $\boldsymbol{\sigma}^{\varepsilon}$ in (15.19) will be stochastic processes.

Define $Y_t = \check{B}_t^T \varepsilon_t$, i.e. Y_t is the price of the foreign zero-coupon bond measured in units of the domestic currency. If we let $\check{\boldsymbol{\sigma}}_t^T$ denote the sensitivity vector of the foreign zero-coupon bond, i.e.

$$d\check{B}_t^T = \check{B}_t^T \left[\dots, dt + \left(\check{\boldsymbol{\sigma}}_t^T\right)^{\top} \, d\boldsymbol{z}_t^{\mathbb{Q}} \right],$$

it follows from Itô's Lemma that the sensitivity vector for Y_t can be written as $\boldsymbol{\sigma}_t^{\varepsilon} + \check{\boldsymbol{\sigma}}_t^T$. 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

$$dY_t = Y_t \left[r_t\, dt + \left(\boldsymbol{\sigma}_t^{\varepsilon} + \check{\boldsymbol{\sigma}}_t^T\right)^{\top} \, d\boldsymbol{z}_t^{\mathbb{Q}} \right].$$

For the domestic zero-coupon bond the price dynamics is of the form

$$dB_t^T = B_t^T \left[r_t\, dt + \left(\boldsymbol{\sigma}_t^T\right)^{\top} \, d\boldsymbol{z}_t^{\mathbb{Q}} \right].$$

According to (15.18), 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

$$dF_t^T = F_t^T \left[-\left(\boldsymbol{\sigma}_t^T\right)^\top \left(\boldsymbol{\sigma}_t^\varepsilon + \check{\boldsymbol{\sigma}}_t^T - \boldsymbol{\sigma}_t^T\right) dt + \left(\boldsymbol{\sigma}_t^\varepsilon + \check{\boldsymbol{\sigma}}_t^T - \boldsymbol{\sigma}_t^T\right)^\top d\boldsymbol{z}_t^{\mathbb{Q}} \right]. \tag{15.20}$$

This is identical to the dynamics of the forward price on a stock, except that the stock price sensitivity vector $\boldsymbol{\sigma}_t^{\text{st}}$ has been replaced by the sensitivity vector for Y_t, which is $\boldsymbol{\sigma}_t^\varepsilon + \check{\boldsymbol{\sigma}}_t^T$. It follows that

$$\begin{aligned} \varepsilon_T = F_T^T = F_t^T \exp\Bigg\{ \int_t^T \left[-\frac{1}{2} \left\| \boldsymbol{\sigma}_u^\varepsilon + \check{\boldsymbol{\sigma}}_u^T - \boldsymbol{\sigma}_u^T \right\|^2 - \left(\boldsymbol{\sigma}_u^T\right)^\top \left[\boldsymbol{\sigma}_u^\varepsilon + \check{\boldsymbol{\sigma}}_u^T - \boldsymbol{\sigma}_u^T\right] \right] du \\ + \int_t^T \left[\boldsymbol{\sigma}_u^\varepsilon + \check{\boldsymbol{\sigma}}_u^T - \boldsymbol{\sigma}_u^T\right]^\top d\boldsymbol{z}_u^{\mathbb{Q}} \Bigg\}. \end{aligned} \tag{15.21}$$

Here $\boldsymbol{\sigma}^\varepsilon$, $\check{\boldsymbol{\sigma}}^T$, 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. It follows from the analysis in Chapter 6 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

$$dF_t^T = F_t^T \left(\boldsymbol{\sigma}_t^\varepsilon + \check{\boldsymbol{\sigma}}_t^T - \boldsymbol{\sigma}_t^T\right)^\top d\boldsymbol{z}_t^T. \tag{15.22}$$

This can also be seen by substituting (15.13) into (15.20).

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, e.g. 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 (6.13) on page 122 we have that

$$\Phi_t^T = \mathrm{E}_t^{\mathbb{Q}}\left[\varepsilon_T\right],$$

where we can insert (15.21). 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

$$\Phi_t^T = F_t^T \exp\left\{ -\int_t^T \boldsymbol{\sigma}^T(u)^\top \left(\boldsymbol{\sigma}^\varepsilon(u) + \check{\boldsymbol{\sigma}}^T(u) - \boldsymbol{\sigma}^T(u)\right) du \right\}.$$

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 $\boldsymbol{\sigma}^T(u) = 0$, i.e. 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, i.e. the option payoff is $\max(\varepsilon_T - K, 0)$. According to the analysis in Chapter 6, 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 (15.13) into (15.19), we obtain

$$d\varepsilon_t = \varepsilon_t \left[\left(r_t - \check{r}_t + (\boldsymbol{\sigma}_t^\varepsilon)^\top \boldsymbol{\sigma}_t^T \right) \, dt + (\boldsymbol{\sigma}_t^\varepsilon)^\top \, d\boldsymbol{z}_t^T \right].$$

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:

$$C_t = B_t^T \, \mathrm{E}_t^{\mathbb{Q}^T} \left[\max\left(F_T^T - K, 0 \right) \right]. \tag{15.23}$$

It follows from (15.22) 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

$$C_t = B_t^T F_t^T N(d_1) - K B_t^T N(d_2), \tag{15.24}$$

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

$$C_t = \varepsilon_t \check{B}_t^T N(d_1) - K B_t^T N(d_2), \tag{15.25}$$

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

$$C_t = \varepsilon_t e^{-\check{y}_t^T (T-t)} N(d_1) - K e^{-y_t^T (T-t)} N(d_2), \tag{15.26}$$

where d_1 can be written as

$$d_1 = \frac{\ln(\varepsilon_t/K) + [y_t^T - \breve{y}_t^T](T-t)}{v_F(t,T)} + \frac{1}{2}v_F(t,T).$$

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 $B_t^T = e^{-r[T-t]}$, $\breve{B}_t^T = e^{-\breve{r}[T-t]}$, and $\boldsymbol{\sigma}^T(t) = \breve{\boldsymbol{\sigma}}^T(t) = 0$. In addition, the sensitivity vector of the exchange rate, i.e. $\boldsymbol{\sigma}^\varepsilon(t)$, 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

$$C_t = \varepsilon_t e^{-\breve{r}[T-t]} N(d_1) - K e^{-r[T-t]} N(d_2), \tag{15.27}$$

where

$$d_1 = \frac{\ln(\varepsilon_t/K) + (r - \breve{r})(T-t)}{\|\boldsymbol{\sigma}^\varepsilon\|\sqrt{T-t}} + \frac{1}{2}\|\boldsymbol{\sigma}^\varepsilon\|\sqrt{T-t},$$

$$d_2 = d_1 - \|\boldsymbol{\sigma}^\varepsilon\|\sqrt{T-t}.$$

This option pricing formula is called the **Garman-Kohlhagen formula**. If we compare with Equation (15.25), 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 $\breve{r}$ in the Garman-Kohlhagen formula (15.27) must be replaced by the zero-coupon yields y_t^T and $\breve{y}_t^T$. 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 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, Lando, and Miltersen (1997) propose a model for exchange rates in a target zone with possible realignments and show how currency options may be priced numerically

[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.

within that model. See also Dumas, Jennergren, and Näslund (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, etc., commodities will typically be valuable as consumption goods or production inputs. This value is modeled in terms of a **convenience yield**, cf. Hull (2003, Chap. 3). 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 modeled. Miltersen and Schwartz obtain Black-Scholes-Merton-type pricing formulas for such options under assumptions similar to those we have applied in this chapter, e.g. 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, e.g. 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, Chap. 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, Chap. 17), who also provide pricing formulas for the case of deterministic volatilities.

Chapter 16

Numerical techniques

Numerical solution of PDE's. See Ames (1977), Johnson (1990), Wilmott, Dewynne, and Howison (1993), Christiansen (1995), Thomas (1995), and Dydensborg (1999). Schwartz (1977) and Brennan and Schwartz (1977) were the first papers applying a finite difference method to price options.

Tree models. First application in finance: Cox, Ross, and Rubinstein (1979), Rendleman and Bartter (1979).

"Born" tree-models of the term structure of interest rates: Ho and Lee (1986), Pedersen, Shiu, and Thorlacius (1989), Black, Derman, and Toy (1990).

Tree-models as approximations to one-factor continuous-time term structure models: Tian (1993), Hull and White (1990b, 1993, 1994a, 1996), Hull (2003, Ch. 23).

Tree-models as approximations to multi-factor continuous-time term structure models: Hull and White (1994b).

See He (1990) for a general approach to approximating an n-factor diffusion model by an $(n+1)$-nomial discrete-time model, i.e. a tree with $n+1$ branches from each node.

Monte Carlo simulation. For models which cannot be formulated as diffusion models with relatively few state variables neither approximating trees or numerical solutions to PDE's are usually applicable in practice. But many problems can be solved using Monte Carlo simulation.

Introduction to Monte Carlo simulation: Hull (2003, Sec. 18.6-18.7).

First application to option pricing: Boyle (1977).

Monte Carlo simulation for assets with more than one exercise opportunity, e.g. American options and Bermuda swaptions: Tilley (1993), Barraquand and Martineau (1995), Boyle, Broadie, and Glasserman (1997), Broadie and Glasserman (1997a, 1997b), Broadie, Glasserman, and Jain (1997), Carr and Yang (1997), Andersen (2000), Longstaff and Schwartz (2001).

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. □

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 $\gamma = -1$)

$$\mathrm{E}[Y] = \mathrm{E}\left[e^X\right] = \exp\left\{m + \frac{1}{2}s^2\right\}. \tag{A.1}$$

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

$$\begin{aligned} \mathrm{Var}[Y] &= \mathrm{E}\left[Y^2\right] - (\mathrm{E}[Y])^2 \\ &= e^{2(m+s^2)} - e^{2m+s^2} \\ &= e^{2m+s^2}\left(e^{s^2} - 1\right). \end{aligned} \tag{A.2}$$

The next theorem provides an expression of the truncated mean of a lognormally distributed random variable, i.e. 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. □

Bibliography

Abramowitz, M. and I. Stegun (1972). *Handbook of Mathematical Functions*. Dover Publications.

Ames, W. F. (1977). *Numerical Methods for Partial Differential Equations* (Second ed.). Academic Press.

Amin, K. I. and R. A. Jarrow (1991). Pricing Foreign Currency Options under Stochastic Interest Rates. *Journal of International Money and Finance 10*, 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*, 141–180.

Andersen, L. (2000). A Simple Approach to the Pricing of Bermudan Swaptions in the Multi-Factor Libor Market Model. *Journal of Computational Finance 3*(2), 1–32.

Andersen, L. and J. Andreasen (2000). Volatility Skews and Extensions of the Libor Market Model. *Applied Mathematical Finance 7*(1), 1–32.

Andersen, T. G. and J. Lund (1997). Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate. *Journal of Econometrics 77*, 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 *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*, 91–96.

Arrow, K. (1970). *Essays in the Theory of Risk Bearing*. North-Holland.

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.

Bakshi, G. S., C. Cao, and Z. Chen (1997). Empirical Performance of Alternative Option Pricing Models. *The Journal of Finance 52*(5), 2003–2049.

Bakshi, G. S. and Z. Chen (1996). Inflation, Asset Prices and the Term Structure of Interest Rates in Monetary Economies. *The 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*, 123–165.

Balduzzi, P., G. Bertola, and S. Foresi (1997). A Model of Target Changes and the Term Structure of Interest Rates. *Journal of Mathematical Economics 39*, 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. *The Journal of Fixed Income 6*(3), 43–53.

Ball, C. A. and W. N. Torous (1983). Bond Price Dynamics and Options. *Journal of Financial and Quantitative Analysis 18*(4), 517–531.

Barraquand, J. and D. Martineau (1995). Numerical Valuation of High Dimensional Multivariate American Securities. *Journal of Financial and Quantitative Analysis 30*, 383–405.

Beaglehole, D. R. and M. S. Tenney (1991). General Solutions of Some Interest Rate-Contingent Claim Pricing Equations. *The Journal of Fixed Income 1*(September), 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.

Bhar, R. and C. Chiarella (1997). Transformation of Heath-Jarrow-Morton Models to Markovian Systems. *The European Journal of Finance 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*, 47–72.

Björk, T. (1998). *Arbitrage Theory in Continuous Time*. 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*, 167–179.

Black, F. and J. Cox (1976). Valuing Corporate Securities: Some Effects of Bond Indenture Provisions. *The Journal of Finance 31*, 351–367.

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.

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.

Bliss, R. R. and D. C. Smith (1997, November). The Stability of Interest Rate Processes. Working paper 97-13, Federal Reserve Bank of Atlanta.

Boudoukh, J., M. Richardson, R. Stanton, and R. F. Whitelaw (1995). A New Strategy for Dynamically Hedging Mortgage-Backed Securities. *The Journal of Derivatives Summer*, 60–77.

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.

Boudoukh, J., R. F. Whitelaw, M. Richardson, and R. Stanton (1997). Pricing Mortgage-Backed Securities in a Multifactor Interest Rate Environment: A Multivariate Density Estimation Approach. *The Review of Financial Studies 10*(2), 405–446.

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*, 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*, 3–39.

Brennan, M. J. and E. S. Schwartz (1977). The Valuation of American Put Options. *The 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*, 133–155.

Brennan, M. J. and E. S. Schwartz (1980). Analyzing Convertible Bonds. *Journal of Financial and Quantitative Analysis 15*, 907–929.

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*, 85–107.

Broadie, M. and P. Glasserman (1997a). Monte Carlo Methods for Pricing High-Dimensional American Options: An Overview. *Net Exposure* (December), 15–37.

Broadie, M. and P. Glasserman (1997b). 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. *The Journal of Derivatives Fall*, 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. *The Journal of Finance 41*, 617–630.

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. *The Journal of Finance 41*, 617–630.

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? *The Journal of Fixed Income 5*(2), 31–52.

Carr, P. and G. Yang (1997, December). Simulating Bermudan Interest Rate Derivatives. Working paper, Morgan Stanley and Open Link Financial.

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. *The Journal of Fixed Income 5*(2), 89–96.

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. *The Journal of Finance 47*, 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? *The 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, N.-f., R. Roll, and S. A. Ross (1986). Economic Forces and the Stock Market. *Journal of Business 59*, 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. *The 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. *The Journal of Fixed Income 3*(3), 14–31.

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, P. O., C. R. Flor, D. Lando, and K. R. Miltersen (2002). Dynamic Capital Structure with Callable Debt and Debt Renegotiation. Working paper, University of Southern Denmark.

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. and B. G. Sørensen (1994). Duration, Convexity, and Time Value. *Journal of Portfolio Management 20*(2), 51–60.

Christiansen, E. (1995, December). *Numerisk Analyse*. Odense Universitets Trykkeri.

Clewlow, L. and C. Strickland (1994). A Note on Parameter Estimation in the Two-Factor Longstaff and Schwartz Interest Rate Model. *The Journal of Fixed Income 3*(4), 95–100.

Cochrane, J. H. (2001). *Asset Pricing*. Princeton University Press.

Collin-Dufresne, P. and R. S. Goldstein (2002). Pricing Swaptions within the Affine Framework. *Journal of Derivatives 10*(1), 1–18.

Constantinides, G. M. (1992). A Theory of the Nominal Term Structure of Interest Rates. *The Review of Financial Studies 5*(4), 531–552.

Courtadon, G. (1982). The Pricing of Options on Default-Free Bonds. *Journal of Financial and Quantitative Analysis 17*, 75–100.

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. *The 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*, 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. and S. A. Ross (1976). The Valuation of Options for Alternative Stochastic Processes. *Journal of Financial Economics 3*, 145–166.

Cox, J. C., S. A. Ross, and M. Rubinstein (1979). Option Pricing: A Simplified Approach. *Journal of Financial Economics 7*, 229–263.

Culbertson, J. M. (1957, November). The Term Structure of Interest Rates. *Quarterly Journal of Economics*, 489–504.

Dai, Q. and K. J. Singleton (2000). Specification Analysis of Affine Term Structure Models. *The Journal of Finance 55*(5), 1943–1978.

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*, 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*, 463–520.

Delbaen, F. and W. Schachermayer (1999). The Fundamental Theorem of Asset Pricing for Unbounded Stochastic Processes. *Mathematische Annalen 312*, 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*, 373–397.

Deng, Y., J. M. Quigley, and R. Van Order (2000). Mortgage Terminations, Heterogeneity and the Exercise of Mortgage Options. *Econometrica 68*(2), 275–307.

Ding, C. G. (1992). Algorithm AS 275: Computing the Non-Central χ^2 Distribution Function. *Applied Statistics 2*, 478–482.

Dothan, M. U. (1978). On the Term Structure of Interest Rates. *Journal of Financial Economics 6*, 59–69.

Dothan, M. U. (1990). *Prices in Financial Markets*.

Duffee, G. R. (1996). Idiosyncratic Variation of Treasury Bill Yields. *The Journal of Finance 51*, 527–551.

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. *The Review of Financial Studies 5*(3), 411–436.

Duffie, D. and M. Huang (1996). Swap Rates and Credit Quality. *The 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., J. Ma, and J. Yong (1995). Black's Consol Rate Conjecture. *The Annals of Applied Probability 5*(2), 356–382.

Duffie, D. and K. Singleton (1999). Modeling Term Structures of Defaultable Bonds. *The Review of Financial Studies 12*(4), 687–720.

Duffie, D. and R. Stanton (1992). Pricing Continuously Resettled Contingent Claims. *Journal of Economic Dynamics and Control 16*, 561–574.

Dumas, B., L. P. Jennergren, and B. Näslund (1995). Realignment Risk and Currency Option Pricing in Target Zones. *European Economic Review 39*, 1323–1344.

Dunn, K. B. and J. J. McConnell (1981a). A Comparison of Alternative Models for Pricing GNMA Mortage-Backed Securities. *The Journal of Finance 36*(2), 471–484.

Dunn, K. B. and J. J. McConnell (1981b). Valuation of GNMA Mortage-Backed Securities. *The 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.

Dydensborg, L. (1999). *Numerical Methods for Option Pricing*. Ph. D. thesis, University of Southern Denmark – Odense University.

Fabozzi, F. J. (2000). *Bond Markets, Analysis and Strategies* (Fourth ed.). Prentice-Hall, Inc.

Fama, E. F. (1981). Stock Returns, Real Activity, Inflation, and Money. *American Economic Review 71*, 545–565.

Fama, E. F. and M. Gibbons (1982). Inflation, Real Returns and Capital Investment. *Journal of Mathematical Economics 9*, 297–323.

Filipović, D. (1999). A Note on the Nelson-Siegel Family. *Mathematical Finance 9*(4), 349–359.

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, October). Coping with the Risk of Interest Rate Fluctuations: Returns to Bondholders from Naive and Optimal Strategies. *Journal of Business 44*, 408–431.

Fisher, M. and C. Gilles (1998). Around and Around: The Expectations Hypothesis. *The 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*, 483–485.

Gallant, A. R. (1987). *Nonlinear Statistical Methods*. Wiley.

Garman, M. B. and S. W. Kohlhagen (1983). Foreign Currency Option Values. *Journal of International Money and Finance 2*, 231–237.

Geman, H. (1989). The Importance of the Forward Neutral Probability in a Stochastic Approach of Interest Rates. Working paper, ESSEC.

Gibbons, M. R. and K. Ramaswamy (1993). A Test of the Cox, Ingersoll, and Ross Model of the Term Structure. *The Review of Financial Studies 6*, 619–658.

Goldstein, R., N. Ju, and H. Leland (2001). An EBIT-Based Model of Dynamic Capital Structure. *Journal of Business 74*(4), 483–512.

Goldstein, R. and F. Zapatero (1996). General Equilibrium with Constant Relative Risk Aversion and Vasicek Interest Rates. *Mathematical Finance 6*, 331–340.

Grabbe, J. O. (1983). The Pricing of Call and Put Options on Foreign Exchange. *Journal of International Money and Finance 2*, 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.

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*, 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*, 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*, 313–316.

He, H. (1990). Convergence from Discrete- to Continuous-Time Contingent Claims Prices. *The 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*, 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.

Hicks, J. R. (1939). *Value and Capital*. Oxford: Clarendon Press.

Ho, T. S. Y. (1992). Key Rate Durations: A Measure of Interest Rate Risks. *The 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. *The Journal of Finance 41*(5), 1011–1029.

Hogan, M. (1993). Problems in Certain Two-Factor Term Structure Models. *The 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.

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. *The 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. *The Journal of Derivatives Fall*, 7–16.

Hull, J. and A. White (1994b). Numerical Procedures for Implementing Term Structure Models II: Two-Factor Models. *The Journal of Derivatives Winter*, 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. *The Journal of Fixed Income 5*(2), 97–102.

Hull, J. and A. White (1996). Using Hull-White Interest Rate Trees. *The Journal of Derivatives Spring*, 26–36.

Hull, J. C. (2003). *Options, Futures, and Other Derivatives* (Fifth ed.). Prentice-Hall, Inc.

Ingersoll, Jr., J. E. (1987). *Theory of Financial Decision Making*. 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*, 423–440.

Jakobsen, S. (1992). *Prepayment and the Valuation of Danish Mortgage Backed Bonds*. Ph. D. thesis, Aarhus School of Business. Working Paper D92-2.

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. *The Journal of Finance 44*(1), 205–209.

Jamshidian, F. (1991, June). Forward Induction and Construction of Yield Curve Diffusion Models. *The 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*, 293–330.

Jarrow, R. A., D. Lando, and S. M. Turnbull (1997). A Markov Model for the Term Structure of Credit Risk Spreads. *The Review of Financial Studies 10*(2), 481–523.

Jaschke, S. R. (1998). Arbitrage Bounds for the Term Structure of Interest Rates. *Finance and Stochastics 2*, 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.

Johnson, C. (1990). *Numerical Solution of Partial Differential Equations by the Finite Element Method*. Cambridge University Press.

Johnston, J. (1984). *Econometric Methods* (Third ed.). McGraw-Hill.

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, New York, USA: 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.

Knez, P. J., R. Litterman, and J. Scheinkman (1994). Explorations Into Factors Explaining Money Market Returns. *The Journal of Finance 49*, 1861–1882.

Krugman, P. R. (1991). Target Zones and Exchange Rate Dynamics. *The Quarterly Journal of Economics 106*, 669–682.

Lando, D. (1998). On Cox Processes and Credit Risky Securities. *Review of Derivatives Research 2*, 99–120.

Langetieg, T. C. (1980). A Multivariate Model of the Term Structure. *The Journal of Finance 35*, 71–97.

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. *The Journal of Finance 49*(4), 1213–1252.

LeRoy, S. F. (1996). Mortgage Valuation Under Optimal Prepayment. *The Review of Financial Studies 9*(3), 817–844.

LeRoy, S. F. and J. Werner (2001). *Principles of Financial Economics*. Cambridge University Press.

Li, A., P. Ritchken, and L. Sankarasubramanian (1995). Lattice Models for Pricing American Interest Rate Claims. *The Journal of Finance 50*(2), 719–737.

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. *The Journal of Fixed Income*, 54–61.

Litzenberger, R. H. and R. Rolfo (1984). An International Study of Tax Effects on Government Bonds. *The 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*, 195–224.

Longstaff, F. A. (1993). The Valuation of Options on Coupon Bonds. *Journal of Banking and Finance 17*, 27–42.

Longstaff, F. A. (2002, December). Optimal Recursive Refinancing and the Valuation of Mortgage-Backed Securities. Working paper, The Anderson School, UCLA.

Longstaff, F. A. and E. S. Schwartz (1992a). Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model. *The 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. *The Journal of Fixed Income 2*(3), 16–23.

Longstaff, F. A. and E. S. Schwartz (1993a). Implementation of the Longstaff-Schwartz Interest Rate Model. *The 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* (July-August), 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". *The Journal of Fixed Income 3*(4), 101–102.

Longstaff, F. A. and E. S. Schwartz (1995). A Simple Approach to Valuing Risky Fixed and Floating Rate Debt. *The Journal of Finance 50*(3), 789–819.

Longstaff, F. A. and E. S. Schwartz (2001). Valuing American Options By Simulation: A Simple Least-Squares Approach. *The Review of Financial Studies 14*(1), 113–147.

Lutz, F. (1940, November). The Structure of Interest Rates. *Quarterly Journal of Economics*, 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.

Marsh, T. A. and E. R. Rosenfeld (1983). Stochastic Processes for Interest Rates and Equilibrium Bond Prices. *The Journal of Finance 38*(2), 635–646.

Marshall, D. (1992). Inflation and Asset Returns in a Monetary Economy. *The Journal of Finance 47*, 1315–1342.

McConnell, J. J. and M. Singh (1994). Rational Prepayments and the Valuation of Collateralized Mortgage Obligations. *The Journal of Finance 49*(3), 891–921.

McCulloch, J. H. (1971). Measuring the Term Structure of Interest Rates. *Journal of Business 44*, 19–31.

McCulloch, J. H. (1975). The Tax-Adjusted Yield Curve. *The Journal of Finance 30*, 811–830.

McCulloch, J. H. (1993). A Reexamination of Traditional Hypotheses about the Term Structure of Interest Rates: A Comment. *The Journal of Finance 48*, 779–789.

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*(Spring), 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. *The Journal of Finance 29*, 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*, 125–144. Reprinted as Chapter 9 in Merton (1992).

Merton, R. C. (1992). *Continuous-Time Finance*. Padstow, UK: Basil Blackwell Inc.

Miltersen, K. R. (1994). An Arbitrage Theory of the Term Structure of Interest Rates. *The 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., K. Sandmann, and D. Sondermann (1997). Closed Form Solutions for Term Structure Derivatives with Log-Normal Interest Rates. *The Journal of Finance 52*(1), 409–430.

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.

Modigliani, F. and R. Sutch (1966, May). Innovations in Interest Rate Policy. *American Economic Review*, 178–197.

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.

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*, 92–97.

Nelson, C. R. and A. F. Siegel (1987). Parsimonious Modeling of Yield Curves. *Journal of Business 60*(4), 473–489.

Ogden, J. (1987). An Analysis of Yield Curve Notes. *The Journal of Finance 42*, 99–110.

Øksendal, B. (1998). *Stochastic Differential Equations* (Fifth ed.). Springer-Verlag.

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.

Pelsser, A. (2000). *Efficient Methods for Valuing Interest Rate Derivatives*. Springer.

Phoa, W. and M. Shearer (1997). A Note on Arbitrary Yield Curve Reshaping Sensitivities Using Key Rate Durations. *The Journal of Fixed Income 7*(3), 67–71.

Piazzesi, M. (2001). An Econometric Model of the Yield Curve with Macroeconomic Jump Effects. Working paper, UCLA and NBER.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1992). *Numerical Recipes in C* (Second ed.). Cambridge University Press.

Rady, S. and K. Sandmann (1994). The Direct Approach to Debt Option Pricing. *Review of Future Markets 13*(2), 461–515.

Rebonato, R. (1996). *Interest-Rate Option Models*. John Wiley & Sons, Inc.

Rendleman, R. and B. Bartter (1979). Two State Option Pricing. *The Journal of Finance 34*, 1092–1110.

Rendleman, R. and B. Bartter (1980). The Pricing of Options on Debt Securities. *Journal of Financial and Quantitative Analysis 15*, 11–24.

Riedel, F. (1999). Heterogeneous Time Preferences and Humps in the Yield Curve – The Preferred Habitat Theory Revisited. Working paper, Humboldt University, Berlin. Forthcoming in the *European Journal of Finance*.

Riedel, F. (2000). Decreasing Yield Curves in a Model with an Unknown Constant Growth Rate. *European Finance Review 4*(1), 51–67.

Ritchken, P. and L. Sankarasubramanian (1995). Volatility Structures of Forward Rates and the Dynamics of the Term Structure. *Mathematical Finance 5*, 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*, 453–475.

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*, 199–204.

Schaefer, S. M. and E. S. Schwartz (1987). Time-Dependent Variance and the Pricing of Bond Options. *The Journal of Finance 42*(5), 1113–1128.

Schwartz, E. S. (1977). The Valuation of Warrants: Implementing a New Approach. *Journal of Financial Economics 4*, 79–93.

Schwartz, E. S. and W. N. Torous (1989). Prepayment and the Valuation of Mortgage-Backed Securities. *The 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.

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. *The Journal of Finance 40*, 319–325.

Shimko, D. C., M. Tejima, and D. R. van Deventer (1993). The Pricing of Risky Debt When Interest Rates are Stochastic. *The 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.

Stambaugh, R. F. (1988). The Information in Forward Rates. Implications for Models of the Term Structure. *Journal of Financial Economics 21*, 41–70.

Stanton, R. (1995). Rational Prepayment and the Valuation of Mortgage-Backed Securities. *The 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.

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*, 285–314.

Tilley, J. A. (1993). Valuing American options in a Path Simulation Model. *Transactions of the Society of Actuaries 45*, 83–104.

Turnbull, S. M. and L. M. Wakeman (1991). A Quick Algorithm for Pricing European Average Options. *Journal of Financial and Quantitative Analysis 26*(3), 377–389.

van Horne, J. C. (2001). *Financial Market Rates and Flows* (Sixth ed.). Prentice-Hall, Inc.

Vasicek, O. (1977). An Equilibrium Characterization of the Term Structure. *Journal of Financial Economics 5*, 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*, 169–196.

Wachter, J. A. (2002). Habit Formation and Returns on Bonds and Stocks. Working paper, Stern School of Business, NYU.

Wang, J. (1996). The Term Structure of Interest Rates in a Pure Exchange Economy with Heterogeneous Investors. *Journal of Financial Economics 41*, 75–110.

Wei, J. Z. (1997). A Simple Approach to Bond Option Pricing. *Journal of Futures Markets 17*(2), 131–160.

Willner, R. (1996). A New Tool for Portfolio Managers: Level, Slope, and Curvature Durations. *The Journal of Fixed Income 6*(1), 48–59.

Wilmott, P., J. Dewynne, and S. Howison (1993). *Option Pricing: Mathematical Models and Computation*. Oxford Financial Press.

Wu, L. (1999, May). Macroeconomic Foundations of Discontinuous Price Movements. Working paper, Graduate School of Business, Fordham University, New York.