THE SHORT INTEREST RATE IN TURKEY – A REGIME SWITCHING MODEL

by

HEIKKI AARRE OLAVI LAIHO

A dissertation submitted to the Graduate Faculty in Economics in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

2003

UMI Number: 3074660

ProQuest Information and Learning Company
300 North Zeeb Road
P.O. Box 1346
Ann Arbor, MI 48106-1346

This manuscript has been read and accepted for the Graduate Faculty in Economics in satisfaction of the dissertation requirement for the degree of Doctor of Philosophy.

November 25, 2002 [signature]

Date Chair of Examining Committee

November 25, 2002 [signature]

Date Executive Officer

Professor Salih N. Neftci

Distinguished Professor Michael Grossman

Professor Thom B. Thurston

Supervisory Committee

THE CITY UNIVERSITY OF NEW YORK

Abstract

THE SHORT INTEREST RATE IN TURKEY –

A REGIME SWITCHING MODEL

by

Heikki Aarre Olavi Laiho

Adviser: Professor Salih N. Neftci

This study examines the behavior of the overnight interbank interest rate in Turkey. First, Principal Component Analysis is conducted on the Turkish bond indices. The results indicate that a traditional one-factor model is not sufficient for describing the dynamics of the yield curve. Next, the time series of the overnight interbank interest rate and its difference are examined, and found to be non-normal, but stationary. The interest rate exhibits mean reversion and volatility clustering, and the Bai-Perron model finds structural breaks in its time series. These characteristics are consistent with a regime switching process. A Markov switching Vasicek model is used to estimate the dynamics of the short rate. Four regimes are found, representing a constant interest rate, random walk, mean reversion, and unusually high and very volatile interest rate. Volatility is inversely related to the level of the interest rate. The estimations also show that modeling volatility correctly is more important than modeling the other parameters. Overall, the regime switching model is an improvement over the traditional Vasicek model.

Acknowledgements

I am indebted to many people for their assistance and encouragement in completing this project. I want to especially thank my adviser Professor Salih N. Neftci for his helpful comments and all the inspiration and encouragement I have received from him. Professor Michael Grossman is one of the truly great people I have met in my life. Without him, I could never have completed this project. Professor Thom B. Thurston has been very helpful and supportive during all the years I have spent in CUNY. I am also very grateful for my wife Deniz, and my son Olavi Salih, as well as my parents.

I also want to use this opportunity to say special thanks to the individuals who have provided many of the used non-commercial software. Jurgen A. Doornik is the author of Oxedit – an advanced statistics engine, and Herman J. Bierens is the developer of Easyreg International, an advanced regression program. The script for the Bai-Perron model is from Jushan Bai and Pierre Perron, and the Markov switching package is from Hans-Martin Krolzig. The program for the BDS test is written by Dee Dechert.

Table of Contents

List of Tables

List of Figures

1 Introduction

Modeling interest rates is difficult. Theories provide little guidance and we know only two things definitely, interest rates have to be positive and they cannot grow unboundedly. Even then, interest rates have been negative - in Switzerland in 1960's. The relationship of the interest rates earned by bonds of all different maturities is called the term structure of interest rates. A good understanding of the term structure of interest rates is essential in any area of finance. It is also often applied in various types of macroeconomic models.

Especially during the last twenty years, a significant amount of interest rate modeling has taken place. This has occurred due to two major reasons. First, the world markets have changed. World trade has grown, foreign exchange and capital markets have become less controlled, and governments have issued more bonds to finance budget deficits. Interest-rate derivatives, whose values are based directly on the interest rate, have become very popular. Other derivative securities, whose values are based on other underlying variables, need interest rates for discounting. For example, options, although known earlier, were first traded on an exchange on April 26, 1973. The place was The Chicago Board Options Exchange, and call options on 16 stocks were included. Now derivative securities are commonplace all over the world.

Second, these changes have been accompanied with the development of theories in the field of finance and economics in general as well as in the area of the term structures of interest rates.

The short interest rate is essential for modeling the whole term structure of interest rates. Many interest rate models are based on the assumption that the whole term structure is driven either only or at least to a very large extent by the short rate.

The purpose of this dissertation is to study the stochastic behavior of the short interest rate in Turkey. The Turkish short rate has not been studied very extensively, and its behavior is not well known. We show that the short rate process can be described using a hidden Markov chain model, and then estimate the model. This study contributes to the knowledge of the short interest rate behavior in Turkey.

This dissertation is organized in the following way. The objective of the study is given in this chapter. In Chapter 2, we provide a general overview of so-called one-factor interest rate models, which assume that the term structure of interest rates is driven by a single factor, which usually is the short rate. This presentation shows the relevance of the short rate in modeling the whole term structure of interest rates. Chapter 3 is an overview of the Turkish economy. This is included because the Turkish economy differs from many others, having had for example, high interest rates and high rate of inflation for decades. Covering essentially years 2000 and 2001, we point out some of the shocks that have occurred and their impact on the interest rate. Chapter 4 is a principal component analysis of the Turkish Bond Market index. This analysis helps in finding out what kind of a model may be suitable for explaining the behavior of the interest rates in Turkey. In Chapter 5, the data, the overnight interbank interest rate, is described and analyzed. Chapter 6, which is the main part of this study, uses a Markov switching Vasicek model to describe the stochastic behavior of the overnight interest rate. Chapter 7 concludes this dissertation.

2 Stochastic One-Factor Interest Rate Models

Models describing the behavior of interest rates are used for pricing bonds and other interest rate derivatives. Many of these models assume that interest rates can be modeled as a stochastic process driven by one underlying factor. In this chapter, we first review the basic interest rate theories. Then we proceed to examine how, in presence of uncertainty, interest rates can be modeled as a stochastic process using partial differential equations or the martingale method.

We proceed to discuss how interest rates and their dynamic behavior can be studied statistically. After that, we describe the method of principal component analysis (PCA), which has been shown in many studies to support the hypothesis that the dynamic process of interest rates can be explained by very few underlying factors. This result is the empirical justification for using one-factor interest rate models, the most popular of which are examined at the end of the chapter.

2.1 Interest Rate Theories

There are three main theories for explaining the relationship between the interest rates of different maturities (we follow Gibson, Lhabitant and Talay 1999). These are the expectation hypothesis, the liquidity preference theory, and the preferred habitat theory.

2.1.1 The Expectations Theory

The expectations theory is based on the assumption that the term structure is determined by the investor's expectations on the future spot rates. The term spot rate refers to the shortest possible interest rate, and it is often called the short rate. Various different versions of the expectations theory exist, but the general idea can be written using continuously compounded interest rate as

$$R(t,T) = \frac{1}{T-t}\left[\int_t^T E_t(r(s))ds\right]. \tag{2.1}$$

R is the yield curve of a zero-coupon bond for different maturities T at time t (see (2.7) and (2.8)). Here r is the spot rate, which is a function of time s. The equation states that at time t the rate of return for a bond of given maturity T is equal to the geometric average of the expected short-term rate from t to T. Many observers have criticized this model for assuming that investors are able to see in the far future, even decades away.

The expectations theory is consistent with any general shape of the yield curve. For example, a U-shaped yield curve is the result of expecting the spot rate to first decline for a while and the start declining. Note that according to this theory a constant spot rate leads to a horizontal yield curve.

There are four different versions of the expectations hypothesis. Cox, Ingersoll, and Ross (CIR, 1981) examined these cases. The naïve expectations hypothesis, also called the yield-to-maturity expectations hypothesis (YTM-EH), asserts that the expected return following any investment strategy for any period is the same. For a given time period the return will be the same if the investor holds any one bond through the period or rolls over a combination of bonds of any maturities. Formally,

$$-\frac{\ln B(t,T)}{T-t} = E\left[\frac{1}{T-t}\int_t^T r(s)ds\right]. \qquad (2.2)$$

where $B(t,T)$ is, at date t, the price of a default-free zero-coupon bond that pays one unit at maturity T.

The local expectations hypothesis (L-EH) states that for any bond on the market the rate of change of the price, i.e. the momentary interest rate, is equal to the risk-free spot rate.

$$E\left[\frac{dB(t,T)}{B(t,T)}\right] = r(t)dt. \qquad (2.3)$$

The third form is called the return-to-maturity expectations hypothesis (RTM-EH), sometimes called the Lutz hypothesis (after Lutz 1940), which states that a strategy of rolling over a set of short term bonds has the same expected return as a longer bond held to maturity. This can be written as

$$\frac{1}{B(t,T)} = E\left[e^{\int_t^T r(s)ds} \mid r(t)\right]. \qquad (2.4)$$

The fourth form is the unbiased expectations hypothesis (U-EH) also called the Malkiel hypothesis. The first paper on the term structure of interest rates using the U-EH was by Fisher (1896). Later Keynes (1930), Hicks (1953) and Malkiel (1966) have held this view, followed by numerous others. Formally, the U-EH is given by

$$\frac{\partial B(t,T)/\partial T}{B(t,T)} = E[r(t)]. \tag{2.5}$$

According to this, the forward rate is the conditional expectation of the future spot rate. For example to get the two-year forward spot rate for today we would have to look what is the percentage rate of the price change of a two-year bond if we increase its maturity by an infinitesimal amount.

CIR (1981) shows that U-EH and YTM-EH are identical in continuous time. It can be shown, using Jensen's inequality, that the other theories above are mutually inconsistent (under uncertainty). Interestingly, under uncertainty, the expectations hypotheses above do not hold if the expectations use the real-world probabilities. Here 'real-world probabilities' refers to the actual 'observed' probability distributions of the spot-rate. The issue about the probability measure used by the expectations operator is discussed below in part 2.3.

2.1.2 The Liquidity Preference Theory

The liquidity preference theory states that investors want to preserve liquidity and invest for short maturities, while borrowers prefer to borrow at fixed rates for long term. This is a natural consequence of risk aversion.

$$R(t,T) = \frac{1}{T-t}\left[\int_t^T E_t(r(s))ds + \int_t^T L(s,T)ds\right]. \tag{2.6}$$

This is an extension of the expectations theory. The first term is like the expectations theory. Then we add another term to describe the liquidity premium, which is increasing in maturity T. Liquidity preference theory predicts that long rates will be higher than the geometric average of expected short rates.

One way of explaining this theory is to think about a financial intermediary who would under these assumptions take short-term deposits to finance fixed-rate long-term loans. This leads to a substantial interest rate risk that can be compensated by raising the long rates higher than the average of the short rates.

2.1.3 The Preferred Habitat Theory

Modigliani and Sutch (1966) explained the term structure of interest rates by what has since been called the preferred habitat theory. This is quite similar to the liquidity preference theory but now the assumption is that investors and borrowers have different specific time horizons. The term structure is again given by equation (2.6) but now the liquidity premium can take different values. Cox, Ingersoll and Ross (1981) found that time horizon has no impact in homogeneous economies. Riedel (1999) agrees that this is the case in homogeneous economies, but shows that if agents are heterogeneous with different time horizons, preferred habitat theory can be used to explain the term structure of interest rates.

A more restricted version of the preferred habitat theory is the market segmentation theory. In this case, participants in the market are strictly confined to some maturity, without a trade-off. Interest rates for different maturities are determined separately for each maturity by their supply and demand as in partial equilibrium analysis.

2.2 The Term Structure of Interest Rates

The term structure of interest rates gives today's interest rate for each different maturity. Two other descriptions are given for this concept. The discount function, or its graph the discount curve, shows the current spot prices for the discount bonds with different maturities. The yield curve gives the yields of coupon-bearing bonds for different maturities. The information given by all three curves is identical and they can be obtained from each other. We use the concepts 'term structure of interest rates' and 'yield curve' throughout interchangeably.

The term structure of interest rates can be defined more formally. Assume that, at time t, zero-coupon default-free bonds exist for all maturities up to date T_{max}. Now the yield for a bond of maturity T is given by the $R(t,T)$ that solves

$$B(t,T) = e^{-R(t,T)(T-t)}. \qquad (2.7)$$

The yield for a bond with maturity T is then given by

$$R(t,T) = \frac{\ln B(t,T)}{T-t}. \tag{2.8}$$

The term structure of interest rates is the whole yield curve $R(t,T)$ at time t as a function of $T \in (t, T_{max}]$. The interest rates $R(t,T)$ given by (2.8) are the spot rates for different maturities at time t. The forward rate at time t for a loan that starts at u and ends at T ($t<u<T$) is

$$F(t,u,T) = \frac{\ln B(t,u) - \ln B(t,T)}{T-u}. \tag{2.9}$$

Then the instantaneous forward rate, or the forward short rate, will be

$$f(t,u) = \lim_{T \to u} F(t,u,T). \tag{2.10}$$

2.3 Expectations and Uncertainty

The various expectations hypotheses have been the tools used to explain the shape of the yield curve. If the expectations theory holds, it should be possible to predict the future interest rates from the slope of the term structure of interest rates. (Actually, we do not need such a strict condition for predicting the spot rates from the yield curve.)

When constructing the term-structure of interest rates, we use default-free bonds, so there is no default risk. Now clearly, the shortest possible interest rate, the instantaneous spot-rate, is risk-free - it is known for certainty and it is instantaneous. If the risk-free instantaneous r were constant until maturity, the discount bond price of a bond paying one unit at time T would be at time t

$$B(t,T) = e^{-r(T-t)}. \tag{2.11}$$

If the instantaneous risk-free spot-rate is not constant but its path is known with certainty, the bond price will be

$$B(t,T) = e^{-\int_t^T r dt}. \tag{2.12}$$

As long as the interest rate is known with certainty and default is not allowed, the value of a bond is known with certainty at any time, and the strong form of the s hypothesis (U-EH and YTM-EH) holds.

If interest rates are stochastic, which is a reasonable assumption, the above formulas have to use expected values. Now $r(t)$ is the expected instantaneous spot-rate at any moment t

$$B(t,T) = E\left[e^{-\int_t^T r(s)ds} \mid I_t \right]. \tag{2.13}$$

We apply the conditional expectations operator since we are using future instantaneous (spot) interest rates. The expectation is conditional on information set I_t, which includes all the information until time t. At current date t, we know, with certainty, the instantaneous spot rate at time t as well as for any date in the past. However, the integral is over future dates and the instantaneous rate at any future date $s > t$ is not known at time t and has to be predicted. According to (2.13) the bond price depends on the whole spectrum of the instantaneous spot rate for the life of the bond, and the yield curve at any time t contains all the available information about the future spot rates (Neftci 1996). In other words, bond prices depend on the whole yield curve. Suppose that, at time t, we run the following linear regression

$$y = \alpha + x\beta + e, \tag{2.14}$$

where $y = f(t,u)$ and $x=r(u)$, with u a date in the future. Here α is the intercept term and β the slope term, while e is the error term in the regression. Like usually in a simple regression, the dependent variable is stochastic while the regressors are deterministic and the expected value of the error term is zero. In other words, the future spot rates are conditional expectations of the current spot rates.

If we estimate the regression coefficients in (2.14), according to the strong form of the expectations theory the slope term β should be one while the intercept term α should be zero. However, empirical research does not confirm these parameter values, which is the problem with the strong form of the expectations hypothesis. Most of the time, yield curves slope upwards at the short end while they tend to flatten out at long maturities. In general, interest rates do not rise on average. Of course, like Fisher (2001) mentions, one could just assume that the investors are wrong on average, but that would not lead to a very good theory. Fisher also mentions the phenomenon called the 'Peso problem'. This refers to investors giving some weight to very large increases to interest rates that have not yet been observed. The 'Peso problem' rises due to the risk of devaluation of the currency.

The introduction of the additional term, liquidity premium, in the liquidity preference and preferred habitat theories above takes us closer to explaining the relationship between the yield curve and expectations. The liquidity premium is also called a risk premium and a term premium. We will call this premium the term premium.

Default-free bonds are not risk-free. Their prices fluctuate and there is uncertainty about the returns. This uncertainty increases with the maturity. The term premium compensates for this uncertainty.

It is important to understand the difference between forward rate and future (instantaneous) short rate. Fisher and Gilles (1998) describe the difference between a forward rate and the expected spot rate for the same time as the term premium, which is the sum of a risk premium and a convexity premium. (Note that sometimes the whole term premium is called the risk premium.) At time t, the future spot rate for a date $u>t$ is a conditional expectation. However, at time u that spot rate is known with certainty. The convexity premium, which is also called the Jensen inequality term, is a result of the fact that the bond price is not a linear function of the interest rate – in other words $E_t(1/r_{t+1}) > 1/E_t(r_{t+1})$. Consider a simple example. We have a two-period discount bond that pays one at the end of the second period. The interest rate of the bond for the first period is known with certainty, but for the second period it will be with equal probability either 10 percent or 20 percent. Now at the end of the first period the expected present value of the bond is

$E_{t+1}(1/r_{t+2})=0.5(1/1.1)+0.5(1/1.2)=0.8712$ which gives a forward rate of $f(0,1,2)=14.78$ percent for the second period. This is less than 15 percent, which is the expected value of the interest rate. No assumption was made about the investor's attitude towards risk, and this holds even if the investor is risk-neutral. However, investors are usually risk-averse and prefer less risky payment to more risky payment of equal expected value. A risk premium has to be added to the expected future spot rate to compensate the risk averse investor for uncertainty.

The weak version of the U-EH states that forward rates are equal to expected spot rates plus a constant that only depends on the forecast horizon. This constant is the term premium, which is the sum of the risk premium and convexity premium. In this case, the regression equation (2.14) would still have a slope coefficient β of unity but now the intercept term α would be positive. For the weak form of the U-EH the term premium has to be a deterministic function of maturity. A sufficient condition is to hold all volatilities constant, which leads to the Gaussian model.

A necessary condition for the U-EH to hold is to have the risk premium and the convexity premium balance each other for all horizons. Fisher and Gilles (1998) discuss the mathematics of this condition. For U.S. data, empirical studies do not in general support the weak form of the U-EH – so obviously the strong form is not supported either. In other words, the hypothesis that $\beta=1$ in regression (2.13) is rejected (see e.g. Campbell and Shiller (1991), Evans and Lewis (1994), Brown and Schaefer (1994), Bekaert, Hodrick and Marshall (1997a)). On the other hand, for many other countries research has often found support for the weak U-EH (see Hardouvelis (1994), Gerlach and Smets (1997), Dahlquist and Jonsson (1995), and Bekaert, Hodrick and Marshall (1999)).

Lanne (1999) observes that while for short maturities the evidence seems to be clearly against the expectations hypothesis, with the longer maturities the empirical results may be influenced by inappropriate estimation procedures. On the other hand, Longstaff (2000) finds that at the (rarely studied) very short end of the term structures the s hypotheses hold well for US Treasury Bill data. For maturities from one week to three months, the longer-term spot rates are almost unbiased forecasts of the future overnight rate. This supports the hypothesis that the term premium is zero.

Bekaert and Hodrick (2001) mention three main reasons that could lead to the rejection of the U-EH. First, the tests to study the EH's may themselves perform poorly in finite samples. Second, investors may make systematic forecast errors. Third, the term premia may not be independent of the interest rates.

The small sample properties of the EH tests for the term structure of the interest rates have been studied by Bekaert and Hodrick (2000), Bekaert, Hodrick and Marshall (1997, 1999) and Valkanov (1998). They note that if the used tests do not behave well in small samples, alternative methods are necessary. Bekaert (2000) compared different tests and found out that the most frequently used test, the Wald test, has the worst small-sample properties.

Another effect that can cause the expectations hypothesis to fail is the mentioned 'Peso effect'. If there is any risk of a 'regime shift', i.e. devaluation, the expectations will be affected and a premium will be added on the interest rates. Now, if we examine a time period during which devaluation did not take place, it looks as if the expectations hypothesis failed. As a matter of fact, because the term premium is based on probabilistic thinking, the expectations hypothesis would also fail during a time period when the devaluation takes place. Even if the expectation is correct in the long run, it would always fail during a short time period.

2.4 History of Stochastic Models in Finance

The first model of stochastic behavior in finance is by Bachelier (1900). He presented a model where stock price changes follow an arithmetic Brownian motion. (In continuous time, this results in a stochastic differential equation that in the field of physics is often called the Langevin equation). Osborne (1959) and Samuelson (1965) followed later. However, these papers did not use the 'arbitrage-free' assumption.

The use of continuous–time stochastic models took off at the very end of 1960's and beginning of 1970's. Intertemporal asset pricing theory was developed mostly by Merton (1969, 1971,1973b). Lucas (1978) presented a general equilibrium approach to asset pricing in discrete time. Then Cox, Ingersoll, and Ross (CIR, 1985b) developed the continuous-time general equilibrium model for the term structure of interest rates.

In pricing derivative securities, the groundbreaking work was done by Black and Scholes (1973) and Merton (1973a). They show that, by holding a portfolio of the underlying stock and a risk-free money-market instrument and continuously rebalancing the portfolio, it is possible to replicate the call option. The literature on risk-neutral pricing started from Cox and Ross (1976a, 1976b) who studied option pricing using different stochastic models. These models are solved as stochastic partial differential equations (PDE). Later, in an innovative paper, Harrison and Kreps (1979) elaborated the idea of risk-neutral valuation, showing that no-arbitrage implies the existence of a risk-neutral probability measure. This leads to the use of martingale methods.

2.4.1 Arbitrage-Free and General Equilibrium Models

An interesting classification of interest rates models is to draw a distinction between arbitrage-free and equilibrium models. The previous ones directly use the arbitrage theorem, which states the conditions under which arbitrage profits cannot exist.

Arbitrage opportunities can arise in two different ways. Arbitrage opportunities of the first kind occur when an investor can earn a positive profit with a zero set-up cost. Arbitrage opportunities of the second kind occur when an investor can earn an immediate profit with no cost at any stage. Arbitrage-free models rule out any arbitrage opportunities of the first kind or the second kind.

General equilibrium models start by describing general equilibrium conditions for the economy, including the utility function of the representative investors. Asset prices and the term structure of interest rates are determined simultaneously as a part of the general equilibrium.

Merton (1975) examined a stochastic growth model, a general equilibrium model, and had the spot rate follow a diffusion process. However, the term structure of interest rates was not studied, because it was not relevant for the paper.

The Black-Scholes (1973) model is based on the idea of no arbitrage. The traditional arbitrage-free interest rate models start from the late 1970's with Vasicek (1977) and Dothan (1978).

Harrison and Kreps (1979) show that the no-arbitrage condition implies the existence of a risk-neutral probability measure under continuous trading. This observation brings the general equilibrium and arbitrage-free approaches closer together.

Back (1997) and CIR (1985b) have drawn attention to the fact that the 'traditional arbitrage-free approach' does not imply the absence of arbitrage opportunities. Back (1997) points out that 'arbitrage-free' in the earlier models means the existence of a risk-neutral probability measure.

This is not surprising, since a model in general would not 'fit' all the observations, like e.g. a regression line would not go through every point. In a similar fashion, a bond price, or interest rate, model may misprice some bonds in the sense that the theoretical prices would not be the observed ones. A later development among the arbitrage-free models is the calibrated models starting from Ho and Lee (1986), which adjust the parameters of the model to fit the data. However, this fitting has to be done repeatedly in time. Sundaresan (2000) and Back (1997) point out that even in the calibrated models lack of arbitrage should be understood in the sense that it guarantees the existence of a risk-neutral probability measure.

2.5 The Partial Differential Equation and Martingale Approaches

A financial contract is a 'derivative security' or a 'contingent claim', if its value at expiration date T is determined exactly by the market price of the underlying cash instrument at time T.

Determining the behavior of the price of a derivative security can be done using two different approaches, which both lead to the same result. The first, and older method, is to assume a stochastic differential equation for the evolution of the price of the underlying asset. Then it is possible to derive a stochastic partial differential equation ((S)PDE) that describes the behavior of the derivative security.

An alternative way of modeling the stochastic behavior of the price of a derivative security is to change the probability distribution of the price so that under the new probability distribution the strict form of the U-EH holds.

For example, a stock option is a derivative security of a traded asset, which is the stock itself. The option can be priced using the no-arbitrage condition and a risk-free portfolio containing the right amounts of the stock and the option.

A bond is like a derivative asset of the interest rate. However, interest rates are not traded securities. Thus, the price of a bond is not a derivative security of a traded asset, and is more difficult to model.

2.5.1 Modeling Interest Rates Using PDE's

In risk-free valuation, a portfolio of traded assets is constructed so that it contains the underlying asset and the derivative asset in such proportions that all the risk is hedged away. Since the bond price is a derivative of the interest rate, and interest rate is not a traded security, there is no underlying security to use for hedging. Risk-free valuation can be applied to bonds by hedging one bond with another of a different maturity.

Assume that the short rate, or instantaneous interest rate, follows a stochastic process

$$dr = u(r,t)dt + w(r,t)dW. \quad (2.15)$$

The term $u(r,t)dt$ is a deterministic drift. The second part, $w(r,t)dW$, is a Wiener process, which is stochastic and can also be called the diffusion term or the innovation term. To understand the stochastic term we need some definitions (see Neftci 1996).

Definition (i)

A process $\{S_t, t>0\}$ is a martingale with respect to a family of information sets I_t and with respect to probability P if for all $t>0$, where t is time, and

(1) S_t is known given I_t, i.e. S_t is I_t-adapted,

(2) unconditional forecasts are finite, $E[S_t]<\infty$, and

(3) $E_t[S_u] = S_t$ for all $u>t$.

I_t is the information set that contains all information up to time t.

This implies that the future variations in S_t are completely unpredictable. Therefore, a martingale is a random variable whose variation is totally unpredictable given the current information set. A martingale is always defined with respect to some probability measure P and information set I_t.

Definition (ii)

A martingale S_t is a continuous martingale if its trajectories in time are continuous in the sense that when $\Delta t \rightarrow 0$, $P(|S_{t+\Delta t} - S_t|) > \varepsilon \rightarrow 0$.

Definition (iii)

A continuous martingale S_t is square integrable if it has a finite second moment for all $t>0$, so $E[S_t]^2 < \infty$. In other words, the process has a finite variance. This actually implies that the stochastic increments of the random variable, $S_{t+\Delta t}$ - S_t can be assumed normally distributed.

Definition (iv)

A Wiener process W_t is a continuous stochastic process that is defined relative to some family of information sets $\{I_t\}$ such that the pair I_t, W_t is a continuous square integrable martingale with $W_0=0$ and $E[(W_t - W_s)]^2 = t\text{-}s$, and $s \leq t$.

Some regularity conditions must be fulfilled by (2.15) to guarantee a unique solution:

(1) u and w must be measurable functions from $R_+ \text{x} R$ to R.

(2) A constant $k_1>0$ exists s.t. the Lipschitz condition holds. Formally, this is

$$\frac{|u(t,x)-u(t,y)|+|w(t,x)-w(t,y)|}{|x-y|}\le k_1 .$$

This condition holds for any function that has a continuous partial derivative.

(3) A constant k_2>0 exists s.t. the growth condition is specified over the interval [0,*T*]

$$\frac{|u(t,x)|^2+|w(t,x)|^2}{(1+|x|)}\le k_2 .$$

The important features of a Wiener process are that

- the process W_t is continuous,

- all the increments of W_t are uncorrelated,

- the mean is zero with $E[W_t]$=0 and $E[\Delta W_t]$=0, and

- the variance of W_t is t, since W_0=0, and variance of $\Delta W_t = \Delta t$.

The Levy Theorem:

Any Wiener process defined relative to a family of information sets I_t is a Brownian motion process.

We can draw a very important conclusion based on the Levy theorem. A Brownian motion has, by assumption, normally distributed increments with zero mean. Thus, we can state that the increments of the Wiener process, ΔW_t, in addition to having zero mean and variance equal to Δt, are normally distributed.

Since the stochastic term depends on *r*, the equation (2.15) is 'multiplicative' and has no meaning unless an interpretation is given to the stochastic term. The stochastic term dW is 'white noise' and extremely random. In fact, it is not smooth, and a derivative does not exist in the traditional sense. The path of *W* is even of unbounded variation so a

Riemann integral does not exist either. An integral of the stochastic term only exists in the 'mean-square' sense of a random sum.

Equation (2.15) can be interpreted in the Ito or in the Stratonovich sense. The Ito method is non-anticipating, while the Stratonovich method is anticipating. Van Kampen (1981) has shown that no physical reason can be used to legitimize the interpretation of the SDE used in pricing assets. In physics, almost always, the Stratonovich method makes more sense. Generally, the Ito convention has dominated in finance, because it, although being older, fits better together with the intuition of the equivalent martingale approach. (If the stochastic process is non-Markovian, the Ito solution is intuitively less appealing). One can always change a Stratonovich solution to an Ito solution and vice versa relatively easily. We will always follow the Ito approach.

At time t the price of a bond of maturity T_1 is

$$B_1 = B_1(r,t,T_1). \tag{2.16}$$

where r is the interest rate. Applying Ito's lemma with respect to r and t to (2.16) gives

$$dB_1 = \frac{\partial B_1}{\partial t}dt + \frac{\partial B_1}{\partial r}dr + \frac{1}{2}w^2\frac{\partial^2 B_1}{\partial r^2}dt. \tag{2.17}$$

This shows how the value of this bond changes in time. Next substitute (2.15) for dr

$$dB = \frac{\partial B_1}{\partial t}dt + u(r,t)\frac{\partial B_1}{\partial r}dt + \frac{1}{2}w^2\frac{\partial^2 B_1}{\partial r^2}dt + w\frac{\partial B_1}{\partial r}dW. \tag{2.18}$$

After dividing both sides by B_1 we get the instantaneous return on the bond at any time t

$$\frac{dB_1}{B_1} = \frac{1}{B_1}\left[\frac{\partial B_1}{\partial t} + u(r,t)\frac{\partial B_1}{\partial r} + \frac{1}{2}w^2\frac{\partial^2 B_1}{\partial r^2}\right]dt + \frac{1}{B_1}\left[w\frac{\partial B_1}{\partial r}\right]dW. \qquad (2.19)$$

Equation (2.19) is the stochastic process for the bond price, and it can be written identically as

$$\frac{dB_1}{B_1} = \mu_1(r,t)dt + \sigma_1(r,t)dW. \qquad (2.20)$$

Note that we have defined

$$\mu_1(r,t) = \frac{1}{B_1}\left[\frac{\partial B_1}{\partial t} + u_1(r,t)\frac{\partial B_1}{\partial r} + \frac{1}{2}w^2\frac{\partial^2 B_1}{\partial r^2}\right] \qquad (2.21)$$

and

$$\sigma_1(r,t) = \frac{1}{B_1}\left[w\frac{\partial B_1}{\partial r}\right]. \qquad (2.22)$$

Recall that we had $B_1 = B_l(r,t,T_1)$. Now we introduce another bond of different maturity T_2:

$$B_2 = B_2(r,t,T_2) \qquad (2.23)$$

This bond has an instantaneous growth rate, analogically to (2.20),

$$\frac{dB_2}{B_2} = \mu_2(r,t)dt + \sigma_2(r,t)dW . \qquad (2.24)$$

Next, we build a portfolio of these two bonds to hedge away all randomness:

$$\Pi = a_1 B_1 + a_2 B_2 . \qquad (2.25)$$

The value of this portfolio evolves according to

$$d\Pi = [a_1\mu_1(r,t) + a_2\mu_2(r,t)]dt + [a_1\sigma_1(r,t) + a_2\sigma_2(r,t)]dX . \qquad (2.26)$$

We can use one bond to hedge the other to eliminate all randomness in Π by choosing a_1 and a_2 so that the last term in brackets goes to zero. Since all randomness has been eliminated, the deterministic part of (2.26) has to grow at the risk-free rate. Now we have two equations

$$[a_1\mu_1(r,t) + a_2\mu_2(r,t)] = r(t), \text{ and}$$

$$[a_1\sigma_1(r,t) + a_2\sigma_2(r,t)] = 0 .$$

We can rewrite these as

$$
\begin{aligned}
a_1[\mu_1(r,t) - r(t)] &= -a_2[\mu_2(r,t) - r(t)] \\
a_1\sigma_1(r,t) &= -a_2\sigma_2(r,t) .
\end{aligned}
\tag{2.27}
$$

Dividing the top equation in (2.27) by the bottom one yields

$$
\frac{\mu_1(r,t) - r(t)}{\sigma_1} = \frac{\mu_2(r,t) - r(t)}{\sigma_2}. \tag{2.28}
$$

The difference between the left-hand and right-hand sides is the maturity of the bonds. Thus, the left-hand side is a function of T_1 but not T_2, and the right-hand side a function of T_2 but not T_1. The equality can hold only if both sides are independent of maturity. This allows us to write for a bond of any maturity μ_B=μ_1= μ_2 and σ_B=σ_1=σ_2 in (2.28), (2.21), and (2.22). Now, using (2.28), we can write a definition

$$
\lambda = \frac{\mu_B(r,t) - r(t)}{\sigma_B}, \tag{2.29}
$$

which is called the market price of risk. Now the instantaneous return on a bond, at time t, can be expressed as

$$
\mu_B(r,t) = r(t) + \lambda(r,t)\sigma_B(r,t). \tag{2.30}
$$

The return on a bond is the sum of the risk-free interest rate and the risk premium. Insert into (2.29) the terms for μ_B and σ_B from (2.21) and (2.22) to get

$$\frac{\partial B}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 B}{\partial r^2} + (u - \lambda w)\frac{\partial B}{\partial r} - rB = 0 \ . \tag{2.31}$$

Equation (2.31) is the (discount) bond pricing equation. The term u-λw is the risk-adjusted drift. The equation can be used to price, not only bonds, but other interest rate derivatives as well. The solution to equation (2.31) is the bond price under probability measure P. Assuming that at maturity the bond pays one unit to the holder, we impose a final condition $B(r,T,T)$=1. We also need two boundary conditions, which depend on the exact form of equation (2.31).

Equation (2.31) is a linear second-order parabolic partial differential equation. It is parabolic, meaning that it is like a diffusion equation. This equation is quite similar to the Black-Scholes (1973) equation. It consists of four parts from left to right: time decay, diffusion, drift, and discounting (Wilmott 1998). Although explicit closed-form solutions to partial differential equations are rare in general, this case is an exception. Linear in this context means that, if we find two solutions to this equation, the sum of them is a solution as well; this leads to solutions using Green's functions.

The solution to (2.31) is of the form

$$B(r,t,T) = E^P\left[e^{-\int_t^T r(s)ds - \frac{1}{2}\int_t^T \lambda^2(s,r(s))ds - \int_t^T \lambda(s,r(s))dW(s)} \mid I_t \right]. \tag{2.32}$$

where the expectation is with respect to the real-world probabilities P.

2.5.2 The Market Price of Risk

In (2.15) the drift rate of the risk-free interest rate is u, but in the bond pricing equation (2.31) the drift rate is $u-\lambda w$. The latter drift rate is called the risk-neutral (instantaneous) interest rate and it follows the stochastic process

$$dr = (u-\lambda w)dt + wdW. \tag{2.33}$$

Bonds are risky in the sense that their prices fluctuate in time. The interest rate that is consistent with bond prices includes the market price for risk and follows (2.33). On the other hand, if we want to model the real world spot rate process, we should set $\lambda=0$, and (2.33) becomes (2.15).

The bond value changes in time by dB, which is given in (2.17). Inserting from (2.31)

$$\frac{\partial B}{\partial t} = rB - (u - \lambda w)\frac{\partial B}{\partial r} - \frac{1}{2}w^2\frac{\partial^2 B}{\partial r^2}$$ into (2.17) gives

$$dB = \left[(rB + \lambda w)\frac{\partial B}{\partial r}\right]dt + w\frac{\partial B}{\partial r}dW. \tag{2.34}$$

This is the increase in the value of a single unhedged bond in time. If B earned the risk-free rate, it would grow at $rBdt$. However, the bond is not risk-free and the compensation for taking risk is

$$dB - rBdt = w\frac{\partial B}{\partial r}(\lambda dt + dW). \tag{2.35}$$

This excess is stochastic because of the presence of the dW term (with E[dW]=0. The compensation for accepting extra risk is λdt per unit of risk dW. This is why the function λ is called the market price of risk. Unfortunately, the market price of risk cannot be observed directly, which is a problem if one tries to find the risk-neutral process.

2.5.3 The Expectations Hypothesis and the Risk Premium

Recall the strong form of the expectations hypothesis:

$$\frac{1}{B(t,T)} = E^P\left[e^{\int_t^T r(s)ds} \mid I_t\right]. \tag{2.36}$$

If λ=0 then μ=r always and the strong form of the expectations hypothesis holds. However, λ is not generally zero.

2.6 The Martingale Approach

An alternative approach for pricing bonds and other derivative securities originates from Harrison and Kreps (1979). This involves changing the probability distribution of the stochastic term so that an expectation with respect to the new probability measure can be taken directly without solving a PDE. It is easy to see how this method leads to the same result as solving the PDE.

Recall the definition of a martingale above in part (2.5.1). Obviously, bond prices are not martingales, because they increase in time, on average. However, it is possible to transform the probability measure P to another probability measure Q, so that under Q the bond prices, discounted by the risk-free rate, become martingales. The Q is called an equivalent martingale measure.

In order to construct a risk-free portfolio or to find an equivalent martingale measure, we need to have complete markets. There exist many different characterizations for a complete market. One way to formalize this is to define markets as complete when asset prices are determined by a unique state price process. In a complete market, a derivative security can be replicated using the underlying security so that the derivative becomes redundant. (Karatzas and Shreve (1991) provides a technical discussion about complete markets.)

Suppose that under the real-world probability distribution of P, r_t is normally distributed

$$r_t \sim N(\mu, \sigma^2) . \tag{2.37}$$

Now we can define the new probability measure Q, under which

$$r_t \sim N(\rho, \sigma^2) \tag{2.38}$$

and the only difference from (2.37) is the drift parameter. With Q the following conditional probability holds:

$$B(t,T) = E^Q\left[e^{-\int_t^T r(s)ds} \mid I_t \right]. \tag{2.39}$$

We see that, under the new probability measure, 'the strong form of the expectations hypothesis holds'. Harrison and Kreps (1979) called Q the equivalent martingale measure, Cox and Ross (1976b) risk-neutral probabilities, and Cox, Ross, and Rubinstein (1979) artificial probabilities.

Technically moving from P to Q is done by using the Girsanov theorem (see Neftci 1996). The Girsanov theorem states that, provided that Q exists and is unique, we can find a new I_t-adapted stochastic process

$$dW^*(t) = dW(t) + dX(t). \tag{2.40}$$

Both stochastic processes $W(t)$ and $W^*(t)$ are Wiener processes. The difference is that $W(t)$ is a martingale under probability measure P while $W^*(t)$ is a martingale under Q. The probability measures are linked by

$$dP = \xi(t,\lambda)dQ\,, \tag{2.41}$$

where $\xi(t,\lambda)$ is the Radon-Nikodym derivative, which is commonly written as

$$\xi = dQ/dP\,. \tag{2.42}$$

(The arguments are not written out in order to simplify notation.)

We examined the interest rate process under P in (2.15):

$$dr = u(r,t)dt + w(r,t)dW. \tag{2.43}$$

We saw in (2.33) above that for the risk-neutral rate we write

$$dr = (u\text{-}\lambda w)dt + wdW. \tag{2.44}$$

We want to express the bond price process as a martingale under probability measure Q when discounted by the risk-free interest rate $r(t)$. In other words, the bond price will grow at

$$E^Q\left[\frac{dB(t,T)}{B(t,T)}\right] = r(t)dt. \tag{2.45}$$

In (2.20) we saw that the bond price grows under P at

$$\frac{dB}{B} = \mu_B(r,t)dt + \sigma_B(r,t)dW. \tag{2.46}$$

Now we can find the new Q process applying the Girsanov theorem (2.40) to (2.46). Substitute $\mu_B = r + \lambda\sigma_B$, and replace dW by dW^* using(2.29) and (2.40) to get

$$\frac{dB}{B} - r(t)dt = \left[\lambda(r,t)\sigma_B(r,t) + r(t) - r(t)\right]dt + \sigma_B(r,t)[dW(t) + X(t)] = 0. \tag{2.47}$$

We know that that under the new probability measure Q, dW^* will be a standard Wiener process. By choosing $dX(t)$ so that the equality in (2.47) holds, it becomes a martingale, and Q is a martingale measure

$$dX(t) = -\frac{\lambda(r,t)\sigma_B(r,t)}{\sigma_B(r,t)} = -\lambda(r,t). \tag{2.48}$$

Applying (2.40) yields

$$W^*(t) = W(t) - \int_0^t \lambda(r,t)ds. \tag{2.49}$$

In (2.49) we have subtracted from the P-process the 'excess return' over the risk-free rate. The Radon-Nikodym derivative is given by

$$\xi(t,\lambda) = e^{\int_0^t \lambda dW^*(t) - \frac{1}{2}\int_0^t \lambda^2 ds}. \tag{2.50}$$

The PDE approach and the martingale approach give identical results. The bond value will be the same if we calculate it using the P process with the adjusted drift, or the Q process with the adjusted probability measure.

2.7 Estimating the Term Structure Statistically

Estimating the term structure of interest rates from cross section data is a widely studied subject. The theoretical term structure is a continuum of default-free interest rates. The rates can be obtained in various ways. They can be e.g. repo rates, swap rates or inter-bank rates (e.g. London Interbank Offered Rate, LIBOR). Traditionally, the term structure is obtained from the discount curve, i.e. from the prices of default-free discount bonds. A discount bond pays certainly, regardless of what happens, the face value at maturity and nothing earlier. The face value is also called the principal or the notional value. These bonds are sold at a price that is cheaper than the face value, thus the name discount bond, so that the holder earns interest. Typically, bonds with short maturities are discount bonds, while bonds of longer maturities are coupon-bearing bonds. Coupon-bearing bonds pay a predetermined amount, coupon payments, at some given dates and a principal payment at maturity. Conceptually, there is no difference between zero-coupon and coupon bonds. Coupon-bearing bonds can be treated as a collection of discount bonds. Each payment is seen as a separate discount bond with a maturity date on the date of the coupon and a face value of the coupon payment. While finding the yield curve, the problem with possible default risk is usually avoided by using rates of instruments with which default is extremely unlikely, like government bonds.

The term structure of interest rates can be calculated from bond prices. However, this is not always convenient. One problem with bonds is that they exist only for some maturities. We may for example have a bond with 85 days until maturity today, but not a bond for 75 days. Then tomorrow the 85-day bond has only 84 days to maturity. In addition, some bonds are not very liquid. Besides, regulations, taxation, and institutional issues can be important.

An alternative approach is to use money-market interest rates. On the interest rate markets, a full selection of maturities is available at all times. For example, an overnight, 7-day, 1-month, 3-month or 12-month interest rates are quoted every trading day. Even then, not all maturities are equally liquid. A common practice is to use the most liquid assets over the different parts of the yield curve (Rebonato 1998). Thus, one curve might contain overnight repo rates, deposit rates, futures rates, and swaps in addition to rates derived from bond prices. The term structure of money-market interest rates can be different from that derived from bond prices. According to Meier (2000), in Switzerland the term structure of the Swiss Euromarket (interbank) interest rates normally lies above the curve derived from the bond prices, and the difference sometimes varies with maturity. The difference, and its variability, cannot be explained by lack of liquidity in the bond market. Meier mentions that this could be the problem with many models that try to match the observed bond prices using money-market interest rates (usually short rates).

The traditional textbook derivation of interest rates is based on discount bonds and easily extended to coupon-bearing bonds. For calculating a term structure, interest rates have to be converted to the same compounding frequency. An investment with a yearly interest rate of 12 percent earns more if the interest is paid at the end of each month rather than at the end of the year. Suppose that the original investment is one unit, so the monthly investor has after one month 1.01 units. He invests this 1.01 for another month and the he rolls over his investment for the whole year. At the end, he has 1.01^{12} which is 1.126825 units. This is more than the 1.12, which is earned by the yearly investor. Often the models in finance use continuous compounding. Suppose that we want to convert a return R_m that is compounded m times per year to a continuously compounded return (with a compounding frequency of infinity). We can use a simple formula to find the rate of return R_c for continuous compounding:

$$R_c = m \ln\left(1 + \frac{R_m}{m}\right). \tag{2.51}$$

To express this return as a continuously compounded interest rate, one has to multiply it by 100 percent. For example a 1-year interest rate of 12 (i.e. a yearly return of 0.12) gives a continuously compounded rate of R_c = 0.113329. For discount bonds (2.51) can be applied easily to get

$$R_i = m_i \ln\left(1 + \frac{(1 - B_i)}{B_i}\right) \tag{2.52}$$

where B_i is the current price of the bond which pays one at maturity. If the time to maturity is $\tau = (T\text{-}t)$, then $m_i = 1/\tau$.

A traditional method for obtaining the term structure from bond prices is called bootstrapping. Table 2.1 uses a simple numerical example to illustrate bootstrapping.

Principal (paid at maturity $)	Time to maturity (years)	Annual coupon (paid every 6 months $)	Current bond price ($)
100	0.25	0	98.0
100	0.5	0	95.2
100	1.0	0	90.0
100	1.5	8	101.0

Table 2.1. An example of the bootstrapping method.

First, take the shortest maturity bond on the market. This bond has 3 months to maturity. We find the 3-month rate from this bond price using (2.52) at

$$R_{0.25} = 4\ln\left(1 + \frac{2}{98}\right) = 0.08081$$

The 6-month and 12-month rates will be

$$R_{0.5} = 2\ln\left(1+\frac{4.8}{95.2}\right) = 0.09838 \qquad \text{and}$$

$$R_{1.0} = \ln\left(1+\frac{10}{90}\right) = 0.10536\,.$$

The next bond is a coupon-bearing bond with coupon payments of \$4 paid at 6, 12 and 18 months. In addition to these, at 18 months the principal of \$100 will be paid. The present value of all the payments has to equal the bond price. Formally

$$PV(Bond) = \sum_{i=1}^{n} C_i e^{-r_i t_i} + Pe^{-r_n t_n} \tag{2.53}$$

The first coupon of C_1 is paid at time t_1=0.5, the second coupon of C_2 at t_2=1.0 from today and so on. At maturity t_n the coupon C_n and principal P are paid. We can insert the rates we calculated before

$4e^{-0.09838\times0.5} + 4e^{-0.10536\times1.0} + 104e^{-R\times1.5} = 101$, which gives

R = -ln(0.89992)/1.5 = 0.0703.

Now we have solved for the yields, and interest rates, for all the given maturities. The relationship between maturity and interest rate, if we have also the longer maturities, can be called the term structure of the interest rates, the discount curve or the zero-coupon yield curve. Essentially, we detached all the coupon payments and treated each of them like a separate discount bond. Using (2.53) we can calculate the bond value as a linear combination of all the payments. Of course, the result will be piecewise linear, a graph with kinks at the maturities and straight lines between them.

Note that the current forward interest rate for a time period starting at time t_1 and ending at t_2 is given by

$$f(0,T_1,T_2) = \frac{r(0,T_2)T_2 - r(0,T_1)T_1}{T_2 - T_1}, \tag{2.54}$$

where $r(0,T_1)$ is the current rate of length T_1, and $r(0,T_l)$ of length T_2. Look at the example above. Using the 6-month rate is 0.09838 and the 12-month rate is 0.10536. We can calculate the current forward rate from 6 to 2 months using

$$f(0,0.5,1.0) = \frac{0.10536 \times 1 - 0.09838 \times 0.5}{1 - 0.5} = 0.11234.$$

It is possible to define a forward function, or a forward curve, that relates the forward rates to different maturities. All these concepts provide the same information and are conceptually identical. However, in practice, the discount function is the most convenient choice because it leads to the use of linear methods.

2.7.1 Fitting the Yield Curve

Once the spot rates for given maturities are calculated, the points have to be connected to get a continuous curve for the term structure of interest rates. Just 'connecting the dots' is the simplest way to do it, but the curve will be piecewise linear with kinks, and we get jumps in the forward rates. The Mitsubishi Finance Risk Directory published in the early 1990's a survey of thirteen companies selling software to price complex derivative securities. Twelve of them used straight-line interpolation (described above) for the term structure of interest rates (see Adams and Deventer 1994). However, much more sophisticated and better methods exist and are used.

A smooth curve can be fitted to a set of points using polynomial interpolation – i.e. finding a polynomial that goes through the given points. Unfortunately, even some smooth functions cannot be approximated well by polynomials. Another problem is that as the number of points becomes high, the degree of the polynomial tends to become very high as well. A similar problem occurs if the intervals between the points are very different.

A solution to keep the approximating function simple, e.g. a low-degree polynomial, is to divide the whole interval into shorter sub-intervals and then separately approximate each segment with a polynomial. If the procedure is carried out so that these segments can be connected to each other so that the interval is covered with a continuous piecewise function (e.g. polynomial), the whole curve is called a spline. The two commonly used splines are the piecewise polynomial form and the B-spline. De Boor (1978) discusses these splines and their mathematical derivation. Another alternative, which is not commonly used anymore, would be to try to find a single function that describes the whole term structure curve.

The seminal papers in the estimation techniques are McCulloch (1971 and 1975). The basic idea is that the price of a bond is equal to the present value of its coupon and principal payments like was seen in (2.52). Then the present value function is parameterized as a cubic spline and the term structure is estimated using simple linear regression. A cubic spline curve fits the points smoothly. Variations of the spline method were studied among others by Schaefer (1981), Vasicek and Fong (1982), Shea (1984), Jordan (1984), Chambers, Carleton and Waldman (1984), Coleman, Fisher and Ibbotson (1992).

Some studies have used exponential splines but according to Shea (1985), they are not necessarily better than polynomial splines. Exponential splines are estimated with nonlinear models, while cubic splines use linear models. Buono, Gregory-Allen and Yaari (1992) conducted Monte Carlo simulations and conclude that when the term structure has a complex shape, cubic splines and least-squares estimation may provide more accurate results than exponential splines. Fisher, Nychka and Zervos (1994) describes a method using smoothing splines with a 'roughness penalty' and shows, using Monte Carlo simulations, that it gives better results than normal cubic splines. Bliss (1996) compares different models for estimating the term structure function and Tanggaard (1997) introduces a non-parametric method.

An important question to address is whether the curve should go through 'all the points' or not. The current practice is to use a smooth curve rather than trying to fit all, possibly illiquid and maybe mis-priced bonds on the curve.

2.8 Statistical Approach to the Dynamics of the Term Structure of Interest Rates

In the previous section we discussed how the spot rates for different maturities can be calculated and combined to give the term structure of interest rates, also called the (zero coupon) yield curve. This is a snapshot and refers to some specific time. The evolution of this curve describes the dynamics of the interest rates. We can think about different forward rates (or spot rates of different maturities) as correlated random variables. The 'closer' the rates are on the maturity axis the higher is the correlation between them. Suppose that we look at two interest rates. One is the one-month spot rate $f(0,0,1)$. The other one is $f(0,5,6)$, which is the current forward rate for a one-month loan that starts five years from today. The first argument is the day of quote, with t=0 as today. The second argument is the beginning and the third is the end of the loan, in months. If we observe these rates in time, each of them would have their own variance. However, an interesting measure for understanding the dynamics would be the covariance (or correlation) between the two rates.

2.8.1 Principal Component Analysis

We will use a simple linear factor decomposition to see how we can benefit from the correlations between interest rates, or yields, of different maturities. This method is called principal component analysis (PCA). It is possible to use yields (or interest rates), or changes in them, as well as forward rates or changes in them.

Suppose that we have n different (instantaneous) forward rates f_i as defined in (2.10) with i=[1,2,3,...,n]. Since the f_i are random variables, we can write a vector

$g \equiv df = \mu dt + \sigma dz$, or in vector form

$$g = \begin{bmatrix} df_1 \\ df_2 \\ \dots \\ df_n \end{bmatrix} = \begin{bmatrix} \mu_1 dt + \sigma_1 dz_1 \\ \mu_2 dt + \sigma_2 dz_2 \\ \dots \\ \mu_n dt + \sigma_n dz_n \end{bmatrix}. \quad (2.55)$$

where μ_i is drift and dz_i is a Wiener process. At any point in time, the forward rates are pairwise correlated, i.e.

$$E[dz_i(t)dz_j(t)] = \rho_{ij}(t)dt, \quad (2.56)$$

where $\rho_{ij}(t)$ is the correlation between the increments in two forward rates i and j. Now this stochastic process consisting of n forward rates completely describes the dynamics of the forward rates and the term structure of interest rates, provided that n is large enough and the rates properly cover the maturity axis.

We can write the covariance matrix Ω for vector g as

$$E[(df-\mu)(df-\mu)'] = \Omega. \quad (2.57)$$

A transformation can be applied to any vector to rotate or stretch it. Now apply the transpose of an nxn matrix A to vector g and get Y

$$Y = A^T g. \tag{2.58}$$

We can think that matrix A consists of column vectors called a_s, each being nx1, which we call factors. Suppose that we choose matrix A so that all the vectors a_s are orthogonal to each other. Now matrix A is said to be orthogonal so that $A^T=A^{-1}$, and we can write

$$Y = A^T g = A^{-1} g, \text{ and}$$
$$AY = AA^{-1} g = g. \tag{2.59}$$

Verbally, this means that the increments in the forward rates, vector g, can be expressed as a linear combination of the new factors. Since Ω is a symmetric nxn matrix, we can always find the orthogonal axes of Ω. Then the covariance matrix of Y will be

$$\Lambda = A^T \Omega A, \tag{2.60}$$

where Λ is a diagonal matrix with diagonal elements that are the characteristic roots (eigenvalues) of Ω. Usually the vectors a_s are normalized so that their Euclidean norm, or length, is equal to one. Each of these vectors a_s has one element a_{is} for each maturity included in the data. Now we can see the individual values a_{is} as weights. The new factors are not correlated with each other (obviously, since the vectors a_s are orthogonal) and their variances are the eigenvalues of the covariance matrix of g. If $\rho(df_i, df_j)<1$ for all $i \neq j$ matrix A has full rank, which is also the number of eigenvalues, but if changes in any forward rates are pairwise perfectly correlated, the number of eigenvalues will be lower.

The new factors are called principal components. We can arrange the components according to the eigenvalues (variances) in descending order from largest to smallest. Then the one with the highest value is called the first principal component, the second highest the second principal component and so on. The first principal component accounts for the largest proportion of the variance, the second principal component the second largest etc. Suppose that all the changes in forward rates are perfectly correlated. Then only one principal component exists, and all changes in interest rates are driven by one factor. If the correlation between the different rates is large, very few factors are enough to explain the variability in interest rates. Respectively, if the correlation is low more factors are needed.

The PCA method often leads to principal components that cannot be easily interpreted, but applying it to the term structure of interest rates is an exception. If the method is applied to changes in yields, an easy interpretation is available. Usually the first principal component consists of elements a_{i1} of approximately the same value. This is interpreted as the level effect of the curve. Any changes in the first principal factor are then 'parallel shifts' having an equal effect on each part of the yield curve.

The second principal component has weights a_{i2}, which have opposite signs but similar magnitudes at each end of the curve. Thus, an increase in this principal component changes the slope of the curve, and this change is described as a 'twist'.

The third principal component generally has weights a_{i3}, which are of one sign at the ends of the curve and have opposite sign in the middle. In addition, the values in the middle are larger than at the ends. This causes the middle of the curve to move to the opposite direction of the ends. This is described as 'curvature', 'butterfly', or 'bow'.

Note that if one uses changes in forward rates, the interpretations apply to the forward curve (which is the term structure of forward rates), but the connection to the yield curve is direct.

Empirical studies carried out on changes (or sometimes levels) in yields or forward rates often show that the correlations between the different maturities are very high, which indicates that using PCA can be very useful. It is common that the first principal component explains 80 to 90 percent of the total variance, and the first three together explain 95 to 99 percent. Results of various empirical studies are mentioned in Wilson (1994). Litterman and Scheinkman (1991) studied weekly yield changes in U.S. data, and found that the first component accounts for about 90 percent of the total variance and the three first ones together to practically all of it. Rebonato (1998) finds 92.17 percent for the first component and 99.10 and 99.71 for the first two and three components respectively using UK data. In Reimers and Zerbs (1999) the first principal component accounts for 93 percent of total variance and the first two together for 99.59 percent for the daily changes of US log yields. Chapman and Pearson (2001) document a figure of 88 percent for the first factor alone, and 99 percent for the first three ones with US data.

Since empirical results indicate that it is possible to describe the dynamics of the whole term structure using very few variables without losing much information, relatively simple models, that are tractable and economical, should explain the behavior of the term structure. The simplest choice is a one-factor model, where all randomness is described by a single variable. This means that the changes in rates of different maturities show perfect instantaneous correlation, but not necessarily that any changes in the term structure have to be parallel moves. Rebonato (1998) notes that all one-factor models can be seen as allocations of weights on different forward rates and on how they change in time. One-factor models will be discussed in the following part.

There is a difference whether PCA is applied to changes in yields or forward rates. If changes in forward rates are chosen the correlation between the forward rates shows only in the stochastic terms, but if changes in interest rates are chosen the drifts which are deterministic are also correlated. The correlation matrices are different, and when studying the correlation structure, one may prefer to use the changes in forward rates. However, the interpretation of the results for changes in yields is easier to apply directly to the yield curve.

2.8.2 Dimensionality

Dimensionality refers to the number of independent underlying variables. We saw above that PCA can justify models of very low dimensionality.

Brown and Schaefer (1994) studied the U.S. Treasury Bond data and show that the degree of decorrelation between n-month forward rates of different maturities is not constant, but decreases as the maturities are further apart from each other. Many empirical studies show similar results. There are two general results to consider. The first observation is that e.g. $\rho[f(0,3,4), f(0,5,6)]$ is higher than $\rho[f(0,3,4), f(0,9,10)]$. All of these forward rates are for 1-month loans, but the second pair is separated by a longer time gap. The second observation is that e.g. $\rho[f(0,2,3), f(0,4,5)]$ is higher than ρ $[f(0,7,8), f(0,9,10)]$. Each of these is a pair of 1-month forward rates, which are two periods apart. However, the second pair is farther away in the future and thus has a lower correlation.

Rebonato and Cooper (1995), and Rebonato (1998) note that the correlation structure of forward rates can generally be described as an exponentially decaying function

$$\rho(T_1, T_2) = e^{-\beta(T_2 - T_1)}, \qquad (2.61)$$

where ρ is the correlation between (instantaneous) forward rates maturing at time T_1 and T_2. Now the whole correlation structure is characterized by the variable β. If (2.61) holds, the loadings of the factors as a function of maturity follow a sine (or cosine) wave. This conforms to the usual interpretation of the principal components. Take for example the first principal component, factor a_{i1}. We have n forward rates so we have i=[1,n], and a_{i1} refers to the loading of the first principal component for the i:th maturity. If we draw a_{i1} as a function of maturity, the path will be a sine wave with frequency zero. This path is only slightly concave, almost horizontal, giving nearly constant factor loadings. If we draw the loading of the second principal component a_{i2} as a function of maturity, it also follows a sine wave, but with frequency ½. This path starts around the lowest point of the sine wave, a negative value, and ends up close to the peak, when it reaches the longest maturity included. This corresponds to the second factor, or slope. Then take the third principal component a_{i3}. This follows a sine wave with frequency 1 and represents the curvature.

This pattern will be repeated for the following principal components with the frequency increasing by ½ from the previous principal component. Here frequency 1 means that over the time span of all maturities exactly one full wave is completed, so if at time $t=0$ the forward rates are given by $f(0,0,T)$, then $T \in (0,u]$, and during the period u-0= u one wave is completed. Technically speaking, the eigenvectors are the basis functions to a Fourier series expansion where each eigenvector represents a higher frequency. The last eigenvector corresponds to the maximum frequency of the series. We see that the typical outcome of the empirical studies, i.e. the shift, slope, curvature etc, can be explained with only one parameter, the rate of decay β in the correlation function (2.61).

One of the benefits of the PCA approach is dimension reduction. This is achieved by using only the m first principal components to explain the total variance. Then only the first m eigenvectors are included, and the remaining ones are set to zero. The total variance explained by the model using only these m variables is then less than the actual observed variance, which includes 'all' variables. Often this loss of variance is small and can be ignored, but in some cases it is important to rescale the variances explained by the m included variables so that the total variance is correct. If the correlation between the first and subsequent forward rates is graphed as a function of maturity, the shape depends on m, the number of included factors. If many factors are included, the curve will be convex with negative slope. However, as the number of included factors decreases, and number of omitted variables increases, the shape of the curve changes and it will become concave first and then convex, leading to an elongated inverted S-shape (sigmoid-shape) (Rebonato 1998). The effect of the number of factors on the correlations between the first and subsequent forward rates is sketched in Figure 2.1. Rebonato (1998) emphasizes that any term structure model that uses very few principal components to describe the evolution of the curve will always show sigmoid-like correlation behavior. This happens because of dimension reduction, and it is not due to the specification of the model. The implication of this is that as the number of included factors decreases, the correlations between the first forward rates are biased upwards.

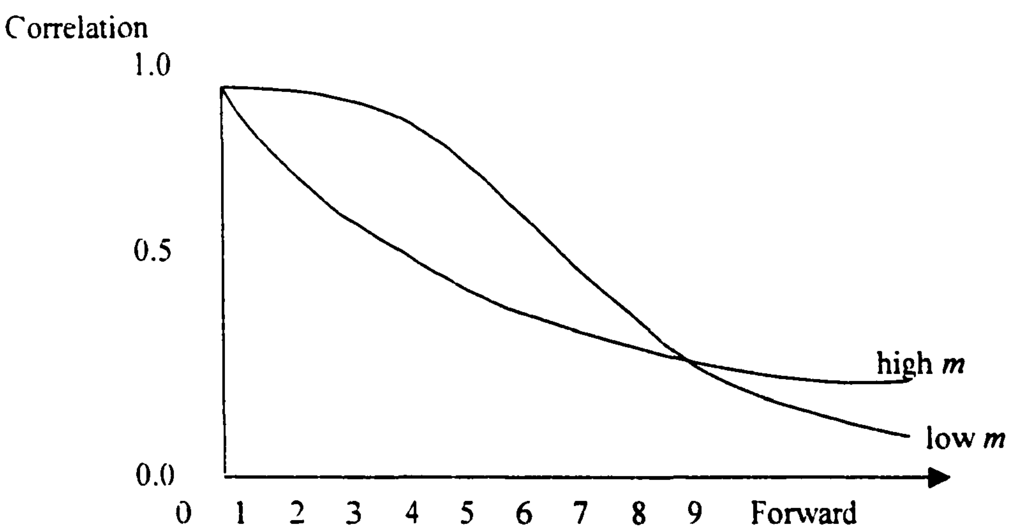

Figure 2.1. The correlation between the first and subsequent forward rates given by using m factors.

2.8.3 Non-Orthogonal Alternatives to Principal Component Analysis

Empirical studies based on principal component analysis (PCA) suggest that the dynamics of the whole term structure of interest rates is driven by very few variables. The main benefit of PCA is dimension reduction, and it is the best linear dimension reduction technique in the mean-square sense (see e.g. Jolliffe 1986). This reduction can be beneficial in simplifying computations, especially if subsequent stages are involved. Another benefit is that the data not contained in the n first components may be noise, and omitting the remaining factors may reduce noise. The third benefit is that in the case of the term structure the three principal components are meaningful and can be visualized as the shift, slope, and curvature.

With PCA it is possible to use either the interest rates (yields) or forward rates, or their changes, as the original data, assuming that the time series is stationary. The method then uses a linear transformation to obtain a diagonal matrix, the eigenvalue decomposition, from the original covariance matrix. Now the whole term structure can be described using few factors, which are orthogonal having zero covariance. The diagonal matrix has the variances of the principal components (eigenvalues) on the diagonal and the off-diagonal elements (covariances) are zero. PCA tries to only explain the diagonal elements and the overall variance, which is the sum of them. Some other methods, e.g. factor analysis, do not assume that the factors are orthogonal, and use covariance matrices where the off-diagonal elements are not zeroes.

Factor analysis is quite similar to PCA. However, given the original variables, PCA determines the factors uniquely (the first factor is chosen to be to the direction of maximum variance), but factor analysis does not. In factor analysis, we can describe the transformation in general terms

$$g = As + z, \tag{2.62}$$

where g is the column vector (nx1) of observed variables. A is an nxk matrix and s is a kx1column vector of k factors that cannot be directly observed. The last term z is a noise vector (nx1). Suppose that the covariance matrix of noise is known and given by

$$\Sigma = E\left(zz^T\right). \tag{2.63}$$

Then we can find the factors by performing PCA on the covariance matrix $\Omega - \Sigma$, where Ω is the covariance matrix of the original observations g.

In factor analysis many different representations and techniques can be used so we can always choose a representation that has the properties we desire.

PCA and factor analysis are both second-order methods in the sense that they do not use any information except what is given by the first two moments, i.e. the mean and the covariance matrix. With a normal distribution, the higher moments do not provide additional information, so these methods are sufficient. Higher-order methods use information that is not provided by the covariance matrix, and are not relevant if the distribution is normal.

Suppose that n random variables $x_1, x_2, x_3,\ldots x_n$ have a joint probability distribution $f(x_1, x_2, x_3, \ldots x_n)$. The random variables are independent if $f(x_1, x_2, x_3, \ldots x_n)$=

$f(x_1)f(x_2)$x...x$f(x_n)$. On the other hand, the random variables are uncorrelated if $E(y_i y_j) - E(y_i)E(y_j) = 0$ when $i \neq j$. Clearly, independence implies that the variables are uncorrelated. Furthermore, if the random variables have a joint normal distribution, zero-correlation and independence are equivalent concepts. However, if the joint probability distribution is not normal, variables can be uncorrelated but not independent.

Our formalization of the statistical approach to the term structure of interest rates is given in (2.55). The original random variables are df_1, df_2, ..., df_n. These forward rate changes have a joint probability distribution, but they are correlated. Because of this, the dimensions can be reduced. The new random variables are the factors. These are uncorrelated, but the question arises whether they are independent. A linear transformation of a normally distributed random variable is normally distributed. Then if the changes in forward rates have a joint probability distribution that is normal, which sometimes is the case, then the factors are normally distributed as well. From this follows, in theory, that the orthogonal factors should be independent in those cases. It is important to keep in mind that PCA does not use information of moments of higher order than two, and it does not explicitly consider autocorrelations or autocovariances. Additionally, it does not use information of the levels of variables and changes in them simultaneously.

Molgedey and Galic (2000) examine interest rates and changes in them and suggest that, even though with PCA the dimensions of the problem are reduced and the intuition is appealing, there is no reason to assume that the dynamics of the system can be represented by orthogonal factors. The principal components of interest rates are orthogonal, but they are not independent, showing relatively persistent autocorrelations and autocovariances. The obtained axes using PCA for interest rates and changes in interest rates are different.

Since in the end PCA uses only the variances of the new factors, the amount of information it uses is small and Molgedey and Galic find that better results can be obtained using a method called independent component analysis (ICA). They apply ICA using variance-covariance matrices of both, the interest rates and interest rate changes, simultaneously. In this case, instead of equation (2.60), we write two equations. The first equation uses the variance-covariance matrix of the interest rates, and the second one the variance-covariance matrix of changes in interest rates, giving

$$\Lambda = W^{-1}\Omega W^{T} \text{ and} \tag{2.64}$$

$$\hat{\Lambda} = W^{-1}\hat{\Omega} W^{T}, \tag{2.65}$$

where (2.64) gives the diagonal matrix of factor variances and (2.65) gives the diagonal matrix of the variances of the changes in factors. Now the factors are not orthogonal (so matrix W is not an orthogonal matrix) and we have to solve

$$\left(\Omega\hat{\Omega}^{-1}\right)W = W\left(\Lambda\hat{\Lambda}^{-1}\right). \tag{2.66}$$

Molgedey and Galic conduct Monte Carlo simulations comparing PCA and ICA and conclude that ICA can describe better the local behavior of the term structure of interest rate, and shows better performance in hedging. These results are not surprising since PCA uses less information to begin with and 'only' decorrelates the data, i.e. uses second-order statistics. ICA uses in addition to second-order statistics also higher order statistics. The resulting decomposition is a non-orthogonal, and depending on the chosen model, linear or nonlinear coordinate system where the new factors are as independent as possible. The factors obtained with ICA do not have a clear intuitive interpretation and, for example, the shift is not described by one factor.

ICA is generally used to extract information when the underlying probability distribution is not known. In the case of the term structure of interest rates, we generally make assumptions about the specific form of the driving function and the underlying probability distribution that governs the interest rates. For example, in a Gaussian model, interest rates are normally distributed by assumption, and PCA having orthogonal factors may be an appropriate choice from the theoretical viewpoint. ICA does not use a priori information.

It is important to understand that the specific factors that are chosen should depend on what the model is designed to do. For example, Dybvig (1988) explains that a two-factor model, with level and slope as the two factors, may be appropriate for examining the variability in bond prices. However, if one intends to price derivative securities whose values depend on bond prices, a better strategy may be to take the level as the first variable and the volatility of the level as the second one.

2.9 One-Factor Models

The results obtained from principal component analysis suggest that in most cases one factor can explain about 90 percent of the variance in the term structure. If this is the case, a model with one independent random variable is to a large degree sufficient for describing the dynamics of the term structure. Following from this, the simplest models used to describe the stochastic process of interest rates have only a single source of randomness. These models are called single-factor, or one-factor, interest rate models. Although this factor could be any variable, or at least any interest rate, generally the short-term interest rate is used. The rate that is chosen as the short rate could be the overnight rate, which would be in theory the shortest rate, but many times it is considered to be too sensitive to microstructure problems like intraday or intraweek volatility, and a 7-day or 30-day rate (e.g. 7-day Eurodollar or one-month LIBOR) is chosen as the short rate. Now the evolution of the whole term structure of interest rates depends only on one random variable, the short rate $r(t)$, and one deterministic variable, time to maturity $(T-t)$. Then the price of any interest rate derivate can be written as a function of the two variables, for example a discount bond price as $B(t,T,r(t))$. Typically, the short rate process is modeled using some sort of random walk, and there is no other source of randomness. The bond price equation becomes a parabolic partial differential equation.

When we price bonds, or any other interest rate derivatives, we have to use the risk-neutral stochastic process. Recall that the process for the risk-neutral short rate was calculated earlier in (2.33) and it was

$$dr = (u-\lambda w)dt + wdW, \qquad (2.67)$$

where dW is a Wiener process. We write $\mu(r,t)=(u-\lambda w)$ for the drift term. We can write (2.66) then, emphasizing that r, w and dW are functions, as

$$dr(t) = \mu(r(t),t) + w(t,r(t)dW(t) \qquad (2.68)$$

The features of this equation were discussed in detail earlier.

Here the only stochastic independent variable is the short interest rate. The drift and volatility terms have to be chosen so that the model has the desired properties. Various different specifications of this model have been presented by different authors.

2.9.1 Special Classes of Short Rate Models

Most of the traditional one-factor models belong to a class of 'affine models', using the expression of Duffie and Kan (1996). In these models μ and w are specified so that the generic form solution for the price of the zero-coupon bond is

$$B(r,t,T) = e^{a(t,T)+rb(t,T)} \qquad (2.69)$$

with the final condition

$$B(r,T,T)=1. \qquad (2.70)$$

Then the term structure of interest rates is an affine function of the short rate:

$$R(t,T) = \frac{a(t,T)}{T-t} - \frac{b(t,T)}{T-t} r(t,T). \qquad (2.71)$$

Note that this is under the risk-neutral process or, using the martingale expression, under probability measure Q.

A subclass of affine models is 'Gaussian models' which can be written as

$$dr(t) = \mu(r(t),t) + w(t,r(t)dW(t) = [\mu_1(t) + \mu_2(t)r(t)]dt + \sigma_1(t)dW(t) . \quad (2.72)$$

In a Gaussian model, the interest rate is normally distributed, while the bond prices are lognormally distributed (under Q).

Another class of affine models is lognormal models in which the logarithm of the short rate $r(t)$ is normally distributed. The problem with these models is that, to find the bond price we need to find the integral of interest rates over time as can be seen in (2.32), and the sum (integral) of normally distributed logarithmic variables is not normally distributed.

2.9.2 Some Desired Features of an Interest Rate Model

The models we have written above are very general since u or μ, and w were not given a specific form. In order to get a model that makes sense intuitively, we need to fulfill some general requirements. Especially with one-factor models, the specification of the exact form of the driving function has strong implications on the properties of the distribution of the interest rates. Some of these properties can be examined based on our experience about the behavior of interest rates.

- Interest rates, unlike stock prices do not have a long-term trend. This rules out the interest rates following a lognormal random walk.

- Interest rates cannot become negative (although it happened for a short period in Switzerland in the 1960's). Note that a normal distribution has tails that reach from minus infinity to plus infinity. If $r(t)$ is normally distributed, negative interest rates are not ruled out, although depending on the specific form of the function their probability could be extremely low.
- If $r(t)$ is very high, it has a tendency to come down rather than to go up even more. Respectively, if $r(t)$ is very low we expect it to go up rather than to go down lower. This is intuitively plausible and is also shown empirically to be the case. Many interest rate models incorporate some kind of mean reversion to show this kind of desired behavior.
- The lower bound should be unattainable in the sense that the short rate should move back towards the mean once it goes close to the boundary.
- Interest rates of different maturities are correlated. The degree of correlation falls faster as we move from the long end towards the short end.
- Volatilities for different maturities are different and shorter maturities have usually higher volatilities.
- Empirical experience shows that volatilities are not homoscedastic. If the interest rate for a given maturity changes, its volatility changes as well.

It is important to keep in mind whether we are modeling the *P*-process in the real world, or the *Q*-process in the risk-neutral world. As Rebonato (1998) mentions, when one moves from one process to the other, the drift term changes and there is no reason why the existence of mean reversion in one world should imply it in the other. In other words, if the real observed interest rates follow mean reversion, this does not necessarily indicate that mean reversion occurs in the risk-neutral world. Most of the models introduce mean reversion into the drift term. Alternatively, the volatility of the short rate can be made decaying. For fitting today's yield curve identical curves can be produced with both methods. In terms of the dynamics, they are very different because they forecast different future volatilities -- and thus interest rate levels as well.

We mentioned earlier that that the bond pricing equation needs a final condition and two boundary conditions. The final condition is given by (2.70) and the two boundary conditions are

$$B(r,t,T) \to 0 \qquad \text{as } r \to \infty \text{, and}$$

$$B(r,t,T) < k \qquad \text{for some upper bound } k>0. \tag{2.73}$$

It is impossible to perfectly capture all the wanted features with any one model, especially if it is a one-factor model. The exact emphasis on the particular characteristics depends on the use of the model. If we want to forecast bond prices, we may use a model that may do very poorly in pricing derivative securities based on bonds (see Rebonato 1998). Often a model that is not very good if judged by its assumptions can provide good forecasts.

2.9.3 Some Well-known One-Factor Models

Various different specific forms, or parameterizations, of interest rate models are based on the generic diffusion process given by (2.68). Some of the well-known models are presented in Table 2.2 (Aït-Sahalia 1996b, Sunderasan 2000).

Note that all the models in Table 2.2 are time-homogeneous, i.e. time-invariant, so the parameters are not a function of time. This means that the term structure is endogenously determined by the short rate. The models cannot be calibrated to especially fit exactly any observed term structure. Even if the model fits the term structure one day, it cannot be 'recalibrated' to fit any future term structure.

$\mu(r)$	$w(r)$		
1. $\beta(\alpha-r)$	σ	S	Vasicek 1977
2. $\beta(\alpha-r)$	$\sigma r^{1/2}$	S	CIR (1985), Brown and Dybvig (1986) Gibbons and Ramaswamy (1993)
3. $\beta(\alpha-r)$	σr	S	Courtadon (1982)
4. $\beta(\alpha-r)$	σr^{λ}	S	Chan et al. (1992)
5. $\beta(\alpha-r)$	$\sqrt{\sigma + \gamma r}$	S	Duffie and Kan (1996)
6. $\beta r(\alpha - \ln r)$	σr	S	Brennan and Schwartz (1979)
7. $\beta r(\alpha - \ln r)$	$\sigma r^{\delta/2}$	S	Marsh and Rosenfeld (1983)
8. $\alpha + \beta r + \gamma r^2$	$\alpha + \gamma r$	S	Constantinides (1992)
9. β	σ	NS	Merton (1973a)
10. 0	σr	NS	Dothan (1978)
11. 0	$\sigma r^{3/2}$	NS	CIR (1980)

Table 2.2. Some well-known one-factor models. S=stationary and NS=nonstationary.

2.9.4 An Example: The Vasicek (1977) Model

One-factor short rate models go back to Merton (1973b) and Vasicek (1977). Vasicek (1977) introduced mean reversion, so it is often used as the simplest specification for the short rate process. From Table 2.2 the Vasicek model can be written (for the P-process) as

$$dr(t) = \beta(\alpha - r(t))dt + \sigma\, dW(t), \tag{2.74}$$

where β, α and σ are positive constants. This is a real-world process that reverts to level α with reversion speed of β. The solution (stochastic integral) to (2.74) is

$$r(t) = \alpha + (r(s) - \alpha)e^{-\beta(t-s)} + \sigma \int_s^t e^{-\beta(t-s)} dW(t'), \quad (2.75)$$

where $s \le t$. The short rate has at time $t \le u$, conditional on current information set I_t, expected value, and variance

$$E_t(r_u \mid I_t) = \alpha + (r(t) - \alpha)e^{-\beta(u-t)} . \quad (2.76)$$

$$Var_t(r_u \mid I_t) = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta(u-t)}) . \quad (2.77)$$

Equation (2.74) gives the short rate process, which does not involve risk. If we want to find bond prices or the term structure of interest rates, we need to include the term premium. Technically speaking we have to switch from the real world P-process to the risk neutral Q-process. For this purpose we need to include the risk premium, or term premium, which Vasicek (1977) assumes to be constant λ. The bond pricing equation (2.31) becomes

$$\frac{\partial B}{\partial t} + \frac{1}{2}\sigma^2 \frac{\partial^2 B}{\partial r^2} + (\beta(\alpha - r(t)) - \lambda\sigma)\frac{\partial B}{\partial r} - rB = 0 \quad (2.78)$$

The PDE solution for this stochastic process will be of the general form shown in (2.69)

$$B(r,t,T) = e^{a(t,T)+br(t,T)}. \tag{2.79}$$

The final condition of the discount bond is $B(T,T)=1$, so we need $a(T,T)=0$ and $b(T,T)=0$. If one substitutes (2.79) into (2.78) and then differentiates the equation with respect to r to get a second equation, it is possible to use the two equations to find $\partial a/\partial t$ and $\partial b/\partial t$. When these are (stochastically) integrated the solution we obtain is

$$b = \frac{1}{\beta}(e^{-\beta(T-t)} - 1) \quad \text{and} \tag{2.80}$$

$$a = \frac{\sigma^2}{4\beta^3}\left(1 - e^{-2\beta(T-t)}\right) + \frac{1}{\beta}\left(\alpha - \frac{\lambda\sigma}{\beta} - \frac{\sigma^2}{\beta^2}\right)\left(1 - e^{-\beta(T-t)}\right) - \left(\alpha - \frac{\lambda\sigma}{\beta} - \frac{\sigma^2}{\beta^2}\right)(T-t). \tag{2.81}$$

Using this result, it is possible to write down the bond price dynamics, and furthermore derive the whole term structure of interest rates

$$R(t,T) = R(t,\infty) + \frac{1-e^{-\beta(T-t)}}{\beta(T-t)}\left(r(t) - R(t,\infty)\right) + \frac{\sigma^2}{4\beta^3(T-t)}\left((1 - e^{-\beta(T-t)}\right)^2, \tag{2.82}$$

where

$$R(t,\infty) = \lim(t \to \infty)[R(t,T)] = \alpha - \frac{\lambda\sigma}{\beta} - \frac{\sigma^2}{2\beta^2} \qquad (2.83)$$

is the yield to infinite maturity. The Vasicek model is able to produce concave, convex, and humped term structures.

2.9.5 Time-Non-Homogeneous Models

Since time-homogeneous models are not calibrated to fit the data, a different class of models (not included in Table 2.2) has emerged. These are time-non-homogeneous, or time-variant models, whose parameters are dependent on time and can be 'forced' to fit the observed term structure on a given date. These models do not have closed-form solutions and have to be solved numerically. The larger the number of parameters that are fitted, the better the fit ex ante, and the less predictable the model becomes. Hull and White (1995) suggest that only one variable in a Markov model should be time-variant. One problem with these models is that the process is not stationary (see Chapter 5 for a definition of stationarity).

The simplest of the fitted one-factor models is the Ho and Lee (1986) model. They choose the drift parameter as time dependent so $\mu=\mu(t)$. Then $\mu(t)$ is chosen to have a functional form that exactly fits the whole term structure on a given date. Most one-factor models can be changed to fitted models relatively easily. For example, Hull and White (1990) extend the Vasicek model to have time-dependent mean reversion so it can be calibrated. The Ho and Lee model is a special case of the Hull and White model. Black, Derman and Toy (1990) present a one-factor model similar to Ho and Lee (1986), but with mean reversion. This model is very popular among practitioners because it can be fitted easily, it prices swaps exactly, it is easy to understand, and implied volatilities can be used as input. However, its assumptions and implications are not clear. Numerous other fitted models, some of which are extensions to the ones mentioned, exist.

Hull and White (1993) present a general one-factor fitted model

$$dr(t) = \big(\alpha(t) - \beta(t)r(t)\big)dt + \sigma(t)r^{\eta}(t)dW(t) \tag{2.84}$$

with risk premium

$$\lambda(r,t) = \lambda r^{\gamma}. \tag{2.85}$$

The risk premium is determined exogenously. This model embeds some other models, e.g. Ho and Lee, as special cases. The parameters $\alpha(t)$, $\beta(t)$ and $\sigma(t)$ are all varying in time and the model can be calibrated exactly to match the term structure of interest rates on a given date.

Eight of the models In Table 2.2 were examined by Chan, Karolyi, Longstaff and Sanders (1992) using one-month Treasury bill yield data. They conclude that when volatility is modeled to depend on the level of interest rates, results are better. However, in general the one-factor models did not do very well. Aït-Sahalia (1996a) uses non-parametric estimation, and rejects all the one-factor models. Duffee (2002) concludes that affine models do poorly at forecasting future changes in US Treasury yields; a random walk outperforms most of them. On the other hand, for example Chapman and Pearson

(2001) argue that any dramatic changes in the short-term interest rates, i.e. bond yields or deposit rates, are captured by the first factor of a PCA. In this case, a one-factor model is appropriate.

3 The Economy of Turkey

Only during 1990's, Turkey started five stabilization programs to solve its persistent problems, high inflation rate, and problems in financing the budget and servicing the debt. None of the programs succeeded. After over twenty years of high inflation, inflation expectations are a major problem. Nominal interest rates have been high and volatile, and real interest rates have been volatile, and often high, as well. Since from 1989 capital flows have been free, the economic performance of the country is very dependent on three interrelated issues, inflation rate, interest rates, and exchange rates.

In the 1990's, consumer price index (CPI) inflation rose from 60% in 1991 to 120% in 1994, and then went back to around 70% at the end, averaging at 78% for the decade.

We start with a brief overview of the Turkish economy in 2000-2001. This limited period provides some understanding of the 'dynamics' of the Turkish economy and helps to understand the behavior of interest rates in Turkey. Since Turkey has had high inflation rate for over two decades, we also discuss briefly the relationship between inflation and interest rates. Finally, we provide some general information on the Turkish bond and repo markets.

3.1 The Turkish Economy in 2000-2001

Due to various reasons – mainly political problems, an increase in oil price, the economic slowdown in Russia and the disastrous earthquakes - the Turkish economy was in serious crisis at the end of 1999. IMF (The International Monetary Fund) responded with emergency assistance. In return, the Turkish government agreed to the 'December 9, 1999 Stabilization Program', with 'upfront fiscal adjustment, structural reforms and firm exchange rate commitment supported by consistent income policies' (IMF 1999). Transparency and accountability, which included 'the cost of credit subsidies of state banks' (IMF 1999), were also mentioned. The belief was that the markets would trust the program enough for it to work.

When the stabilization program became effective, overnight interest rates declined sharply at the beginning of January 2000, while the volatility increased and remained high for a period of about a month. The devaluation rate of Turkish Lira in U.S. dollars fell gradually (see e.g. Demiroz (2001)).

However, from the beginning, the government was very slow to implement structural reforms and privatization was not going ahead. At the same time, the Turkish Lira (TRL) was appreciating and the current account deficit was growing.

Turkish banks have had a long-lasting practice of 'carry-trade' - borrowing foreign currency on international markets, and then using that money for buying Turkish government bonds and offering loans in Turkey. The banks would also borrow short-term on the domestic markets. This strategy was profitable especially in 1998-1999 when real interest rate in Turkey was high. The new stabilization program included a crawling peg system for keeping the exchange rate under control, but this encouraged the 'carry-trade'. All the banks had massive imbalance between their foreign currency assets and liabilities.

During the year 2000, numerous banks had been taken over by the Banking Regulation and Supervision Agency (BRSA). Finally, on November 20, 2000 Demirbank was not able to acquire short-term loans to finance its portfolio of USD 4.5 to 7.5 billion government bonds on the money market. The bond market was paralyzed, foreign creditors did not want to give credit to Turkish banks, and foreign investors pulled out from the country. Foreign currency outflow was over USD 1 billion per day, and on December 1, 2000 the overnight (simple annual) interest rate was 873 percent. Although the markets stabilized, the Central Bank lost USD 6.5 billion of its reserves in only two weeks. With the aid of IMF and World Bank, the country recovered quickly and by the third week of January 2001, interbank rates were down to 35 percent, which was the level before the crisis. However, at the same time the bond yields remained at 50 to 60 percent, about 20 percent above the pre-crisis level.

Increasing political uncertainty in February 2000 caused again an increase in interest rates. The banking sector was struggling, structural reforms were not implemented, and privatization was delayed.

On February 19, 2001, the government had scheduled to raise USD 5 billion with bonds to cover part of a USD 7 billion debt that was to mature the following day. Due to a political crisis - an argument between the President and the Prime Minister - the interest rates for the 1-month TRL as well as the 4-month USD bonds turned out very high. The banks had their borrowing in short maturities and lending in long maturities, so they were hurt by the high short-term rates. Because of lack of confidence, foreign reserves were flowing out and the overnight interbank interest rate went up to 4000 percent; the highest overnight repo was even higher at 7500 percent. In two days, the Turkish lira depreciated over 36 percent against the USD. On February 22, 2001, the peg system on the exchange rate ended, and the Turkish Lira started floating.

On May 15, 2001, a new agreement was reached with the IMF. As a result, both the currency and the interest rates recovered slightly, but the bond yields were high, reflecting a real interest rate of 40 percent in July 2001. The government extended the average maturity of its debt from under 6 months to over 6 years by swapping them from TRL to other currencies. The yields on the currency bonds were between 14 and 15 percent. Since foreign investors did not want to hold TRL denominated bonds, the Turkish banks were buying them with a premium. The TRL bond yields were in mid-2001 about 30 percentage points above interbank rates.

The summer of 2001 was not politically stable, which led to many problems in the economy and an increase in interest rates. The total number of banks taken over by the BSRA rose to eighteen. This caused both the IMF and World Bank to show serious concern over the financial aid programs and to suspend aid payments to Turkey.

The events of September 11, 2001 had a negative impact on the already suffering Turkish economy, and the GDP, measured in USD, fell in 2001 by 25%, from USD 200 billion to USD 150 billion. The rehabilitation of the private and public banks had cost at least USD 35 billion, with indirect costs much more.

Due to the declining figures through 2000, at the beginning of 2001 consumer price inflation (CPI) was below 40 percent and the wholesale price inflation (WPI) at about 30 percent. During 2001, both figures rose steadily and at the end of the year they were at 68.5 and 88.6 percent respectively. During the last few months of the year, the Turkish Lira appreciated against the main currencies. This combined with the high inflation rate reduced competitiveness.

After the February 2001 crisis, the interbank overnight interest rate declined steadily and stabilized finally at about 60 percent, where it remained for the last three months of the year 2001. This was still 15 to 20 percent higher than the level in January 2001. The banking sector was facing the problem of high proportion, 20 percent, of non-performing loans.

3.2 Inflation and the Term Structure

Turkey has had very high inflation since late 1970's. Expected inflation and its volatility, as well as default risk and maturity of the loan, affect the interest rates. Extensive literature has been written about the relationship between inflation and the term structure of interest rates. Investors, as well as borrowers, are interested in the real interest rate, which consists of the nominal interest rate and the effect of the inflation. This is quite uncomplicated if the inflation rate is known with certainty. If investors are risk averse, they have to be compensated for uncertainty. Uncertainty about the level of inflation increases 'inflation risk', the risk that given the nominal interest rate until the maturity of a loan, an increase in inflation will reduce the real interest rate. The inflation risk is then part of the term premium of the term structure of interest rates.

The simplest theory about inflation and interest rates is called the 'Fisher effect' after Irvin Fisher (1907). The real interest rate is assumed independent of the expected inflation rate. Consequently, the nominal interest rate changes accordingly with the expected inflation rate holding the real interest rate constant. Fisher effects do not hold well in empirical studies in the short run, but they do better in the long run. This occurs especially when actual, rather than expected inflation is used. Expectations may not adjust immediately, but with a lag.

Another theory, the 'liquidity effect', predicts that expansionary monetary policy accelerates inflation while it at the same time increases the money supply, thus causing a fall in the short-term interest rates.

Mundell (1963) and Tobin (1965) made a point, later known as the 'Mundell-Tobin effect', that an increase in the inflation rate can lower the real interest rate when holders of wealth reduce their money holdings and consumption. Unlike the liquidity effect, the Mundell-Tobin effect is persistent.

The 'Peso problem' was mentioned in Chapter 2. It refers to the probability of a sudden devaluation, or inflation, causing the interest rates to have an excessive premium to compensate for this risk. In a country with high inflation risk, the 'Peso effect' could cause the real interest rates to be high. High inflation rate is generally associated with high volatility of nominal and real interest rates.

While the term structure of (nominal) interest rates gives the (nominal) spot rates for different maturities with certainty, the respective real interest rates are stochastic and their expected values can be estimated. The ex post real rates can be observed at the time of maturity, but they do not necessarily give accurate information about the expected real rates. A better alternative then is to model the inflation expectations explicitly. However, it is important to realize that modeling expected inflation is not required when a model of the stochastic process of the evolution of the nominal interest rates is written. If we look back at any form of the expectations hypothesis, we see that the information set I_t includes all information that is available at time t. All information about past inflation, and thus inflation expectations, are included in I_t. Using the PDE approach, inflation expectations are in the drift and volatility terms, and with the martingale approach, they affect the probability measure Q. In fact, some studies use the term structure to predict future inflation. Among others, Fama (1990) and Mishkin (1990) show that the term structure of interest rates can be used to predict future inflation, and as the maturity of the interest rate gets longer, this relationship becomes even stronger. Some other writers (e.g. Tzavalis and Wickens 1996) conclude that the term structure of interest rates does not contain information about the behavior of the future inflation.

Friedman (1977) and Holland (1984) argue that inflation risk increases with the level of inflation. This result has been confirmed with Turkish data (Berument and Güner 1997, Kaya and Neyapti 2001, and Kirmanoğlu 2001). Missale and Blanchard (1994) report that as the debt burden of a government increases, a positive dependence is found between maturity and interest rate. As a result, the government shortens the maturity of the debt in order to take advantage of the lower interest rate. Berument and Malatyalı (2001) show that with Turkish data an increase in inflation risk shortens the maturity of the debt issues and increases interest rates. Furthermore, an increase in the level of inflation, controlling for the risk, lowers real interest rates.

Keeping other factors constant, an increase in inflation level, inflation volatility or maturity increases the inflation risk and the inflation premium. Another important issue is liquidity. Turkish data shows that often bond yields are higher than the corresponding money market rates, which indicates an extra liquidity premium on the bonds.

Most interest rate models are developed using U.S. data. These models are do not necessarily describe the behavior of the Turkish term structure of interest rates very well.

The Turkish economy experiences a crisis every few years and has very high and volatile inflation.

3.3 The Turkish Bond Market

The only organized market for buying and selling bonds and for repo and reverse repo transactions in Turkey is the Bonds and Bills Market (BBM) that operates under the Istanbul Stock Exchange (ISE). In addition to transactions in Turkish Lira (TRL), Turkish bonds in major foreign currencies are included. We discuss only bonds issued in TRL as well as TRL interest rates.

In year 2000, the total value of the trade within the BBM was USD 4.5 billion (TRL 728.5 quadrillion), of which 77 percent were repo and reverse repo transactions. In addition, the off-exchange transactions totaled USD 8.2 billion in year 2000. The daily average value of repo and reverse repo transactions at the Bonds and Bills Market increased from under USD 0.5 billion in 1995 to USD 3.5 billion in 2000. During 2001 the value dropped gradually towards under USD 1.4 billion per day in the last three months, showing an average of USD 2.5 billion for the year 2001.

The average maturity of traded government bonds increased from only 90-150 days during the earlier half of 1990's to almost a year in 1999-2000. However, in 2001 the average maturity of Treasury borrowing went down to less than 150 days.

The average length of the repo and reverse repo contracts has varied between 1.5 and 3.6 days. Vast majority of these transactions are overnight repos (or reverse repos) with a maturity of one day.

Şahinbeyoğlu and Yalçin (2000) have estimated simple interest rates for different maturities using the BBM daily bond price data between July 1991 and March 1999. The results are shown in Table 3.1. If the results are transformed to continuously compounded yields, the yield curve is generally downward sloping, and even more so in the first half of 1990's.

Year	1-month	3-month	6-month	9-month	12-month
1991(July-Dec)	80.5	86.0	81.2	77.5	75.7
1992	87.2	92.5	88.6	80.5	75.5
1993	83.7	85.5	86.6	87.6	85.8
1994	176.8	175.8	171.2	123.8	123.6
1995	107.1	117.3	122.8	114.7	107.4
1996	118.3	126.5	131.5	120.7	113.4
1997	92.2	101.2	102.5	101.4	100.2
1998	73.1	104.0	110.5	107.7	101.4
1999 (Jan-Mar)	84.2	120.7	119.6	121.6	119.8

Table 3.1. Average interest rates implied by the bond prices during 1990's. (Source: Şahinbeyoğlu and Yalçin 2000.)

One should be very cautious when looking at yearly data, because the fluctuations during a year are very high. Especially, the volumes of trades for long maturities are often quite low compared to short maturities.

3.4 The Interbank Market

Another important interest rate in Turkey is the interbank rate. By far, the most interesting one is the overnight interbank rate. This rate is the benchmark rate used for Turkey in many international comparisons. for example by the OECD, while for other countries the three-month interbank rate is used. The overnight interbank rate is discussed and analyzed in Chapter 5.

4 A Principal Component Analysis of the Turkish Bond Indices

Istanbul Stock Exchange (ISE) Bonds and Bills Market has since the beginning of 1996 reported daily Government Debt Securities Indices (GDSI) for Turkish Lira denominated bond prices. Two types of indices are used, the price index and the performance index. The price index is quoted each trading day for a hypothetical bond that has day after day exactly the same number of days to maturity. The performance index, on the other hand, also takes into account the fact that on each successive day the time to maturity decreases by one day. We only examine the price index and do not discuss the performance index. From 1996, the indices have been quoted on every trading day for three maturities, 30 days (1 month), 91 days (3 months) and 182 days (6 months). The base date of the index is December 25 to 29, 1995. On this date, the hypothetical bond for each of the three maturities is given a value of 100. From the beginning of 2001 a family of new indices with four maturities, 182 days (6 months), 273 days (9 months), 365 days (12 months) and 456 days (15 months), was introduced. The base date for these new indices is January 2, 2001. Note that the 182-day bond is included in both indices.

We use the principal component analysis (PCA) to study whether the evolution of the term structure of interest rates is driven by a low number of factors. In order to do this, we first convert the changes in the price indices into changes in yields and forward rates. Our results show that the first factor explains between 65 and 71 percent of the variance, and the first two factors together between 90 and 95 percent. The usual interpretation of the principal factors as the level, slope, and curvature is shown when changes in yields are used. The low percentage of the first factor may partly be due to the method by which the GDSI index is calculated.

4.1 Calculating the Bond Price Index

The procedure used by the ISE to calculate the GDSI is explained in Kona (1996). The steps for finding the price index for the 'short' maturities of 30, 91 and 182 days re following:

1. Because typically during a trading day the same bonds are sold at different prices, an average price for each traded bond is calculated for each trading day. All discount bonds that mature in 182 days or less are included. On a given day, the (weighted) average price of bond i is

$$\overline{P}_i = \frac{(P_{i1}V_{i1} + P_{i2}V_{i2} + \ldots + P_{in}V_{in})}{V_{i1} + V_{i2} + \ldots + V_{in}}, \tag{4.1}$$

where $j=1 \ldots n$ is an index that refers to a realized trade with price P_{ij} and total face value (volume) V_{ij}. For the trading day, the number of trades of bond i is n.

2. The yield for the day is calculated from $\overline{P}_i$ for each bond i using

$$Y_i = \left(\frac{1}{\overline{P}_i} - 1\right) \times \frac{365}{D_i}, \tag{4.2}$$

where D_i is the number of days to maturity.

3. The zero-coupon yields Y_i for different maturities D_i are points in the space d x y – as a graph the horizontal axis is the number of days to maturity and the vertical axis is the yield. In order to fit the yield curve, i.e. connect the dots in a meaningful way, some function has to be chosen. The GDSI applies a simple linear regression using ordinary least squares (OLS), so we write

$$Y_i = \alpha + D_i\beta + u_i, \quad (4.3)$$

where i=1...M, and M is the number of different bonds (maturities).

4. All the data, pairs (D_i, Y_i) are inserted into equation (4.3) and the OLS parameters $\hat{\alpha}$ and $\hat{\beta}$ are estimated.
5. The 'predicted yield' for each maturity $\hat{y}_{d^*}$, with d^*={30, 91, 182} is then calculated from the estimated regression line.
6. The estimated yield is used to calculate the price of the hypothetical bond using the bond price formula. This is done for each maturity d^*.
7. To get the price index for the particular bond on a given day, the price is divided by the base date price (on December 25-29, 1995) and multiplied by 100. The price index is calculated for each maturity d^*.

The result of the steps 1 to 7 is the bond price index for each maturity d^*={30, 91, 182}, which is published for each trading day by ISE.

The newer price index is calculated exactly the same way, but including only bonds with maturities from 182 days to 456 days. The base date is January 2, 2001 and $d^*=\{182, 273, 365, 456\}$.

4.2 Calculating Forward Rates from the Price Indices

At time t the price of a zero-coupon bond that matures at time T_1 is given by

$$B(t,T_1) = e^{-R(t,T_1)(T_1-t)}. \tag{4.4}$$

This bond pays one unit at time T_1 and nothing at any other time. Its lifetime until maturity is T_1-t. We can solve equation (4.4) to get the discount bond yield

$$R(t,T_1) = \frac{-\ln B(t,T_1)}{T_1 - t}. \tag{4.5}$$

This is at time t the yield of a (zero-coupon) bond with time T_1-t to maturity. Identically, this is the continuously compounded interest rate for this maturity. Suppose that, also at time t, we have another discount bond with maturity at T_2, which is later than T_1. The lifetime of this bond is longer with T_2-t > T_1-t. The price of this bond is given by

$$B(t,T_2) = e^{-R(t,T_2)(T_2-t)} = e^{-R(t,T_1)(T_1-t)}e^{-f(t,T_1,T_2)(T_2-T_1)}. \tag{4.6}$$

where $f(t,T_1,T_2)$ is at time t the forward rate for the period from T_1 to T_2. We can insert (4.4) into (4.6) and solve for the forward rate

$$f(t,T_1,T_2) = \frac{-\ln[B(t,T_2)/B(t,T_1)]}{(T_2-T_1)}. \tag{4.7}$$

An infinitesimal change in the forward rate df can now be approximated by

$$\Delta f(t,T_1,T_2) = f(t,T_1,T_2) - f(t-\Delta,T_1-\Delta,T_2-\Delta) =$$

$$\frac{-\ln[B(t,T_2)/B(t,T_1)]}{(T_2-T_1)} + \frac{\ln[B(t-\Delta,T_2-\Delta)/B(t-\Delta,T_1-\Delta)]}{(T_2-T_1)}. \tag{4.8}$$

Respectively, a small change in the spot rate maturing at T_1 (equation (4.4)) will be a special case of (4.8):

$$\Delta R(t,T_1) = R(t,T_1) - R(t-\Delta,T_1-\Delta) = \tag{4.9}$$

$$\Delta f(t,t,T_1) = f(t,t,T_1) - f(t-\Delta,t-\Delta,T-\Delta_1) =$$

$$\frac{-\ln B(t,T_1)}{T_1-t} + \frac{\ln B(t-\Delta,T_1-\Delta)}{T_1-t}.$$

The method can be extended to the following forward rates. However, the ISE data does not show any bond prices, but only price indices. In the indexed data all four different maturities are given a value of 100 on the base date, so we cannot know the relative prices of two bonds of different maturities. Instead of trying to first calculate the forward rates and then the evolution of the forward rates, another approach has to be taken.

We begin with the price index. At time t look at two bonds, one is a bond of T_1-t to maturity and the other of T_2-t to maturity. The price index starts at time a, which is before time t. The two indices can be written as

$$B(t,T_1) = I_1(t) \times B(a,T_1 - (t-a)) \quad \text{and} \tag{4.10}$$

$$B(t,T_2) = I_2(t) \times B(a,T_2 - (t-a)).$$

For simplicity, we will write $B(a,(T_i - (t-a))) = B_i(a)$. We can write the change in the yield, given by equation (4.9), using the index instead of the bond price

$$\Delta R(t,T_1) = \frac{-\ln[I_1(t) \times B_1(a)]}{T_1 - t} - \frac{-\ln[I_1(t-\Delta) \times B_1(a)}{T_1 - t}$$

$$= \frac{-\ln[I_1(t) - \ln[B_1(a)]}{T_1 - t} + \frac{+\ln[I_1(t-\Delta)] + \ln[B_1(a)]}{T_1 - t},$$

which simplifies to

$$\Delta R(t,T_1) = \frac{-\ln[I_1(t)/I_1(t-\Delta)]}{T_1 - t}. \tag{4.11}$$

A vector composed of the changes in yields (interest rates) of different maturities can be written as

$$h = \begin{bmatrix} \Delta R_1 \\ \Delta R_2 \\ \cdots \\ \Delta R_n \end{bmatrix}. \tag{4.12}$$

We can use the same method to find the forward rates. First, substitute (4.10) into (4.7) so

$$\Delta f(t, T_1, T_2) = f(t, T_1, T_2) - f(t-\Delta, T_1 - \Delta, T_2 - \Delta) =$$

$$\frac{-\ln[I_2(t) \times B_2(a) / I_1(t) \times B_1(a)]}{T_2 - T_1}$$

$$+ \frac{\ln[I_2(t-\Delta) \times B_2(a) / I_1(t-\Delta) \times B_1(a)]}{T_2 - T_1}$$

$$= \frac{-\ln[I_2(t)] - \ln[B_2(a)] + \ln[I_1(t)] + \ln[B_1(a)]}{T_2 - T_1}$$

$$+ \frac{\ln[I_2(t-\Delta)] + \ln[B_2(a)] - \ln[I_1(t-\Delta)] - \ln[B_1(a)]}{T_2 - T_1}$$

$$= \frac{-\ln[I_2(t)] + \ln[I_1(t)]}{T_2 - T_1} + \frac{\ln[I_2(t-\Delta)] - \ln[I_1(t-\Delta)]}{T_2 - T_1}$$

$$= \frac{\ln[I_1(t) / I_1(t-\Delta)] - \ln[I_2(t) / I_2(t-\Delta)}{T_2 - T_1}. \tag{4.13}$$

The forward rates can be written in vector form

$$g = \begin{bmatrix} \Delta f_1 \\ \Delta f_2 \\ \dots \\ \Delta f_n \end{bmatrix}. \tag{4.14}$$

4.3 Estimation

In chapter 2 we explained how changes in forward rates (or other variables) can be used with PCA to find the principal components driving the term structure of interest rates. Alternatively, forward rate levels, spot rate levels, or changes in spot rates can be used. We will first use the changes in forward rates. Write vector g as

$$g = \begin{bmatrix} df_1 \\ df_2 \\ \dots \\ df_n \end{bmatrix} = \begin{bmatrix} \mu_1 dt + \sigma_1 dz_1 \\ \mu_2 dt + \sigma_2 dz_2 \\ \dots \\ \mu_n dt + \sigma_n dz_n \end{bmatrix}. \tag{4.15}$$

where the change in each forward rate follows a Wiener process and the changes in different rates are correlated. For each trading day, we get a new vector g. The PCA finds the orthogonal axis so that the first principal component explains as much of the variance as possible. The method was explained in chapter 2.

The data is the ISE Bonds and Bills Market indices (GDSI) for 30, 91, 182, 273, 365, and 456 days. The data used for the three shorter maturities is from the old index and for the three longer ones from the new index.

The observations are reported for each trading day. However, since we use at time t for the calculations observations at both time t and t-1, we need to pay attention to the weekends (and other holidays). In order to avoid any weekend effects, the first trading day after a non-trading day (e.g. a Monday) is always excluded.

We perform PCA first with time series from January 2, 2001 to September 28, 2001. (The bond indices for the longer maturities start from January 2, 2001). This time series contains, excluding any day that is the first day after a holiday, 146 observations (n=146). To see the potential effect of the February 2001 crisis, we repeat the analysis for a shorter period, from March 12, 2001 to September 28, 2001, with 111 observations (n=111). In Chapter 3, the proceedings of the Turkish economy were discussed briefly. In the beginning of 2001, the country was recovering from the November 2000 crisis and stabilizing. Then came the February 19, 2001 crisis after which the economy recovered largely, but declined steadily towards the end of the year.

These price indices are transformed into changes in forward rates as described above in (4.10) to (4.14). The used forward rate matrix is

$$\begin{bmatrix} R(t,t,+91) \\ f(t,t+91,t+182) \\ f(t,t+182,t+273 \\ f(t,t+273,t+365) \\ f(t,t+365,t+456) \end{bmatrix}.$$

Note that the first rate is the three-month spot rate, which is a special case of a forward rate. All the rates are for three months, which is 91 or 92 days.

4.4 Empirical Results

Standard PCA is applied to the changes in forward rates. The correlation matrix is shown in table 4.1. It is easy to see from Table 4.1 that the first forward rate, the spot rate, shows lower cross-correlations than the other rates. The p-values for these indicate that they are not statistically as significant as the other values. For the other cross-correlations, the typical pattern can be seen in the table. All the changes in forward rates show positive correlation with each other. The correlation is higher between rates that are close to each other than between ones that are far apart.

The magnitude of the correlation coefficients between the original variables is important for the reliability of the results. Generally, it is desirable that the correlations between all the variables exceed 0.3 if the PCA is to be used. The values for the first rate are lower than this.

4.4.1 Tests for Statistical Significance

A commonly used test to examine the fit of a PCA model is Bartlett's Test for Sphericity. It finds the overall significance of the correlation matrix by testing the null hypothesis that the correlation matrix is an identity matrix. If the correlation matrix is an identity matrix, each variable is perfectly correlated (by assumption) with itself but not with other variables. In order to be able to use PCA, we need to be able to reject the null hypothesis and have the original variables sufficiently correlated. In this sample. Bartlett's test gave a result - chi-square 1915 with 10 degrees of freedom - which is statistically significant at p<0.0005. The conclusion is that the data set shows sufficient overall cross-correlations for PCA to be applied.

Rate	*R(0,91)*	*F(91,182)*	*f(182,273)*	*f(273,365)*	*f(365,456)*
R(0,91)	1.000	0.071 (.196)	0.332	0.244 (.002)	0.178 (.016)
f(91,182)	0.071 (.196)	1.000	0.673	0.641	0.608
f(182,273)	0.332	0.673	1.000	0.991	0.972
f(273,365)	0.244 (.002)	0.641	0.991	1.000	0.995
f(365,456)	0.178 (.016)	0.608	0.972	0.995	1.000

Table 4.1. The correlation matrix of the forward rates. Non-zero p-values (p>0.005) are shown in parentheses.

Another common test used with PCA to measure if there is adequate correlation between the variables is the Kaiser-Meyer-Olkin (KMO) Measure of Sampling Adequacy. It compares the magnitude of observed correlation coefficients to partial correlation coefficients. Partial correlation is the correlation between variables while controlling for changes in any third variable. We look, for example, at *f*(182,273) and *f*(365,456) controlling for any changes in the other variables, and calculate the controlled 'partial correlation'. The KMO's idea is then to compare this 'partial correlation' with the original, uncontrolled, 'observed correlation'. If there is no difference, the conclusion is that the other variables have no effect. At the other extreme, the partial correlation can be zero, and any correlation between the variables is spurious due to e.g. a common cause or an intervening variable. The idea is that the partial correlations should not be very large if one is to expect distinct factors to emerge from factor analysis. In the KMO test for the whole sample, the sum of squared correlations is divided by the sum of squared partial correlation, so the result has to be between zero and one. Usually a value of 0.5 or more is considered to be adequate. We find a value of 0.533 for the KMO test, which is sufficient. This result implies that the correlations are relatively compact and meaningful factors can be obtained.

4.4.2 Contributions of Different Factors

The contributions of the principal components to the overall explained variance are shown in Table 4.2. One can see that the first principal component explains 70.838 percent of the variance in the changes of forward rates and the first two components together 90.139percent. These figures are slightly lower than results from other empirical studies (see Chapter 2). The eigenvalue of each component is the total variance explained by that component. The 'weights' of the first four principal components are shown in Table 4.3. The weights given are the changes in the variables, i.e. the forward rates of different maturities, that are needed to produce the amount of variance that is explained by each principal component. If we take the values of the column titled 'First Principal Component' and add up their squares, we get exactly this first eigenvalue (allowing for rounding errors). The shown weights are not scaled, so that the relative importance of each factor can be seen more directly.

Principal Component	Eigenvalue	Explained Variance (%)	Total Variance Explained Cumulatively (%)
1	3.542	70.838	70.838
2	0.732	19.300	90.139
3	0.483	9.662	99.800
4	0.009892	0.198	99.998
5	0.00008912	0.001782	100.000

Table 4.2. Percentage of total variance explained by the principal components (n=146).

Unscaled Loadings	First Principal Component	Second Principal Component	Third Principal Component	Fourth Principal Component
1	0.313	0.939	0.144	0.01257
2	0.750	-0.267	0.605	0.009873
3	0.992	0.03998	-0.09876	-0.07427
4	0.983	-0.04069	-0.178	-0.0003597
5	0.965	-0.09624	-0.235	0.06419

Table 4.3. The weights of the variables needed to explain the total variance in forward rate changes (n=146).

Note that we used the forward rates, so the principal components do not apply directly to the yield curve. However, we can make the following observations.

i. The first principal component has approximately equal effect on the longer rates, but is weaker for the first one and slightly weaker for the second one.

ii. The second principal has a strong effect at the short end.

iii. The third principal component has its strongest effect on the second forward rate.

iv. The fourth principal component has a very small effect.

Each column in Table 4.3 is an eigenvector, whose Euclidean length is the eigenvalue of the respective component. By scaling the eigenvectors so that each one has a length of one, we get what were earlier in Chapter 2 called loadings. Table 4.3 shows that the short end of the term structure behaves slightly differently. The second component is pronounced with the shortest forward rate. From the percentages in Table 4.2 one would expect that two factors would explain quite well the behavior of the forward rates, and even one-factor models could be considered.

Many studies apply PCA directly to the changes in yields. We did this in order to find results that can be more readily applied to the yield curve. The changes in yields are calculated using (4.11) and (4.12). All the yields provided by the data were used: 30, 91, 182, 273, 365, and 456 days. The correlations between these rates are shown in Table 4.4. It is easily seen, that the shortest rate shows low correlation with the longer rates, but many of these have p-values, which indicate that they are not statistically significant. Now the KMO Measure has a value of 0.436, which is lower than desired, but Bartlett's test is statistically significant at the $p<0.0005$ level.

The eigenvalues and the percentages of total variance explained by the principal components are shown in Table 4.5. The first principal component explains 68.3 percent of the total variance, which is a low figure. However. the first two and three components explain 94.9 percent and 99.8 percent respectively. This indicates that probably two factors is a good choice for most purposes. The unscaled loadings of the factors are shown in Table 4.6.

Rate	$\Delta R(30)$	$\Delta R(91)$	$\Delta R(182)$	$\Delta R(273)$	$\Delta R(365)$	$\Delta R(456)$
$\Delta R(30)$	1.000	0.904	0.225 (0.003)	0.192 (0.010)	0.108 (0.096)	0.053 (0.264)
$\Delta R(91)$	0.904	1.000	0.620	0.528	0.442	0.379
$\Delta R(182)$	0.225 (0.003)	0.620	1.000	0.839	0.790	0.746
$\Delta R(273)$	0.192 (0.010)	0.528	0.839	1.000	0.989	0.969
$\Delta R(365)$	0.108 (0.096)	0.442	0.790	0.989	1.000	0.995
$\Delta R(456)$	0.053 (0.264)	0.379	0.746	0.969	0.995	1.000

Table 4.4. The correlation matrix of yield changes. Non-zero p-values ($p>0.0005$) are shown in parentheses ($n=146$).

Principal Component	Eigenvalue	Explained Variance (%)	Total Variance Explained Cumulatively (%)
1	4.098	68.295	68.295
2	1.595	26.582	94.877
3	0.293	4.891	99.768
4	0.01355	0.226	99.994
5	0.0003241	0.005402	100.000
6	0.00001766	0.0002944	100.000

Table 4.5. Percentage of total variance explained by the principal components using yield changes ($n=146$).

The factor loadings in Table 4.6 can be clearly interpreted in the traditional way. The shift shown by the first factor is stronger at the long end of the yield curve. The second factor shows the change in the slope and the third factor the change in curvature. Using the changes in yields gives low percentage for the first principal component as well as some low correlations, but otherwise the results are very typical.

Unscaled Loadings	First Principal Component	Second Principal Component	Third Principal Component
1	0.378	0.912	0.160
2	0.700	0.711	-0.05732
3	0.892	-0.04317	-0.451
4	0.971	-0.206	-0.07723
5	0.944	-0.297	0.143
6	0.914	-0.354	0.187

Table 4.6. The weights of the variables needed to explain the total variance in the yield changes (n=146).

4.5 Did the Dynamics Change in February 2001?

An interesting question is whether the results are approximately the same if some other time period is studied. The data for the long maturities of the ISE bond index is only available from the beginning of 2001. In order to see if the February 2001 crisis would influence our conclusions, we carry out the PCA done above using a shorter time series that extends from March 12, 2001 to September 28, 2001. Because of the lagged variables, again days after holidays are excluded, so we have 111 observations (n=111). Note that no trading was carried out between from March 3 to March 11, so the last trading day of the previous dataset was March 2, 2001. This second data set is a period of steady recovery and the turbulence of late February is over.

The results are different, but not dramatically. This is what one would expect because the difference in the data sets is small. The problem is that the data set is rather small in the first place, and now it is even smaller. The correlations using the changes in yields are shown in Table 4.7. We can see that the correlation between the changes in the 30-day rate and the 182-rate is almost zero. This is quite unusual, but the changes of the 30-day rate show very low correlation with the changes in the three longest rates as well. Note that the p-values for the low correlations may justify rejecting these findings as not being statistically significant.

	$\Delta R(30)$	$\Delta R(91)$	$\Delta R(182)$	$\Delta R(273)$	$\Delta R(365)$	$\Delta R(456)$
$\Delta R(30)$	1.000	0.946	0.090 (0.175)	0.287 (0.001)	0.233 (0.007)	0.197 (0.019)
$\Delta R(91)$	0.946	1.000	0.407	0.492	0.436	0.394
$\Delta R(182)$	0.090 (0.175)	0.407	1.000	0.717	0.687	0.660
$\Delta R(273)$	0.287 (0.001)	0.492	0.717	1.000	0.988	0.968
$\Delta R(365)$	0.233 (0.007)	0.436	0.687	0.988	1.000	0.995
$\Delta R(456)$	0.197 (0.019)	0.394	0.660	0.968	0.995	1.000

Table 4.7. The correlation matrix of yield changes. Non-zero p-values ($p>0.0005$) are shown in parentheses ($n=111$).

The contributions of the principal components to the overall explained variance are shown in Table 4.8. We see that the first principal component explains only 66.0 percent of the total variance, and the first two and three together explain 91.9 and 99.6 percent respectively. Hence, the percentages of the first principal components have dropped slightly. This may be expected, since from February 22, the exchange rate was floated and changes in yields may be less pronounced.

The KMO Measure is a borderline value of 0.464, but Bartlett's Test shows an acceptable value with p<0.0005. The loadings of the factors, and thus their interpretations remain practically the same as earlier with the longer time series.

Principal Component	Eigenvalue	Explained Variance (%)	Total Variance Explained Cumulatively (%)
1	3.961	66.010	66.010
2	1.554	25.899	91.909
3	0.462	7.692	99.601
4	0.02368	0.395	99.996
5	0.0002439	0.004065	100.000
6	0.00001384	0.0002307	100.000

Table 4.8. Percentage of total variance explained by the principal components using yield changes (n=111).

We also repeated the analysis using the changes in the same forward rates, which were used earlier. These results (not shown here) were quite similar, with one, two and three principal components explaining 64.7, 91.7, and 99.8 percent of the total variance respectively.

4.6 Conclusions

In this chapter, we have used the principal component analysis (PCA) on the Government Debt Security (GDSI) index published by the Istanbul Stock Exchange (ISE) Bills and Bonds Market. The indices are published separately for the maturities of 30, 91, 182, 273, and 365 days. The indices are obtained using a linear regression. One regression is done including bonds with maturities of 182 days and less, and another one with bonds with maturities of 182 days and more. From the first regression, the predicted values of the yields for maturities of 30, 91 and 182 days are found, and the corresponding price indices are calculated. The price indices for maturities of 182, 273 and 365 days are found exactly the same way using the second regression.

We have shown how the price indices can be used to calculate changes in yields and forward rates, which can then be used to analyze the term structure of interest rates applying PCA. We found that 64.7 to 70.8 percent of the total variance can be explained by the first principal component, 90.1 to 94.9 percent by the first two, and 99.6 percent or more by the first three principal components. This basically confirms the general fact known since Litterman and Scheinkman (1991), that three principal components, often referred to as the 'level', 'slope', and 'curvature', are sufficient to explain the variability in bond yields. Our results are similar to other reported findings, but with lower percentage of the variance explained by the first principal component. This indicates that a model with a low number of factors, i.e. low dimensionality, might be able to explain the dynamics of the term structure. However, in light of our results, two sources of randomness rather than one should be used. A closer look shows that the first principal component, which corresponds to the overall level of interest rates, although it explains quite well the changes in the yield curve in general, is not as pronounced at the short end (for the short maturities).

We have also examined whether the February 2001 crisis makes a difference. This was done carrying out the analysis first for a longer time series from January 2 to September 28, 2001, and then repeating it for a shorter time series from March 12 to September 28, 2001. The difference between these results is slight. In the shorter series the first principal component explains slightly less of the variations than in the longer one.

It is important to emphasize that our data consists of indices that are estimated from the realized bond prices using a linear regression. Any changes in a regression line are totally explained by either changes of the intercept or the slope. Here the yield curve we use consists of two straight regression lines. Obviously this simplification does affect the results, but exactly how is difficult to know. An interesting question is, whether this issue is important when data obtained from any fitted yield curve is used. On the other hand using original daily bond prices might reflect microstructure problems, like intra-day volatility and the bid-ask bounce. Besides, the number of different maturities of bonds in the Turkish market is low and, on a given day, relatively few bonds might be actively traded. Since maturities are not far apart, generally, choosing a linear regression to fit the yield curve may be well founded. Kona (1996) states that these indices represent the actual prices very well in general, but points out that since regression is 'averaging' the method has a tendency of underestimating the shifts of the yield curve. The implication to our results is that the percentage found for the first principal component may be an underestimate.

Another important finding is that the results are sensitive to the exact choice of the variables in the model. We saw that using changes in yields leads to different results from using changes in forward rates. This happens because these variables lead to different axes in decomposition. From the theoretical viewpoint, forward rates have the advantage that they are not correlated with each other in a deterministic way as the yields are. If one wants to study the decorrelation of the rates of different maturities, this may be an advantage. On the other hand using changes in yields leads to a straightforward interpretation of results on the yield curve, and the rates being more correlated may be even beneficial with PCA.

A third issue is the difference between the changes at the short and long ends, as shown by the second principal component. As conditions change, the term premium probably changes. Although we have not examined the term premium here, the changes in it would be important to understand.

While interpreting these results, it is important to keep in mind the nature of PCA, which was discussed briefly in Chapter 2. Like other methods in the multiple general linear hypothesis (MLGH) family, PCA makes many of the same assumptions as a multiple regression. If the relationships between the variables are nonlinear, using PCA does not give good results. One should also keep in mind that the tests of significance rely on normality of variables.

5 The Distribution of the Turkish Short Rate

In modeling the dynamics of the term structure of interest rates, it is essential to understand the behavior of the short rate. In general, the time series of the short rate tends to show mean reversion, persistence and heteroscedasticity. Moreover, the probability distribution, as is the case with other financial returns, usually has heavy tails and may be very skewed.

With interest rates, and financial returns in general, the observed mean reversion does not usually occur at a constant rate, but can depend on time or the level of interest rates. Typically, interest rates are persistent, staying at relatively high or low levels for prolonged periods rather than returning to the middle range quickly. During these times mean reversion may be very slow or non-existent. Interest rates are usually heteroscedastic and exhibit volatility clustering, which means that persistent periods of high volatility are followed by prolonged periods of low volatility.

Many different methods have been used to deal with these characteristics of the short rate. Essentially the issue breaks down into two important questions. First, what kind of a model is suitable, and second, how should the model be estimated. The traditional one-factor models explained in Chapter 2 have two inherent problems - changes in interest rates are perfectly correlated, and the shapes of the yield curves are restricted to certain alternatives. Besides, these models do not fit empirical data well (see Chapter 2). In Chapter 4, we applied the Principal Component Analysis to the Turkish Government Debt Securities Indices (GDSI), and the results suggest that a traditional single-factor model is not sufficient for modeling the dynamics of this yield curve.

We start this chapter with a brief discussion of different approaches to estimating a short rate model. Then we examine the Turkish overnight interest rate, and find it leptokurtic and skewed. The time series is clearly (weakly) stationary and autocorrelated, following an AR(1) (or AR(2)) process. The results are consistent with mean reversion. We also find volatility clustering. In the end of the chapter, we see that the Turkish data exhibits structural breaks, which are consistent with 'switching' between regimes. We conclude that the dynamic process driving the short rate in Turkey may follow a Markov switching model.

5.1 Nonparametric Estimation

The traditional dynamic interest rate models start by modeling the short-term interest rate as a Wiener process. Then the whole term structure of interest rates, also called the yield curve, is derived from that process. In the simple versions, the dynamics of the yield curve are assumed to be driven by a single underlying factor, the short rate, so these models are called one-factor models, or single-factor, models. These models are not able to explain many characteristics of the data, e.g. nonlinear mean reversion, kurtosis, and skewness, well. The models that use two or more factors as inputs are able to explain more complicated shapes of yield curves. Various different models have been presented with different drift and volatility structures.

An alternative approach is the Heath, Jarrow and Morton (HJM, 1992) framework that was a major breakthrough in modeling the dynamics of the forward rate curve. Starting with the term structure on a given day, the no-arbitrage condition is applied and the drift and volatility are chosen so that a good fit with the original term structure is obtained. The HJM model gives the evolution of all the forward rates, and thus the whole term structure at all times. The fit can be as good as desired, but this is achieved by having time-dependent parameters (so the model is time non-homogeneous). HJM is an approach rather than a specific theory, and the other models can be written within the HJM framework.

The different interest rate models try to exhibit the essential characteristics of the observed behavior of interest rates. Understanding the process of the short rate is the most profound issue when modeling interest rates. The two components that need to be studied are the drift and volatility. The HJM (as well as the Hull and White 1990) method is widely used and successful in pricing complex derivatives, but it takes the bond prices as given, and then finds the risk-neutral drift to fit the prices. The relationship between the drift and volatility is an outcome of the 'fitting' and does not require attention. Any other fitted models share this problem (Jones 2001).

In the traditional spot-rate models the drift and volatility structures are exogenous. Then, given the empirical data, the values of the parameters of the model are estimated. Many of these models are designed so that they have nice mathematical properties, e.g. time-homogeneity or closed form solutions.

Another approach, which has grown in popularity during the recent years, is non-parametric estimation. The idea is to avoid 'a priori' restrictions that may distort the analysis. In the simplest case, rather than starting with a given parameterization or trying to fit a curve, the researcher tries to find the drift and diffusion functions of the short rate so that they are consistent with the observed data. Aït-Sahalia (1996a and 1996b) estimate the drift parametrically and the volatility nonparametrically, while Stanton's (1997) model is totally nonparametric. Although these authors use different data, both find essentially zero mean reversion for a wide range of interest rates. In Stanton (1997), mean reversion is not observed when interest rates are below 15 percent, but very strong mean reversion is found for higher rates. Aït-Sahalia (1996b) reports zero mean reversion except for very high or very low interest rates. A similar finding is in another non-parametric study by Conley, Hansen, Luttmer and Scheinkman (CHLK 1997). Jiang and Knight (1997) find this non-constant mean reversion for Canadian data. This non-constant, or non-linear, mean reversion leads to the rejection of most popular models, which assume linear drift.

Pritsker (1998) mentions the following four problems that arise with non-parametric estimation.

1. They need very large amounts of data. For example, hundred years of daily observations may not be enough.
2. Most results are based on asymptotic properties and the small (i.e. finite) sample properties are not well known.
3. Their behavior with persistency is not well known. (Interest rates are usually persistent; they spend extended time periods at high values and respectively at low values rather than return to the middle range quickly.)
4. The methods require a selection of bandwidth, but the results of bandwidth selection are not well known for finite non-iid data.

The first three points are valid even with parametric methods if the time series has a near unit root, in other words is non-stationary (Ball and Torous 1996). Chapman and

Pearson (2000) report that in both Aït-Sahalia (1996b) and Stanton (1997) small-sample bias causes bias towards finding non-linearity in the drift. Jones (2001) suggests that problems arise due to the assumption that the process is assumed stationary, which leads to too restricted functional forms. While the methods in Aït-Sahalia (1996a, 1996b), and CHLK (1997) require stationarity, the one suggested by Stanton (1997) needs only a 'recurrent process' (Bandi and Phillips 1999).

Jones (2001) also points out that in finite time series the commonly used estimators, like ordinary least squares and maximum likelihood, produce biased estimators and this bias is not well known, except in very simple models.

5.2 Non-Normality of Financial Time Series

Usually with time series of financial returns the data are not normally distributed. This is especially obvious with high-frequency data. The returns usually show (excess) kurtosis, or 'fat tails', and often skewness, or asymmetry. This shape can be the result of a normal distribution having changing volatility or jumps, and it is consistent with non-linear drift. Various models have been used to model this non-normality. For example, kurtosis has been modeled using, among other distributions, the Student's t-distribution (Bollerslev 1987), a generalized error distribution (Nelson 1991), non-parametric density (Engle and Gonzales-Rivera 1991), the generalized (skewed) Student's t-distribution (Hansen 1994), and non-central gamma distribution (Harvey and Siddique 1999).

Any non-normal distribution can be modeled using various parametric and non-parametric methods. Jackwerth (1999) reviews the alternative methods in the context of option pricing, and this classification is quite useful here as well. The two broad classes of methods are parametric and nonparametric methods, with various subclasses.

5.2.1 Parametric Estimation of Non-Normal Returns

The parametric estimation methods are divided into generalized distribution methods, expansion methods, and mixture distributions.

A normal or a lognormal distribution is fully characterized by two parameters, the mean, and the variance. Generalized distributions use additional parameters to account for skewness and kurtosis. Many commonly known distributions can be seen as special cases of generalized distributions. An example is generalized beta functions, which include, for example, the lognormal, gamma and exponential distributions (Bookstaber and McDonald 1987).

Expansion methods use a procedure similar to a Taylor expansion. First, a common simple (e.g. normal) distribution is chosen, and then terms are added to obtain more complicated shapes. It is possible to use, for example, a hypergeometric distribution or various polynomials for the expansions. Jondeau and Rockinger (1999, 2001) examine Gram-Charlier and Edgeworth expansions, which have the advantage that skewness and kurtosis can be entered as parameters. However, both methods can only be applied if the amount of kurtosis and skewness are within certain limits. Most financial time series data, including the one we use, exceed these limits.

Another alternative is to think that each observation is drawn with some probability from one of many simple distributions, each with different parameters. A typical procedure is to mix several different normal distributions. Each distribution has some probability of being the source of the observation and various switching mechanisms exist for choosing the one distribution from which an observation is drawn. This switching could occur e.g. as an unexpected jump or a gradual shift, and it can depend on history. The resulting mixture distribution, even if normal distributions are used, allows for kurtosis and skewness.

5.2.2 Nonparametric Estimation of Non-Normal Returns

Non-parametric methods, which were discussed briefly earlier, can be divided into three groups – kernel methods, entropy methods, and curve fitting methods.

Kernel density estimation is based on the idea of 'drawing a histogram' of the given data. However, a 'standard' histogram has two important drawbacks. First, it shows jumps when moving from one interval to another. Second, each observation in an interval has the same weight regardless of its location in the interval. Kernel density estimation provides a method to deal with these two issues by finding a continuous smooth 'weighting function', called a kernel. The kernel can be any positive function (that integrates to one), but often the standard normal distribution is used. Both theory and practice show that the choice of the kernel function is not very important to the performance of this method (Epanechnikov 1969). Selecting the 'window width', i.e. the width of the intervals, is a controversial issue. If it is high, the variance of the estimate is reduced but bias is increased, and vice versa.

Entropy densities are based on the entropy principle, which produces the density that involves the smallest amount of prior information, given the moments. With our data, the kurtosis and skewness are too high for using this method successfully. A good discussion about entropy density estimation is in Jondeau and Rockinger (2002).

Curve-fitting methods use polynomial splines. The basic idea is the same as was discussed earlier in the context of fitting yield curves.

5.3 The Three Major Models: GARCH, Jump Diffusion, and Regime Switching

The traditional models that are used to explain the non-normal distribution of the short interest rate (and other time series) assume a (conditionally) normal distribution, which is subjected to a process that makes it unconditionally non-normal. These are GARCH models, jump diffusions, and regime switching models.

The GARCH (general autoregressive conditional heteroscedasticity) model is an extension of the ARCH (autoregressive conditional heteroscedasticity). ARCH was developed by Engle (1982) to explain why variance in a time series clusters, in other words variance during a period depends on the variance of the previous period. In ARCH, the unconditional variance is constant, but the conditional variance of the error term is written as a linear function of the square of the previous error term. Bollerslev (1986) extended this method to GARCH, which allows a more general lag structure for the error terms.

Jump diffusions are models with a stochastic diffusion process and occasional jumps. The diffusion is a basic Wiener process with a normally distributed stochastic term. The jump component is often a Poisson process. The variable is normally distributed, i.e. follows the diffusion, on the condition that there is no jump. The seminal work using a jump diffusion is by Merton (1976). The research is too numerous to mention, but a good recent paper is Das (2002).

In regime switching models, the stochastic process can switch between two or more regimes. Typically, these models are Markov switching diffusions, which use a Markov chain to choose between two or more diffusion processes. The advantage of regime switching models over jump diffusions is that they can easily explain persistence of volatility. Markov switching diffusions will be discussed in Chapter 6 in detail. Another popular model, the normal mixture distributions, is a special case of the Markov switching model.

5.4 Data

We use the Turkish overnight interbank interest rate data provided by the Central Bank of Turkey. Each observation is the weighted average of the transactions during the trading day, where the weights are the volumes of the transactions. The first day of observations is July 1, 1996, and the last day is December 31, 2001. All the rates are recorded as the simple annual interest rate. We follow the standard procedure of excluding weekends and holidays, so that, e.g. Monday becomes the next day after Friday. This practice is based on empirical studies like French and Roll (1986), which does not find a clear weekend effect for money market instruments. Thus for a normal weekend, the Friday value is actually the three-day interest rate until Monday. Strictly speaking, the daily value is the simple annual interest rate for the interbank loan, which matures on the next trading day. The number of observations is 1384 for the overnight interest rate r_t, and 1383 for the first difference in the overnight interest rate $\Delta r_t = r_t - r_{t-1}$. The number of trading days per year is shown in Table 5.1.

Year	Number of Trading Days
1996	130 (July – December)
1997	252
1998	250
1999	249
2000	252
2001	251
Total Mean	1384 251.6

Table 5.1. Number of trading days per year of the overnight interbank market from July 1, 1996 to December 31, 2001.

The time series for r_t is shown in Figure 5.1 For the sake of readability, values exceeding 200 percent are cut out. The cut observations, eight days altogether, which occurred during the November 2000 and February 2001 crises, are listed in Table 5.2.

Visual inspection of Figure 5.1 indicates that the mean of the interest rate remained at approximately 70 percent from mid-1996 until the end of 1999. Long periods of low volatility occurred from the latter half of 1998 until the end of 1999. All of year 2000 volatility was higher, but the mean was lower at approximately 40 percent, although rising towards the end of the year. The November 2000 crisis, as well as the immediate recovery, can be seen easily. In year 2001, the February crisis is shown as a very high peak, but the economy recovered quite quickly, the interest rates declined gradually stabilizing at about 60 percent level, and volatility went very low. One may be surprised to see in Figure 5.1 that the mean level and volatility of the overnight rate are not necessarily positively related to each other: during year 2000 the mean was at the lowest level, but volatility is high. This is not unusual; for example, Ball and Torous (1999) find that although interest rate volatility and level are often correlated with each other, this dependency is not always present and is difficult to estimate. (A brief overview of the Turkish economy covering years 2000 and 2001 is in Chapter 3.)

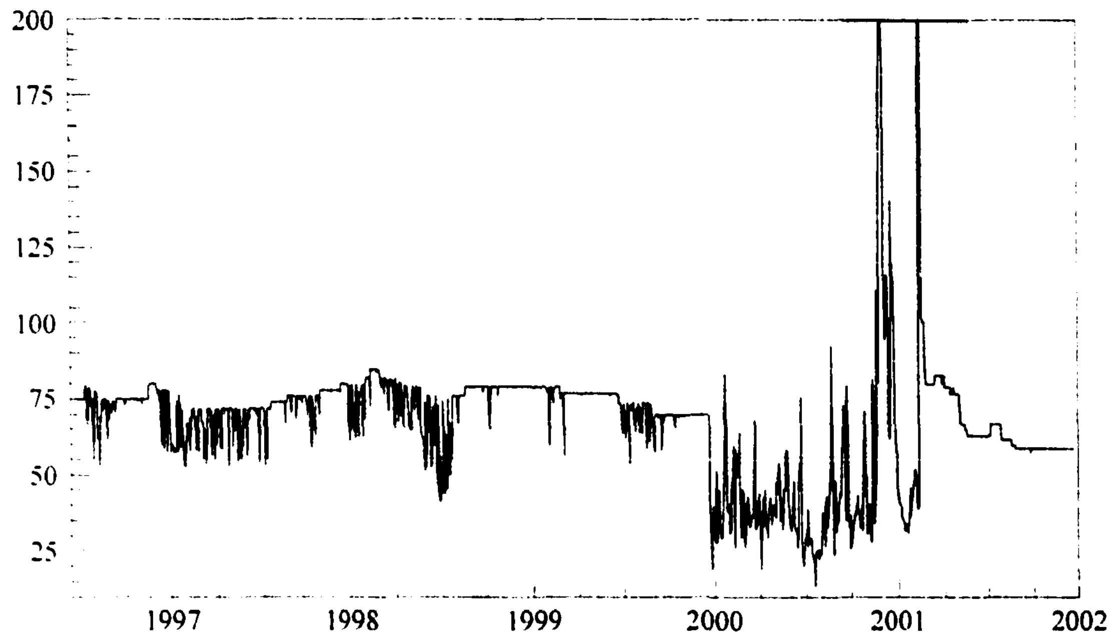

Figure 5.1. The time series of the overnight interbank rate from July 1, 1996 to December 31, 2001. Values exceeding 200 percent not shown.

Table 5.2 shows all observations in our data, which exceed 200 percent. For some purposes, unusual values like these may be considered outliers, and are excluded from the data. On the other hand, if we, and more importantly 'the market', believe that similar rates may occur again with even a small probability, they may be relevant for modeling interest rates and cannot be excluded. The 'Peso problem' that shapes the term structure, as was mentioned in Chapter 2, is based on events like these.

November-December 2000		February 2001	
Day	Interest rate	Day	Interest rate
November 30	315.92	February 20	2057.74
December 1	873.13	February 21	4018.58
December 4	782.46	February 22	1195.28
December 5	363.21	February 23	568

Table 5.2. Dates and overnight interest rates excluded from graph 5.1. All values in excess of 200 percent are shown.

Descriptive statistics for both r_t and Δr_t are shown in Table 5.3. The mean interest rate for the period is 73.135 percent and the mean for its first difference, i.e. change from the previous trading day, is –0.0115. The variances for both variables are very high and they both exhibit very strong positive (excess) kurtosis. One might conclude from the difference that the time series shows a slight negative trend. However, this just represents the difference between the last and first observations. Besides, even if a trend is present, it would be stochastic rather than deterministic and should not be included in a model as a time trend.

Statistic	*r*	Δ*r*
Mean	73.13545	-0.011518
Variance	16436.830	12386.13
Kurtosis	698.7469	451.2468
Skewness	24.57256	-3.695541
Minimum	13.6	-2823.3
Maximum	4018.58	2014.08
N	1384	1383

Table 5.3. Summary statistics for the interest rate and its first difference.

5.5 Normality

It is possible to test if the observations are drawn from a normal distribution using statistical tests for normality. A relatively simple test is the Bera-Jarque (BJ) statistic, which is a form of Wald test. The null hypothesis is that the data are normally distributed. The test uses the estimates of skewness (τ) and kurtosis (κ) and has the formula

$$BJ = n\left[\left(\frac{\tau^2}{6}\right)+\left(\frac{(\kappa-3)^2}{24}\right)\right]. \tag{5.1}$$

Note that the kurtosis of a normal distribution is equal to 3. The BJ statistic is asymptotically chi-squared with 2 degrees of freedom. Bai and Ng (2001) argue that kurtosis cannot be precisely estimated, and BJ test is not a good measure for describing the tail behavior, unless the tails are thin. However, the null hypothesis of the BJ statistic assumes normal distribution so the test can be used to reject normality.

The BJ statistic for the interest rates is over 27 million and for the interest rate differences over 11 million, so neither distribution is normal. (Note that for a p-value of 0.000001 the chi-square value is less than thirty). We also conducted the Kolmogorov-Smirnov normality test for r_t, and normality was clearly rejected (not shown).

5.6 Stationarity

A time series can be described as a stochastic process by specifying the joint distribution of the random variables for any finite set of times. In discrete time we can write $\{y_1, y_2, y_3, \ldots, y_n\}$, where the random variables can be, for example, daily observations of the overnight interest rate.

A time series is stationary if its stochastic properties are invariant with respect to time. A stationary time series has a tendency to return to its mean, it has a finite variance and shocks are transitory.

Definition (i)

A time series is covariance-stationary, or weakly stationary, if:

(1) $E(y_t) = \mu$ for all t, and

(2) $E(y_t - \mu)(y_{t-j} - \mu) = \gamma_j$ for all t and any j.

Condition (1) states that the mean is independent of date, and condition (2) that covariances only depend on the lag, not the date itself. If j=0, it follows from (2) that the variance is independent of date. Usually if a time series is said to be 'stationary' it is weakly stationary. Another form is strict stationarity, which requires that the whole joint distribution is independent of date. While strict stationarity is a very strong assumption and cannot be tested easily, for a Gaussian process, which is uniquely defined by its first two moments, weak stationarity implies strict stationarity. Respectively if a non-Gaussian process is (weakly) stationary, its distribution does not necessarily remain unchanged over time, but the mean, variance and autocovariances remain constant.

If a time series has a trend, it cannot be stationary, but can often be converted to a stationary one by detrending it. If this is the case, the process is called trend-stationary. If the process for y_t is not stationary, but the process for its differences y_t-y_{t-1} is stationary, the series is said to be difference-stationary.

Definition (ii)

A time series y_t is integrated of order n, written $y_t \sim I(n)$ if the stochastic part is non-stationary, but becomes stationary after differencing a minimum of n times. A random walk is an integrated process of $I(1)$, so if y_t follows a random walk y_t is not stationary, but the difference y_t-y_{t-1} is.

Consider a simple example. Assume a stochastic process

$$y_t = b + ct + \alpha y_{t-1} + \varepsilon_t. \tag{5.2}$$

where ε_t is the stochastic error term, which is white noise, ε_t=iid(0,σ^2). The process starts at $y_0=b$ at time t=0. If c>0 the series has a deterministic trend upwards and if c<0 it has a deterministic trend downwards. A necessary condition for stationarity is that c=0. Assume that c=0. Then if α=1 the system is said to have a unit root and it becomes a random walk with $\Delta y_t = \varepsilon_t$. The first difference $\Delta y_t = y_t - y_{t-1}$ is now stationary. This process is integrated of order one, i.e. $I(1)$. If we have $c \neq 0$, the system can be described as $I(1)$+trend, and Δy_t is trend-stationary.

5.6.1 The Augmented Dickey-Fuller Unit Root Test

The tests commonly used for examining stationarity of a time series are called unit-root tests. In these tests, the null hypothesis is that the time series has a unit root, i.e. it is a random walk. If the null hypothesis is rejected, the alternative hypothesis of stationarity is accepted. We examine the stationarity of the short rate r_t as well as the first difference Δr_t using various tests. We start with the Augmented Dickey-Fuller (ADF) test. The auxiliary model can be written as

$$y_t - y_{t-1} = \alpha y_{t-1} + \beta_1(y_{t-1} - y_{t-2}) + \ldots + \beta_p(y_{t-p} - y_{t-p-1}) + \beta_{p+1} + \varepsilon_t, \tag{5.3}$$

where ε_t is white noise and β_{p+1} is a constant intercept (i.e. mean). The test uses p lags of the random variable y_t, and all the differences are first differences, i.e. for one period. The first observation that is included in the calculation is $t=p+2$. The null hypothesis is that a unit root exists, so H_0 is that $\alpha=0$, and H_1 is that the process is stationary with $\alpha<0$ and a positive mean. The test statistic is the Student's t-value. In this test, like many others, the number of lags can be chosen, and a general guideline is to choose p so that autocorrelation in the errors is just removed, so the error terms will be white noise. The limiting distributions of the test statistics depend on the autocorrelation of the errors. If the number of lags is too low, the errors will be autocorrelated and the practice of comparing the calculated statistic to the critical value may not be justified. If too many lags included the test will lack power, and it will be excessively difficult to reject the null hypothesis. An objective method for choosing the number of lags is to use the Akaike Information Criterion (AIC)

$$AIC(s) = \ln \hat{\sigma}_s^2 + \frac{2s}{N}, \tag{5.4}$$

where s is the number of parameters in a model, N is the sample size, and the variance $\hat{\sigma}_s^2$ is the estimated variance with s parameters. The first term measures the fit of the model. The second term in the formula penalizes for adding parameters. Adding parameters improves the fit, ceteris paribus. The number of lags that minimizes the AIC will be chosen.

The AIC finds the optimum number of lags for r_t at 14 and for Δr_t at 30. With the recommended numbers of lags, the unit root is rejected at $p<0.000001$ in both cases. The t-values for α are –8.8317 for r_t and –10.8860 for Δr_t. The conclusion is that the process is stationary without a trend.

We also carried out three other unit-root tests, which are the ADF test with linear trend, the Phillips-Perron test, and the Bierens nonparametric test with nonlinear trend. The results are shown in Appendix 1.

5.6.2 KPSS Test for Stationarity

It is also possible to test for the stationarity of a time series using stationarity, rather than the unit root, as the null hypothesis. We use the KPSS test introduced by Kwiatkowski, Phillips, Schmidt, and Shin (1992). The null hypothesis H_0 is that the process is stationary without a trend:

$$y_t = c + \varepsilon_t, \qquad (5.5)$$

where c is a constant and ε_t is a zero-mean stationary process. The alternative hypothesis H_1 is that the process has a unit root without a trend:

$$y_t = y_{t-1} + \varepsilon_t. \qquad (5.6)$$

The KPSS test is based on the idea that, in a unit root series, variance increases over time, but with stationary series it does not. The KPSS test can be written as a Lagrange Multiplier test

$$LM = \sum_{t=1}^{T} \frac{S_t^2}{\hat{\sigma}_\varepsilon^2}, \tag{5.7}$$

where S_t^2 is the partial sum of the residuals of an OLS regression, and $\hat{\sigma}_\varepsilon^2$ is the estimated error variance of the null hypothesis model. The $\hat{\sigma}_\varepsilon^2$ is estimated using a Newey-West type variance estimator of the long-run variance with truncation lag m. No strict guidelines exist for the choice of m, but many authors propose m=7. Experimenting with different values of m shows that in clear cases like ours, the results are not sensitive to the choice of m.

The hypothesis that the process is stationary without trend is accepted with both variables, the interest rate and the difference. The asymptotic distribution of the test statistic is a Brownian bridge and its upper tail values (critical values) can be found e.g. in KPSS (1992). The estimated test statistic for r_t is 0.0940 (m=7), while the critical values at 5 and 10 percent level are 0.463 and 0.347 respectively. The estimated test statistic for Δr_t is 0.0029 (m=7), while the critical values are the same. We accept stationarity for both time series. We also conducted the KPSS test with the null hypothesis that the process is trend stationary and the alternative hypothesis that it is a unit root process with a trend (not shown). The results remained almost exactly same.

Earlier we rejected the unit root hypothesis with p<0.00001 for the interest rate as well as the interest rate difference using the ADF test. The results of the KPSS stationarity test also support the conclusion that the overall time series for both r_t and Δr_t are stationary, without a trend. This result is not surprising. Interest rates (r_t) are usually assumed to be mean reverting with some long run mean. Furthermore, if r_t is stationary, the first difference Δr_t is stationary as well.

5.7 Autocorrelation and Mean Reversion

In the case of a univariate autoregressive (AR) time series, the own lags of the variable can be used as explanatory variables. The order of the AR process is given by the number of lags n and denoted as AR(n). The general simple form with one lag (AR(1)) is

$$y_t = c + \alpha y_{t-1} + \varepsilon_t, \tag{5.8}$$

where the constant term c gives the trend. Note that, for stationarity we need c=0.

A stationary time series is always mean reverting. The speed of mean reversion is determined by the autocovariances. If the autocovariances are large, the series is persistent and mean reversion is slow. On the other hand, if the autocovariances are small, mean reversion occurs quickly. In the extreme case if α=0, autocovariance is zero, the process is only 'white noise' and mean reversion is instant. At the other extreme α=1, that is a unit root, and the process is a random walk with no mean reversion.

Autocorrelations for the time series can be estimated using (5.10). For r_t the formula gives a $\rho(r_t, r_{t-1})$=0.6235 and $\rho(r_t, r_{t-2})$=0.2198, while the following autocorrelations are much lower. The first-order autocorrelation for the Δr_t time series is only 0.0360, which is a very low value, but the second-order autocorrelation is –0.3400, and also the higher-order autocorrelations are larger than the first one.

We test for the statistical significance of autocorrelations using the Box-Pierce (1970) test. The test statistic is calculated as

$$Q = T\sum_{n=1}^{k} \rho(n)^2 \text{, where} \tag{5.9}$$

T = sample size

k = order to which calculation is done

Here $\rho(n)$ is the nth order sample autocorrelation calculated as

$$\rho(n) = \frac{\sum_{t=n+1}^{T} y_t y_{t-n}}{\sum_{t=1}^{T} y_t^2}, \tag{5.10}$$

where t is the index of the period -- the number (date) of the observation. The test statistic Q belongs to the family of Lagrange multiplier tests and is asymptotically distributed as chi-square with k degrees of freedom. The null hypothesis is that the first k autocorrelations of a covariance, i.e. weakly, stationary time series are zero. If the test statistic is above the critical value, the null hypothesis is rejected, and it may make sense to try to model the time series as an AR process of order n or less.

We apply the Box-Pierce statistic to the time series of r_t and Δr_t. For the interest rate, r_t the test statistic $Q(1)$ for AR(1) is 538.01, while the critical value at 5% level is

3.84, so the null hypothesis of independence is rejected ($p<0.000005$), and the series is an AR(1) process. The higher order autocorrelations are also statistically very significant. Since $Q(2)=604.91$, $Q(3)=611.87$ and $Q(4)=612.57$, terms of order three and higher do not add much information, so the AR(1) or AR(2) form would be a sufficient AR-type description of the process. For the interest rate difference Δr_t, the test statistic for AR(1) is $Q(1)=1.79$, which has a p-value of 0.18032. Thus, we cannot conclude that the process for Δr_t is AR(1). However, for higher order autocorrelations the test statistics are $Q(2)=160.05$, $Q(3)=184.66$, and $Q(4)=189.70$, which are statistically significant. The results indicate that the time series of the interest rate follows an AR(1) (or AR(2)) process and exhibits mean reversion. This conforms to the unit root and stationarity test results.

5.8 Volatility Clustering

A simple, although not very robust, method for finding volatility clustering is to apply the Box-Pierce test to the squared returns. We conduct this test using the squares of r_t. The first-order autocorrelation is 0.3269 and the Box-Pierce statistic $Q(1)$ takes a value of 147.90, which is statistically significant with $p<0.000005$. Higher order autocorrelations are small and they do not increase the value of Q, which means that AR(1) presentation for the volatility is sufficient. This result indicates the presence of volatility clustering.

5.9 Structural Breaks

An interesting question concerning any time series is if the parameter values of the intercept, slope, and volatility change over time and if these changes occur suddenly. Another issue is that even though the data over a long period is stationary, it is possible that there are subperiods with unit roots. In order to answer these kinds of questions, one will want to estimate, if possible, the existence, number, and dates of structural breaks in the time series. Numerous models exist for estimating a single structural break. The classical procedure is called the Chow (1960) test - break the time series into two subperiods and apply the F-test to see if the parameter values are equal. The weakness of this test is that the break date has to be known. Although Chow (1960) suggested that the breakdate should be chosen using statistical tests, it was not until early 1990's when methods for actually estimating the dates were presented by Andrews (1993) and Andrews and Ploberger (1994) among others. A system for finding multiple structural breaks has been introduced by Bai and Perron (1998, 2002).

Bai and Perron (2002) estimate structural changes in the mean of the U.S. three-month real interest rate (using the CPI as the deflator) with quarterly data. The method is rather complicated and it is not explained here, but the interested reader is referred to the two papers by Bai and Perron (1998 and 2002).

We analyze the daily Turkish overnight interest rate with the Bai-Perron model. It is important to keep in mind that the data was found stationary earlier. High-frequency financial data may exhibit frequent jumps with overall mean reversion. This may make it difficult to estimate structural breaks. Essentially, a structural break is a 'permanent shift'. A problem arises, for example, if the time series has two breaks so that the value of the parameter returns to its original value after the second break. In this case, a test for one break may fail to reject the null hypothesis of no breaks, while a test for two breaks may reject the null hypothesis and accept the alternative hypothesis of two breaks. If the breaks occur frequently, the Bai-Perron method may or may not find a structural break, depending on the case, but estimating break dates is not easy. In addition, the Bai-Perron model requires each segment to contain a minimum number of consecutive observations, which may be quite large.

To avoid some of the problems due to the high frequency of the observations, the frequency of the data is reduced from daily to weekly observations. The weekly value is calculated simply as the (unweighted) average of the daily data for the week. This gives us 287 weekly observations that cover the whole period from the beginning of July 1996 to the end of December 2001. The estimation is done following the instructions of Bai and Perron (2002). The maximum number of breaks is five. We allow for different variances of the residuals across segments as well as serial autocorrelation of errors. For estimating the critical values accurately, the minimum length of a segment is set at 43; obviously, this is restrictive in our case.

The first test is for the existence of an unknown number of breaks. This is done using the UDmax and WDmax tests by Bai and Perron (2002). The tests use a null hypothesis of zero breaks against the alternative hypothesis of an unknown number of breaks. The estimated value of the UDmax test is 149.4120, which is much higher than the critical value at p=0.01 level (12.3700) so the conclusion is that the data includes structural breaks. The result of the WDmax test is very similar with a test value of 167.4937 and a critical value of 13.83 (p=0.01).

Next, we test for the number of breaks. This procedure is done using a SupF test (Andrews 1993) with a null hypothesis of no breaks against the alternative hypothesis of a fixed number of breaks. The test is conducted separately for each number of breaks. The estimated statistics, the respective critical values for p=0.01, and the conclusions are given in Table 5.4.

Test	Estimated statistic	Conclusion H_0 = no breaks
0 versus 1 break	1.5600 (CV=7.04; p=0.1)	Accept H_0
0 versus 2 breaks	149.4120 (CV=9.36;p=0.01)	Reject H_0
0 versus 3 breaks	101.722 (CV=7.60;p=0.01)	Reject H_0
0 versus 4 breaks	76.7317 (CV=6.19:p=0.01)	Reject H_0
0 versus 5 breaks	69.7028 (CV=4.91;p=0.01)	Reject H_0

Table 5.4. The SupF test statistics for zero versus a known number of structural breaks.

All estimations in Table 5.4 are with a null hypothesis of no breaks. Interestingly, if the alternative hypothesis is for one break only, the null hypothesis is accepted, but if the alternative hypothesis is for two (or more) breaks, the null hypothesis is clearly rejected. This can happen with a true model where the economy is first in one regime, and then switches temporarily to another one before returning back to the earlier regime. This phenomenon can make it difficult to estimate the accurate number of breaks with the Bai-Perron model. The regime switching model discussed in Chapter 6 is consistent with this kind of behavior.

We also conducted the SupF test stepwise testing for 0 versus 1 break, 1 versus 2 breaks, 2 versus 3 breaks etc. The test showed that two breaks is significantly better than one, but fitting more than two breaks in the time series does not improve the result.

Bai and Perron (2002) also provide the tools for estimating the break dates. Note that for the estimations above the minimum length of a segment was set at 43 observations in order to estimate the critical values. If we reduce the minimum length of a segment to 10 days, the model finds the following breaks - in order of importance:

1. Week ending November 17, 2000.
2. Week ending December 31, 1999.
3. Week ending February 23, 2001.

The first break corresponds to the November 2000 banking crisis, the second one to the introduction of the stabilization program following the December 1999 crisis, and the third break to the February 2001 crisis (see Chapter 3). Note that these are the three first (or most important) breaks suggested by the model for the whole period from July 1996 to December 2001.

5.9.1 Structural Breaks and Unit root tests

Since Perron (1989, 1990) it has been known that if a time series fluctuates stationarily around a trend that has a structural break, the unit root tests may indicate the existence of a unit root, even though the time series is stationary on the condition that the jump is known. Christiano (1992) as well as Banerjee, Lumsdaine and Stock (1992), and Hamori and Tokihisa (1997) report the opposite problem; when there is a single structural break in trend or variance the standard unit root test is rejected too easily (and stationarity is found).

Researches have developed various unit root tests in the presence of structural breaks. The assumptions vary in terms of the number of breaks and whether the break date is known a priori or not. For example, Perron (1989, 1990) designed a unit root test that allows for one structural break with a known date by adding dummy variables to the augmented Dickey-Fuller test. Tests designed by Lumsdaine and Papell (1997), and Clemente, Montanes, and Reyes (1998) allow for two structural breaks with an unknown date.

It is important to keep in mind that structural breaks in time series do interfere with the unit root and stationarity tests, and the direction of the error may not be clear. We return to this question in the context of Markov switching model in Chapter 6.

5.10 Conclusions

We have shown in this chapter that the Turkish overnight interbank interest rate has a heavy-tailed and very skewed distribution. The time series is clearly covariance stationary and AR(1), which indicates mean reversion. In addition to high volatility, we saw volatility clustering.

The short rate time series contains structural breaks. The estimated dates of the breaks match the times of shocks in the Turkish economy, which were described briefly in Chapter 3.

A standard one-factor model, as the ones described in Chapter 2, is not able to produce the results we have seen in this chapter. However, allowing for structural breaks, or jumps, makes it possible to explain these findings. In order to model these jumps in a meaningful fashion, it is important to define well the process that generates them. The most obvious alternatives are jump diffusions, mixing distributions and regime switching models. In addition to jumps, we found volatility clustering, and visual inspection of Figure 5.1 suggests persistence in means and volatility. A good candidate for capturing these features is a regime switching model. In the next chapter, we will use a Markov switching model to explain the behavior of the Turkish short rate.

6 A Markov Switching Model of the Short Rate

Regime switching models have been widely used in economic time series since the seminal work by Sims (1980) and Hamilton (1988, 1989). The technique is based on older ideas, like the theory of Markov chains developed in Baum and Petrie (1966) and Baum, Petrie, Soules and Weiss (1970). The hidden Markov chain models are based on Heller (1965), and Blackwell and Koopmans (1975), while the normal mixture distributions, which are a special case of regime switching models, have their origins in Pearson (1894).

Regime switching models forecast interest rates well (Gray 1996, and Ang and Bekaert 2002a) and often they do better than single-regime models in empirical studies (e.g. Cecchetti, Lam and Mark (1993), and Garcia (1998)). Numerous studies have used regime switches to study the behavior of interest rates. Some examples are Hamilton (1988), Lewis (1991), Sola and Driffill (1994), Evans and Lewis (1995), Garcia and Perron (1996), Gray (1996), Li and Xu (2000), Bekaert, Hodrick and Marshall (2001), Ang and Bekaert (2002a, 2002b) and Hong, Li and Zhao (2002).

The regime switching model can explain the excess kurtosis and skewness of the short rate that is found in many studies. Moreover, regimes of random walk behavior (unit root) can be combined with stationary regimes while the whole process remains weakly stationary (Ang and Bekaert (2002a), and Holst, Lindgren, Holst and Thuvesholmen (1994)). The nonlinear mean reversion found, among others, in Aït-Sahalia (1996b) and Stanton (1997) can be a result of a regime switching process. Applying the model to U.S. data, Gray (1996) as well as Bekaert, Hodrick and Marshall

(2001) find high mean reversion with high interest rates, but random walk at low interest rate levels. Hong, Li and Zhao (2002) find a high interest rate regime with high volatility, and strong mean reversion, and another regime with low volatility and near random-walk behavior. Ang and Bekaert (2002b) use time-dependent switching probabilities to show how the nonlinearities in drift and volatility can be generated by a model with regime switches. Li and Xu (2000) make a similar observation.

We use a regime switching autoregressive Hidden Markov Model version of the Vasicek (1977) model to explain the behavior of the overnight interbank interest rate in Turkey. This particular model is chosen partly because of its tractability. The Vasicek model is Gaussian, so the interest rates are normally distributed, and our model leads to a distribution that is a Markov switching mixture of normal distributions. Another reason in favor of this choice is that a Gaussian model is essentially as good as any other affine model (see Chapter 2). With regime switching one of the limitations of the Gaussian models is lifted since the historical (that is, regime-unconditional) short rate does not need to be normally distributed anymore. However, the problem of potentially negative interest rates remains. Hansen and Poulsen (2000) use a regime switching Vasicek model for calculating prices of bonds and interest rate derivatives. Our model is quite different from theirs. We examine the process of the short rate dynamics in the 'real world' while they use martingale measures. Using the concepts introduced in Chapter 2 we examine the P-process and theirs is a Q-process. Our model allows for changes in all the relevant parameters and four regimes while the Hansen-Poulsen model allows for changes in the mean between two regimes. Moreover, our regime switching is governed by a hidden Markov chain and theirs by a Poisson process.

The time series for the Turkish overnight interbank interest rate, which was analyzed in Chapter 5, covers the period from July 1, 1996 to December 31, 2001. We find four regimes for the short rate, one is mean reverting, two are (near) unit root regimes, and one captures a low number of extreme values. The mean reversion occurs at low rates, while at high rate the interest rate becomes close to constant. Note that, even in a unit-root regime there is a positive probability of switching to a mean reverting regime, so mean reversion is still present through the switch. Volatility is higher at low than at high rates. The results indicate that regime switches may occur frequently. The estimated regime-conditional and regime-unconditional residuals are (nearly) normally distributed. However, the regime-unconditional standardized residuals are not independent, so it is possible that the model is not able to reproduce nonlinearities to the extent they are present in the data. Overall, the regime switching Vasicek model is a major improvement over the standard Vasicek model.

6.1 A Markov Switching Vasicek Model

The Vasicek (1977) model, which is one of the best-known one-factor models, was discussed in Chapter 2. We use a relatively simple extension of the Vasicek model with many regimes. The model is a univariate Markov switching autoregressive process of the short rate similar to the one in Ang and Bekaert (2002b). The short rate process can be written as a simple discrete time AR(1) process

$$r_t = \nu(s_t) + \rho(s_t) r_{t-1} + \sigma(s_t)\varepsilon_t, \qquad (6.1)$$

where

$\nu(s_t)$ = intercept

$\rho(s_t)$ = lag coefficient

$\sigma(s_t)$ = volatility

s_t = regime, state

and the error term is iid ε_t~N(0,1). The parameters of the model depend on the regime s_t. Equation (6.1) can alternatively be written for the difference as

$$\Delta r_t = \nu(s_t) - (1 - \rho(s_t))r_{t-1} + \sigma(s_t)\varepsilon_t . \tag{6.2}$$

The (continuous-time) Vasicek (1977) model was written in Chapter 2 as

$$dr(t) = \beta(\alpha - r(t))dt + \sigma dW , \tag{6.3}$$

which has a discrete time first-order Euler approximation

$$\Delta r_t = \beta(\alpha - r_{t-1})\Delta t + \sigma\varepsilon_t\sqrt{\Delta t} . \tag{6.4}$$

It is possible to use an exact discretization rather than the approximation. However, since we use overnight data the discretization error is negligible. Discretization is an important issue if one wishes to calculate the prices of derivative securities rather than just model the short rate dynamics.

Equation (6.4) is, conditional on the regime s_t, a discrete time Vasicek model. The mean reverts to level α at rate β. We estimate the parameters ν, ρ and σ using the form in (6.1) and then calculate the Vasicek parameters from $\beta = 1-\rho$ and $\alpha = \nu / \beta$.

This model is a doubly stochastic process. The first stochastic process is the Markov chain, which chooses the state, i.e. the regime, according to transition probabilities. The second stochastic layer is the stochastic Vasicek process, which gives the output probability distribution for each state.

6.2 The Hidden Markov Chain

The realization, or choice, of the regime $s_t \in \{1,2,\ldots M\}$ in is governed by transition probabilities

$$p_{ij} = p(s_{t+1} = j \mid s_t = i) . \qquad (6.5)$$

where M is the number of states. This is the probability that the regime changes from i to j during the period t. These transition probabilities follow a Markov chain.

6.2.1 Important Definitions

The following definitions relate to the properties of Markov chains and are very important.

Definition (i)

A Markov chain is an integer time process $\{s_t, t \geq 0\}$ where each random variable s_t, with $t \geq 1$, depends on the history $s_{t-1}, s_{t-2}, \ldots$, only through the most recent variable s_{t-1}. With this Markov property, we have $p(s_t = j \mid s_{t-1}) = p(s_t = j \mid s_{t-1}, s_{t-2}, \ldots s_0)$ for all states j.

The time series vector of the short rate is observed up to r_t at time t, but the actual regime s_t is unobservable at all times and follows a hidden, discrete-state, homogeneous, irreducible and ergodic M state Markov process with the transition matrix

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix}. \tag{6.6}$$

Since the states s_t are mutually exclusive and collectively exhaustive the row sums add up to 1 so

$$\sum_{j=1}^{M} p_{ij} = 1 \text{ for all } i. \tag{6.7}$$

The assumptions of irreducibility and ergodicity are essential for the stability and other properties of the model.

Definition (ii)

If the transition probabilities of a Markov chain do not change with time it, is called a homogeneous Markov chain and $p_{ij}^{(t)} = p_{ij}\ \forall i, j, t$.

Definition (iii)

A Markov chain is said to be irreducible if from every state we are able to reach every other state of the Markov chain with a non-zero probability $\exists t$ s.t. $p_{ij}^{(t)} > 0 \quad \forall i, j$. Note that an irreducible matrix does not have to be a positive matrix (with all elements strictly positive), but each state has to be reachable from all states directly or through other states.

Definition (iv)

A state i of a Markov chain has period d if d is the largest integer such that $p_{ii}^{(t)} = 0$ whenever t is not divisible by d. If d=1 then state i is called aperiodic. Consider an example from Gallager (1996). Assume a discrete stochastic system with four states. Suppose that from state 1 it is possible to move directly to states 2 or 4, from state 2 directly to states 3 or 1, from state 3 directly to states 2 or 4, and from state 4 directly to states 3 or 1. If the system starts from state 2 at t=0, it will be in 1 or 3 for t odd and in 2 or 4 for t even. The process in this example is clearly periodic with d=2.

Definition (v)

A Markov chain is called ergodic if the largest eigenvalue of matrix P is unity and all the other eigenvalues are of magnitude less than one (so they lie inside the unit circle).

More intuitively, a stochastic process (not necessarily a matrix) is said to be ergodic for the mean if the mean converges to a given value μ as $T \to \infty$. Ergodicity can be defined respectively for higher moments. Ergodicity is about asymptotic convergence, while stationarity is about time-invariance. Ergodicity is a stronger concept and implies stationarity, but even a strictly stationary process is not always ergodic. For a weakly stationary process, a sufficient condition for ergodicity is that the sum of the absolute values of autocovariances is finite. Any irreducible, finite-state, aperiodic Markov chain is ergodic.

We can define the eigenvalue problem as the solution to the system $Pv = \lambda v$, where P is the $N \times N$ transition matrix, v is the $N \times 1$ eigenvector, and the eigenvalues are the values of λ. This equation can be written as $(\lambda I - P)v = 0$, and if λ is an eigenvalue, then there is a solution with $v \neq 0$. Thus, the solution to the eigenvalue problem is the roots λ that solve $|\lambda I - P| = 0$.

Theorem

Let P be the transition probability matrix of an ergodic Markov chain. Then there exists a unique stationary probability vector ξ^* such that $P'\xi^* = \xi^*$. This vector ξ^* gives the unconditional probabilities of the different regimes. The probabilities ξ^* are called ergodic probabilities. (The proofs for the theorem are in Gallager (1996).)

6.2.2 The Stochastic Process of the Short rate

The full dynamic process of the short rate can now be described as

$$r_t = \left\{ \begin{array}{ll} \mu(s_1) + \rho(s_1) r_{t-1} + \sigma(s_1)\varepsilon_t & \text{if } s_t = 1 \\ \vdots & \\ \mu(s_M) + \rho(s_M) r_{t-1} + \sigma(s_M)\varepsilon_t & \text{if } s_t = M \end{array} \right\}. \quad (6.8)$$

The parameters are assumed to remain constant within a regime, but allowed to be different for each regime. This still allows heteroscedasticity of the historical (regime-unconditional) process. This type of a regime switching model can be an alternative to ARCH and GARCH models. Krolzig (1997) shows that regime-unconditional heteroscedasticity can be a result of constant volatility within a regime but having different values for different regimes in means, volatilities, or autoregressive parameters. Some studies have chosen to allow stochastic volatility within regimes using, for example the ARCH (Cai 1994) or GARCH (Gray 1996, Ball and Torous 1999, and Ang and Bekaert 2002b) models. However, like Gray (1996) notes, the simple regime switching model is able to demonstrate most of the effects of stochastic volatility by itself. We assume that volatility is only regime-dependent, i.e. constant within a given regime.

We adopt the notation where the first lagged observation, which is the first observation we have, occurs at t=0, and the first term in the AR representation refers to t=1. We can write the probability distribution of the interest rate r_t conditional on the regime and the history of the interest rate. At time t, the available history is given by the permutation (sequence) $R_{t-1}=(r_0, r_1, r_2, r_3, \ldots, r_{t-1})$. The conditional density of the short rate, given that regime m occurs, is then

$$r_t \mid (s_t = m, R_{t-1}) \sim N(\bar{r}_m, \sigma_m). \tag{6.9}$$

In each regime the interest rate is normally distributed (by assumption). However, with the regime switches the unconditional distribution of the interest rate will not be normal. A sufficient condition for the unconditional distribution of r_t to show skewness and excess kurtosis is to have at least two regimes with different means.

The regime-conditional probability distribution of r_t can be written for the whole system over the M regimes as a vector. The information about the realization of the state is given by vector

$$\xi_t = \begin{bmatrix} I(s_t = 1) \\ \vdots \\ I(s_t = M) \end{bmatrix}. \tag{6.10}$$

where I is an $M\times 1$ indicator function $(\iota_1, \iota_2, \ldots, \iota_M)'$. If state j is realized, then $\iota_j(s_t=j)=1$ and $\iota_i(s_t=i\neq j)=0$. Note that at all times one state is realized, but the true state, and thus vector ξ_t, is always unobservable. We can then write the probability densities of the short rate as a vector conditional on the true regime

$$\eta_t = \begin{bmatrix} p(r_t \mid \xi_t = \iota_1, R_{t-1}) \\ \vdots \\ p(r_t \mid \xi_t = \iota_M, R_{t-1}) \end{bmatrix}. \tag{6.11}$$

As time passes from t to t+1, the transition between regimes follows the equation

$$\xi_t = P'\xi_{t-1}. \tag{6.12}$$

The conditional probability distribution of r_t, given the realized information up to time t-1, can be written as

$$p(r_t \mid \xi_{t-1}, R_{t-1}) = \eta_t' P' \xi_{t-1}. \tag{6.13}$$

The realized regimes cannot be observed at any time so the available information set can not include ξ_t or ξ_{t-1}. However, beliefs of the state vector are formed based on information set R_{t-1}. This inference can be written as a vector of the discrete conditional probabilities of being in different regimes

$$\hat{\xi}_{t|t-1} = \begin{bmatrix} \Pr(s_t = 1 \mid R_{t-1} \\ \vdots \\ \Pr(s_t = M \mid R_{t-1} \end{bmatrix} = \begin{bmatrix} \Pr(\xi_t = \iota_1 \mid R_{t-1} \\ \vdots \\ \Pr(\xi_t = \iota_M \mid R_{t-1} \end{bmatrix} \tag{6.14}$$

where $\hat{\xi}_{t|t-1}$ is the (inferred) probability distribution of $\hat{\xi}_t$ conditional on the available information set R_{t-1}. ('Pr' refers to a discrete probability.) The probability density of r_t, conditional on knowing the regime, was given in (6.9). When the regime is unobserved, but a result of inference, the conditional probability density is

$$p(r_t \mid R_{t-1}) = \sum_{m=1}^{M} p(r_t \mid \xi_{t-1} = \iota_m, R_{t-1}) \Pr(\xi_{t-1} = \iota_m \mid R_{t-1}) = \eta_t' P' \hat{\xi}_{t-1|t-1} \qquad (6.15)$$

At time t we observe the information set R_{t-1}. The regime is unknown at all times, but from the known information we infer an estimate of the state, or regime $\hat{\xi}_{t-1|t-1}$, at t-1. The inferred regime is an $M \times 1$ vector with probabilities assigned for different regimes. This vector is conditional on the information at time t-1. In the last part of (6.15) the inferred vector is pre-multiplied by the transition matrix P to give us the expected conditional probabilities of the regimes at time t, and the conditional probability densities of the short rate are multiplied by them.

6.2.3 The Markov Property

It is easy to understand that the process of the short rate is Markovian conditional on the history of regimes:

$$p(r_t \mid r_{t-1}, \xi_t) = p(r_t \mid R_{t-1}, \xi_t, \xi_{t-1}, \dots, \xi_0) \qquad (6.16)$$

Thus, if the current regime and the observation r_{t-1} are given, the history of the regimes or the short rate does not provide additional information. However, the regime-unconditional process of the regimes is not a Markov process

$$p(r_t \mid r_{t-1}) \neq p(r_t \mid R_{t-1}). \tag{6.17}$$

This happens because the history of the short rate is used for finding the probability distribution of the states ξ_t.

6.3 Filtering and Smoothing

The model we use combines the hidden Markov chain process for regime switching with an autoregressive model (which would be a vector-autoregression in a multivariate case). Estimating Markov switching models is generally rather complicated because in addition to the parameters $\mu(s_t)$, $\rho(s_t)$ and $\sigma(s_t)$, the transition probabilities in matrix P have to be estimated. The estimation procedure, including the filter and the smoother, is from Krolzig (1996, 1997), where the details can be found. We explain the main features of the model.

The estimation is carried out using the Expectation Maximization (EM) method and a likelihood function. The EM algorithm was originally used in Dempster, Laird and Rubin (1977), but Hamilton (1990) adapted it for the Markov switching model. Each iteration of the EM algorithm estimation involves two steps. In the expectation step, the probabilities of the hidden states are estimated, and in the maximization step, parameters of the model are estimated using the probabilities from the expectation step. The iterative process alternates then between these two steps. For finding the likelihood function, one needs to find a way to extract the optimum estimates for the transition probabilities. We will first have a look at two different procedures, filtering and smoothing, which are needed for finding the regime probabilities. They are also used for finding the likelihood function.

6.3.1 The Choice of the Information Set

The data in the sample is given as daily observations of the short rate. The inference of the regime probabilities, i.e. the sequence $\{\xi_t\}_{t=1}^{T}$, can be constructed using different information sets τ. $\xi_{t|\tau}$ stands for predicted probabilities when $\tau < t$, filtered probabilities when $\tau = t$, and smoothed probabilities when $t < \tau \leq T$. With full-sample smoothed probabilities all the information in the sample is used and $\tau = T$. The commonly used ones are the filtered and smoothed probabilities.

6.3.2 Filtering

Filter methods are used to find the probability distribution of unobserved variables using the available information. The filter we use is from Krolzig (1997). The basic idea is from Hamilton (1989), although it is based on earlier work by others.

Hamilton's (1989) filter is an algorithm, which calculates the optimum value of ξ_{t+1} given the available information set R_{t-1} as shown in (6.14). The filtering algorithm finds the joint probability density of ξ_t and r_t conditional on the observed past time series. At time t, the information set R_{t-1} consists of the observed interest rates up to the previous period t-1. At time t, before r_t is observed, the prior probability of the state ξ_t conditional on the information set R_{t-1} is given by

$$\Pr(\xi_t \mid R_{t-1}) = \sum_{\xi_{t-1}} \Pr(\xi_t \mid \xi_{t-1}) \Pr(\xi_{t-1} \mid R_{t-1}). \tag{6.18}$$

This is called the 'prior' probability because it is not adjusted for the observation r_t at time t. The right-hand side is the summation over all regimes for the probabilities of different regimes at time t-1, and conditional on information up to t-1. We write the conditional probability density of r_t as

$$p(r_t \mid R_{t-1}) = \sum_{\xi_t} p(r_t, \xi_t \mid R_{t-1}) = \sum_{\xi_t} \Pr(\xi_t \mid R_{t-1}) p(r_t \mid \xi_t, R_{t-1}). \tag{6.19}$$

The first equality follows from the fact that a marginal probability can be written as the sum of the joint probabilities of mutually exclusive and collectively exhaustive subsets. Now the summation is over the regimes for probabilities at time t.

The posterior distribution is obtained from the Bayes's rule

$$\Pr(\xi_t \mid R_t) = \frac{p(r_t \mid \xi_t, R_{t-1}) \Pr(\xi_t, R_{t-1})}{p(r_t \mid R_{t-1})}. \tag{6.20}$$

This distribution is 'posterior' because it is adjusted for the observation of the current period t, so it is conditional on the information set R_t rather than R_{t-1}.

We can write, using the η_t from (6.11) the conditional density of the interest rate

$$p(r_t \mid R_{t-1}) = \eta' \hat{\xi}_{t|t-1} = \mathbf{1}'_M (\eta_t \circ \hat{\xi}_{t|t-1}). \tag{6.21}$$

where $\mathbf{1}_M$ is a $M \times 1$ vector of ones as all elements and the symbol $\circ$ is used as the operator for Hadamard (sometimes called the Schur) product (elementwise multiplication). Then the rule for updating the estimate for the regime vector becomes

$$\hat{\xi}_{t|t} = \frac{\eta_t \circ \hat{\xi}_{t|t-1}}{\mathbf{1}'_M (\eta_t \circ \hat{\xi}_{t|t-1})}. \tag{6.22}$$

The filtered regime probabilities $\hat{\xi}_{t|t}$ are updates of $\hat{\xi}_{t|t-1}$ as new information arrives, in other words when r_t is observed.

6.3.3 Smoothing

The disadvantage of the described filtering method and the estimator (6.22) is that the information set contains only past information, i.e. at time t only information set R_t is used. The best estimate is found by always using the information provided by the full sample rather than only the past observations. This inference, which uses all the available information, is called smoothing.

Hamilton (1988, 1989) introduced a smoothing algorithm that uses the full sample, but it is rather complicated. Kim (1994) developed a smoothing algorithm that is much simpler. When using full sample smoothing we add the information that was missing in the filtering above as $R_{t+1 \bullet T} = (r_{t+1}, r_{t+2}, \ldots, r_T)$. Having available the full sample, we can iterate backwards until t starting from the last period filtering estimate $\hat{\xi}_{T|T}$ and using

$$\Pr(\xi_t \mid R_T) = \sum_{\xi_{t+1}} \Pr(\xi_t, \xi_{t+1} \mid R_T) = \sum_{\xi_{t+1}} \Pr(\xi_t \mid \xi_{t+1}, R_T) \Pr(\xi_{t+1} \mid R_T). \tag{6.23}$$

Assume that the information set includes the full sample and we have a probability distribution over all the states ξ_{t+1} at time $t+1$. Then, according to the last part in identity (6.19), we can write the distribution for ξ_t as a sum of mutually exclusive and collectively exhaustive joint probabilities. Next, we examine the two terms in (6.23). Suppose that r_t and ξ_{t+1} depend on ξ_t, but not on the history of the regimes (see Definition 1). Then we can write

$$\Pr(\xi_t \mid \xi_{t+1}, R_T) = \Pr(\xi_t \mid \xi_{t+1}, R_t, R_{t+1 \bullet T}) = \frac{p(R_{t+1 \bullet T} \mid \xi_t, \xi_{t+1}, R_t) \Pr(\xi_t \mid \xi_{t+1}, R_t)}{p(R_{t+1 \bullet T} \mid \xi_{t+1}, R_t)}$$

$$= \Pr(\xi_t \mid \xi_{t+1}, R_t). \tag{6.24}$$

We can write respectively

$$\Pr(\xi_t, \xi_{t+1} \mid R_T) = \Pr(\xi_t \mid \xi_{t+1}, R_t) \Pr(\xi_{t+1} \mid R_T) = \frac{\Pr(\xi_t \mid R_t) \Pr(\xi_{t+1} \mid \xi_t, R_t)}{\Pr(\xi_{t+1} \mid R_t)} \Pr(\xi_{t+1} \mid R_T)$$

$$= \frac{\Pr(\xi_t \mid R_t) \Pr(\xi_{t+1} \mid \xi_t)}{\Pr(\xi_{t+1} \mid R_t)} \Pr(\xi_{t+1} \mid R_T). \tag{6.25}$$

Equations (6.24) and (6.25) can be inserted in (6.23). This gives, using matrix notation

$$\hat{\xi}_{t|T} = (P(\hat{\xi}_{t+1|T} \varnothing \hat{\xi}_{t+1|t})) \circ \hat{\xi}_{t|t}. \tag{6.26}$$

where the symbols $\oslash$ and $\circ$ denote elementwise division and multiplication. Note that P is the transition matrix and we wrote earlier in (6.12) that $\xi_t = P'\xi_{t-1}$.

The filtering algorithm calculates the estimate $\hat{\xi}_{t|t}$ for all values of t, so the last output of filtering is $\hat{\xi}_{T|T}$. By inserting this into the smoothing formula (6.26) we can get the estimate $\hat{\xi}_{T-1|T}$, then use it to calculate $\hat{\xi}_{T-2|T}$, and so on, until we reach the beginning at t=1.

Applying filtering and smoothing gives the filtered and smoothed probability distributions of the regimes, $\hat{\xi}_{t|t}$ and $\hat{\xi}_{t|T}$.

6.4 Maximum Likelihood Estimation

The estimation of the model is done with the EM algorithm using the maximum likelihood method. The parameters of the likelihood function are in vector $\lambda(\theta', \varphi', \xi_0')$. The parameters of the AR(1) model (6.1) for each state are given by θ, the vector of transition probabilities by φ and the initial regime by ξ_0. Before examining the estimation procedure, we will have a brief look at the potential identification problem of this model.

6.4.1 The Identification Problem

In order to use the maximum likelihood estimation procedure, the model has to be identified, at least locally. First, the states have to be identifiable – except for the trivial case of switching their labels. Following Leroux (1992) and Krolzig (1997) this type of an identifiability problem should not arise if the parameter vectors for the M regimes θ_1, $\theta_2 \ldots \theta_M$ are distinct, and the Markov chain, and thus vector φ, is irreducible and aperiodic, and has a unique stationary distribution ξ^*, that is the ergodic probabilities.

Another identifiability requirement has to do with the behavior of the distributions within regimes. If the densities in the regimes are Gaussian with iid errors, this condition is always met. Technically, this condition is not met if a specific mixture distribution can be achieved with many different mixtures. In our specification, the theoretical model is regime-conditionally Gaussian, and fulfills this condition.

6.4.2 The Likelihood Function

The likelihood function can be derived in various ways and it can be written as

$$\begin{aligned} L(\lambda \mid R_T) = p(R_T \mid \lambda) &= \int p(R_T, \xi \mid \lambda) d\xi \\ &= \int p(R_T \mid \xi, \theta) \Pr(\xi \mid \varphi, \xi_0) d\xi . \end{aligned} \tag{6.27}$$

Here the integration is a summation of all possible states over the whole time series, so $\xi = \xi_T \otimes \xi_{T-1} \otimes \ldots \xi_1$, where $\otimes$ is the Kronecker product. We can insert

$$p(R_T \mid \xi, \theta) = \prod_{t=1}^{T} p(r_t \mid \xi_t, R_{t-1}, \theta) \text{, and} \tag{6.28}$$

$$p(\xi \mid \varphi, \xi_0) = \prod_{t=1}^{T} \Pr(\xi_t \mid \xi_{t-1}, \varphi) . \tag{6.29}$$

In addition to the maximum likelihood function, the given constraints are

$$P1_M = 1$$
$$1'_M \xi_0 = 1 \quad (6.30)$$
$$\varphi \geq 0, \sigma \geq 0, \xi_0 \geq 0.$$

This simply means that the probabilities have to be nonnegative and add up to one, and the parameter values should be non-negative. Combining the constraints with the likelihood function gives a constrained log-likelihood function, which is the Lagrangean

$$\ln L^c(\lambda) = \ln L(\lambda \mid R_t) - \kappa_1(P1_M - 1_M) - \kappa_2(1'_M \xi_0 - 1)\,. \quad (6.31)$$

The first-order conditions of the Lagrangean $\partial L^c(\lambda)/\partial\theta' = \partial L^c(\lambda)/\partial\varphi' = \partial L^c(\lambda)/\partial\xi_0{}' = 0$, which are assumed to be well behaving so that an interior solution exists, give the normal equations of the maximum likelihood estimator. The derivation of the first-order conditions is in Krolzig (1997). With respect to θ we get

$$\sum_{t=1}^{T} \hat{\xi}_{t|T}(\lambda)\left[\frac{\partial \ln \eta_t(\theta)}{\partial\theta}\right] = 0 \quad (6.32)$$

where η_t is the conditional density of r_t as given in (6.10). That is

$$\ln \eta_t = \begin{bmatrix} -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_1^2 - \frac{1}{2\sigma_1^2}u_{1t}'(\theta_1)u_{1t}(\theta_1) \\ \vdots \\ -\frac{T}{2}\ln(2\pi) - \frac{1}{2}\ln\sigma_M^2 - \frac{1}{2\sigma_1^2}u_{Mt}'(\theta_M)u_{Mt}(\theta_M) \end{bmatrix} \tag{6.33}$$

since the distribution of r_t is normal conditional on the regime. The parameters θ_i are the Vasicek parameters, given the regime, and the regime-conditional residuals u are a function of them. T is the number of observations

The second first-order condition, which is with respect to φ, gives

$$\varphi = [\hat{\xi}^{(2)}(\lambda)]\oslash[1_M \otimes \hat{\xi}^{(1)}(\lambda)] \tag{6.34}$$

where $\hat{\xi}^{(1)} = \sum_{t=0}^{T-1}\hat{\xi}_t^{(1)}$, and $\hat{\xi}_t^{(1)} = \xi_t$. Respectively, we have $\hat{\xi}^{(2)} = \sum_{t=1}^{T-1}\hat{\xi}_t^{(2)}$, with $\hat{\xi}_t^{(2)} = \xi_t^{(1)} \otimes \xi_{t-1}^{(1)}$.

The third first-order condition is with respect to the initial state ξ_0 and yields the solution

$$\xi_0 = \hat{\xi}_{0|T}(\lambda) \tag{6.35}$$

In equation (6.35) the right-hand side is a function of ξ_0. The initial state can be found by using the smoother to find $\hat{\xi}_{0|T}$ and by assuming that ξ_0 is initially equal to the ergodic probabilities ξ^*.

Because the first-order conditions include the conditional regime probabilities, the constraints are non-linear, and numerical iteration has to be used in maximizing the likelihood function.

6.4.3 The EM algorithm

The Expectation Maximization (EM) algorithm includes two steps, the expectation step, and the maximization step.

The expectation step uses the filtering (6.22) and smoothing (6.26) algorithms to estimate the unobserved states ξ_t. The conditional probabilities $\Pr(\xi \mid R, \lambda^{j-1})$ (j is the index of the iteration) are calculated iteratively. First, the initial values of the parameters are given in λ^1 and the estimated conditional regime probabilities will be given by $\hat{\xi}$. Then this $\hat{\xi}$ is used in the maximization step to obtain estimated parameter vector λ.

The smoothed probabilities $\hat{\xi}_{t|T}(\lambda)$ from the preceding expectation step are inserted into the maximization step. Then the parameters λ are estimated so that the first-order conditions (6.32), (6.34) and (6.35) of the likelihood function are fulfilled. The result of this maximization step, the estimated parameter vector λ, is then used as an input in the expectation step for estimating the filtered and smoothed probabilities.

Each iteration of the algorithm goes through the expectation and maximization steps. Every time when the maximization step is entered, the parameter vector λ is found using iterations.

6.4.4 The Maximization Step

The expectation step gives the estimated smoothed probabilities, which will be used as the regime probabilities in the maximization step. The maximization step can be seen as a regular (generalized least squares) regression of (6.1). Each observation includes the values r_t, r_{t-1} and regime s_t. The smoothed probabilities are used as the realization weights of the regimes.

The regression equation depends on the parameter restrictions of the model. The regressions and estimators for the different cases are explained in Krolzig (1997). We will consider here the case where all parameters in equation (6.1) are allowed to be different across regimes. In this case, the regression equation becomes

$$y = \sum_{m=1}^{M} \left(\Xi_m \overline{X}\right)\beta_m + u, \qquad (6.36)$$

where

$$u = N(0,\Omega), \text{ and} \qquad (6.37)$$

$$\Omega = \sum_{m=1}^{M} \Xi_m \sigma_m^2 .$$

The Vasicek model we use involves only one lag so vectors y and $\overline{X}$ will be

$$y=\begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_T \end{bmatrix} \qquad \overline{\mathrm{X}}=\begin{bmatrix} 1 & r_0 \\ 1 & r_1 \\ \vdots & \vdots \\ 1 & r_{T-1} \end{bmatrix}. \tag{6.38}$$

Note that we have labeled the first observation date of the time series as t=0, so that t=1 is the date of the second observation. Matrix Ξ_m is diagonal matrix, which has the ξ_t for regime m on the diagonal (see 6.10), so it works like an indicator function

$$\Xi_m=\begin{bmatrix} \xi_{m1} & 0 & \cdots & 0 \\ 0 & \xi_{m2} & \cdots & 0 \\ \vdots & \vdots & \ddots & 0 \\ 0 & 0 & 0 & \xi_{mT} \end{bmatrix}. \tag{6.39}$$

The variance conditional on the regime σ_m^2 is a scalar. The (generalized least-squares) estimator for β_m is given as

$$\tilde{\beta}_m=\left(\left(\overline{\mathrm{X}}'\hat{\Xi}_m\overline{\mathrm{X}}\right)^{-1}\overline{\mathrm{X}}'\hat{\Xi}_m\right)y. \tag{6.40}$$

The matrix indicating the true states Ξ_m is unknown, so the realizations of the regimes have to be replaced using the smoothed probabilities. If the regimes are known, the vector ξ_t is given by the indicator function $I(s_t)$. Since the regimes are not known, ξ_t has to be replaced by $\hat{\xi}_t$, which is the estimate given by the smoothed probabilities $\hat{\xi}_{t|T}$, so that we have

$$\hat{\Xi}_m = \begin{bmatrix} \hat{\xi}_{m1|T} & & & \\ & \hat{\xi}_{m2|T} & & \\ & & \ddots & \\ & & & \hat{\xi}_{mT|T} \end{bmatrix}. \tag{6.41}$$

The conditional variance for each regime is estimated separately as

$$\tilde{\sigma}_m = T_m^{-1}\tilde{U}'_m \hat{\Xi}_m \tilde{U}_m, \tag{6.42}$$

where

$$T_m^{-1} = \text{tr}(\hat{\Xi}_m) = \sum_{t=1}^{T} \xi_{mt}\text{, and} \tag{6.43}$$

$$\tilde{U}_m = y - \overline{X}\tilde{\beta}_m'. \tag{6.44}$$

In (6.44), the value that is predicted for regime i is subtracted from the observed value to get the regime-conditional residual for each date. These daily residuals are added together using the smoothed probabilities (6.41) as weights and divided by the 'number of days in the regime'. The 'number of days in the regime' is the sum of the smoothed probabilities for regime i over all dates. The regression equations and estimators for other specifications and different parameter restrictions can be found in Krolzig (1997).

6.5 Model Specification

We have introduced the Markov switching model, which is a doubly stochastic process. The first process is the hidden Markov chain model and the second process is the Vasicek model. There is no test for finding out whether the data is generated by a Markov switching process. However, as was discussed at the end of the previous chapter, the stationarity, non-normality, volatility clustering, as well as the autocorrelation structure and presence of structural breaks found in the data indicate that this type of a model may be suitable.

As a benchmark, we estimate the Vasicek model without regime switching. After that, we will examine two important questions. First, we will look at the influence of the number of regimes on the estimation results. Second, we will examine which parameters should be allowed to vary. Since the two most important features of all short rate models are heteroscedasticity and mean reversion, the parameters relating to them are of special interest. The model we estimate is equation (6.1) with regime switches, which follow a hidden Markov chain (so we get 6.8).

6.5.1 A Single-Regime Vasicek Model

Applying the OLS estimation to the Vasicek model (6.1) delivers parameter values $v = 27.52830$ with a t-value of 8.862 ($p<0.000005$), and $\rho = 0.62349$ with a t-value of 29.636 ($p<0.000005$). These values correspond to the following discrete Vasicek specification for Δr

$$\Delta r_t = 0.37651(73.11439 - r_{t-1})\Delta t + 100.34435\varepsilon\sqrt{\Delta t} \qquad (6.45)$$

where we can write $\Delta t=1$. The residuals show high skewness (18.485) and kurtosis (525.583), and the Bera-Jarque test for the normality of the error terms gives a statistic of over 15 million ($p<0.000001$), which clearly shows non-normality. The Breusch-Pagan test for homoscedasticity indicates heteroscedasticity with a statistic of 102758 ($p<0.000001$), which is distributed as a chi-square(1) under the null hypothesis. These results indicate that the test statistics (t-values) for the estimation are not reliable. Only adjusting for heteroscedasticity by using White's heteroscedasticity consistent variance lowers the t-values to 1.238 ($p=0.21566$) for ν, and to 1.881 ($p=0.06003$) for ρ.

The estimated volatility is high at $\sigma=100.34$, so the probability of negative interest rates is very high. The Akaike information Criterion is 12.0573. We will use this value as a benchmark for the multiple-regime models.

6.5.2 The Number of Regimes

A major attraction of regime switching models is that periods of random walks (unit root) can be included in a historically weakly stationary time series. Typically, one finds at least a regime with mean reversion and another one with a unit root, or close to it. Many empirical studies using nonparametric estimation (see Chapter 5), as well as some using regime switching (see beginning of this chapter) have found nonlinear mean reversion of the short rate. Often mean reversion is strong at high interest rates, while at lower rates one may see random walk behavior. In some studies mean reversion is found again if the rate becomes very low. Another observation is volatility clustering. Some regimes have higher volatility than others do, and high volatility is often associated with high levels of interest rate. The number of regimes should be sufficient, but if the number of regimes increases, two things happen. First, the number of parameters increases, and second, some of them may contain very few observations.

In order to see the effect of the number of regimes on the model, we estimated equation (6.1) letting all the variables, the intercept ν, the AR(1) parameter ρ, and the volatility σ to be different across regimes. When testing for the number of regimes many standard test statistics cannot be used due to the presence of nuisance parameters, so we choose to use the Akaike Information Criterion (AIC) (see Chapter 5) as a test statistic. The AIC results from estimating equation (6.1) with different numbers of regimes are shown in Table 6.1.

Regimes	1	2	3	4	5
AIC	12.0573	6.7716	3.6257	3.3408	3.2363

Table 6.1. The Akaike Information Criterion for different numbers of regimes with all parameters (ν, ρ and σ) varying across regimes.

The first result (in Table 6.1) with one regime is just the normal OLS regression. Introducing two regimes improves the results considerably and three regimes are even better. Increasing the number of regimes to four or five provides only a slight decrease in the AIC. The final decision on the exact number of regimes can only be made after examining the results of the estimations. Increasing the number of regimes improves the fit of the model in the sample, but a more complicated model may not be practical for extending the results beyond the used data.

6.5.3 Parameter Restrictions

The estimations above allowed all the parameters to be different in different regimes. This need not be the case; it is possible to restrict some of the parameters to have the same value in all states. For example, one can assume that only the intercept is different in different states, while other parameters remain constant. Table 6.2 shows the results of different parameter restrictions with two regimes. Two regimes are chosen in order to avoid problems arising from the large number of parameters. The names of the models refer to different parameter restrictions and are from Krolzig (1997). MS refers to Markov switching and the other letters indicate the varying parameters. So, I = intercept ν, A = autoregression parameter ρ, and H = heteroscedasticity with parameter σ. The results indicate that the most important feature to include in the model is heteroscedasticity, since the models with varying σ do much better than the other ones. Of much lesser importance seems to be the mean reversion, which is captured by ρ, while changes in the intercept are least important. Since the differences between the three models with varying volatility are 'small', one needs to look at the actual estimation results to be able to evaluate the meaning of this finding. When reading these results one should keep in mind that AIC imposes a penalty as the number of parameters increases.

Name	MSIAH	MSIA	MSI	MSA	MSH	MSAH
varying constant	ν, ρ, σ	ν, ρ σ	ν ρ, σ	ρ ν, σ	σ ν, ρ	P, σ ν
AIC	6.7716	10.9630	12.0616	11.0932	6.7751	6.7715

Table 6.2. The Akaike Information Criterion for different parameter restrictions.

6.5.4 Comparison of Three and Four Regimes

The conclusion from the tables (6.1 and 6.2) above is that we want to have three or four regimes and varying heteroscedasticity. The earlier discussion about nonlinear mean reversion suggests that, a priori, varying autoregression should be included as well. This is also essential for finding a unit root regime. The Turkish overnight interest rate shows very large jumps occasionally (see Figure 5.1), so we also include the varying ν, thus ending up with the MSIAH model with all parameters varying. The estimated parameter values are given for three regimes in Table 6.3

At first glance, the results with three regimes show a unit-root regime (1), a mean reverting regime (2), and a very high volatility regime (3). A closer look at regime 1 shows that this is rather a regime of almost constant interest rate. These results are shown as the parameters of the Vasicek model ($\Delta r_t = \beta(\alpha - r_{t-1})\Delta t + \sigma\varepsilon_t\sqrt{\Delta t}$) in Table 6.4.

Table 6.5 shows the estimates of the parameters with four regimes. Now there are essentially two unit root regimes (1 and 2), a mean reverting regime (3), and a very high volatility, or jumpy, regime (4), which can also be called an outlier regime. Regime 4 occurs only very rarely with only 14 observations in the sample of 1384.

Regime	ν	P	σ	n
1	0.0133 (0.0272)	0.9998 (0.0004)	0.049011	589.5
2	8.3393 (1.2300)	0.8608 (0.0199)	8.6144	770.2
3	229.8035 (0.0002)	0.5308 (0.1500)	733.29	24.3

Table 6.3. The estimated parameter values, their standard errors (), and the estimated number of observations per regime. M=3. AIC=3.6257.

Regime	B	a	σ
1	(*)	(*)	0.049011
2	0.1392	59.9088	8.6144
3	0.4692	489.7772	733.29

Table 6.4. The estimated parameters of the Vasicek model. M=3.(*) indicates (near) unit root behavior with no mean reversion.

The estimation results can be written again in the form of a Vasicek equation. These parameters are shown in Table 6.6. The model with four regimes (M=4) is intuitively more appealing and it fits the data better with AIC=3.3048 compared to the AIC=3.6257 with three regimes. The disadvantage of increasing regimes from three to four is that the number of parameters increases (with this MSIAH specification) from 15 to 24.

Regime	V	ρ	σ	n
1	0.0104 (0.0174)	0.9998 (0.0002)	0.039811	548.8
2	0.3981 (0.5861)	0.9945 (0.0081)	1.5996	309.3
3	14.6627 (1.8199)	0.7426 (0.0300)	11.226	512
4	450.2739 (0.0003)	0.4386 (0.1913)	933.90	14

Table 6.5. The estimated parameter values, their standard errors (), and the estimated number of observations per regime. M=4. AIC=3.3408

Regime	B	a	σ
1	(*)	(*)	0.039811
2	(0.0055) (*)	(72.3818) (*)	1.5996
3	0.2574	56.9646	11.226
4	0.5614	802.0554	933.90

Table 6.6. The estimated parameters of the Vasicek model. M=4. (*) indicates (near) unit root behavior with no mean reversion.

6.6 Estimation Results

We have chosen the model with all parameters different across regimes. The order of autoregression is set at one. which gives us a Vasicek model. and the number of regimes is four. Table 6.5 shows the regimes that are the result of the estimation process and the respective Vasicek parameters are shown in Table 6.6. We comment on the estimated results based on Tables 6.5 and 6.6 for each regime individually.

- In regime 1. the interest rate has a unit root. The autoregression parameter is ρ=0.9998 with a standard error σ_ρ=0.0002. This indicates that unit root cannot be rejected at 10 percent significance level. The volatility parameter. which is the standard error of the regression, is σ= 0.039811, and is very small. The intercept is ν=0.0104 which is relatively small and statistically not significantly different from zero. In this regime the interest rate is essentially constant.
- Regime 2 is basically a unit root regime with ρ=0.9945 and σ_ρ=0.0081 since the unit root cannot be rejected at 10 percent level. Using the estimated parameter values gives mean reversion towards 72.38 percent. but it is not based on statistically significant parameter values. This regime has higher volatility than the first one at σ= 1.5996 percent.
- Regime 3 exhibits quite moderate mean reversion (β=0.2574) towards 56.965 percent rate. The volatility is quite high in this regime with 11.226 percent.
- Regime 4 has extremely high volatility at 933.90 percent with moderate mean reversion (β=0.5614) towards 802.0554 percent level. However. the number of observations in this regime is very low. thus giving a very low probability that this state occurs. and making the estimates unreliable. This regime could rather be

characterized as a jump regime. Sometimes a regime like this is called an outlier regime. Whether these observations should be excluded from the data as outliers is a difficult question. Our belief is that, although unusual, the regime can still occur again, albeit with a very low probability, so we do not exclude these observations.

We also estimated the smoothed regime probabilities for all the dates and all four regimes $\left\{\hat{\xi}_{t|T}\right\}_{t=0}^{T}$, which is a matrix with a column referring to each regime. From this matrix, it is possible to find the expected regime classification by choosing for each t the regime with the highest probability (not shown). The smoothed regime probabilities for all four regimes are shown in Appendix 2. We also calculated the corresponding means, or 'expected values', for the interest rate for each regime using these smoothed probabilities. This is calculated, for regime m as

$$\bar{r}_m = \bar{r} \mid (s_t = m, R_T) = \frac{\sum_{t=0}^{T} \Pr(\xi_t = \iota_m \mid R_T) r_t}{\sum_{t=0}^{T} \Pr(\xi_t = \iota_m \mid R_T)} \tag{6.46}$$

The conditional means for the regimes are shown in Table 6.7. It is important to keep in mind that the regimes have different autocorrelation parameters, and one should be cautious when reconstituting the regime-conditional distributions this way. The values in Table 6.7 rather give information on how the regimes are associated with this sample. The results, excluding the rare jumps, indicate that at high interest rates the system is in a random walk regime, while the regime with low interest rates shows mean reversion. Additionally as interest rates are lower, volatility increases. In order to understand the process better we need to look at the transition probabilities, given by the estimated matrix $\hat{P}$, which are shown in Table 6.8. State *i* is on the left side, and state *j* at the top.

Regime	1	2	3	4
Mean	72.2	66.6	58.3	790.6

Table 6.7. The conditional means of the regimes (calculated using the smoothed probabilities).

i \ *j*	1	2	3	4
1	0.8648	0.1082	0.02700	0.00000000003954
2	0.2186	0.5718	0.2097	0.00000000005513
3	0.01317	0.1408	0.8397	0.006345
4	0.00003018	0.06331	0.1693	0.7674

Table 6.8. The estimated regime transition probabilities.

In the transition probabilities, we see persistence: once the economy is in some regime, that same regime is the most likely state for the next period. Regime 1 is most persistent, and if the state changes it is most likely to change to regime 2 which is also a unit root regime, but with higher, although relatively low, volatility. If in regime 2, the system is most likely to stay there, but if it moves away, it is as likely to go to regime 1 or regime 2. If the economy is in the mean reverting regime (3), it stays there with a high probability (p_{33}=0.8397). Regime 4 can be, practically speaking, attained except obviously from itself, only from regime 3. The moves from regimes are very likely to occur through the adjacent regime (using our ordering). Looking at volatility, this means that the system moves from low to high volatility, and vice versa, through the regimes in the middle. We can also think that any mean reversion in volatility is shown as a move from regime 3 to 2, and then 1. The assumptions of the Markov chain were discussed earlier. The very low probabilities for getting into regime 4 are not desirable, but even if the number of regimes is reduced to three, one of them will only include the outliers like regime 4 here. Due to the ergodicity of matrix P we can estimate the unconditional probabilities of the regime realizations. These are the ergodic probabilities of the regimes ξ^*, which are shown in table 6.9.

Regime	n (total=1384)	ξ^*	Duration (days)
1	548.7	0.3971	7.40
2	309.3	0.2233	2.34
3	512.0	0.3696	6.24
4	14.0	0.0101	4.30

Table 6.9. The estimated number of days spent in each regime (n), the ergodic probabilities (ξ^*), and the duration spent in a regime continuously.

The ergodic probabilities can be interpreted as the relative frequencies for being in different regimes. The Turkish spot rate spends 37.0 percent of the time in the mean-reverting regime with low rates and high volatility, 39.7 percent in the (almost) constant regime, and 22.3 percent in the random walk regime with low volatility. The outlier regime occurs only 1 percent of the time with our data.

For every date, the model estimates the probabilities of being in each of the regimes. The outlier regime, regime 4, has the highest probability of all regimes only on the dates November 28 – December 6, 2000, December 21, 2000, and February 20 – 26, 2001. Two of these periods relate to crises explained in Chapter 3; they were also found as structural breaks in Chapter 5 using the Bai-Perron model. The third period, the day of December 21, 2000, could be related to the crisis few weeks earlier, but at that moment IMF approved loans for Turkey.

The smoothed probabilities for the regimes for each day in the sample are shown as graphs in Appendix 2. The estimated average duration of a regime is short, so changes in probabilities, and thus switches of regimes, occur frequently. Since the regimes cannot be observed directly, this duration refers to the time that the regime has been the most probable one continuously. For example, if there were two regimes, and regime 1 has a probability of 0.51 on all odd days, and regime 2 has probability of 0.51 on all even days, the average duration is one day for both regimes.

6.7 Evaluating the Model

The Markov switching model is a great improvement over the (single-regime) Vasicek model; this is shown by the reduction in volatility and the AIC. We saw earlier with two regimes (in Table 6.2) that allowing all parameters to vary provided practically no improvement over a model in which only volatility is different across regimes. In order to find out whether this is the case with four regimes, we estimated a four-regime MSH model; that is keeping all the other parameters except volatility σ constant. This model performs well, but not as well as the MSIAH version with an AIC=3.3912 vs. 3.3408. (Both Hannan-Quinn Information Criterion and Schwartz Information Criterion select the MSIAH specification over the MSH model.) Additionally, the MSH model showed slightly less normality and more autocorrelation of the residuals. Interestingly, the MSH model turns out to be a random walk with an autocorrelation parameter ρ=0.9998 (standard error of 0.0002). The number of days spent in each regime, the ergodic probabilities, and even the regime probabilities for individual days are almost identical to the ones in the MSIAH model. The standard errors, or volatilities, are similar as well at 0.039425, 1.5381, 11.804, and 1029.4 for regimes 1, 2, 3, and 4, respectively. It is not surprising that a random walk with only varying volatility does so well. Vast majority of research in the area considers modeling volatility more important than modeling the drift. Ball and Torous (1999) find that in short rate models with stochastic volatility, volatility estimates are not very sensitive to the drift specification. Duffee (2002) shows that a random walk is superior to most affine short rate models with different drift specifications. Hong, Li, and Zhao (2002) emphasize the importance of modeling conditional heteroscedasticity and conclude that a random walk model without drift outperforms a linear drift model. Another interesting point is raised by Li and Xu (2002); because regime switching models capture some of the effects of heteroscedasticity, the level effects are diminished.

The model can be checked by examining the residuals. Ideally, we want to find them iid, and normally distributed. There are various ways for measuring the residuals, depending on which information set is used for the regime probabilities, how the residual is standardized and measured, etc. The estimation of the smoothed residuals was shown in (6.42)-(6.44). The smoothed residuals are defined as:

$$u_t = y_t - E[y_t \mid \xi_t = \hat{\xi}_{t|T}, R_{t-1}; \lambda = \tilde{\lambda}]. \qquad (6.47)$$

For a four-regime AR(1) model the smoothed residuals for regime j can be calculated using the smoothed probabilities

$$\tilde{u}_{jt} = \sum_{i=1}^{4} \{[r_t - E[r_t \mid s_t = j, s_{t-1} = i, R_{t-1}]]\Pr(s_t = j \mid s_{t-1} = i, R_T\}. \qquad (6.48)$$

The residuals in each regime have (almost) zero mean, but their standard deviations are different. On a given date, more than one regime may occur with positive probability. We can standardize these regime-conditional residuals by dividing them by their estimated standard deviations:

$$\bar{u}_{jt} = \frac{\tilde{u}_{jt}}{\tilde{\sigma}_j}. \qquad (6.49)$$

To get an estimate for the regime-unconditional residuals we calculate the weighted average of the regime-conditional residuals using the smoothed probabilities as weights

$$\bar{u}_t = \sum_{j=1}^{4} \bar{u}_{jt} \Pr(s_t = j \mid R_T) \tag{6.50}$$

to get what are called the 'standardized smoothed residuals'.

The standardized smoothed residuals are shown in Figure 6.1 with a curve of the normal distribution with the same standard deviation superimposed on the graph. A QQ-plot of the same residuals is shown in Figure 6.2. A visual inspection of the graphs indicates that the standardized smoothed residuals are normally distributed. The skewness is –0.01304 and kurtosis 3.773389, which indicates nearly normality.

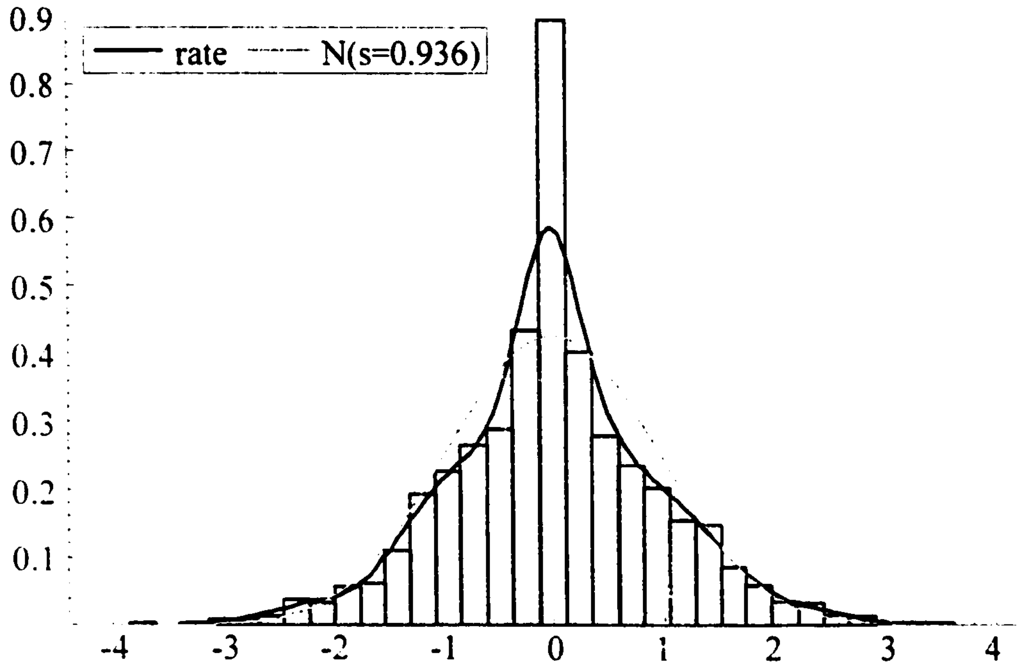

Figure 6.1. The unconditional probability density distribution of the standardized smoothed residuals. A normal density curve with the same standard deviation is superimposed on the graph.

It is important to note that the residuals shown in Figures 6.1 and 6.2 are weighted averages of standardized distributions. If one calculates the residuals for each day by subtracting the expected value from the observed value, the result is very different. One problem with the described standardizing procedure is that it tends to make the residuals 'behave better' than they do without it. This may lead to a bias when evaluating the model by checking the residuals.

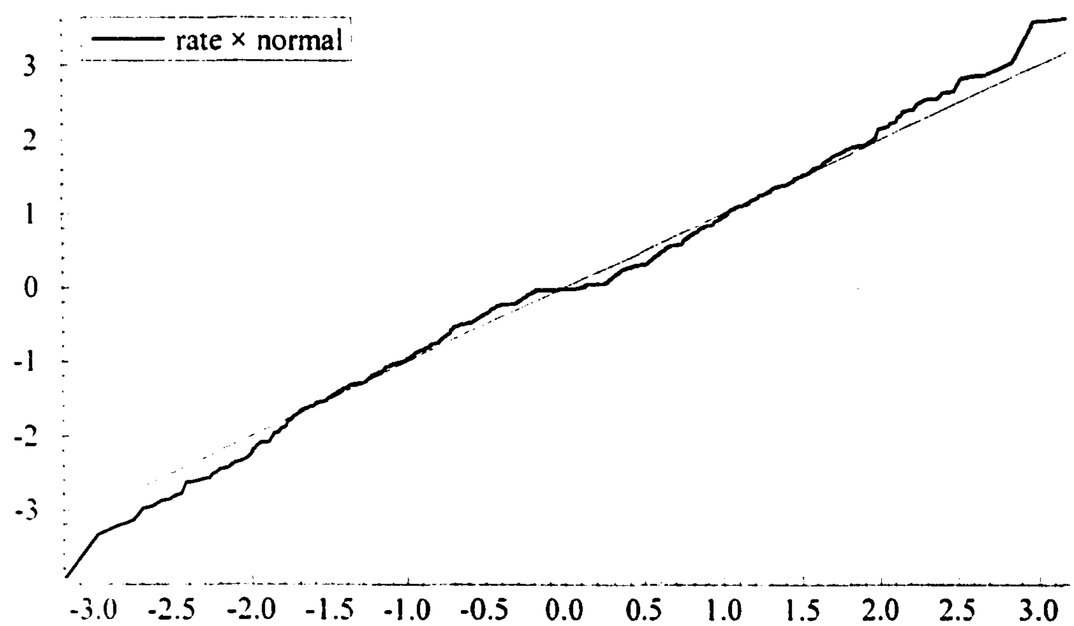

Figure 6.2. The QQ-plot of the standardized smoothed residuals.

In a well-specified model, the residuals should not only be normal, but also show no autocorrelation. The autocorrelation (ACF) and partial autocorrelation (PACF) functions of the standardized residuals for different lags are shown in Figure 6.3. Ideally, autocorrelations of the residuals should be close to zero. The shown pattern of the autocorrelations is quite common, and may indicate that the estimated timing of the switches is not perfect. It could be tempting to analyze the autocorrelations using e.g. the Box-Pierce statistic (which gives a value of Q(1)=28.95 rejecting 'no autocorrelation'). This is not recommended for two reasons - the residuals are standardized, and the model is not linear. Thus, the asymptotic distribution of the test statistic is not known. We have however included in Figure 6.3 the horizontal lines at ± 0.053. This value corresponds to two standard deviations of a normally distributed sample ACF and PACF with a standard deviation of $1/\sqrt{T}$. However, the first-order autocorrelation is less than –0.15, which may not be too high, especially in a regime switching model.

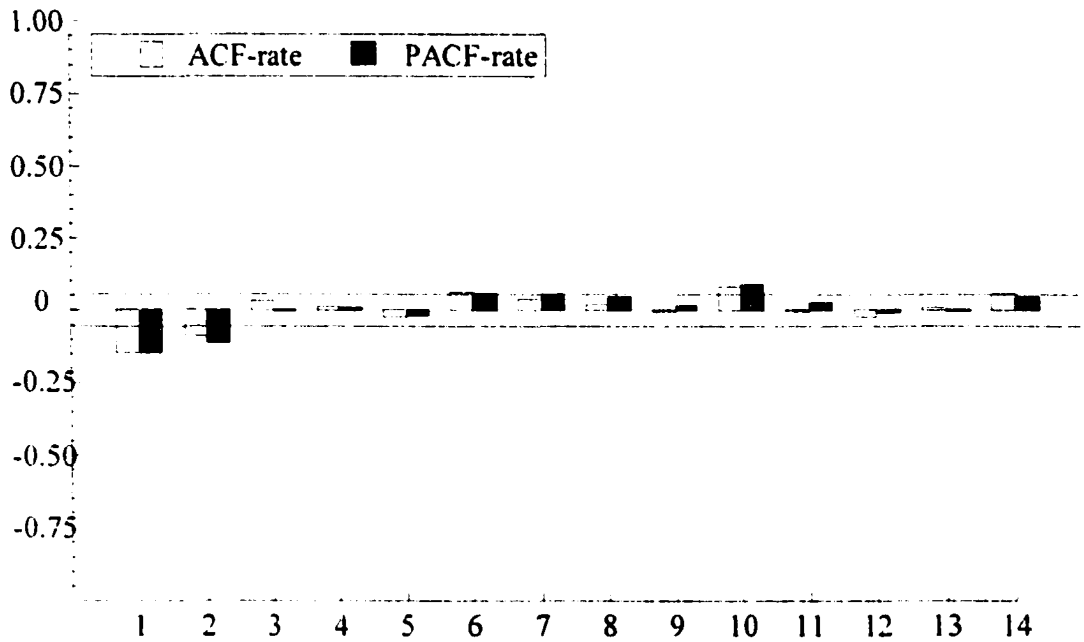

Figure 6.3. Autocorrelation (ACF) and partial autocorrelation (PACF) functions of the standardized smoothed residuals for different lags. The added horizontal lines are at +/- 0.053.

A more general approach to examining the residuals is to test if they are iid. This can be done using the correlation integral of Grassberger and Procaccia (1983a, 1983b), which measures the frequency by which patterns in data are repeated. The null hypothesis, that the data is generated by an iid stochastic process. If the correlation integral finds a high number of repetitive patterns, the null hypothesis is rejected.

The BDS test by Brock, Dechert, and Scheinkman (1987) and Brock, Dechert, Scheinkman, and LeBaron (1996) is based on this idea. A short description of the test is given in Appendix 3. A major advantage of this test is that is that it only assumes that the distribution is iid under the null, higher moments of the distribution do not even need to exist. Although tests with higher power exist for specific alternatives, the BDS test has the benefit that the alternative hypothesis need not be specified. It has power against virtually all kinds of linear and nonlinear stochastic and deterministic systems.

We apply the BDS test to the standardized residuals. The choice of the model parameters (see Appendix 3), which is subjective, is studied in Kanzler (1999), which includes the tables with the critical values of the DBS statistic for various parameter values and sample sizes. It is recommended that the test be carried out using several different parameter values. If the distribution is truly iid, the test statistic should be close to zero. The results of our estimations using ε/σ values of 1.0, 1.5, and 2.0 are shown in Table 6.10 (σ is the standard deviation of the estimated standardized residuals). Note that the BDS statistic has an asymptotic standard normal distribution N(0,1). The BDS test clearly rejects the null hypothesis that the distribution of the standardized smoothed residuals is iid.

ε/σ	m=2	m=4	m=6
1.0	7.37	10.20	11.77
1.5	7.64	10.25	11.16
2.0	7.43	9.46	9.54

Table 6.10. The BDS statistics for the time series of the standardized smoothed residuals.

An interesting question is whether the regime-conditional distributions are Gaussian. We calculated for each regime i the conditional skewness ($\tau_i=(\tau|s=i)$) and kurtosis ($\kappa_i=(\kappa|s=i)$) using (following Campbell 2001)

$$\tau_i = \frac{1}{\hat{\sigma}_i^3 \sum_{t=0}^{T} \Pr(s_t = i \mid T)} \sum_{t=0}^{T} \Pr(s_t = i \mid T)(\tilde{u}_{it} - \bar{u}_i)^3, \text{and} \tag{6.51}$$

$$\kappa_i = \frac{1}{\hat{\sigma}_i^4 \sum_{t=0}^{T} \Pr(s_t = i \mid T)} \sum_{t=0}^{T} \Pr(s_t = i \mid T)(\tilde{u}_{it} - \bar{u}_i)^4. \tag{6.52}$$

where $\bar{u}_i$ is the regime-conditional mean of the residuals, which should be zero. The Bera-Jarque (BJ) statistic's weakness is that it is accurate only if the distribution has thin tails (see Chapter 5), so the results are shown for reference only. Note that for the 'sample size' for each regime, one has to use

$$T_i = \sum_{t=0}^{T} \Pr(s_t = i \mid T) . \tag{6.53}$$

The results for the four regimes are shown in Table 6.11. The residual distributions in regimes 1, 2, and 3 are practically symmetrical - and regime 4 is the 'jump regime'. While regime 1 is clearly heavy-tailed, regimes 2 and 3 are only moderately so. With typical significance levels, the BJ statistics suggests that normality be rejected in all regimes. However, the results are a vast improvement over the single-regime Vasicek model, whose residuals have skewness of 18.486 and kurtosis of 525.583.

We analyze here the shapes of the distributions of the estimated residuals, not the actual distributions of the interest rates. We have given above the informal conditional means of the distributions. If one wants to find the informal variances of the regime-conditional interest rates one should use equations (2.76) and (2.77). However, deriving the regime-conditional distributions in a switching model with varying autocorrelations is not necessarily a good practice.

	1	2	3	4
Skewness	0.05347	0.05521	-0.00855	1.87056
Kurtosis	5.11938	3.79649	3.97915	5.47712
Bera-Jarque (p-value)	102.8 (<0.000005)	8.3 (0.016)	20.5 (0.00004)	11.7 (0.0029)

Table 6.11. The regime-conditional skewness and kurtosis.

We showed earlier in Chapter 5 that the our time series is weakly stationary based on various unit-root and stationarity tests. However, these tests are not necessarily reliable in our model. Generally, with a Markov switching process, a sufficient condition for stationarity is that the transition matrix is ergodic and all the submodels, or regimes, are stationary (Krolzig 1997). Karlsen (1990) finds a sufficient condition for weak stationarity, which according to Warne (1996) may actually be a necessary condition. The condition, in our model, is that the largest eigenvalue of the matrix

$$B^* = \begin{bmatrix} \rho_1^2 p_{11} & \rho_2^2 p_{12} & \rho_3^2 p_{13} & \rho_4^2 p_{14} \\ \rho_1^2 p_{21} & \rho_2^2 p_{22} & \rho_3^2 p_{23} & \rho_4^2 p_{24} \\ \rho_1^2 p_{31} & \rho_2^2 p_{32} & \rho_3^2 p_{33} & \rho_4^2 p_{34} \\ \rho_1^2 p_{41} & \rho_2^2 p_{42} & \rho_3^2 p_{43} & \rho_4^2 p_{44} \end{bmatrix} \tag{6.54}$$

is less than one. Inserting the estimated transition probabilities (Table 6.8) and autocorrelation parameters (Table 6.5) into (6.53) gives us four eigenvalues with the largest one of 0.938. Thus, the regime switching process is weakly stationary. This could be, of course concluded from the ergodicity of the probability matrix together with the fact that all the autocorrelation parameters are less than one. Yang (2000) also derives a formal condition for the weak stationarity of a Markov switching process.

6.8. The Term Structure with Markov Switching

The number of regime switching short rate models is quite numerous, but few papers extend the model to the yield curve and the ones, which do generally deal with simple cases. Naik and Lee (1997) examine the term structure with regime switching volatility. Hansen and Poulsen (2000) derive a closed-form solution for bond prices in a two-regime Vasicek model using the equivalent martingale measure. Switching between regimes is generated by a Poisson process, and only the mean reversion level of the interest rate is allowed to change. They show that this model can produce non-parallel shifts of the term structure as well as changes in the shape of the curve. Elliott and Mamon (2000a) also derive the term structure of a Vasicek model with the mean reversion level parameter following a Markov chain. In another paper, they (Elliott and Mamon 2000b) derive the yield curve generally from the assumption that the short rate is a function of a Markov chain. Dai and Singleton (2002) examine some of the mathematics behind the term structure as well. Very few empirical studies have been carried out on term structures using regime switching.

6.9 Conclusions

In this chapter, we have examined the Turkish overnight interbank interest rate using a regime switching autoregressive Hidden Markov Model version of the Vasicek (1977) model. We first estimated the single-regime Vasicek model as a point of reference. The volatility was so high that the probability of negative interest rates is a serious problem. Additionally, the estimated residuals were skewed and leptokurtic.

We proceeded to estimate the regime switching model with various numbers of regimes and different parameter restrictions. The best results were achieved by letting all the parameters, level, and rate of mean reversion, as well as volatility vary across regimes. The optimum number of regimes was found to be four. One of the regimes occurs very rarely, being a jump or 'outlier regime'. Two regimes are (nearly) random walks, both of them having low volatilities. (Note that mean reversion is present in this model always through regime shifts.) One of them is practically speaking a regime of constant interest rate. The remaining regime shows mean reversion and high volatility. The interest rate spends only 37 percent of time in the mean reverting regime. The average time spent continuously in any regime is very short, so regime switches occur frequently. This makes it very difficult to predict the short rate, but is not necessarily a problem if one tries to understand the behavior of the term structure. At low interest rates volatility and mean reversion are high. At higher interest rates, the process follows a random walk with lower volatility. Volatility was found to be the most important single variable in the model. Letting only volatility vary leads to a model that is only slightly inferior to the chosen one where all parameters are allowed to vary.

We examined the model by checking the behavior of the regime-conditional and unconditional residuals. The regime-conditional distributions of the residuals are close to normal, symmetrical but with fat tails. For checking the unconditional residuals, the standardized smoothed residuals are used. This means that first the regime-conditional residuals are divided by the standard deviation of the regime, and then added up together over the regimes using the smoothed probabilities as weights. The standardized smoothed residuals were found to be normally distributed, but slightly autocorrelated and not iid. It is difficult to assess the importance of this finding, because the properties of regime switching models are not well known and these residuals are standardized in the way mentioned above. We think that the model is not able to produce the nonlinearities to the extent they are present in the data.

The properties of regime switching models are not well known and there are important issues to consider. A well-known problem is the convergence difficulties of the likelihood function due to the presence of local extreme points. Typically, one tries to overcome this by experimenting with different values. Most commonly, the overall model is assessed by checking the residuals. The typical statistical tests (e.g. Bera-Jarque, Box-Price and Likelihood Ratio tests) give statistics which are numerical, thus giving the impression of accuracy, but their behavior in nonlinear models is often not well known, and they should not be used.

A different approach to testing the model is to use the estimated model to create data using simulations. Then the simulated data, e.g. its conditional and unconditional moments or densities, can be compared to the sample. One can also try how well the model that is estimated using the sample, can explain out-of-sample.

7 Concluding Remarks

In this dissertation, we have studied the behavior of the Turkish overnight interbank interest rate. The Turkish economy is less stable than most others are. The inflation rate and interest rates are often high and volatile, and depreciation of the currency has been fast. Most studies on interest rate behavior use well-known data and research on the Turkish interest rate seems to have a macroeconomic perspective. Consequently, the Turkish short has not been studied extensively and its characteristics are not well known.

In Chapter 2, we have seen the importance of the short interest rate for the whole term structure of interest rates. However, one-factor interest rate models have often not performed well in empirical studies. In Chapter 4, we used the principal component analysis (PCA) to study the number of factors that drive the Turkish yield curve. We found that a single factor seems to explain less of the variation of the forward rates (and yields) than what is found in most other studies using data from other countries. This result indicates that a traditional one-factor model is not sufficient for explaining the behavior of the Turkish yield curve. It is possible that the outcome is partly due to the structure of the Turkish bond price index used as the data. However, PCA itself is a second-order method, and does not use information beyond the first two moments of the data. Thus, it is not able to capture the nonlinearities, which we feel are present in the Turkish market.

In Chapter 5, we found that the Turkish overnight interbank interest rate distribution is very skewed and heavy tailed. It also shows high volatility and volatility clustering. The time series is weakly stationary, indicating mean reversion. The Bai-Perron (1998) test found brakes, potentially ones that are caused by a Markov switching process.

In Chapter 6, we applied the normal one-regime Vasicek model to the data. The results were poor. The residuals were very skewed and leptokurtic and the volatility was so high that negative interest rates are very possible. Then we applied a hidden Markov chain model to the data. The idea is that the short rate always follows the Vasicek model, but the economy switches, following a Markov chain, between different states, and the Vasicek parameters can be different in each of them. The states cannot be observed but an inference of the state is made by observing the (past) history of the short rate. The regime switching model performed much better than the single-regime one. We found four regimes, two of which were (near) unit root and very low volatility. A third regime was mean-reverting with higher volatility and the fourth one an outlier regime, which includes the few observations of very high interest rates. The volatility of the daily changes in the interest rate decreased as the interest rate got higher. Modeling volatility turned out to be the most important aspect of this model. It is important to keep in mind that even in a unit-root regime, there is a positive probability of switching to a mean reverting regime, so the mean reversion is always present to some extent.

We evaluated the regime switching model by checking the estimated residuals. The regime-conditional residuals were close to normal, but with fat tails. The overall residuals were represented by standardized smoothed residuals. These standardized residuals were normally distributed, but slightly autocorrelated and the BDS test indicated that they were not independent.

This dissertation raises many questions, many of which have been mentioned already in various other studies. There is much more research to do concerning the Turkish short rate and term structure of interest rates. Applying different known and elsewhere tested, short rate and term structure models to the Turkish data would be useful. Another interesting question is the term premium and the differences between bond yields and other interest rates.

Most naturally, the estimated model should be tested using out-of-sample data to see how well it performs. Generally, regime switching models perform well in-sample and poorly out-of-sample. Another approach is to run simulations using the estimated model to see what kind of distributions it creates, and whether the model can be used for forecasting. These simulations could be extended to discovering the term structure of interest rates and then seeing if the result is a good match with the term structure data. Another route for finding the term structure would be to derive a model explaining the term structure directly rather than only the spot rates, and then test it on empirical data. In general, the behavior of the regime switching model, choosing its specifications, stability, testing etc., is not well known.

The PCA raises some questions as well. First, since much of recent research has addressed the nonlinearity issue of interest rates, one may want to ask if PCA is an appropriate tool for studying them at all. PCA is used in the term structure literature because of its intuitive appeal, but other methods are able to detect dependencies that are more complicated. One example, which has been used on interest rates by Molgedey and Galic (2000), is the Independent Component Analysis (ICA). On interest rate data, it is possible to use, for example, both interest rate levels and changes in interest rates as inputs at the same time.

The amount of theoretical and empirical research on the stochastic behavior interest rates is growing at a very fast pace. This was clearly observed during this study; new studies emerge on almost weekly basis. They may provide new approaches to the issues discussed in this dissertation in a very short period.

Appendix 1: Additional Unit Root Tests

The Augmented Dickey-Fuller Test with Linear Trend

The Augmented Dickey-Fuller (ADF) test with a linear trend has a null hypothesis that the process has a unit root and an alternative hypothesis that it is linear trend stationary (LTS). The auxiliary model is

$$\begin{aligned} y_t - y_{t-1} &= \alpha y_{t-1} + \beta_1(y_{t-1} - y_{t-2}) + \ldots + \beta_p(y_{t-p} - y_{t-p-1}) \\ &\quad + \beta_{p+1} + \beta_{p+2}t + \varepsilon_t. \end{aligned} \tag{A1.1}$$

The second line shows the intercept and trend-terms in addition to the white noise. The null and alternative hypotheses are the same as in the ADF test in presented in the text (Chapter 5). The AIC is minimized with the same number of lags as in the first test, 15 for r_t and 62 for Δr_t. The t-values with these lags are −8.8377 for r_t and −11.2134 for Δr_t with $p<0.000001$. Again, the unit root is clearly rejected.

The Phillips-Perron Test

The Phillips-Perron (PP) test is more general than the ADF tests, because it allows for heteroscedasticity. The auxiliary model for the zero-drift case is written

$$y_t = \beta + \alpha y_{t-1} + \varepsilon_t \, . \tag{A1.2}$$

where the error term is a zero-mean stationary process. Under the null hypothesis the process y_t is a unit root with zero drift (H_0:α=1), while the alternative hypothesis is that the process is a zero-trend stationary process (H_1:α<1). In ADF tests, the requirement that the errors are white noise is solved by adding enough lags. The PP test achieves this goal by estimating the variance and then adjusting the DF test statistic itself. Generally, the (limiting) distribution of the Dickey-Fuller (DF) statistic is

$$\sigma_\varepsilon^2 = \lim_{T\to\infty} T^{-1}\sum_{j=1}^{T} E\left[\left(\sum_{i=1}^{t}\varepsilon_i\right)_j^2\right]. \qquad \text{(A1.3)}$$

A sufficient condition for using the DF statistic is that the variance of the statistic is equal to this. A complete description of the test can be found in Phillips and Perron (1988). The number of lags, which is used for estimating the adjustment of the DF statistic, is usually chosen to be low. We conduct the PP test with the null hypothesis of unit root with drift and the alternative hypothesis that the process is linear trend stationary. With seven lags the unit root is rejected for both r_t and Δr with p<0.000001.

The Bierens Non-Parametric Test with Nonlinear Trend

Bierens (1997) introduces a non-parametric ADF test for the unit root with the alternative hypothesis that the process is stationary with a nonlinear trend. We use the model that is called by Bierens the auxiliary model. It can be written in a general form as

$$y_t - y_{t-1} = \alpha y_{t-1} + \beta_1(y_{t-1} - y_{t-2}) + \ldots + \beta_p(y_{t-p} - y_{t-p-1}) +$$
$$+ \beta_{p+1} + \beta_{p+2}P_t(1) + \ldots + \beta_{p+m+1}P_t(m) + \varepsilon_t \,. \tag{A1.4}$$

where p determines the number of lags included. The second line shows the adjustment for the drift. The coefficient β_{p+1} is a constant. The terms $P(t,k)$ are m detrended Chebyshev time polynomials with standardized time (see Bierens 1997). Under the null hypothesis, y_t is a process with unit root and a linear drift. We have

$$H_0: \alpha = \beta_{p+2} = \beta_{p+3} = \ldots = \beta_{p+m+1} = 0 \tag{A1.5}$$

The condition $\alpha = 0$ is the standard requirement for having a unit root. The other conditions imply that the coefficients of the terms in the polynomial (which is a nonlinear drift) are to be zero. However, since $\beta_{p+1}=0$ is not required, under the null hypothesis linear trend is allowed. If the null hypothesis is rejected, the true process may be stationary with a linear or nonlinear trend, but it may still be a unit root process with a nonlinear trend. Under the alternative hypothesis, y_t is a nonlinear trend stationary process with $\alpha < 0$. Different test statistics can be used. The t-value of α is of course interesting under the alternative hypothesis. The condition for the null hypothesis (A1.5) can be tested as a joint hypothesis.

First, we find the optimum number of lags by minimizing the AIC assuming that H_0 holds. We choose the degree of the polynomial at m=2 and use p lags to find the OLS estimates and t-values of all the parameters of the auxiliary model. Additionally, we conduct a joint F-test for the null hypothesis.

The test is explained using the r_t series. According to AIC, the optimum number of lags p is zero. Applying OLS to the model with p=0 and m=2 gives the OLS estimates and t-values shown in table A1.1

Parameter	Estimated value	t-value	Regressor
α	-0.37664	-17.887	r_{t-1}
β_1	27.53816	8.859	r_t-r_{t-1}
β_2	0.87928	0.326	$P(1)$
β_3	0.19772	0.073	$P(2)$

Table A1.1. The estimated parameters under the alternative model. (p=0, m=2.)

The conclusion is that $\alpha<0$ with t=-17.887 and p<0.000001, so the process r_t is stationary. The last test is a joint F-test for H_0: $\alpha=\beta_2=\beta_3=0$. The test statistic $F_{1384,3}=106.646$ with p<0.00005. The conclusion is that r_t is not a unit root process with a linear drift, but stationary.

We also conducted the same tests for Δr_t using p=33, which is given by the AIC, and m=2. The result is t=-10.466 for α under H_1 with p<0.000005. The joint F-test uses H_0: $\alpha = \beta_{33} = \beta_{34} = 0$ and yields $F_{1383,33}$= 36.511 with p<0.000005. Again, the results indicate stationarity.

Appendix 2: The Smoothed Residuals

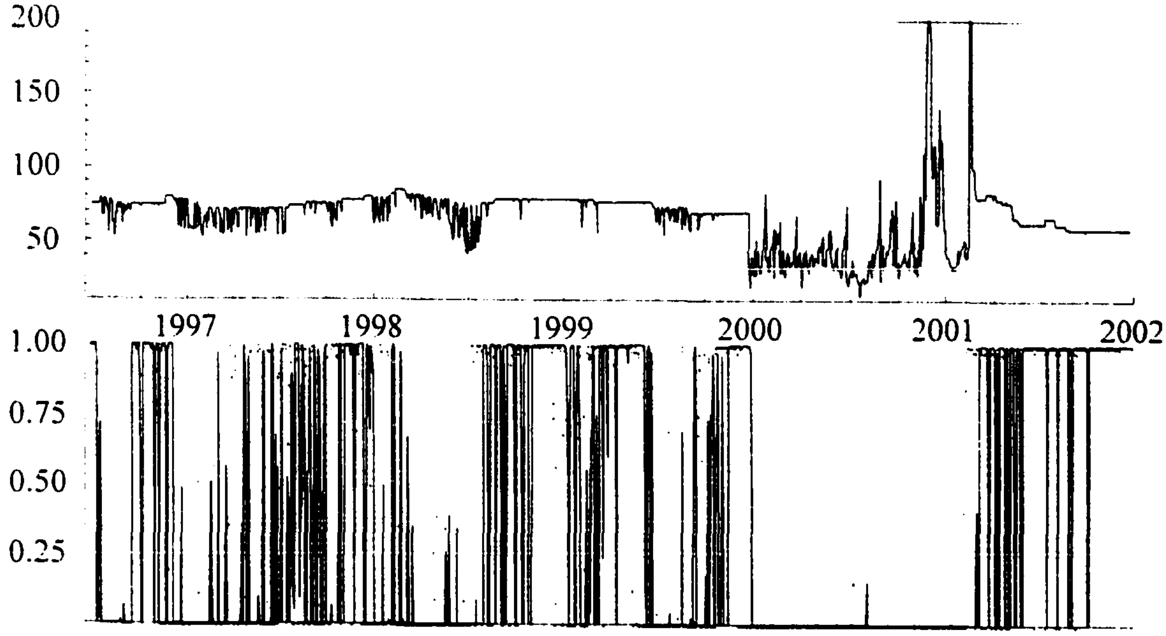

Figure A2.1. The smoothed probability of regime 1 and the short rate.

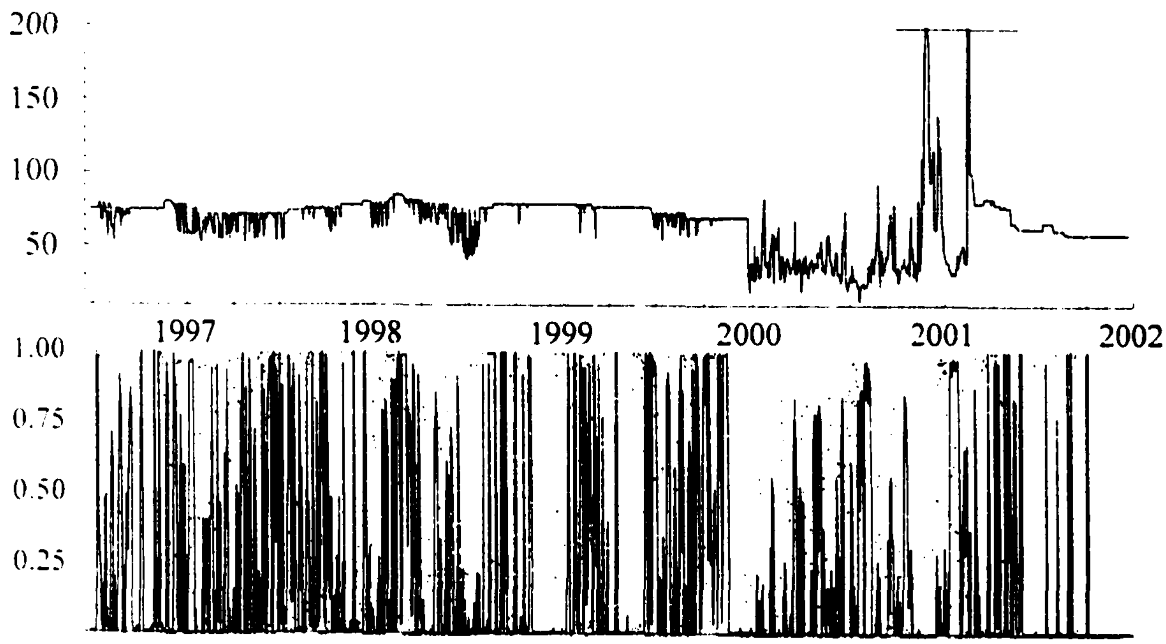

Figure A2.2. The smoothed probability of regime 2 and the short rate.

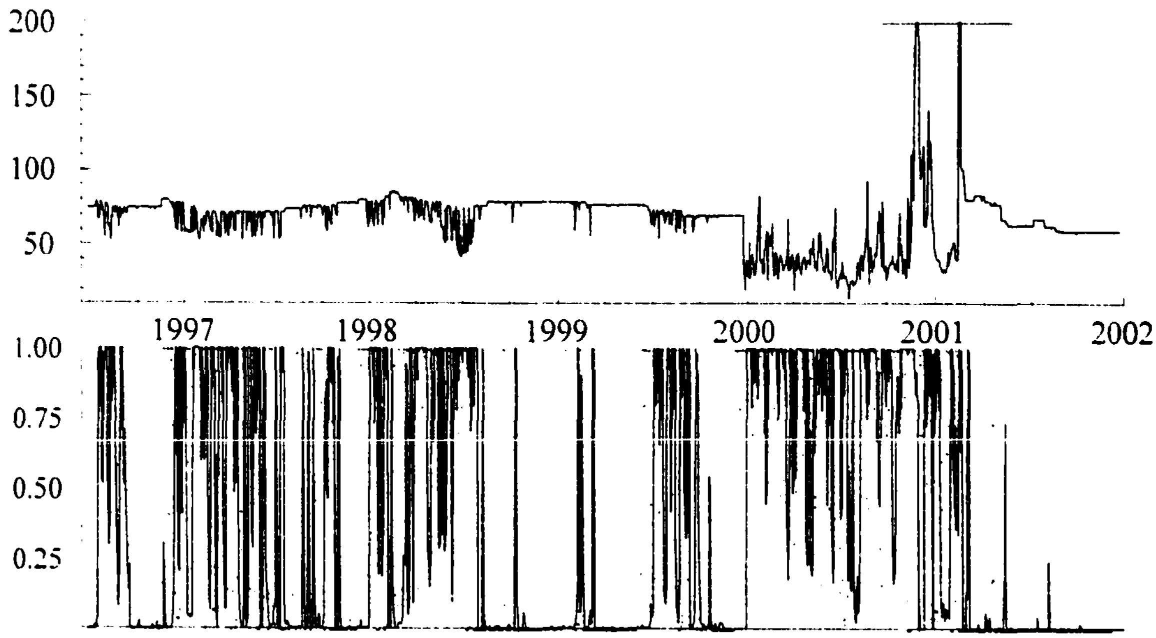

Figure A2.3. The smoothed probability of regime 3 and the short rate.

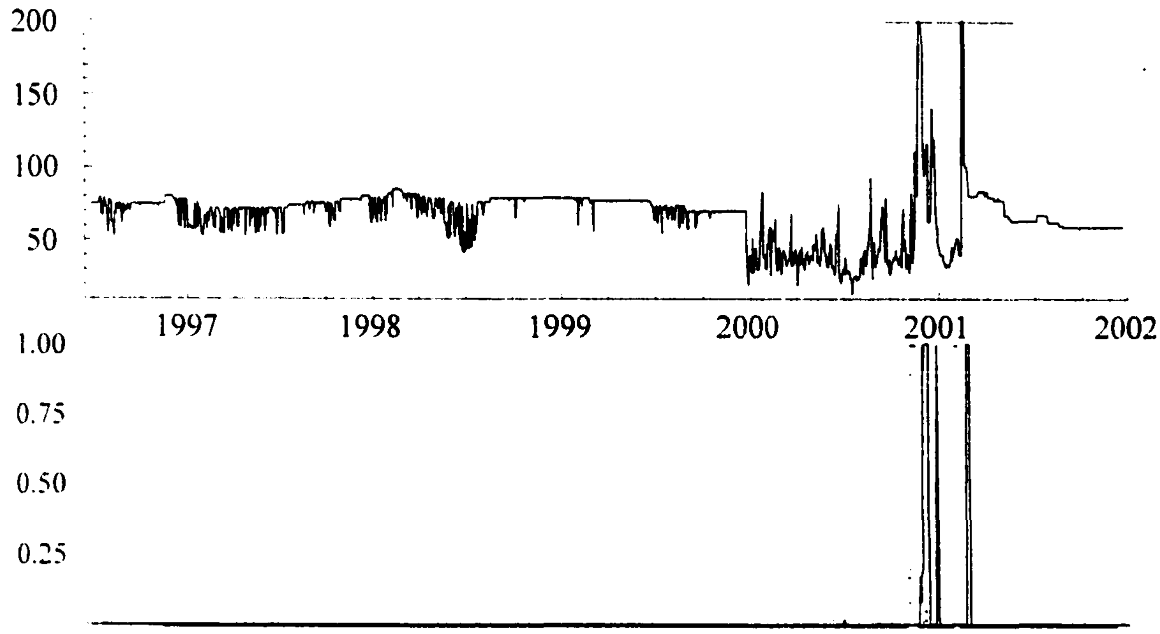

Figure A2.4. The smoothed probability of regime 4 and the short rate.

Appendix 3: The BDS Test

The mathematical derivation of the BDS (or BDSL) statistic is given in Brock, Dechert, Scheinkman, and LeBaron (1996). A good discussion of it, accompanied with tables for critical values, is in Kanzler (1999), which this brief outline follows.

Given a time series x, which follows some distribution F, we can choose an arbitrary distance ε so that

$$0 < \varepsilon < \max(x) - \min(x). \qquad \text{(A3.1)}$$

We can define the probability that a pair of two observations X_i and X_j lie within the distance ε of each other

$$P_1 \equiv P(|X_i - X_j| \le \varepsilon, \qquad \text{(A3.2)}$$

where i and j are integers referring to observations in the time series. The probability of a history of two observations being within ε of each other is

$$P_2 \equiv (|X_i - X_j| \le \varepsilon, |X_{i-1} - X_{j-1})| \le \varepsilon \qquad \text{(A3.3)}$$

If the time series is iid, then

$$P_2 = P_1^2. \qquad \text{(A3.4)}$$

This relationship can be extended to dimension m, an m-history, where

$$P_m \equiv (|X_i - X_j| \leq \varepsilon, |X_{i-1} - X_{j-1}| \leq \varepsilon, \ldots, |X_{i-m} - X_{j-m-1}| \leq \varepsilon \qquad \text{(A3.5)}$$

and

$$P_m = P_1^m . \qquad \text{(A3.6)}$$

The BDS test is based on this idea with a null hypothesis of

$$H_0 : P_m = P_1^m . \qquad \text{(A3.7)}$$

If the data is iid. H_0 cannot be rejected.

The correlation integral $c_{m,n}(\varepsilon)$ is used for estimating the probability P_m. In the sample consisting of n observations, the correlation integral (here in summation form for discrete time) for a chosen dimension m is calculated as

$$c_{m,n}(\varepsilon) \equiv \frac{2}{(n-m+1)(n-m)} \sum_{s=m}^{n} \sum_{t=s+1}^{n} \prod_{j=0}^{m-1} I_\varepsilon(X_{s-j}, X_{t-j}) , \qquad \text{(A3.8)}$$

where I_ε is a Heaviside function which takes value 1 if the points are within ε of each other and zero otherwise

$$I_\varepsilon(X_i, X_j) = 1 \quad \text{if} \quad |X_i - X_j| \le \varepsilon \text{, and 0 otherwise.} \tag{A3.9}$$

The BDS statistic is then

$$w_{m,n}(\varepsilon) = \sqrt{m-n+1}\left[\frac{c_{m,n}(\varepsilon) - (c_{1,n-m+1}(\varepsilon))^m}{\sigma_{m,n}(\varepsilon)}\right]. \tag{A3.10}$$

To find the standard deviation in the denominator of the bracketed term, the variance of the term in the numerator is estimated as

$$\sigma_{m,n}^2(\varepsilon) \equiv 4\left[k^m + 2\sum_{j=1}^{m-1} k^{m-j}c^{2j} + (m-1)^2 c^{2m} - m^2 k c^{2m-2}\right], \tag{A3.11}$$

where the correlation integral of dimension one is

$$c \equiv c_{1,n}(\varepsilon). \tag{A3.12}$$

The parameter k is the probability of three observations lying within distance ε of each other

$$k_n(\varepsilon) \equiv \frac{2}{n(n-1)(n-2)} \sum_{t=1}^{n} \sum_{s=t+1}^{n} \sum_{r=s+1}^{n} [I_\varepsilon(X_t, X_s) I_\varepsilon(X_s, X_r) + I_\varepsilon(X_t, X_r) I_\varepsilon(X_r X_s) + I_\varepsilon(X_s, X_t) I_\varepsilon(X_t, X_r)] \tag{A3.13}$$

The statistic has, for all m and n, an asymptotic distribution of

$$\lim_{n\to\infty} w_{m,n}(\varepsilon) \sim N(0,1) \quad \text{(A3.14)}$$

Then, asymptotically the critical values can be found from the standard normal distribution

Before starting, the researcher has to choose the parameters ε and m. This affects the results, so caution is required.

Bibliography

Adams, K. J. and van Deventer D. R. 1994, 'Fitting yield curves and forward rate curves with maximum smoothness', *Journal of Fixed Income*, 52-56.

Ahn, D. and Gao, B. 1999, 'A parametric nonlinear model of term structure dynamics', *Review of Financial Studies* 12, 721-762.

Aït-Sahalia, Y. 1996a, 'Nonparametric pricing of interest rate derivative securities,' *Econometrica* 64, 527-560.

Aït-Sahalia, Y. 1996b, 'Testing continuous-time models of the spot interest rate', *Review of Financial Studies*, 9, 385-426.

Aït-Sahalia, Y. 2002, 'Maximum-likelihood estimation of discretely-sampled diffusions: A closed-form approximation approach," *Econometrica*, 70, 223-262.

Alexander, C. 2001, '*Market Models: A Guide to Financial Data Analysis*', John Wiley & Sons Ltd, UK.

Andrews, D. W. K. 1993, "Tests for parameter instability and structural change with unknown change point', *Econometrica*, 61, 825-856.

Andrews, D. W. K. and Ploberger, W. 1993, 'Optimal tests when a nuisance parameter is present only under the alternative', *Econometrica*, 62, 1383-1414.

Ang, A. and Bekaert, G. 2002a, 'Regime switches in interest rates', *Journal of Business and Economic Statistics*', 20, 163-182.

Ang, A. and Bekaert, G. 2002b. 'Short rate nonlinearities and regime switches', *Journal of Economic Dynamics and Control*, 26, 1243-1274.

Bachelier, L. 1900, *Théorie de la Spéculation*, Annales de l'Ecole Normale Superiure (trans. *Random Character of Stock Market Prices*).

Back, K. 1997, 'Yield curve models; A mathematical review' in I. Nelken, ed.: *Option Embedded Bonds*, Irwin professional publishing, Chicago, USA.

Bai, J. and Ng, S. 2001, *Tests for Skewness, Kurtosis, and Normality for Time Series Data*, Working Paper, Boston College, Boston, USA.

Bai, J. and Perron, P. 1998, 'Estimating and testing linear models with multiple structural changes', *Econometrica*, 66, 47-78.

Bai, J. and Perron, P. 2002, 'Computation and analysis of multiple structural change models', *Journal of Applied Econometrics*, forthcoming.

Bali, T. G. 2000, 'Testing the empirical performance of stochastic volatility models of the short term interest rate' *Journal of Financial and Quantitative Analysis*, 35, 191-215.

Ball, C. A. and Torous, W. N. 1996, 'Unit roots and the estimation of interest rate dynamics', *Journal of Empirical Finance*, 3, 215-238.

Ball C. A. and Torous, W. N. 1999, 'The stochastic volatility of short-term interest rates: Some international evidence', *Journal of Finance*, 54, 2339-2359.

Bandi, F. and Phillips, P. C. B. 1999, *Econometric Estimation of Diffusion Models*, Working Paper, Yale University, New Haven, USA.

Banerjee, A., Lumsdaine, R. L. and Stock, J.H. 1992, 'Recursive and sequential tests of the unit-root and trend-break hypotheses: Theory and international evidence', *Journal of Business and Economic Statistics*, 10, 271-288.

Baum, L. E. and Petrie, T. 1966, 'Statistical inference for probabilistic functions of finite state Markov chains', *Annals of Mathematical Statistics*, 37, 1554-1563.

Baum, L. E., Petrie, T., Soules, G. and Weiss, N. 1970, 'A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains', *Annals of Mathematical Statistics*, 41, 164-171.

Bekaert, G. and Hodrick, R. J. 2001, 'Expectations hypotheses tests', *Journal of Finance*, 56, 1357-1394.

Bekaert, G., Hodrick, R. J. and Marshall D. A. 1997a, 'On biases in the tests on the expectations hypotheses of the term structure of interest rates', *Journal of Financial Economics*, 44, 309-348.

Bekaert, G., Hodrick, R. J. and Marshall D. A. 2001, 'Peso problem explanations for term structure anomalies', *Journal of Monetary Economics*, 48, 241-270.

Berument, H and Güner, N. 1997, 'Inflation, inflation risk and interest rates: A case study for Turkey', *Metu Studies in Development*, 24, 3, 319-327.

Berument, H. and Malatyalı, K. 2001, 'Determinants of Interest Rates in Turkey', forthcoming: *Russian and East European Finance and Trade*.

Bierens, H.J. 1997, 'Testing the unit root hypothesis against nonlinear trend stationarity, with an application to the price level and interest rate in the U.S', *Journal of Econometrics*, 81, 29-64.

Bierens, H. J. and Guo, S. 1993, 'Testing stationarity and trend stationarity against the unit root hypothesis', *Econometric Review*, 12,1-32.

Black, F. and Scholes, M. S. 1973, 'The pricing of options and corporate liabilities', *Journal of Political Economy*, 81, 637-659.

Black, F., Derman, E. and Toy, W. 1990 'A one-factor model of interest rates and its application to treasury bond options', *Financial Analysts Journal*, 46, 33-39.

Blackwell, E. and Koopmans, L. 1975, 'On the identifiability problem for functions of finite Markov chains', *Annals of Mathematical Statistics*, 28, 1011-1015.

Bliss, R. R. 1996, *Testing Term Structure Estimation Methods*, Working Paper no 96-12a, Federal Reserve Bank of Atlanta.

Bollerslev, T. 1986, 'Generalized autoregressive conditional heteroscedasticity', *Journal of Econometrics*, 31, 307-327.

Bollerslev, T. 1987, 'A conditionally heteroscedastic time series model for speculative prices and rates of return', *Review of Economics and Statistics*, 69, 542-547.

Bookstaber, R. and McDonald, J. 1987, 'A general distribution for describing security price returns', *Journal of Business*, 60, 401-424.

Box, G. E. P. and Pierce G. A. 1970, 'Distribution of residual autocorrelations in autoregressive integrated moving average time series models', *Journal of the American Statistical Association*, 65, 1509-1526.

Brennan, M. J. and Schwartz, E. 1979, 'A continuous-time approach to the pricing of bonds', *Journal of Banking and Finance*, 3, 133-155.

Brock, W. A., Dechert, W. D. and Scheinkman, J. A. 1987, '*A Test for Independence Based on the Correlation Dimension*' Working Paper 9520, Social Systems Research Institute, University of Wisconsin, Madison, USA.

Brock, W. A., Dechert, W. D., Scheinkman, J. A. and LeBaron, B. 1996, 'A Test for Independence Based on the Correlation Dimension' *Econometric Reviews*, 15, 197-235.

Brown, S. and Dybvig, P. 1986, 'The empirical implications of the Cox, Ingersoll and Ross theory of the term structure of interest rates', *Journal of Finance*, 41, 617-630.

Brown, R. H. and Schaefer, S. M. 1994, 'The term structure of real interest rates and the Cox, Ingersoll and Ross Model', *Journal of Financial Economics* 35, 3-42.

Buono, M., Gregory-Allen, R. B. and Yaari, U. 1992, 'The efficacy of term structure estimation techniques: A Monte Carlo study', *Journal of Fixed Income* 52-63.

Cai, J. 1994, 'A Markov model of switching-regime ARCH', *Journal of Business and Economic Statistics*', 12, 309-316.

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 Shiller, R. J. 1989, 'The dividend-price ratio and expectations of future dividends and discount factors', *Review of Financial Studies*, 1, 195-228.

Campbell, J. Y. and Shiller, R. J. 1991, 'Yield spreads and interest rate movements: a bird's eye view', *Review of Economic Studies*, 58, 495-514.

Campbell, S. D. 2001, '*Specification Testing and Semi-Parametric Estimation of Regime-Switching Models: An Examination of the US Short-Term Interest Rate*, Manuscript, University of Pennsylvania.

Cecchetti, S. J., Lam, P. and Mark, N. C. 1993, 'The equity premium and the risk-free rate: matching the moments', *Journal of Monetary Economics*, 31, 21-46.

Chambers, D.W., Carleton, W. and Waldman, D. 1984, 'A new approach to estimation of term structure of interest rates', *Journal of Financial an Quantitative Studies*, 19, 3, 233-252.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders, A. B. 1992, 'An empirical comparison of alternative models of the short-term interest rate', *Journal of Finance*, 47, 1209-1227.

Chapman, D. A. and Pearson, N. D. 2000, 'Is the short rate drift actually nonlinear?', *Journal of Finance*, 55, 355-388.

Chapman, D. A. and Pearson, N. D. 2001, 'Recent advances in estimating term-structure models', *Financial Analysts Journal*, 57, 77-95.

Chow, G. C. 1960, 'tests of equality between sets of coefficients in two linear regressions', *Econometrica*, 28, 591-605.

Christiano, L. 1992, 'Searching for a break in GNP', *Journal of Business and Economic Statistics*, 10, 237-250.

Coleman, T. S., Fisher, L. and Ibbotson, R. 1992, 'Estimating the term structure of interest rates from data that include the prices of coupon bonds', *The Journal of Fixed Income*, 85-116.

Conley, T. G., Hansen, L. P., Luttmer, E. G. J. and Scheinkman, J. A. 1997, 'Short-term interest rates as subordinated diffusions', *Review of Financial Studies*, 10, 525-578.

Constantinides, G. 1992, 'A theory of the nominal term structure of interest rates', *The Review of Financial Studies*, 5, 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., Ingersoll, J. E. and Ross, S. A. 1980 'An analysis of variable rate loan contracts', *Journal of Finance* 35, 389-403

Cox, J. C., Ingersoll, J. E. and Ross, S. A. 1981, 'A re-examination of traditional hypotheses about the term structure of interest rates." *Journal of Finance* 36, 769-799

Cox, J. C., Ingersoll, J. E. and Ross, S. A. 1985b, 'A theory of the term structure of interest rates' *Econometrica*, 53, 385-407.

Cox, J. C. and Ross, S. A. 1976a, 'A survey of some new results in financial options pricing theory', *Journal of Finance*, 31, 382-402.

Cox, J. C. and Ross, S. A. 1976b, 'The valuation of options for alternative stochastic processes', *Journal of Financial Economics*, 3, 144-166

Cox, J. C., Ross, S. A. and Rubinstein M. 1979, 'Option pricing: A simplified approach' *Journal of Financial Economics*, 7, 229-264.

Dahlquist, M. and Jonsson, G. 1995, 'The information in Swedish short maturity forward rates', *European Economic Review*, 39, 1115-1131.

Dai, Q. and Singleton, K. 2002, 'Term structure modeling in theory and reality', *Review of Financial Studies*, forthcoming.

Das, S. R. 2002, 'The surprise element: jumps in interest rates', *Journal of Econometrics*, 106, 27-65.

de Boor, C. 1978, *A Practical Guide to Splines*, Springer, New York, USA.

Demiroz, D. M. 2001, Codependent Volatility Patterns in daily Foreign Exchange and Interest Rates. A Case Study of Turkey.

Dempster, A. P., Laird, N. M. and Rubin, D. B. 1977, 'Maximum likelihood estimation from incomplete data via the EM algorithm', *Journal of the Royal Statistical Society*, 39, Series B, 1-38.

Dothan, U. 1978, 'On the term structure of interest rates', *Journal of Financial Economics*, 7, 229-264.

Duffee, G. M. 2002, 'Term premia and interest rate forecasts in affine models', *Journal of Finance*, 57, 405-443.

Duffie, D. 1996, *Dynamic Asset Pricing Theory*, Princeton University Press, Princeton, N.J., USA.

Duffie, D and Kan, R. 1996, 'A yield-factor model of interest rates' *Mathematical Finance*, 6, 379-406.

Dybvig, P. H. 1988, *Bond and Bond Option Pricing Based on the Current Term Structure*, Working Paper, Washington University, St. Louis, USA.

Elliott, R. J. and Mamon, R. S. 2000a, '*Term Structure of a Vasicek Model with a Markovian Mean Reverting Level*, Working Paper 2000-10, Department of Statistics and Actuarial Science, University of Waterloo, Canada.

Elliott, R. J. and Mamon, R. S. 2000b, '*A Complete Yield CurveDescription of a Markov Interest Rate Model*, Working Paper 2000-11, Department of Statistics and Actuarial Science, University of Waterloo, Canada.

Engle, R. F. 1982, 'Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation', *Econometrica*, 50, 987-1007.

Engle, R. F. and Gonzalez-Rivera, G. 1991, 'Semiparametric ARCH models', *Journal of Businesss and Economic Statistics*, 9, 345-360.

Epanechnikov, V. 1969, 'Nonparametric estimates of multivariate probability density', *Theory of Probability and Applications*, 14, 153-158.

Evans, M. D. D. and Lewis, K. K. 1994, 'Do stationary risk premia explain it all?', *Journal of Monetary Economics*, 33, 285-318.

Evans, M.D.D. and Lewis, K. K. 1995, 'Do long-term swings in the dollar affect estimates of the risk premia?', *Review of financial Studies*, 8, 709-742.

Fama, E. F. 1990, 'Term structure forecasts of interest rates, inflation and real returns', *Journal of Monetary Economics*, 25, 59-76.

Fisher, I. 1896, *Appreciation and Interest*, Macmillan, New York, USA.

Fisher, I. 1907, *The Rate of Interest*, Macmillan, New York, USA.

Fisher, M. 2001, *Forces that Shape the Yield Curve: Parts 1 and 2*, Working Paper no 2001-3, Federal Reserve Bank of Atlanta, Atlanta, USA.

Fisher, M. and Gilles, C. 1998, 'Around and around: the expectations hypothesis', *The Journal of Finance*, 53, 365-383.

Fisher, M., Nychka, D. and Zervos, D. 1994, *Fitting the Term Structure of Interest Rates with Smoothing Splines*, Working Paper, Finance and Economics Discussion Series, Federal Reserve Board, Washington D.C., USA.

French, K. R. and Roll, R. 1986, 'Stock return variances: the arrival of information and the reaction of traders', *Journal of Financial Economics*, 17, 5-26.

Friedman, M. 1977, 'Nobel lecture: Inflation and unemployment', *Journal of Political Economy*, 81, 451-472.

Fuller, W.A. 1996, *Introduction to Statistical Time Series*, John Wiley and Sons inc., New York, USA.

Gallager, R. G. 1996, *Discrete Stochastic Processes*, Kluwer Academic, London, UK.

Garcia, R. 1998, 'Asymptotic null distribution of the likelihood ratio test in Markov switching models', *International Economic Review*, 39, 763-788.

Garcia, R. and Perron, P. 1996, 'An analysis of the real interest rate under regime shifts', *The Review of Economics and Statistics*, 78, 111-125.

Gençay, R. and Selçuk, F. 2001, *Overnight Borrowing, Interest Rates and Extreme Value Theory*, Working Paper, Bilkent University, Ankara, Turkey.

Gerlach, S. and Smets, F. 1997, 'The term structure of Euro rates: some evidence in support of the expectations hypothesis', *Journal of International Money and Finance*, 16, 305-321.

Gibbons, M. and Ramaswamy, K. 1993, 'A test of the Cox, Ingersoll and Ross model of the term structure', *The Review of Financial Studies* 6, 619-632.

Gibson, R, Lhabitant, F.-S. and Talay, D 1999, *Modeling the Term Structure of Interest Rates: A Review of the Literature*, HEC, Lausanne Working Paper 9801.

Goldfeld, S. M. and Quandt, R. E. 1973, 'A Markov model for switching regressions', *Journal of Econometrics*, 1, 3-16.

Grassberger, P. and Procaccia, I. 1983a, 'Estimating the Kolmogorov entropy from a chaotic signal', *Physical Review*, 28A, 25-91.

Grassberger, P. and Procaccia, I. 1983b, 'Measuring the strangeness of strange attractors', *Physica*, 9D, 189-208.

Gray, S. F. 1996, 'Modeling the conditional distribution of interest rates as a regime-switching process', *Journal of Financial Economics*, 42, 27-62.

Hamilton, J. D. 1988, 'Rational-expectations econometric analysis of changes in regime. An investigation of the term structure of interest rates', *Journal of Economic Dynamics and Control*, 12, 385-423.

Hamilton, J. D. 1989, 'A new approach to the economic analysis of nonstationary time series and the business cycle', *Econometrica*, 57, 357-384.

Hamilton, J. D. 1990, 'Analysis of time series subject to changes in regime', *Journal of Econometrics*, 45, 39-70.

Hamilton, J. D. 1994, *Time Series Analysis*, Princeton University Press, Princeton, NJ, USA.

Hamori, S. and Tokihisa, A. 1997,'Testing for a unit root in the presence of a variance shift', *Economics Letters*, 57, 245-253.

Hansen, A. T. and Poulsen, R. 2000, 'A simple regime switching term structure model', *Finance and Stochastics*, 4, 409-429.

Hansen, B. 1994, 'Autoregressive conditional density estimation', *International Economic Review*, 35, 705-730.

Hansen, L. P. 1982, 'Large sample properties of generalized method of moments estimators', *Econometrica*, 50, 1029-1054.

Hardouvelis, G. 1994, 'The term structure spread and future changes in long and short rates in the G7 countries: is there a puzzle', *Journal of Monetary Economics*, 33, 255-284.

Harrison, J. M. and Kreps, D. 1979, 'Martingales and arbitrage in multiperiod securities markets', *Journal of Economic Theory*, 20, 381-408.

Harrison, J. M. and Pliska, S. R. 1981, 'Martingales and stochastic integrals in the theory of continuous trading', *Stochastic Processes and Their Applications*, 11, 215-260.

Harvey, C. R. and Siddique, A. 1999, 'Autoregressive conditional skewness', *Journal of Financial and Quantitative Analysis*, 34, 465-487.

Heath, D. Jarrow, R. and Morton, A. 1992, 'Bond pricing and the term structure of interest rates: A new methodology for contingent claims valuation," *Econometrica*, 60, 77-105.

Heller, A. 1965, 'On stochastic processes derived from Markov chains', *Annals of Mathematical Statistics*, 36, 1268-1291.

Hicks, J. R. 1953, *Value and Capital*, Oxford University Press, London, UK.

Ho, T. S. Y. and Lee S.-B. 1986, 'Term structure movements and the pricing of interest rate contingent claims', *The Journal of Finance*, 41, 1011-1029.

Holland, A. S. 1984, 'Does higher inflation lead to more uncertain inflation?', *Federal Reserve Bank of St. Louis Review*, USA, 66, 15-26.

Holst, U., Lindgren, G., Holst, J. and Thuvesholmen, M. 1994, 'Recursive estimation in switching autoregressions with a Markov regime', *Journal of Time Series Analysis*', 15, 489-506.

Hong, Y., Li, H. and Zhao, F. 2002, *Out-of-Sample Performance of Spot Interest Rate Models*, Working Paper, Cornell University, New York, USA.

Hull, J. 1993, *Options, Futures and other Derivative Securities*, Prentice-Hall, second edition, New York, USA.

Hull, J. and White A. 1990, 'Pricing interest rates derivative securities', *Review of Financial Studies*, 3, 573-592.

Hull, J. and White, A. 1993, 'Bond option pricing based on a model for the evolution of bond prices', *Advances in Futures and Options Research*, 6, 1-13,

Hull, J. and White, A. 1995 'A note on the models of Hull and White for pricing options on the term structure: Response' *The Journal of Fixed Income* 5, 97-102.

Jarrow, R. A. 1996, '*Modeling Fixed Income securities and Interest Rate Options*, McGraw-Hill, New York, USA.

Jiang, G. J. and Knight J. L. 1997, 'A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model', *Econometric Theory*, 13, 615-45.

Jolliffe, I. T. 1986, *Principal Component Analysis*, Springer-Verlag, New York, USA.

Jondeau, E. and Rockinger, M. 2001, 'Estimating Gram-Charlier expansions under positivity constraints', *Journal of Economic Dynamics and Control*, 10, 1457-1483.

Jondeau, E. and Rockinger, M. 2002, 'Entropy densities with an application to autoregressive conditional skewness and kurtosis', *Journal of Econometrics*, 106, 119-142.

Jones, C. S. 1998, *Bayesian Estimation of Continuous-Time Finance Models*, Working Paper, University of Rochester, New York, USA.

Jones, C. S. 2001, *Nonlinear Mean Reversion in the Short-Term Interest Rate*, Working Paper, Simon School of Business, University of Rochester, New York, USA.

Jordan, J. V. 1984, 'Tax effects in term structure estimation', *Journal of Finance*, 39, 393-406.

Kanzler, L. 1999, '*Very Fast and Correctly Sized Estimation of the BDS Statistic*, Manuscript, Christ Church and Department of Economics, University of Oxford, Oxford, UK.

Karatzas, I. and Shreve, S. E. 1991, *Brownian Motion and Stochastic Calculus*, Springer-Verlag, New York, USA.

Karlsen, H. 1990, '*A Class of Non-Linear Time Series Models*', PhD Dissertation, Department of Mathematics, University of Bergen, Bergen, Norway.

Kaya, N. and Neyapti, B. 2001, '*Inflation and inflation uncertainty in turkey*', *Yapı Kredi Economic Review*, December.

Keynes, J. M. 1930 *Treatise on Money, volumes 1-2*, MacMillan, London, UK.

Kim, C.-J. 1994, 'Dynamic linear models with Markov-switching' *Journal of Econometrics*, 60, 1-22.

Kirmanoğlu, H. 2001. 'Is there inflation-growth trade-off in the Turkish economy?', Paper presented at CEA annual meeting, McGill University, Montreal, Canada.

Kona, G. 1996, *The Bond Markets in Turkey and a Suggestion for a Bond Price Index*, Master's Thesis, Bosporus University, Istanbul, Turkey.

Krolzig, H.-M. 1996, *Statistical Analysis of Cointegrated VAR Processes with Markovian Regime Shifs*', Discussion Paper 25, Humboldt-Universität zu Berlin, Germany.

Krolzig, H-M. 1997, *Markov-Switching Vector Autoregressions. Modeling, Statistical Inference, and Application to Business Cycle Analysis*, Springer-Verlag, Berlin, Germany.

Kwiatkowski, D., Phillips, P.C.B., Schmidt, P., and Shin, Y. 1992, 'Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic series have a unit root', *Journal of Econometrics*, 54, 159-178.

Lanne, M. 1999, 'Near unit roots and the predictive power of yield spreads for changes in long-term interest rates', *Review of Economics and Statistics*, 81, 393-398.

Leroux, B. G. 1992. 'Maximum-likelihood estimation for hidden Markov models', *Stochastic Processes Applications*, 40, 127-143.

Lewis, K. K. 1991, 'Was there a peso problem in the US term structure of interest rates 1979-1982?', *International Economic Review*, 32, 159-173.

Li, H. and Xu, Y. 2000, *Regime Shifts and Short Rate Dynamics*, Working Paper, Cornell University, New York, USA.

Litterman, R. and Scheinkman, J. 1991,'Common factors affecting bond returns', *Journal of Fixed Income*, 3, 54-61.

Longstaff, F. A. 2000 'The term structure of very short-term rates: New evidence for the expectations hypothesis', *Journal of Financial Economics*, 58, 397-415.

Lucas, R. E. 1978, 'Asset prices in an exchange economy', *Econometrica* 46, 1429-45.

Lumsdaine, R. and Papell, D. H. 1997, 'Multiple trend breaks and the unit root hypothesis', *Review of Economics and Statistics*, 212-218.

Lutz, F. A. 1940, 'The structure of interest rates', *Quarterly Journal of Economics*, 36-63.

Malkiel, B. 1966, *The Term Structure of Interest Rates: Expectations and behavior Patterns*, Princeton University Press, Princeton, USA.

Marsh, T. and Rosenfeld, E. 1983, 'Stochastic processes for interest rates and equilibrium bond prices', *Journal of Finance*, 38, 635-645.

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', *Journal of Finance*, 30, 811-830.

Meier, I. 2000, *Estimating the Term Structure of Interest Rates and the Pricing of Interest Rate Derivatives*, Studienzentrum Gerzensee, Switzerland.

Merton, R. C. 1969, 'Lifetime portfolio selection under uncertainty: The continuous time case', *Review of Economics and Statistics* 51, 247-257.

Merton, R. C. 1971, 'Optimum consumption and portfolio rules in a continuous time model', *Journal of Economic Theory*, 3, 373-413.

Merton, R. C. 1973a, 'Theory of rational option pricing', *Bell Journal of Economics* 4, 141-183.

Merton, R. C. 1973b, 'An intertemporal capital-asset pricing model', *Econometrica*, 41, 867-87.

Merton, R. C. 1974, 'On the pricing of corporate debt: The risk structure of interest rates', *The Journal of finance* 29, 449-470.

Merton, R. C. 1975, 'An asymptotic theory of growth under uncertainty', *Review of Economic Studies*, 42, 375-393.

Merton, R. C. 1976, 'Option pricing when the underlying stock returns are discontinuous', *Journal of Financial Economics*, 5, 125-144.

Mishkin, F. S. 1990, 'What does the term structure tell us about future inflation?', *Journal of Monetary Economics*, 25, 77-95.

Missale, A. and Blanchard, O. J. 1994, 'The debt burden and debt maturity', *American Economic Review*, 309-319.

Modigliani, F. and Sutch, R. C. 1966, 'Innovations in interest rate policy', *American Economic Review*, 56, 178-197.

Molgedey, L. and Galic, E.2000, 'Extracting factors for interest rate scenarios', *European Physics Journal B*, 20, 517-522.

Montanes, A. and Reyes, M. 1998, 'Effect of a shift in the trend function on Dickey-Fuller unit root tests' *Ecnometric Theory*, 14, 355-363.

Mundell, R. A. 1963, 'Inflation and real interest', *Journal of Political Economy*.

National Bank of Greece S. A. 2000-2001, *South Eastern Europe & Mediterranean Emerging Market Economies Bulletin(s)*.

Neftci, S. N. 1996, *An Introduction to the Mathematics of Financial Derivatives*, Academic Press, New York, USA.

Nelson, C. R. and Siegel, A. F. 1987, 'Parsimonious modeling of yield curves', *Journal of Business*, 60, 474-489.

Nelson, D. B. 1991, 'Conditional heteroscedasticity in asset returns: a new approach', *Econometrica*, 59, 347-370.

Osborne, M. F. M. 1959, 'Brownian motion in the stock market', *Operations Research*.

Pagan, A. R. Hall, A. D., and Martin, V. 1996, 'Modeling the term structure' in Maddala, G. S. and Rao, C. R. (eds.), *Handbook of Statistics* 14, 91-118.

Pearson, K. 1894, 'Contributions to the mathematical theory of evolution', *Philosophical transactions of the Royal Society*, 185, 71-110.

Perron, P. 1989, 'The great crash, the oil price shock, and the unit root hypothesis', *Econometrica*, 57, 1361-1401.

Perron, P. 1990, 'Testing for a unit root in a time series with a changing mean', *Journal of Business and Economic Statistics*, 8, 153-162.

Perron, P. and Vogelsang, T. J. 1992, 'Nonstationarity and level shifts with an application to purchasing power parity', *Journal of Business and Economic Statistics*, 10, 301-320.

Phillips, P.C.B. and Perron, P. 1988, 'Testing for a unit root in time series regression', *Biometrica*, 75, 335-346.

Pritsker, M. 1998, 'Nonparametric density estimation and tests of continuous time interest rate models', *Review of Financial Studies*, 11, 449-487.

Rebonato, R. and Cooper, I. 1995, 'Limitations of simple two-factor interest rate models', *Journal of Financial Engineering*, 5, 1-16.

Rebonato, R. 1998, *Interest Rate Option Models*, Wiley and Sons, 2 ed. UK.

Reimers, M. and Zerbs, M. 1999, 'A Multi-Factor Statistical Model for Interest Rates', *Algo Research Quarterly*, 2, 53-63.

Rendleman, R. J. and Bartter, B. J. 1980, 'The pricing of options on debt securities', *Journal of Financial and Quantitative Analysis*, 15, 11-24.

Riedel, F. 1999, *Heterogeneous Time Preferences and Interest Rates – The Preferred Habitat Theory Revisited*, Working Paper on EconWPA (web-site).

Şahinbeyoğlu, G. and Yalçin, C. 2000, *The Term Structure of Interest Rates: Does it Tell About Future Inflation?*, Discussion Paper, Research Department, Central Bank of Turkey, Ankara, Turkey.

Said, S.E. 1991, 'Unit root test for time series data with a linear time trend', *Journal of Econometrics*, 47, 285-303.

Said, S.E. and Dickey, D. A. 1984, 'Testing for unit roots in autoregressive moving average of unknown order', *Biometrika*, 71, 599-607.

Samuelson, P. 1965, 'Rational theory of warrant prices', *Industrial Management Review*, 6, 13-31.

Schaefer, S. 1981, 'Measuring a tax-specific term structure of interest rates in the market for British government securities', *The Economic Journal*, 91,415-438.

Schlögl, E. and Sommer, D. 1998, 'Factor models and the shape of the term structure', *Journal of Financial Engineering*, 7.

Shea, G. 1984, 'Pitfalls in smoothing interest rate term structure data: Equilibrium models and spline approximations', *Journal of Financial and Quantitative Studies*, 19, 253-269.

Shea, G. 1985, 'Interest rate term structure estimation with exponential splines: A note', *Journal of Finance*, 40, 319-325.

Sims, C. A. 1980, 'Macroeconomics and reality', *Econometrica*, 48, 1-48.

Sola, M. and Driffill J. 1994, 'Testing the term structure of interets rates using a vector autoregression with regime switching', *Journal of Economic Dynamics and Control*, 18, 601-628.

Stanton, R. 1997, 'A nonparametric model of term structure dynamics and the market price of interest rate risk', *Journal of Finance* 52, 1973-2002.

Sundaresan, S. M. 2000, 'Continuous-time methods in finance: A review and an assessment', *Journal of Finance*, 55, 1569-1622.

Tanggaard, C. 1997, 'Non-parametric smoothing of yield curves', *Review of Quantitative Finance and Accounting*, 9, 251-267.

Tobin, J. 1965, 'Money and economic growth', *Econometrica*, 33, 671-684.

Tzavalis, E. and Wickens, M. 1996, 'Forecasting inflation from the term structure', *Journal of Empirical Finance*, 3, 103-122.

Valkanov, R. 1998, *The Term Structure with highly Persistent Interest Rates*, Manuscript. Princeton University, N.J., USA.

van Kampen, N. G. 1981, 'Ito versus Stratonovich', *Journal of Statistical Physics*, 24, 175-187.

Vasicek, O. 1977, 'An equilibrium characterization of the term structure' *Journal of Financial Economics* 5, 177-188.

Vasicek, O and Fong, G. 1982, 'Term structure estimation using exponential splines', *Journal of Finance*, 38, 339-348.

Warne, A. 1996, *Autocovariance Functions and Maximum Likelihood in a VAR Model under Markov Switching*, Mimeo, Institute for International Economic Studies, Stockholm University, Sweden.

Wilmott, P. 1999, '*Derivatives: The Theory and Practice of Financial Engineering*', John Wiley and Sons, West Sussex, England, UK.

Wilson, T. 1994, 'Debunking the myths', *Risk*, 7, 67-73.

Yang, M. 2000, 'Some properties of vector autoregressive processes with Markov-switching coefficients', *Econometric Theory*, 16, 23-43.