Affine Interest Rate Models - Theory and Practice

Ausgeführt am
Institut für Wirtschaftsmathematik
der Technischen Universität Wien

unter der Anleitung von
ao. Univ.-Prof. Dr. Josef Teichmann

durch
Christa Cuchiero
Hainzenbachstraße 25
4060 Leonding

_________________________  _______________________
Datum                                      Unterschrift
Abstract

The aim of this diploma thesis is to present the theory as well as the practical applications of affine interest rate models. On the basis of the general theory established by Duffie and Kan, we put emphasis on affine models whose state variables have - in contrast to their theoretical abstract definition - a reasonable economic interpretation. Starting from the very first term structure models, namely the Vasicek and the Cox-Ingersoll-Ross model, we describe in sequel two- and more-factor models that have appeared in literature. By means of the Vasicek model we exemplify the calibration to market yields as well as to market cap volatilities.
However, our main focus are affine yield factor models developed by Duffie and Kan, which allow to relate the state variables to yields with different maturities. We show how to calibrate a two-factor version of this model to market data. The results are promising since the model fits the market yields from different dates very well while the parameters remain nearly constant.
## Contents

1 Introduction to Interest Rate Theory 1  
   1.1 Definitions and Notations ................................. 1  
       1.1.1 Short-Term Interest Rate .......................... 1  
       1.1.2 Zero-Coupon Bonds and Spot Interest Rates ............ 2  
       1.1.3 Forward Rates .................................... 3  
       1.1.4 Interest Rate Swaps ................................ 4  
   1.2 No-Arbitrage Pricing ...................................... 6  
1.3 Factor Models of the Term Structure .......................... 7  
   1.3.1 Dynamics under $\mathbb{P}^*$ ........................... 9  
   1.3.2 The Bond Price as Solution of a PDE .................... 12

2 Affine Models 13  
   2.1 Theory of Affine Factor Models ............................ 13  
       2.1.1 Specification of the State Variable Process .......... 14  
       2.1.2 Affine Stochastic Differential Equations .............. 17  
       2.1.3 Ricatti Equations .................................. 18  
   2.2 Types of Affine Models .................................... 19  
       2.2.1 Gaussian Affine Models .............................. 19  
       2.2.2 CIR Affine Models .................................. 20  
       2.2.3 The Three-Factor Affine family ....................... 21  
   2.3 Classification of Affine Models ............................ 22  
       2.3.1 A Canonical Representation .......................... 22  
       2.3.2 Invariant Transformations and Equivalent Models .... 24

3 Examples of Affine Models 26  
   3.1 Examples of One-Factor Affine Models ....................... 26  
       3.1.1 The Extended Vasicek Model .......................... 27  
       3.1.2 The Extended CIR Model .............................. 29  
   3.2 Examples of Multi-Factor Affine Models ..................... 31  
       3.2.1 The Longstaff and Schwartz Two-Factor Model .......... 32  
       3.2.2 The Central Tendency as Second Factor ............... 33
CONTENTS

3.3 Economic Models ........................................... 35
  3.3.1 The General Framework ............................... 36
  3.3.2 IS - LM Framework .................................... 37
3.4 Non-Affine Models - Consol Models ......................... 38
3.5 Criteria for Model Selection ............................... 40

4 Calibration and Estimation ................................ 42
  4.1 Obtaining a Data Set ..................................... 43
    4.1.1 Market Data for the Current Yield Curve .......... 43
    4.1.2 Market Data for Bond Options ..................... 44
    4.1.3 Which Market Rate should be used for the Short-Term Rate? ..................... 47
  4.2 Calibration to Current Market Data ....................... 47
    4.2.1 Calibrating the Vasicek Model to the Current Term Structure .................. 47
    4.2.2 Calibrating the Vasicek Model to Cap Volatilities .................. 49
    4.2.3 Calibrating the Hull-White Extended Vasicek Model ..................... 52
  4.3 Historical Estimation ................................... 53
    4.3.1 Maximum Likelihood Method ......................... 54
    4.3.2 General Method of Moments ......................... 56

5 Affine Yield-Factor Models ................................ 58
  5.1 General Affine Yield-Factor Model ....................... 58
  5.2 A Two-Factor Affine Model of the "Long"- and the Short-Term Rate .................. 60
    5.2.1 Deterministic Volatility ........................... 61
    5.2.2 Calibrating the Deterministic Volatility Model to the Current Term Structure .................. 61
    5.2.3 Stochastic Volatility ............................... 64
    5.2.4 Calibrating the Stochastic Volatility Model to the Current Term Structure .................. 67
    5.2.5 Conclusion ........................................ 71

A Numerical Methods for Calibration ........................ 72
  A.1 Trust-Region Methods for Nonlinear Minimization .......... 72
    A.1.1 Box Constraints .................................... 74
    A.1.2 Nonlinear Least-Squares ............................ 74
List of Figures

4.1 Market- vs. Vasicek Model Yields  ........................................... 49
4.2 Vasicek Model Parameters over Time  ....................................... 50
4.3 Market- vs. Vasicek Model Caps  ............................................. 51
4.4 Hull-White Model Calibration  .................................................. 53

5.1 Market- vs. Model Yields, 2-Factor Deterministic Volatility
       Model  .................................................................................. 65
5.2 Parameters over Time, 2-Factor Deterministic Volatility Model  66
5.3 Market- vs. Model Yields, March 2006, 2-Factor Deterministic
       Volatility Model  .................................................................. 67
5.4 Parameters over Time, 2-Factor Stochastic Volatility Model  .. 70
5.5 Market- vs. Model Yields, 2-Factor Stochastic Volatility Model 70
List of Tables

4.1 Parameters for the Vasicek Model, Calibration to Yields ... 48
4.2 ATM Cap Volatilities ........................................ 51
4.3 Parameters for the Vasicek Model, Calibration to Caps ... 51
4.4 Parameters for the Hull-White Model, Calibration to Caps 53
4.5 Maximum Likelihood Parameters for the Vasicek Model ... 55

5.1 Parameters for the 2-Factor Deterministic Volatility Model . 63
5.2 Residuals for the 2-Factor Deterministic Volatility Model . 63
5.3 Monthly Parameters for the 2-Factor Deterministic Volatility Model ........................................... 64
5.4 Parameters for the 2-Factor Stochastic Volatility Model ... 68
5.5 Residuals for the 2-Factor Stochastic Volatility Model ... 69
5.6 Monthly Parameters for the 2-Factor Stochastic Volatility Model 69
Chapter 1

Introduction to Interest Rate Theory

Although the concept of interest rates seems to be something natural that everybody knows to deal with, the management of interest rate risk, i.e. the control of changes in future cash flows due to fluctuations in interest rates is an issue of great complexity. In particular, the pricing and hedging of products depending in large part on interest rates create the necessity for mathematical models.

This chapter covers the basic definitions and concepts of interest rate theory. The first part focuses on the different kinds of interest rates, whereas in the other sections mathematical basics for interest rate modeling are presented. The approach is similar to Brigo and Mercurio [3] including supplements of Musiela and Rutkowski [14] and Björk [1].

1.1 Definitions and Notations

1.1.1 Short-Term Interest Rate

The first concept that is introduced is the notion of a bank (savings) account representing a risk-free security which continuously compounds in value at a risk-free rate, namely the instantaneous interest rate (also referred as short-term interest rate).

Definition 1.1. Short-term rate, Bank account. Let \( r(t) \) denote the short-term rate for risk-free borrowing or lending at time \( t \) over the infinitesimal time interval \([t, t + dt]\). \( r(t) \) is assumed to be an adapted process on a
CHAPTER 1. INTRODUCTION TO INTEREST RATE THEORY

filtered probability space \((\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0 \leq t \leq T^*})\) for some \(T^* > 0\) with almost all sample paths integrable on \([0, T^*]\). \(B(t) = B(t, \omega)\) is defined to be the value of the bank account at \(t \geq 0\) that evolves for almost all \(\omega \in \Omega\) according to the differential equation

\[
\frac{dB(t)}{B(t)} = r(t) \, dt \quad \text{with} \quad B(0) = 1.
\]

Consequently

\[
B(t) = \exp \left( \int_0^t r(s) \, ds \right) \quad \text{for all} \quad t \in [0, T^*].
\]

By means of \(B(t)\), two amounts of currency which are available at different times can be related. In fact, in order to have one unit of cash at time \(T\) one has to invest the amount \(1/B(T)\) at the beginning. At time \(t > 0\) the value of this initial investment constitutes \(B(t)(1/B(T))\), which leads to the following definition.

**Definition 1.2. Stochastic discount factor.** The stochastic discount factor \(D(t, T)\) is the value at time \(t\) of one unit of cash payable at time \(T > t\) and is given by

\[
D(t, T) = \frac{B(t)}{B(T)} = \exp \left( - \int_t^T r(s) \, ds \right).
\]

### 1.1.2 Zero-Coupon Bonds and Spot Interest Rates

**Definition 1.3. Zero-coupon bond.** A zero-coupon bond of maturity \(T\) is a financial security paying one unit of cash at a prespecified date \(T\) in the future without intermediate payments. The price at time \(t \leq T\) is denoted by \(P(t, T)\). Obviously, \(P(T, T) = 1\) for all \(T \leq T^*\).

**Remark.** It is assumed that the price process \(P(t, T)\) follows a strictly positive and adapted process on a filtered probability space \((\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0 \leq t \leq T^*})\), where the filtration \(\mathcal{F}_t\) is again the \(\mathbb{P}\)-completed version of the filtration generated by the underlying Brownian motion.

Note that there is a close relationship between the zero-coupon bond price \(P(t, T)\) and the stochastic discount factor \(D(t, T)\). Actually, \(P(t, T)\) corresponds to the expectation of \(D(t, T)\) under the risk-neutral probability measure, as we will see in the next section (equation 1.21). If \(r\) is deterministic, then \(D\) is deterministic as well and necessarily \(D(t, T) = P(t, T)\).

---

1. \(\mathcal{F}_t\) is the \(\mathbb{P}\)-completed (i.e. it contains all sets of null probability with respect to \(\mathbb{P}\)) filtration generated by a standard Brownian motion in \(\mathbb{R}^n\).

2. \(T^*\) is the fixed horizon date for all market activities.

3. \(P(t, T)\) is the discount factor at time \(t\) for cashflows occurring at time \(T\).
Definition 1.4. Continuously-compounded spot interest rate or yield on a zero-coupon bond. The continuously-compounded spot interest rate \( R(t, T) \), also referred as yield on the zero-coupon bond \( P(t, T) \), is the constant rate at which an investment of \( P(t, T) \) units of cash at time \( t \) accrues continuously to yield one unit of cash at maturity \( T \), i.e. \( \exp(R(t, T)(T - t))P(t, T) = 1 \). Hence
\[
R(t, T) = -\ln \frac{P(t, T)}{T - t}.
\]

Remark. The short-term rate \( r(t) \) is obtained as limit of \( R(t, T) \), that is
\[
r(t) = \lim_{T \to t^+} R(t, T) = \lim_{T \to t^+} -\frac{\ln P(t, T)}{T - t}.
\]

Definition 1.5. Simply-compounded spot interest rate. The simply-compounded spot interest rate \( L(t, T) \) is the constant rate at which an investment has to be made to produce one unit of cash at maturity \( T \), starting from \( P(t, T) \) units of cash at time \( t \), when accruing is proportional to the investment time. In formulas,
\[
L(t, T) = \frac{1 - P(t, T)}{(T - t)P(t, T)}.
\]

Remark. The market LIBOR and EURIBOR rates are simply-compounded rates, whose day-count convention is ”Actual”/360. This means that the year is assumed to be 360 days long and the corresponding year fraction is the actual number of days between two dates divided by 360.

1.1.3 Forward Rates
Forward rates are interest rates that can be locked in today for an investment in a future time period. Their values can be derived directly from zero-coupon bond prices. Define \( f(t, T, S) \) to be the continuously-compounded forward rate at time \( t \) for the expiry time \( T \) and maturity time \( S \). We must have
\[
\exp(R(t, S)(S - t)) = \exp(R(t, T)(T - t)) \exp(f(t, T, S)(S - T)),
\]
so that
\[
f(t, T, S) = \frac{1}{(S - T)} \ln \frac{P(t, T)}{P(t, S)},
\]
since otherwise arbitrage would be possible.
CHAPTER 1. INTRODUCTION TO INTEREST RATE THEORY

Analogous to the instantaneous short-term rate the instantaneous forward rate \( f(t, T) \) at time \( t \) for the maturity \( T \) is defined by

\[
f(t, T) = \lim_{S \to T^+} f(t, T, S) = -\frac{\partial \ln P(t, T)}{\partial T},
\]

(1.9)

so that we also have

\[
P(t, T) = \exp \left( -\int_t^T f(t, u)du \right).
\]

(1.10)

Beside the continuously compounded forward rate, a simply-compounded forward rate can be defined as well.

**Definition 1.6. Simply-compounded forward interest rate.** The simply compounded forward interest rate at time \( t \) for the expiry \( T \) and maturity \( S \) is denoted by \( F(t, T, S) \) and is defined by

\[
F(t, T, S) = \frac{1}{(S - T)} \left( \frac{P(t, T)}{P(t, S)} - 1 \right).
\]

(1.11)

**Remark.** Expression (1.11) can be derived from a forward rate agreement. This is a contract where at maturity \( S \), a fixed payment based on a fixed rate \( K \) is exchanged against a floating payment based on the rate \( L(T, S) \). Formally, at time \( S \) one receives \((S - T)K\) units of cash and pays the amount \((S - T)L(T, S)\). The value of the contract at time \( S \) is therefore

\[(S - T)(K - L(T, S)).\]

Discounting this value to time \( t \) leads to

\[
\]

(1.12)

The simply-compounded forward rate \( F \) is now the fixed rate that must be inserted for \( K \) to render the contract fair at time \( t \), so that the contract value (1.12) is 0 at time \( t \).

1.1.4 Interest Rate Swaps

An interest rate swap (IRS) is a contract that exchanges interest payments between two differently indexed legs, of which one is usually fixed whereas the other one is floating. When the fixed leg is paid and the floating leg is received the interest rate swap is termed payer IRS and in the other case receiver IRS.
 CHAPTER 1. INTRODUCTION TO INTEREST RATE THEORY

The present value at time $t = T_0$ of borrowing one unit of cash at a fixed rate $K$ with coupons paid at times $T_i$, $i = 1, \ldots, n$ and with $\tau_i = T_i - T_{i-1}$ is

$$PV(\text{fixed leg}) = \sum_{i=1}^{n} P(t, T_i)\tau_i K + P(t, T_n).$$ (1.13)

and the present value at time $t = T_0$ of a stream of floating rate cashflows is

$$PV(\text{floating leg}) = \sum_{i=1}^{n} P(t, T_i)\tau_i L(T_{i-1}, T_i) + P(t, T_n).$$ (1.14)

Thus, the present value at time $t = T_0$ of a payer IRS is given by

$$PV(\text{Payer IRS}) = \sum_{i=1}^{n} P(t, T_i)\tau_i (L(T_{i-1}, T_i) - K).$$ (1.15)

For simplicity we have assumed that the tenors of the floating and fixed legs are the same.\footnote{Indeed, a typical interest rate swap in the market has a fixed leg with annual payments and a floating leg with quarterly or semiannual payments.}

**Definition 1.7.** Swap rate. The swap rate is the rate $S$ that must be inserted for $K$ in equation (1.13) in order to have

$$PV(\text{fixed leg}) = PV(\text{floating leg}).$$ (1.16)

By simplifying the value of the floating leg to

$$PV(\text{floating leg}) = \sum_{i=1}^{n} P(t, T_i)\tau_i L(T_{i-1}, T_i) + P(t, T_n)$$

$$= \sum_{i=1}^{n} P(t, T_i)\tau_i \frac{1 - P(T_{i-1}, T_i)}{\tau_i P(T_{i-1}, T_i)} + P(t, T_n)$$

$$= \sum_{i=1}^{n} \left( \frac{P(t, T_i)}{P(T_{i-1}, T_i)} - P(t, T_i) \right) + P(t, T_n)$$

$$= \sum_{i=1}^{n} (P(t, T_{i-1}) - P(t, T_i)) + P(t, T_n) = 1,$$ (1.17)

and by using equation (1.13) we can express the swap rate $S(t, T_n)$ in terms of bonds prices

$$S(t, T_n) = \frac{1 - P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)}. (1.18)$$
We have now only regarded swaps starting at time \( t = T_0 \). If, however, \( t < T_0 \) cashflows are exchanged starting at a future time, rather than immediately and we have a forward start swap. So the value of the floating leg must be discounted and is therefore \( P(t, T_0) \). Consequently the value of the payer forward start swap is

\[
P(t, T_0) - \sum_{i=1}^{n} P(t, T_i)\tau_i K - P(t, T_n).
\]

This is 0 when \( K \) is the forward start swap rate \( S(t, T_0, T_n) \),

\[
S(t, T_0, T_n) = \frac{P(t, T_0) - P(t, T_n)}{\sum_{i=1}^{n} \tau_i P(t, T_i)}.
\]  

(1.19)

### 1.2 No-Arbitrage Pricing

The absence of arbitrage opportunities between all bonds with different maturities and the bank account is the fundamental economic assumption which will be introduced in this section and which all further considerations are based on.

**Definition 1.8. Arbitrage-free family of bond prices.** A family \( P(t, T) \), \( t \leq T \leq T^* \), of adapted processes is called an arbitrage-free family of bond prices relative to \( r \) if the following conditions hold:

i) \( P(T, T) = 1 \) for all \( T \in [0, T^*] \) and

ii) there exists a probability measure \( \mathbb{P}^* \) on \( (\Omega, \mathcal{F}_{T^*}) \) equivalent\(^6\) to \( \mathbb{P} \), such that for all \( t \in [0, T] \) the discounted bond price

\[
\tilde{P}(t, T) = D(0, t)P(t, T) = \frac{B(0)}{B(t)} P(t, T) = \frac{P(t, T)}{B(t)}
\]

is a martingale under \( \mathbb{P}^* \).

Any probability measure \( \mathbb{P}^* \) that satisfies the required conditions of definition 1.8 is named *martingale measure* for the family \( P(t, T) \). Actually, definition 1.8 is based on the general result that the existence of an equivalent martingale measure implies the absence of arbitrage opportunities in a standard market model. As \( \tilde{P}(t, T) \) follows a martingale under \( \mathbb{P}^* \), we have:

\[
\tilde{P}(t, T) = \mathbb{E}_{\mathbb{P}^*}(\tilde{P}(T, T)|\mathcal{F}_t) \quad \text{for} \quad t \leq T.
\]

\(^6\)\( \mathbb{P} \) and \( \mathbb{P}^* \) are equivalent measures, if \( \mathbb{P}(A) = 0 \iff \mathbb{P}^*(A) = 0 \) for every \( A \in \mathcal{F}_{T^*} \).
Therefore,
\[
D(0, t)P(t, T) = \mathbb{E}_{\mathbb{P}^*}(D(0, T)P(T, T) | \mathcal{F}_t) = \mathbb{E}_{\mathbb{P}^*}(D(0, T) | \mathcal{F}_t),
\]
which leads to the following expression for the bond price
\[
P(t, T) = D(0, t)^{-1}\mathbb{E}_{\mathbb{P}^*}(D(0, T) | \mathcal{F}_t)
= \exp \left( \int_0^t r(s)ds \right) \mathbb{E}_{\mathbb{P}^*} \left( \exp \left( -\int_0^T r(s)ds \right) | \mathcal{F}_t \right)
= \mathbb{E}_{\mathbb{P}^*} \left( \exp \left( -\int_t^T r(s)ds \right) | \mathcal{F}_t \right) = \mathbb{E}_{\mathbb{P}^*}(D(t, T) | \mathcal{F}_t). \tag{1.21}
\]
Thus, \( P(t, T) \) corresponds to the expectation of the stochastic discount factor \( D(t, T) \) under \( \mathbb{P}^* \). So we directly obtained the unique no-arbitrage price for bonds, which is again a special case of the general no-arbitrage price associated with an attainable contingent claim \( H \) given by
\[
\pi_t = \mathbb{E}_{\mathbb{P}^*}(D(t, T)H | \mathcal{F}_t). \tag{1.22}
\]

1.3 Factor Models of the Term Structure

As we have seen in the previous section, the zero-coupon bond price is given by the following expression:
\[
P(t, T) = \mathbb{E}_{\mathbb{P}^*} \left( \exp \left( -\int_t^T r(s)ds \right) | \mathcal{F}_t \right). \tag{1.23}
\]
So, whenever we can characterize the distribution of \( \exp(-\int_t^T r(s)ds) \), we are able to compute bond prices. The general idea of a factor model for the yield curve is to suppose that there is a Markov process \( X \) valued in some open subset \( D \subset \mathbb{R}^n \) such that, for any times \( t \) and \( T \), the market value \( P^M(t, T) \) of a zero-coupon bond at time \( t \) is given by \( g(X_t, \tau) \), where \( \tau = T - t \) and \( g \in C^{2,1}(D \times [0, \infty)) \).

To start with we assume that \( X \) follows an \( n \)-dimensional Itô process under the actual probability \( \mathbb{P} \), i.e.
\[
dX_t = \mu(X_t, t)dt + \sigma(X_t, t)dW_t, \tag{1.24}
\]
where \( W_t \) is a standard \( \mathbb{P} \)-Brownian motion in \( \mathbb{R}^n \). In order to guarantee the existence of a unique solution, \( \mu : D \times [0, T^*] \to \mathbb{R}^n \) and \( \sigma : D \times [0, T^*] \to \mathbb{R}^{n \times n} \) must be measurable functions that satisfy the following conditions,
\[
|\mu(x, t)| + |\sigma(x, t)| \leq C_1(1 + |x|) \quad x \in D, t \in [0, T^*],
|\mu(x, t) - \mu(y, t)| + |\sigma(x, t) - \sigma(y, t)| \leq C_2|x - y| \quad x, y \in D, t \in [0, T^*],
\tag{1.25}
\]
for some constants $C_1$ and $C_2$, where $|\sigma|^2 = \sum |\sigma_{ij}|^2$ (see Øksendal [15]).

**Remark.** Note that (1.24) is a short form of the following integral representation

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

(1.26)

We will now consider *time-homogeneous* Itô processes, meaning that the functions $\mu$ and $\sigma$ only depend on $X$ and not on $t$, i.e.

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t,$$  

(1.27)

where $W_t$ is again a standard $\mathbb{P}$-Brownian motion in $\mathbb{R}^n$ and where $\mu$ and $\sigma$ satisfy the conditions of (1.25), which in this case can be simplified to

$$|\mu(x) - \mu(y)| + |\sigma(x) - \sigma(y)| \leq C|x - y| \quad x, y \in D.$$  

(1.28)

In the following we will briefly explain an important property of these processes, namely the *Markov property*.

**Definition 1.9. Markov process.** The stochastic process $(X_t, t \in T^*)$ is called Markov if for every $n$ and $t_1 < t_2 < \ldots < t_n$,

$$\mathbb{P}(X_{t_n}|X_{t_{n-1}}, \ldots, X_{t_1}) = \mathbb{P}(X_{t_n}|X_{t_{n-1}}).$$  

(1.29)

**Theorem 1.1. The Markov property for Itô processes.** Let $X^x_t$ be a *time-homogeneous* Itô process of the form

$$dX^x_t = \mu(X^x_t)dt + \sigma(X^x_t)dW_t, \quad X^x_0 = x,$$  

(1.30)

where $\mu$ and $\sigma$ satisfy the conditions of (1.28) and let $f$ be a bounded Borel function from $\mathbb{R}^n$ to $\mathbb{R}$. Then, for $t, s \geq 0$

$$\mathbb{E}(f(X^x_{t+s})|\mathcal{F}_s) = \mathbb{E}(f(X^y_s))|_{y=X^x_t}.$$  

(1.31)

This Markov property can now be used for calculating the price of the zero-coupon bond $P(t,T)$, given by (1.23).

**Proposition 1.2. Let $X^z_t$ be of form (1.30) and the short rate process $r$ defined as $r(t) := R(X^z_t)$ where $R : D \to \mathbb{R}$, then there exists a measurable function $g : D \times \mathbb{R}_{\geq 0} \to \mathbb{R}$, such that

$$P(t,T) = g(X^z_t, T-t) = \mathbb{E}_{\mathbb{P}_x}\left( \exp\left(-\int_0^{T-t} R(X^y_x)ds\right) \right)|_{y=X^z_t}. $$  

(1.32)
CHAPTER 1. INTRODUCTION TO INTEREST RATE THEORY

Proof. The result follows directly from the fact that $P(t, T)$ is given by

$$
\mathbb{E}_{\mathbb{F}}\left( \exp\left(-\int_{t}^{T} R(X_s)ds\right) \bigg| \mathcal{F}_t \right)
$$

and by applying theorem (1.1) to this expression.

Remarks. To be consistent with (1.5) the function $R$ defining the short rate process $r$ must be of the form

$$
R(x) = \lim_{\tau \to 0^{+}} -\frac{\ln g(x, \tau)}{\tau}, \quad x \in D,
$$

where $\tau = T - t$.

We are interested in choices for $(g, \mu, \sigma)$ that are compatible, in the sense that (1.32) with (1.33) is fulfilled. One important class of compatible models are affine models whose main properties will be described in section 2.1.

1.3.1 Dynamics under $\mathbb{P}^*$

We are now interested in the dynamics of (1.24) under an equivalent martingale measure $\mathbb{P}^*$ for the bond market. Note that for a fixed maturity date $T$, the zero-coupon bond price $P(t, T)$ is a function of $X_t$ and $t$, i.e. $P(t, T) = G(X_t, t)(\equiv g(X_t, T - t)), t \leq T$. By applying Itô’s formula we get

$$
dP(t, T) = \alpha(X_t, t)dt + \beta(X_t, t)dW_t, \tag{1.34}
$$

$$
\alpha(X_t, t) = \frac{\partial G(X_t, t)}{\partial t} + \frac{\partial G(X_t, t)}{\partial X}\mu(X_t, t) + \frac{1}{2}tr\left(\frac{\partial^2 G(X_t, t)}{\partial X^2}\sigma(X_t, t)\sigma'(X_t, t)\right),
$$

$$
\beta(X_t, t) = \frac{\partial G(X_t, t)}{\partial X}\sigma(X_t, t),
$$

where $\frac{\partial G}{\partial X} = (\frac{\partial G}{\partial X_1}, \ldots, \frac{\partial G}{\partial X_n})$. Furthermore we consider a self-financing portfolio consisting of $\phi^1$ units of the zero-coupon bond $P^1$ and $\phi^2$ units of another zero-coupon bond $P^2$ (i.e. with another maturity date $T$), whose value process is given by

$$
V_t(\phi) = \phi^1 P^1(t, T^1) + \phi^2 P^2(t, T^2). \tag{1.35}
$$

As $\phi$ is self-financing strategy, we have

$$
dV_t(\phi) = \phi^1 dP^1(t, T^1) + \phi^2 dP^2(t, T^2) \tag{1.36}
$$
In order to have a risk-less portfolio we chose \( \phi_i \) for every considered bond \( H \). Hence, the term
\[
d dV_i = (\phi_i^1 \alpha^1(X_t, t) + \phi_i^2 \alpha^2(X_t, t))dt + (\phi_i^1 \beta^1(X_t, t) + \phi_i^2 \beta^2(X_t, t))dW_t.
\]
In order to have a solution for \( \phi_i \) that the following equality holds for each component
\[
d dV_i = (\phi_i^1 \alpha^1(X_t, t) + \phi_i^2 \alpha^2(X_t, t))dt = r(t)(\phi_i^1 P^1(t, T^1) + \phi_i^2 P^2(t, T^2))dt.
\]
Thus, we get a linear system of equations whose unknowns are \( \phi_i^1 \) and \( \phi_i^2 \):
\[
\phi_i^1 (\alpha^1(X_t, t) - r(t)P^1(t, T^1)) + \phi_i^2 (\alpha^2(X_t, t) - r(t)P^2(t, T^2)) = 0,
\]
\[
\phi_i^1 \lambda_i^1(X_t, t) + \phi_i^2 \lambda_i^2(X_t, t) = 0 \quad \text{for all } i. \tag{1.37}
\]
In order to have a solution for \( \phi_i^1 \) and \( \phi_i^2 \) that is different from 0 it is necessary that the following equality holds for each component \( i \) of the vector \( \beta \):
\[
\frac{\alpha^1(X_t, t) - r(t)P^1(t, T^1)}{n \lambda_i^1(X_t, t)} = \frac{\alpha^2(X_t, t) - r(t)P^2(t, T^2)}{n \lambda_i^2(X_t, t)}. \tag{1.38}
\]
Hence, the term
\[
\lambda_i(t) = \frac{\alpha(X_t, t) - r(t)P(t, T)}{n \lambda_i(X_t, t)} = \frac{\alpha(X_t, t) - r(t)}{n \lambda_i(X_t, t)} \tag{1.39}
\]
is invariant for every considered bond \( P(t, T) \), i.e. it does not depend on \( T \).

**Remark.** Equation (1.34) expresses the bond-price dynamics in terms of the short-term rate \( r \), with \( \alpha/P \) being the return and \( \beta/P \) the volatility of the bond. For this reason \( \lambda = (\lambda_1, \ldots, \lambda_n) \) can be interpreted as follows: The difference \( \alpha(X_t, t)/P(t, T) - r(t) \) represents the difference in returns with respect to the risk-less bank account. By dividing by \( (\beta(X_t, t)/P(t, T)) \), we divide by the riskiness of the zero-coupon bond. That is why \( \lambda \) is referred to as risk premium, market price of risk or as proposed by Brigo and Mercurio [3] as “excess return with respect to a risk-free investment per unit of risk”.

If \( \lambda(t) \) satisfies Novikov’s condition, i.e.
\[
\mathbb{E}_\mathbb{F}\left(\exp\left(\frac{1}{2} \int_0^T \lambda(s) \lambda'(s) ds\right)\right) < \infty, \tag{1.40}
\]
we are able to define a probability measure \( P^\ast \) equivalent to \( P \) by the Radon-Nikodym derivative
\[
\frac{dP^\ast}{dP} \Big|_{F_t} = \exp \left( - \int_0^t \lambda(s)dW_s - \frac{1}{2} \int_0^t \lambda(s)\lambda'(s)ds \right), \quad \mathbb{P}\text{-a.s.} \tag{1.41}
\]
Then, in view of Girsanov’s theorem\(^7\), the process
\[
W^\ast_t = W_t + \int_0^t \lambda(s)'ds \quad \text{for all} \quad t \in [0, T^\ast] \tag{1.42}
\]
follows a standard Brownian motion under \( P^\ast \).

**Remark.** Due to the construction of \( \lambda \), where we used the no-arbitrage argument, \( P^\ast \) satisfies the conditions of definition 1.8 and is therefore an equivalent martingale measure. More specifically, the dynamics of the discounted bond price under \( P \) are given by
\[
d\tilde{P}(t,T) = \exp \left( - \int_0^t r(s)ds \right) P(t,T) \tag{1.43}
\]
By moving from \( P \) to \( P^\ast \) using (1.42) we have
\[
dW^\ast_t = dW_t + \lambda(t)'dt,
\]
which leads to
\[
d\tilde{P}(t,T) = \exp \left( - \int_0^t r(s)ds \right) \beta(X_t,t)dW^\ast_t. \tag{1.44}
\]
Thus, as \( \tilde{P}(t,T) \) can be represented as a stochastic integral under \( P^\ast \), it is a martingale and \( P^\ast \) therefore an equivalent martingale measure.

Now we have all necessary tools to state the following result that shows the behavior of \( X \) under \( P^\ast \).

**Proposition 1.3.** Let \( P(t,T) \) be an arbitrage-free family of bond prices and assume that \( X \) follows an Itô process under the actual probability \( P \), as specified by (1.24). Then for any equivalent martingale measure \( P^\ast \) of definition 1.8, whose Radon-Nikodym derivative is given by (1.41) the process \( X \) satisfies under \( P^\ast \)
\[
\begin{align*}
dX_t &= \mu^\ast(X_t,t)dt + \sigma(X_t,t)dW^\ast_t \quad \text{with} \tag{1.45} \\
\mu^\ast(X_t,t) &= \mu(X_t,t) - \sigma(X_t,t)\lambda(t)'. \tag{1.46}
\end{align*}
\]

\(^7\)See Øksendal [15] for details.
Proof. From (1.42) we know that \(dW_t = dW^*_t - \lambda(t)'dt\) holds. Combining this expression with (1.24) we immediately get
\[
dX_t = \mu(X_t, t)dt + \sigma(X_t, t)(dW^*_t - \lambda(t)'dt).
\]

Remark. It is essential to assume that the function \(\lambda\) is sufficiently regular, so that the SDE (1.45) admits a unique global strong solution.

1.3.2 The Bond Price as Solution of a PDE

If we replace in equation (1.39) \(\alpha\) and \(\beta\) by their definition we obtain
\[
\frac{\partial G(X_t, t)}{\partial t} + \frac{\partial G(X_t, t)}{\partial X}(\mu(X_t, t) - \sigma(X_t, t)\lambda(t)') + \\
+ \frac{1}{2} tr\left(\frac{\partial^2 G(X_t, t)}{\partial X^2}\sigma(X_t, t)\sigma'(X_t, t)\right) - rG(X_t, t) = 0.
\]

By using (1.46) the above expression becomes
\[
\frac{\partial G(X_t, t)}{\partial t} + \frac{\partial G(X_t, t)}{\partial X}\mu^*(X_t, t) + \\
+ \frac{1}{2} tr\left(\frac{\partial^2 G(X_t, t)}{\partial X^2}\sigma(X_t, t)\sigma'(X_t, t)\right) - rG(X_t, t) = 0 \tag{1.47}
\]
with the terminal condition \(P(T, T) = G(X_T, T) = 1\). It follows from the Feynman-Kac\(^8\) formula that under mild technical assumptions the risk-neutral valuation formula for bond prices \(E_{\mathbb{P}^*}(\exp(-\int_t^T r(s)ds)|\mathcal{F}_t)\) solves PDE (1.47).

---

\(^8\)See Øksendal [15] for details.
Chapter 2

Affine Models

In this chapter we study the theory of affine models which rank among the most popular models in theory and practice and of which many examples have been investigated. For instance, the very first term structure model, the Vasicek model, is an affine model. Other popular models such as Cox, Ingersoll, Ross (CIR), Hull and White or Longstaff and Schwartz are also of this type. Their popularity is due to their tractability and to their flexibility, because there are often explicit solution for bond prices and bond option prices.

Affine models were investigated amongst others by Duffie and Kan [8] who developed a general theory. Dai and Singelton [6] have provided a classification and have established the most general representative example of each class of affine models.

In the following sections we explain the general theory and properties, whereas important examples and practical applications are examined in chapter 3. Furthermore we give an overview of the three main families of affine models that have appeared in the literature. Complementary to this, we describe the Dai and Singelton [6] classification. We follow the approach of James & Webber [11] and Cairns [4] considering in particular the papers of Duffie and Kan [8] and Dai and Singelton[6].

2.1 Theory of Affine Factor Models

To start with we consider a model with \( n \) state variables

\[ X(t) = (X_1(t), \ldots, X_n(t))^\prime, \]

which are valued in some open subset \( D \subset \mathbb{R} \) and follow a \( n \)-dimensional Itô process described by (1.45) (i.e under \( \mathbb{P}^* \)), whose exact form will be specified
by theorem 2.2. Affine multi-factor models are characterized by the fact that
the zero-coupon bond prices can be written in the form

$$P(t, T) = \exp \left( A(t, T) + \sum_{i=1}^{n} B_i(t, T)X_i(t) \right)$$

$$= \exp(A(t, T) + B(t, T)'X(t)) \quad (2.1)$$

with

$$B(t, T) = (B_1(t, T), \ldots, B_n(t, T))'.$$

Concerning the yields $R(t, T)$, it is obvious that they have to be of the
following form

$$R(t, T) = -A(t, T) \frac{T - t}{T - t} - B(t, T)'X(t). \quad (2.2)$$

Considering remark (1.33) the short-term rate $r(t)$ must be the limit of
$R(t, T)$ for $T$ going to $t$. Thus, we can express $r(t)$ by means of (2.2), i.e.

$$r(t) = R(X(t)) = g + h'X(t), \quad (2.3)$$

where $g$ and $h$ are constants:

$$g = \lim_{T\to t^+} -\frac{A(t, T)}{T - t},$$

$$h' = \lim_{T\to t^+} -\frac{B(t, T)'}{T - t}.$$

**Remarks.** The model is time-homogeneous if $X(t)$ is time-homogeneous and
the functions $A(t, T)$ and $B(t, T)$ are functions of $\tau = T - t$. In the following
we will restrict our considerations to this time-homogeneous case.
The function $g(X(t), T-t)$ specified by proposition (1.2) is here consequently
given by

$$g(X(t), \tau) = \exp(A(\tau) + B(\tau)'X(t)). \quad (2.4)$$

### 2.1.1 Specification of the State Variable Process

The next proposition and the following theorem specify the conditions that
the functions $\mu^*$ and $\sigma$ of SDE (1.45)$^1$ must satisfy under the assumption
that $(g, \mu^*, \sigma)$ is a compatible model in the sense of remark (1.33), where $g$
given by (2.4).

$^1$In contrast to equation (1.45) $\mu^*$ and $\sigma$ only depend on $X$ and not on $t$, as we only
consider the time-homogeneous case.
Proposition 2.1. Suppose that \((g, \mu^*, \sigma)\) is a compatible term structure factor model with functions \(\mu^*\) and \(\sigma\) of SDE (1.45). If \(g\) is of form (2.4) and there exist maturities \(\tau_1, \ldots, \tau_N\) for \(N = 2n + (n^2 - n)/2\) such that the \(N \times N\) matrix \(C(\tau_1, \ldots, \tau_N)\), whose \(i\)th row is of the form\(^2\)

\[
c(\tau_i)' = (c_1(\tau_i), \ldots, c_n(\tau_i), c_n+1(\tau_i), c_{n+2}(\tau_i), c_{n+3}(\tau_i), \ldots, c_N(\tau_i))
\]

\[
= \left( B_1(\tau_i), \ldots, B_n(\tau_i), \frac{B_1(\tau_i)^2}{2}, B_1(\tau_i)B_2(\tau_i), B_1(\tau_i)B_3(\tau_i), \ldots, \frac{B_n(\tau_i)^2}{2} \right),
\]

is non-singular, then \(\mu^*, \sigma\sigma'\) and \(r\) are affine.

Proof. From (2.3) we already know that \(r\) is affine. In order to show that \(\mu^*\) and \(\sigma\sigma'\) are affine functions we have to calculate the derivatives of \(g(X(t), \tau)\) and insert them in PDE (1.47). The partial derivatives of \(G(X(t), t) \equiv g(X(t), \tau)\) are given by

\[
\frac{\partial G}{\partial t} = -\frac{\partial g}{\partial \tau} = g(X(t), \tau) \left( -\frac{\partial A(\tau)}{\partial \tau} - \frac{\partial B(\tau)'}{\partial \tau} X(t) \right),
\]

\[
\frac{\partial G}{\partial X_i} = \frac{\partial g}{\partial X_i} = g(X(t), \tau) B_i(\tau),
\]

\[
\frac{\partial^2 G}{\partial X_i \partial X_j} = \frac{\partial^2 g}{\partial X_i \partial X_j} = g(X(t), \tau) B_i(\tau) B_j(\tau).
\]

Consequently, by (1.47) and (2.3)

\[
g(X(t), \tau) \left( -\frac{\partial A(\tau)}{\partial \tau} - \frac{\partial B(\tau)'}{\partial \tau} X(t) + B(\tau)\mu^*(X(t)) + 
\right.

\[
\left. + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n B_i(\tau) B_j(\tau) \sigma_i(X(t)) \sigma_j(X(t))' - (g + h'X(t)) \right) = 0,
\]

(2.6)

where \(\sigma_i\) denotes the \(i\)th row of the matrix \(\sigma\). Since \(g(X(t), \tau)\) is strictly positive valued, (2.6) is equivalent to

\[
\sum_{i=1}^n B_i(\tau) \mu^*_i(X(t)) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n B_i(\tau) B_j(\tau) \kappa_{ij}(X(t)) = 
\]

\[
\frac{\partial A(\tau)}{\partial \tau} + \frac{\partial B(\tau)'}{\partial \tau} X(t) + g + h'X(t),
\]

:= a(X(t), \tau)

(2.7)

\(^2\)Regard the proof for the construction of this matrix.
where \( \kappa_{ij}(X(t)) = \sigma_i(X(t))\sigma_j(X(t))' \). Under a mild non-degeneracy condition this differential equation implies that \( \mu^* \) and \( \sigma^* \) are affine functions. In order to see this, note at first that the right side of (2.7), which we will denote \( a(\cdot, \tau) \), is affine for each fixed \( \tau \). Define now a function \( H : D \to \mathbb{R}^N \) for \( N = 2n + (n^2 - n)/2 \)

\[
H(x) := (\mu_1^*(x), \mu_2^*(x), \ldots, \kappa_{11}(x), \kappa_{12}(x), \ldots, \kappa_{nn}(x))',
\]

where only the \( \kappa_{ij}(x) \) with \( i \leq j \) are included. Now the aim is to show that each component of \( H \) is affine in \( x \). Equation (2.7) can be written as a system of equation in \( \tau \) and \( x \) of the form

\[
c(\tau)'H(x) = a(x, \tau),
\]

where \( c : \mathbb{R}_{\geq 0} \to \mathbb{R}^N \). For example, \( c_1(\tau) = B_1(\tau)(\text{the coefficient of } H_1(x)) \), while \( c_{n+1}(\tau) = B_1(\tau)^2/2 \) (the coefficient of \( \kappa_{11}(x) \)). By extending (2.8) to \( N \) maturities \( \tau_1, \ldots, \tau_N \) we get

\[
C(\tau_1, \ldots, \tau_N)H(x) = \begin{pmatrix} a(x, \tau_1) \\ a(x, \tau_2) \\ \vdots \\ a(x, \tau_N) \end{pmatrix},
\]

where \( C(\tau_1, \ldots, \tau_N) \) is the \( N \times N \) matrix whose \( i \)-th row is \( c(\tau_i)' \). So, if \( \tau_1, \ldots, \tau_N \) can be chosen, such that \( C(\tau_1, \ldots, \tau_N) \) is non-singular, then there is a unique solution \( H(\cdot) \) to (2.9), which is a linear combination of affine functions (since the right side is a vector of affine functions), and is therefore affine (compare with Duffie and Kan [8]).

**Remark.** Of course, for arbitrary distinct non-zero maturity times \( \tau_1, \ldots, \tau_N \) the matrix \( C(\tau_1, \ldots, \tau_N) \) is non-singular except for \( (B(\tau_1), \ldots, B(\tau_N)) \) in a closed subset of measure zero of \( \mathbb{R}^{Nn} \).

**Theorem 2.2. Duffie and Kan, 1996 [8].** If \( P(t, T) = P(t, t + \tau) = g(X(t), \tau) \) can be written in the form (2.4) then the process for \( X(t) \) must solve the following SDE

\[
dX(t) = (\alpha + \beta X(t))dt + SD(X(t))dW_t^*,
\]

where \( W_t^* \) is a \( n \)-dimensional Brownian motion under \( \mathbb{P}^* \), \( \alpha = (\alpha_1, \ldots, \alpha_n)' \) is a constant vector, \( \beta = (\beta_{ij}) \) and \( S = (s_{ij}) \) are constant \( n \times n \) matrices and, finally, \( D(X(t)) \) is a diagonal matrix of the form

\[
D(X(t)) = \begin{pmatrix} \sqrt{\gamma_1}X(t) + \delta_1 \\ \vdots \\ 0 \\ \sqrt{\gamma_n}X(t) + \delta_n \end{pmatrix},
\]
where $\delta_i \in \mathbb{R}$ and $\gamma_i = (\gamma_{i1}, \ldots, \gamma_{in})' \in \mathbb{R}^n$.

Proof. The proof of this theorem is based on proposition (2.1). If $\sigma \sigma'$ is affine in $x$, then under non-degeneracy conditions and a possible re-ordering of indices $\sigma(X(t)) = SD(X(t))$, where $D(X(t))$ is of form (2.11).


2.1.2 Affine Stochastic Differential Equations

As indicated by the last theorem, the affine class of term structure models seems to be well behaved. Now we formulate the conditions on the coefficients of equation (2.10) under which there is indeed a unique strong solution to the SDE. In order to assure this, there are two problems to cope with: Firstly, the diffusion function $\sigma(X(t)) = SD(X(t))$ is not Lipschitz continuous and secondly $\gamma'_i X(t) + \delta_i$ must be non-negative for all $i$ and $t$. The open domain $G$ with non-negative volatilities is

$$G = \{ x \in \mathbb{R}^n : \gamma'_i x + \delta_i > 0, \ i \in \{1, \ldots, n\} \}. \quad (2.12)$$

In effect, to guarantee the existence of a solution $X$, it is necessary to assume that for each $i$ the volatility process $\gamma'_i X(t) + \delta_i$ has as sufficiently strong positive drift on the $i$-th boundary segment $\partial G_i = \{ x \in G : \gamma'_i x + \delta_i = 0 \}$. The following theorem states the conditions which guarantee the existence of a unique solution to (2.10) that remains in $G$.

**Theorem 2.3.** If the following conditions hold for all $i$,

i) if $\gamma'_i x + \delta_i = 0$, then $\gamma'_i (\beta x + \alpha) > \left(\gamma'_i S S' \gamma_i\right)/2$ for all $x \in \mathbb{R}^n$,

ii) if $(\gamma'_i S)_j \neq 0$, then $\gamma'_i x + \delta_i = \gamma'_j x + \delta_j$ for all $j$,

then there exists a unique (strong) solution $X$ in $G$ to the stochastic differential equation (2.10) with (2.11) and (2.12). Moreover, for this solution $X$ and for all $i$, we have $\gamma'_i X(t) + \delta_i > 0$ for all $t$ almost surely.

Proof. See Duffie and Kan [8]

Remarks. Both conditions of theorem (2.3) are designed to ensure strictly positive volatility. As they are not generally satisfied, they are a significant restriction on the model.

For a state process $X(t)$ satisfying the conditions of this theorem, there is always a strictly positive non constant short rate process $r(t)$ given by (2.3).
2.1.3 Ricatti Equations

In this section we show how to obtain two differential equations for $A(\tau)$ and $B(\tau)$, which are known as Ricatti equations. At first, we can now write (2.7) in the form

$$-\frac{\partial A(\tau)}{\partial \tau} - \frac{\partial B(\tau)'}{\partial \tau} X(t) + B(\tau)'(\alpha + \beta X(t)) +$$

$$+ \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} B_i(\tau)B_j(\tau)\sigma_i\sigma_j' - (g + h'X(t)) = 0,$$

where $\sigma_i$ denotes the $i^{th}$ row of the matrix $SD(X(t))$, i.e

$$\sigma_i = (s_{i1}\sqrt{\gamma_{1}^{'}X(t)} + \delta_{1},... ,s_{in}\sqrt{\gamma_{n}^{'}X(t)} + \delta_{n}).$$

This equation is affine in $x$ and if an affine relationship of the form $a + bx = 0$ for all $x$ in some non-empty open set, then $a = 0$ and $b = 0$. Therefore we have

$$-\frac{\partial A(\tau)}{\partial \tau} + B(\tau)'\alpha + \frac{1}{2} B(\tau)'S\text{diag}(\delta)S'B(\tau) - g = 0, \quad (2.13)$$

$$-\frac{\partial B_k(\tau)}{\partial \tau} + B(\tau)'\beta_k + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{l=1}^{n} B_i(\tau)B_j(\tau) s_{il}s_{jl}\gamma_{lk} - h' = 0, \quad (2.14)$$

for $k = 1,...,n$ with boundary conditions

$$A(0) = 0 \quad \text{and} \quad B(0) = 0. \quad (2.15)$$

$\beta_k$ is the $k^{th}$ column of the matrix $\beta$ and $\gamma_{lk}$ is the $k^{th}$ component of the vector $\gamma_l$. The boundary conditions follow from the fact that $P(T,T) = g(X(T),0) = 1$, which implies that $A(0) + B(0)'X(T) = 0$. Since $T$ is arbitrary, $A(0) + B(0)'x = 0$ must hold for all $x$ in $D$, which yields (2.15).

**Remarks.** There is a non-trivial issue of the existence of finite solutions to Ricatti equations, since the coefficients are not Lipschitz continuous. Solutions exist on the whole time domain for special cases. For any particular given case they exist up to some given time $T > 0$, which is due to the local Lipschitz continuity of the coefficients.

In general, the Ricatti equations need to be solved numerically, although in the case of Vasicek and CIR they are explicitly solvable.
2.2 Types of Affine Models

Commonly used affine models can be conveniently separated into three main types:

- Gaussian affine models
- CIR affine models
- Three factor affine family

Remark. This categorization is distinct from the classification provided by Dai and Singleton, which is a complete canonical mathematical classification (see section 2.3).

We consider now each of the three categories in turn. Note that the dynamics of \( X \) are always specified under the martingale measure \( \mathbb{P}^* \).

2.2.1 Gaussian Affine Models

All time-homogeneous Gaussian models are based on the following general model. Let \( X(t) = (X_1(t), \ldots, X_n(t))' \) evolve according to the following SDE

\[
dX(t) = (\alpha + \beta X(t))dt + \Sigma dW_t^*,
\]

where \( \alpha \in \mathbb{R}^n \), \( \beta, \Sigma \in \mathbb{R}^{n \times n} \), and \( W_t^* \) is a \( n \)-dimensional Brownian motion under \( \mathbb{P}^* \). As specified by (2.3) the short rate \( r(t) \) is a function

\[
r(t) = g + h'X(t)
\]

of the state variables \( X(t) \), where \( g \in \mathbb{R} \) and \( h \in \mathbb{R}^n \).

Remark. If all \( h_i \) are non-zero, it is possible to rescale the \( X_i(t) \) and assume that the \( h_i = 1 \) for all \( i \) without loss of generality.

In order to solve SDE (2.16) we apply the following theorem.

**Theorem 2.4. Variation of Constants.** Let \( B \) be some real valued \( n \times n \) matrix, \( a, \sigma_1, \ldots, \sigma_d \in \mathbb{R}^n \) and \( (W_t)_{t \geq 0} \) a \( d \)-dimensional Brownian motion on \((\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{0 \leq t \leq T^*})\). Then the solution of the SDE

\[
dX(t) = (a + BX(t))dt + \sum_{i=1}^d \sigma_i dW_i(t), \quad X(0) = x \in \mathbb{R}^n
\]

is given by

\[
X(t) = e^{Bt}x + \int_0^t e^{B(t-s)}ads + \sum_{i=1}^d \int_0^t e^{B(t-s)}\sigma_i dW_i(t).
\]

(2.19)
Remark. For any real-valued matrix $C$, $e^C = \sum_{i=0}^{\infty} \frac{C^i}{i!}$.

So, in our case $X(t) = e^{\beta t}X(0) + \int_0^t e^{\beta(t-s)}\alpha ds + \int_0^t e^{\beta(t-s)}\Sigma dW^*_s$, is the unique solution of SDE (2.16) and $r(t)$ is therefore given by

$$r(t) = g + h'e^{\beta t}X(0) + h' \int_0^t e^{\beta(t-s)}\alpha ds + h' \int_0^t e^{\beta(t-s)}\Sigma dW^*_s. \tag{2.20}$$

Remark. The matrix $\beta$ has a spectral decomposition $\beta = \beta_R \Lambda \beta_L$, where $\Lambda = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is the diagonal matrix of eigenvalues of $\beta$. $\beta_L$ and $\beta_R$ are the matrices of left and right eigenvectors respectively. The columns of $\beta_R$ can be scaled in a way which ensures that $\beta_R \beta_L = I$, so that $\beta^i$ can be easily calculated by $\beta_R \Lambda^i \beta_L$. Concerning the eigenvalues we have to require that the real parts of them are negative, so that $X(t)$ is stationary. This ensures that $\exp(\beta t)$ tends to zero as $t$ tends to infinity.

Equation (2.20) implies that $r(t)$ is normally distributed with mean and variance given respectively by

$$\mathbb{E}_{\mathbb{P}_*}(r(t)) = g + h'e^{\beta t}X(0) + h' \int_0^t e^{\beta(t-s)}\alpha ds, \tag{2.21}$$

$$\text{var}_{\mathbb{P}_*}(r(t)) = \int_0^t (h'e^{\beta(t-s)}\Sigma \Sigma' e^{\beta'(t-s)}h) ds. \tag{2.22}$$

Remark. One major drawback of gaussian models is the positive probability of negative interest rates, which is incompatible with no-arbitrage and the presence of cash.

The bond prices $P(t, T)$ can now be derived by computing the expectation $\mathbb{E}_{\mathbb{P}_*}(\exp(-\int_0^T r(s)ds)|\mathcal{F}_t)$. We will calculate this precisely for the one-dimensional Vasicek Model in section 3.1.1. Another way would be solving the PDEs (2.13) and (2.14), where $S = \Sigma$ and $\gamma_i = 0$ and $\delta_i = 1$.

### 2.2.2 CIR Affine Models

A model is CIR (Cox, Ingersoll, Ross) affine if all state variables $X(t) = (X_1(t), \ldots, X_n(t))'$ are independent processes of the one factor CIR-type. Thus for $i = 1, \ldots, n$,

$$dX_i(t) = (\alpha_i - \beta_i X_i(t))dt + \sigma_i \sqrt{X_i(t)}dW^*_i(t), \tag{2.23}$$
where $W^*_1(t), \ldots, W^*_n(t)$ are independent standard Brownian motions. For simplicity we assume here that the short-term rate is then defined as

$$r(t) = \sum_{i=1}^{n} X_i(t),$$

i.e. $g$ and $h$ of equation (2.3) are assumed to be 0 and $h_i = 1$ for all $i$ respectively. As the model is affine in every factor, the bond price $P_t(t,T)$ for the $i^{th}$ one-factor CIR-process can be written in the form $\exp(A_i(t) + B_i(t)X_i(t))$. Due to the independence of factors it is therefore immediate to derive the following formula. The price $P(t,T)$ at time $t$ of a zero coupon bond with maturity $T - t$ is explicitly given by

$$P(t,T) = \mathbb{E}_{P^*} \left( \exp \left( - \int_t^T r(s) ds \right) \bigg| \mathcal{F}_t \right) = \mathbb{E}_{P^*} \left( \exp \left( - \int_t^T \sum_{i=1}^{n} X_i(s) ds \right) \bigg| \mathcal{F}_t \right) = \prod_{i=1}^{n} \mathbb{E}_{P^*} \left( \exp \left( - \int_t^T X_i(s) ds \right) \bigg| \mathcal{F}_t \right) = \prod_{i=1}^{n} P_i(t,T) = \exp \left( \sum_{i=1}^{n} (A_i(t) + B_i(t)X_i(t)) \right).$$

### 2.2.3 The Three-Factor Affine family

This family represents models that mix Gaussian and CIR type state variables. The motivation for that stems from the desire to have a stochastic mean and a stochastic volatility. The three processes are:

1. The short rate process:

$$dr(t) = \kappa(\mu(t) - r(t))dt + \sqrt{\sigma(t)}dW^*_r(t)$$  \hspace{1cm} (2.26)

2. The drift process:

$$d\mu(t) = \beta(\gamma - \mu(t))dt + \eta \mu(t)dW^*_\mu(t),$$

where $\phi = 0, \frac{1}{2}$.  \hspace{1cm} (2.27)

3. The volatility process:

$$d\sigma(t) = \delta(\alpha - \sigma(t))dt + \lambda \sqrt{\sigma(t)}dW^*_\sigma(t)$$

As an example we will mention the BDFS three-factor model in section 3.2.2.
2.3 Classification of Affine Models

In order to enable affine models to be compared and classified Dai and Singleton [6] have provided a general framework where affine models can be classified according to

1. The number of state variables, and
2. How many of the state variables appear in the volatility matrix.

Since a model could be represented by a transformation of variables in two apparently different ways, it is necessary to find a general representative of each class, so that every affine model can be seen as special case of a general type. This section follows the treatment of Dai and Singleton.

2.3.1 A Canonical Representation

Dai and Singleton base their classification on processes for $X(t)$ that are specified in the form

$$
\text{d}X(t) = \kappa^*(\theta^* - X(t)) \text{d}t + SD(X(t)) \text{d}W^*_t,
$$

under $\mathbb{P}^*$, where $\kappa^*$ is a constant $n \times n$ matrix and $\theta^*$ is a constant $n$-dimensional vector. $S$ and $D(X(t))$ have the same form as in equation (2.10).

In order to obtain also an affine model under the measure $\mathbb{P}$ it is assumed that the market price of risk is given by

$$
\Lambda(t) = D(X(t))\lambda,
$$

where $\lambda \in \mathbb{R}^n$. Then, the process (2.29) evolves under $\mathbb{P}$ according to

$$
\text{d}X(t) = \kappa(\theta - X(t)) \text{d}t + SD(X(t)) \text{d}W_t,
$$

where, after proposition 1.3, $\kappa = \kappa^* - S\Phi$ and $\theta = \kappa^{-1}(\kappa^*\theta^* + S\psi)$. The $i^{th}$ row of the matrix $\Phi$ is $(\lambda_1, \gamma_{i1}, \ldots, \lambda_n, \gamma_{in})$ and the $i^{th}$ component of the vector $\psi$ is given by $\lambda_i \delta_i$. As specified by (2.3) the short rate is set to be $r(t) = g + h'X(t)$. The model under $\mathbb{P}$ is therefore specified by the structure

$$
\Lambda = (g, h, \kappa, \theta, S, \Gamma, \delta, \lambda),
$$

where $\Gamma = (\gamma_1, \ldots, \gamma_n) \in \mathbb{R}^{n \times n}$ is the matrix of coefficients on $X(t)$ in $D$ and $\delta$ the vector whose components are $\delta_i$.

---

3. The coefficients $\alpha$ and $\beta$ in (2.10) are now given by $\alpha = \kappa^* \theta^*$ and $\beta = -\kappa^*$.

4. $\Lambda(t)$ corresponds to $\lambda(t)'$ of equation (1.46).
CHAPTER 2. AFFINE MODELS

Let \( m = \text{rank}(\Gamma) \), i.e. \( m \) is the number of state variables that appear in the matrix \( D \). Using this index each \( n \)-factor model can be classified into one of \( n + 1 \) subfamilies based on its value of \( m \), since \( m \) can range from 0 to \( n \). \( \mathcal{A}_m(n) \) denotes the class of \( n \)-factor affine models with index \( m \).

**Definition 2.1. Canonical Representation.** Let \( m = \text{rank}(\Gamma) \), then for each \( m \), we partition \( X(t) \) as \( X' = (X'_1, X'_2) \), where \( X_1 \) is a \( m \)-dimensional vector and \( X_2 \) is a \((n - m)\)-dimensional vector. The canonical representation of \( \mathcal{A}_m(n) \) is defined as the special case of equation (2.31) with

\[
\kappa = \begin{pmatrix} \kappa_{1,1} & 0 \\ \kappa_{2,1} & \kappa_{2,2} \end{pmatrix},
\]

(2.32)

for \( \kappa_{1,1} \in \mathbb{R}^{m \times m} \), \( \kappa_{2,1} \in \mathbb{R}^{(n-m) \times m} \), \( \kappa_{2,2} \in \mathbb{R}^{(n-m) \times (n-m)} \) if \( m > 0 \) and for \( m = 0 \), \( \kappa \) is either upper or lower triangular, and

\[
\theta = \begin{pmatrix} \theta_1 \\ 0 \end{pmatrix},
\]

(2.33)

\[
S = I_{n \times n},
\]

(2.34)

\[
\delta = \begin{pmatrix} 0 \\ 1 \end{pmatrix},
\]

(2.35)

\[
\Gamma = \begin{pmatrix} I_{m \times m} & \Gamma_{1,2} \\ 0 & 0 \end{pmatrix},
\]

(2.36)

where \( I \) is the identity matrix, \( \theta_1 \in \mathbb{R}^m \), \( 1 \in \mathbb{R}^{(n-m)} \) and \( \Gamma_{1,2} \in \mathbb{R}^{m \times (n-m)} \).

**Remark.** Consistently with equation (2.12) parametric restrictions must be imposed. Basically \( \gamma_i X(t) + \delta_i \) must be strictly positive for all \( i \) and \( t \). Thus, it is the same requirement as that of Duffie and Kan. However as the process is regarded under \( \mathbb{P} \), \( D \) also affects the drift parameters. Consequently it is also necessary to constrain \( \kappa \) and \( \theta \) (see Dai and Singleton [6]).

**Definition 2.2. Equivalence class \( \mathcal{A}_m(n) \).** \( \mathcal{A}_m(n) \) is defined as the set of all affine models that are nested special cases of the canonical model or of any equivalent model obtained by invariant transformations of the canonical model.\(^5\)

**Remarks.** For instance, there are obviously two classes of one-factor models. On the one hand \( \mathcal{A}_0(1) \), the class of the Vasicek model, where the state variable does not appear in the volatility matrix and on the other hand \( \mathcal{A}_1(1) \), the class of the CIR model, whose volatility is modeled by a square root

\(^5\)These invariant transformations are formally defined in the next section.
process of the state variable. Both models are described in detail in chapter 3. This classification allows to place existing models that have appeared in the literature in a general context, enabling us to determine when a given model is over-specified, which means, for instance, that the original formulation of the model imposes too many restrictions on parameters. Such facts may not be apparent in the original specification. We dwell on that in section 3.2.1 when dealing with the Longstaff and Schwartz model.

2.3.2 Invariant Transformations and Equivalent Models

Invariant transformation are transformations of the state variables and parameter vectors in ways that leave the bond prices unchanged. More precisely, two models are equivalent in the sense that they generate identical bond prices for all interest rate instruments if they can be transformed into one another by a sequence of operations enumerated in the following list. We consider the model, specified by \( \mathcal{A} = (g, h, \kappa, \theta, S, \gamma_i, \delta_i, i \in \{1, \ldots, n\}, \lambda) \), to which equivalent models can be obtained by means of the subsequent invariant transformations.

1. **Permutation** \( T_P \): If \( \pi \) is a permutation of \( \{1, \ldots, n\} \) and if \( \mathcal{A}_P \) is the model obtained by permuting the elements of \( X(t) \) and the components of the parameters accordingly, then \( \mathcal{A}_P \) is equivalent to \( \mathcal{A} \).

2. **Brownian motion rotation** \( T_O \): \( T_O \) rotates a vector of independent Brownian motions into another vector of independent Brownian motions by using a \( n \times n \) orthogonal matrix \( O \), i.e., \( O' = O^{-1} \), that commutates with \( D \). Then \( T_O W_t = OW_t \) and

\[
\mathcal{A}_O = (g, h, \kappa, \theta, SO', \gamma_i, \delta_i, i \in \{1, \ldots, n\}, O\lambda)
\]

are the Brownian motions and the transformed equivalent model, respectively. Note that the state vector is not affected.

3. **Diffusion rescaling** \( T_R \): This transformation rescales the parameters of \( D \) and \( \lambda \) by the same constant. So, for any \( n \times n \) non-singular matrix \( R \),

\[
\mathcal{A}_R = (g, h, \kappa, \theta, SR^{-1}, R_{ii}^2\gamma_i, R_{ii}^2\delta_i, i \in \{1, \ldots, n\}, R\lambda)
\]

is the transformed model. The state vector and the Brownian motions are not modified. Such rescaling is possible because only the combinations \( SDDS' \) and \( SDD\lambda \) enter the zero-coupon bond price equation.
4. **Invariant affine transformation** $T_A$: It is defined by a non-singular $n \times n$ matrix $L$ and a $n$-dimensional vector $\vartheta$, such that the state vector and the equivalent model are given by

$$
T_A X(t) = L X(t) + \vartheta, \\
\Lambda_A = (g - h' L^{-1} \vartheta, L^{-1} h, \\
L \kappa L^{-1}, \vartheta + L \theta, L S, \{\tilde{\gamma}_i, \tilde{\delta}_i, i \in \{1, \ldots, n\}\}, \lambda),
$$

where $\tilde{\gamma}_i = L^{-1} \gamma_i$ and $\tilde{\delta}_i = \delta_i - \gamma'_i L^{-1} \vartheta$. The Brownian motions are not changed. These transformation are possible because of the affine structure of the models.
Chapter 3

Examples of Affine Models

Beyond doubt the best-known examples of affine models are the Vasicek and the Cox, Ingersoll Ross (CIR) one-factor models. Beside these we also investigate examples of multi-factor models, namely the Longstaff and Schwartz and the Balduzzi, Das, Foresi and Sundaram (BDFS) models. We also describe an economic model of affine form that is based on the general macro-economic IS-LM relationship. The reason why we have chosen to present these examples is on the one hand their popularity and their historic importance, and on the other hand the fact that all of them allow an economic interpretation. At the end of the chapter we will conclude that affine models range among the most appropriate models for practical purposes because of their tractability and flexibility as well as the possibility of an useful economic interpretation. We compare them with consol-rate-models, which were developed to take economic principles into account and which, however, entail some difficulties.

The dynamics are always specified under the martingale measure $\mathbb{P}^*$ if the opposite is not mentioned explicitly.

3.1 Examples of One-Factor Affine Models

A number of widely used one-factor models have the general (Hull-White) form

$$dr(t) = (\alpha(t) - \beta(t)r(t))dt + \sigma(t)r(t)^\gamma dW^*_t, \quad r(0) = r_0,$$

(3.1)

for some constant $\gamma \geq 0$ and where $\alpha, \beta, \sigma: \mathbb{R}_+ \to \mathbb{R}$ are locally bounded functions. $\alpha/\beta$ can be interpreted as the mean reversion level, whereas $\beta$ is
the mean reversion rate.\footnote{The model is often written in the form }{dr(t) = (\alpha(t) - \beta(t)r(t))dt + \sigma(t)dW_t^*}, \quad r(0) = r_0. \tag{3.2}

To explicitly solve this equation, let us denote \(l(t) = \int_0^t \beta(u)du\). By applying a slightly different version of theorem (2.4) for the time-dependent functions \(\alpha, \beta\) and \(\sigma\), we find the following explicit expression for \(r(t)\),

\[
r(t) = e^{-l(t)}\left(r_0 + \int_0^t e^{l(s)}\alpha(s)ds + \int_0^t e^{l(s)}\sigma(s)dW^*_s\right). \tag{3.3}
\]

Properties of the Classical Vasicek Model

If all parameters in (3.2) are constants we get the classical Vasicek model. More precisely, \(r\) is defined as the unique strong solution of the SDE

\[
dr(t) = (\alpha - \beta r(t))dt + \sigma dW_t^*, \quad r(0) = r_0, \tag{3.4}
\]

where \(\alpha, \beta\) and \(\sigma\) are strictly positive constants. This SDE is known as mean-reverting Ornstein-Uhlenbeck process. In this case \(r(t)\) is given by the following equation

\[
r(t) = \frac{\alpha}{\beta} + (r_0 - \frac{\alpha}{\beta})e^{-\beta t} + \sigma e^{-\beta t} \int_0^t e^{\beta s}dW^*_s. \tag{3.5}
\]

(3.5) implies that \(r(t)\) is normally distributed with mean and variance given respectively by

\[
\mathbb{E}_\mathbb{P}^*(r(t)) = \frac{\alpha}{\beta} + (r_0 - \frac{\alpha}{\beta})e^{-\beta t} \tag{3.6}
\]

\[
\text{var}_\mathbb{P}^*(r(t)) = \sigma^2 e^{-2\beta t} \int_0^t e^{2\beta s}ds = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t}) \tag{3.7}
\]

Remarks. Considering (3.6), it is obvious that \(r\) is mean reverting, since the expected rate tends, for \(t\) going to infinity to \(\frac{\alpha}{\beta}\), which can be regarded as a long term average rate.
The bond prices \( P(t, T) \) can now be derived by computing the expectation \( \mathbb{E}_P(\exp(-\int_t^T r(s)ds) | \mathcal{F}_t) \). Using the Markov property for Itô processes (Theorem 1.1) this expression is equivalent to

\[
\mathbb{E}_P\left( \exp\left( - \int_0^{T-t} r(s)ds \right) \right) \bigg|_{y=r(t)}.
\]

Hence, at first we calculate \( (\int_0^{T-t} r(s)ds) \bigg|_{y=r(t)} \):

\[
(\int_0^{T-t} r(s)ds) \bigg|_{y=r(t)} = \frac{\alpha}{\beta} (T-t) + (r(t) - \frac{\alpha}{\beta}) \frac{1 - e^{-\beta(T-t)}}{\beta} + \sigma \int_0^{T-t} \int_0^s e^{\beta(u-s)} dW_u^* ds
\]

\[
= \frac{\alpha}{\beta} (T-t) + (r(t) - \frac{\alpha}{\beta}) \frac{1 - e^{-\beta(T-t)}}{\beta} + \sigma \int_0^{T-t} \left( \int_u^{T-t} e^{\beta(u-s)} ds \right) dW_u^*
\]

\[
= \frac{\alpha}{\beta} (T-t) + (r(t) - \frac{\alpha}{\beta}) \frac{1 - e^{-\beta(T-t)}}{\beta} + \sigma \int_0^{T-t} 1 - e^{\beta(u-(T-t))} du.
\]

As \( (\int_0^{T-t} r(s)ds) \bigg|_{y=r(t)} \) is obviously a gaussian variable we can now easily calculate the expectation

\[
P(t, T) = \mathbb{E}_P\left( \exp\left( - \int_0^{T-t} r(s)ds \right) \right) \bigg|_{y=r(t)} = \exp \left[ \mathbb{E}_P\left( - \int_0^{T-t} r(s)ds \right) \bigg|_{y=r(t)} \right] + \frac{1}{2} \text{var}_P\left( - \int_0^{T-t} r(s)ds \bigg|_{y=r(t)} \right).
\]

with

\[
\mathbb{E}_P\left( - \int_0^{T-t} r(s)ds \bigg|_{y=r(t)} \right) = -\left( \frac{\alpha}{\beta} (T-t) + (r(t) - \frac{\alpha}{\beta}) \frac{1 - e^{-\beta(T-t)}}{\beta} \right),
\]

\[
\text{var}_P\left( - \int_0^{T-t} r(s)ds \bigg|_{y=r(t)} \right) = \frac{\sigma^2}{\beta^2} \int_0^{T-t} (1 - e^{\beta(u-(T-t))})^2 du. \quad (3.9)
\]

So, the bond price can be expressed in the affine form

\[
P(t, T) = e^{A(T-t) - B(T-t)r(t)},
\]
where
\[ B(T - t) = \frac{1 - e^{-\beta(T-t)}}{\beta}, \]
\[ A(T - t) = -\frac{\alpha}{\beta}(T - t) - \frac{1 - e^{-\beta(T-t)}}{\beta} + \frac{\sigma^2}{2\beta^2} \int_0^{T-t} (1 - e^{\beta(u-(T-t))})^2 du \]
\[ = (B(t,T) - (T - t)) \left( \frac{\alpha}{\beta} - \frac{\sigma^2}{2\beta^2} \right) - \frac{\sigma^2}{4\beta} B(t,T)^2, \quad (3.10) \]
which could also be found by solving (2.13) and (2.14).

### 3.1.2 The Extended CIR Model

Extended CIR models have \( \gamma = 1/2 \) in equation (3.1), so that the short rate process is
\[ dr(t) = (\alpha(t) - \beta(t)r(t))dt + \sigma(t)\sqrt{r(t)}dW_t^r. \quad (3.11) \]

When \( \alpha, \beta \) and \( \sigma \) are strictly positive constants we have the classical CIR model, which was developed by Cox, Ingersoll and Ross. Due to the square-root term in the diffusion coefficient, \( r \) can only take positive values which is a major advantage over the Vasicek model.

**Distribution of the Classical CIR Model**

For deducing the distribution of \( r \) we take the following approach which allows to obtain the dynamics of form (3.11) with constant coefficients by a transformation of a \( d \)-dimensional Gaussian process. Let \( (W_t^i)_{i=1,...,d} \) be a \( d \)-dimensional Brownian motion and for \( i = 1, \ldots, d \) we define \( X_i \) as solution of the SDE
\[ dX_i(t) = -\frac{1}{2} \beta X_i(t) dt + \frac{1}{2} \sigma dW_t^i. \quad (3.12) \]

Applying Theorem (2.4) we have
\[ X_i(t) = X_i(0)e^{-\frac{1}{2}\beta t} + \frac{1}{2} \sigma e^{-\frac{1}{2}\beta t} \int_0^t e^{\frac{1}{2}\beta s} dW_s^i. \]
We now define \( r(t) \) as \( \sum_{i=1}^{d} X_i^2(t) \). Hence the dynamics of \( r \) equal

\[
\begin{align*}
    dr(t) &= \sum_{i=1}^{d} 2X_i(t) dX_i(t) + \frac{1}{2} \sum_{i=1}^{d} 2\sigma_i^2 dt \\
    &= \left( -\beta \sum_{i=1}^{d} X_i^2(t) + d\frac{\sigma_i^2}{4} \right) dt + \sigma \sum_{i=1}^{d} X_i(t) dW_i^*(t) \\
    &= \left( d\frac{\sigma_i^2}{4} - \beta r(t) \right) dt + \sigma \sum_{i=1}^{d} X_i(t) dW_i^*(t).
\end{align*}
\]

Setting \( \frac{2W_t^*}{r(t)} = \sum_{i=1}^{d} \int_0^t X_i(s) \sqrt{r(s)} dW_i^*(s) \) is a standard Brownian motion. This follows by Lévy’s Characterization Theorem, since this expression is a martingale with respect to the standard Brownian filtration, whose quadratic variation equals \( t \).

**Proposition 3.1.** \( r(t)/\rho \) has a non-central \( \chi^2 \) distribution under \( \mathbb{P}^* \) with \( d = 4\alpha/\sigma^2 \) degrees of freedom and non-centrality parameter \( \lambda \), where

\[
\begin{align*}
    \rho &= \text{var}_{\mathbb{P}^*}(X_i(t)) = \frac{\sigma_i^2}{4\beta}(1 - e^{-\beta t}), \\
    \lambda &= \frac{4\beta r_0}{\sigma^2(e^{\beta t} - 1)}.
\end{align*}
\]

**Proof.** In order to see this, we write

\[
\begin{align*}
    \frac{r}{\rho} &= \sum_{i=1}^{d} \left( \frac{X_i}{\sqrt{\rho}} \right)^2 \\
    &= \sum_{i=1}^{d} (z_i + \delta_i)^2,
\end{align*}
\]

where \( z_1, z_2, \ldots, z_d \) are independent and identically distributed standard normal random variables and \( \delta_i = m_i/\sqrt{\rho} \), where

\[
    m_i = \mathbb{E}_{\mathbb{P}^*}(X_i(t)) = X_i(0)e^{-\frac{1}{2}\beta t}.
\]
It is well known that \( \sum_{i=1}^{d} z_i^2 \) has a (central) \( \chi^2 \) distribution with \( d \) degrees of freedom. Hence, \( r/\rho \) has a non-central chi-squared distribution with \( d \) degrees of freedom and non-centrality parameter
\[
\lambda = \sum_{i=1}^{d} \delta_i^2.
\]
If we assume that \( X_i(0) = \ldots = X_{d-1}(0) = 0 \) and \( X_d(0) = \sqrt{r_0} \), then
\[
\delta_i = 0, \quad i \in \{1, \ldots, d-1\} \quad \text{and} \quad \delta_d = \frac{\sqrt{r_0}e^{-\frac{1}{2} \beta t}}{\sqrt{\rho}}
\]
and therefore \( \lambda = r_0 e^{-\beta t}/\rho \).

Similar to the Vasicek’s model the bond price \( P(t, T) \) can be derived by solving PDEs (2.13) and (2.14). In this case the bond price is given by
\[
P(t, T) = e^{A(T-t) - B(T-t)r(t)},
\]
where
\[
A(T-t) = \frac{2\alpha}{\sigma^2} \ln \left( \frac{2\gamma e^{(\gamma+\beta)(T-t)/2}}{(\gamma+\beta)(e^{\gamma(T-t)} - 1) + 2\gamma} \right),
\]
\[
\gamma = \sqrt{\beta^2 + 2\sigma^2},
\]
\[
B(T-t) = \frac{2(e^{\gamma(T-t)} - 1)}{(\gamma+\beta)(e^{\gamma(T-t)} - 1) + 2\gamma}.
\]

### 3.2 Examples of Multi-Factor Affine Models

A major drawback of one-factor affine models is the implication that interest rates with different maturities are perfectly correlated. Bearing in mind that \( R(t, T) \) can be expressed by (2.2) it is obvious that the correlation between two yields \( R(t, T_1) \) and \( R(t, T_2) \),
\[
Corr(R(t, T_1), R(t, T_2)) = Corr\left( -\frac{A(t, T_1)}{T_1 - t} - \frac{B(t, T_1)}{T_1 - t} r(t), -\frac{A(t, T_2)}{T_2 - t} - \frac{B(t, T_2)}{T_2 - t} r(t) \right),
\]
equals 1. This means that a shock to the interest rate curve at time \( t \) is transmitted equally through all maturities, which can easily be avoided by including a second factor.
In the following we will give examples of multi-factor models whose factors have an economic interpretation. At first, we will describe the Longstaff and Schwartz model, where the second factor is identified with the volatility of the short rate. Longstaff and Schwartz is a well known two-factor model which has a great deal of flexibility, achieving good calibration to a variety of term structures. Then we deal with a model developed by Balduzzi, Das and Foresi (BDF), where the mean reversion level is chosen to be the second factor. An extension of this approach is the three factor BDFS (Balduzzi, Das, Foresi, Sundaram) model.

### 3.2.1 The Longstaff and Schwartz Two-Factor Model

Longstaff and Schwartz consider an interest rate model where the short rate \( r(t) \) is obtained as a linear combination of two basic processes as follows:

\[
\begin{align*}
    dY_1(t) &= \alpha_1(\mu_1 - Y_1(t))dt + \sqrt{Y_1(t)}dW_1^*(t), \\
    dY_2(t) &= \alpha_2(\mu_2 - Y_2(t))dt + \sqrt{Y_2(t)}dW_2^*(t), \\
    r(t) &= c_1Y_1(t) + c_2Y_2(t),
\end{align*}
\]

(3.17)

where all parameters have positive values and \( W_1^* \) and \( W_2^* \) are independent. Our first aim is to show that the model is essentially a two-factor CIR model, as described in section 2.2.2. We set

\[
X_1(t) = c_1Y_1(t), \quad X_2(t) = c_2Y_2(t),
\]

so that

\[
r(t) = X_1(t) + X_2(t).
\]

It is immediate to check that

\[
\begin{align*}
    dX_1(t) &= \alpha_1(c_1\mu_1 - X_1(t))dt + \sqrt{c_1}\sqrt{X_1(t)}dW_1^*(t), \\
    dX_2(t) &= \alpha_2(c_2\mu_2 - X_2(t))dt + \sqrt{c_2}\sqrt{X_2(t)}dW_2^*(t).
\end{align*}
\]

Since \( X_1 \) and \( X_2 \) describe one factor CIR processes, we see that the Longstaff and Schwartz model can be interpreted as two-factor CIR model.

Via a change of variable it can be expressed as a stochastic-volatility model, so that the state variables also have an economic interpretation. By defining a process \( V(t) = c_1^2Y_1(t) + c_2^2Y_2(t) \) we get the following dynamics for
CHAPTER 3. EXAMPLES OF AFFINE MODELS

33

\[ dr(t) = c_1 dY_1(t) + c_2 dY_2(t) \]

\[ = \alpha_1 c_1 (\mu_1 - Y_1(t))dt + c_1 \sqrt{Y_1(t)} dW_1^*(t) \]

\[ + \alpha_2 c_2 (\mu_2 - Y_2(t))dt + c_2 \sqrt{Y_2(t)} dW_2^*(t) \]

\[ = \left( \alpha_1 c_1 \mu_1 + \alpha_2 c_2 \mu_2 - \frac{(\alpha_1 c_2 - \alpha_2 c_1) r(t) + (\alpha_2 - \alpha_1) V(t)}{c_2 - c_1} \right) dt \]

\[ + \sqrt{\frac{c_1 (c_2 r(t) - V(t))}{c_2 - c_1}} dW_1^*(t) + \sqrt{\frac{c_2 (V(t) - c_1 r(t))}{c_2 - c_1}} dW_2^*(t). \]

The reason for formulating the model in this way is the fact that \( V(t) dt \) is the instantaneous variance of \( r(t) \). The use of \( r(t) \) and \( V(t) \) rather than \( Y_1(t) \) and \( Y_2(t) \) allows to express the dynamics of \( r(t) \) by means of level and volatility.

**Remark.** Referring to the Classification of Dai and Singleton, which we discussed in section 2.3, it is worth mentioning that the Longstaff and Schwartz model belongs to the class \( A_2(2) \), since two state variables appear in the volatility matrix of process (3.17). However it is over-specified, because there are only six parameters that could be chosen, whereas the corresponding canonical model has nine parameters.

### 3.2.2 The Central Tendency as Second Factor

In the model developed by Balduzzi, Das and Foresi [16] the second factor is identified with the central tendency of the short term rate. This approach is based on the fact that the behavior of short-term rates indicates fluctuations around a time-varying rest level - the central tendency - that changes stochastically over time.

In other words, by regressing the future short term rate \( r(t+1) \) on the current short-term rate and current longer-term yields, they found evidence for the fact that the future short-term rate does not only depend on the current one (as implied by one-factor models, where the short rate is implicitly assumed to follow an autoregressive process), but also on longer-maturity yields. Generally, the negligence of this fact is one of the reasons why one-factor models are insufficient.

Indeed, they could show that \( \beta_2 \), the coefficient of the long-term yields of the following regression

\[ r(t+1) = \beta_0 + \beta_1 r(t) + \beta_2 R(t, \tau) + error(t+1), \]
where $R(t, \tau)$ is a vector of yields with maturities $\tau = 1, 2, 3, 4$ years, is significant for both, the regression in levels and first differences. On account of these findings Balduzzi, Das and Foresi developed the following model.

The behavior of the short-term rate, the first factor, is described by the SDE
\[
dr(t) = \kappa(\theta(t) - r(t))dt + \sqrt{\sigma_0^2 + \sigma_1^2 r(t)}dW_r(t),
\]
where $\kappa, \sigma_0, \sigma_1$ are constants, $W_r$ is a standard Brownian motion under the actual measure $\mathbb{P}$ and $\theta$ is the central tendency, toward which the short term rate reverts and which evolves according to the SDE
\[
d\theta(t) = \left( m_0 + m_1 \theta(t) \right)dt + \sqrt{s_0^2 + s_1^2 \theta(t)}dW_\theta(t),
\]
where $m_0, m_1, s_0$ and $s_1$ are constants and $W_\theta$ is a standard Brownian motion under $\mathbb{P}$, independent from $W_r$. Under the assumption that the market price of risk is given by\(^3\)
\[
\Lambda(t) = D\lambda,
\]
where $\lambda \in \mathbb{R}^n$ and
\[
D = \begin{pmatrix}
\sqrt{\sigma_0^2 + \sigma_1^2 r(t)} & 0 \\
0 & \sqrt{s_0^2 + s_1^2 \theta(t)}
\end{pmatrix},
\]
so that\(^4\)
\[
D\Lambda(t) = \begin{pmatrix}
\lambda_{r0} + \lambda_{r1} r(t) \\
\lambda_{\theta0} + \lambda_{\theta1} \theta(t)
\end{pmatrix},
\]
we immediately get an affine model under $\mathbb{P}^*$, as our processes can be written in form (2.10),
\[
\begin{pmatrix}
\frac{dr(t)}{d\theta(t)}
\end{pmatrix} = \begin{pmatrix}
-\lambda_{r0} & -\kappa - \lambda_{r1} \\
-\lambda_{\theta0} & m_1 - \lambda_{\theta1}
\end{pmatrix} \begin{pmatrix}
r(t) \\
\theta(t)
\end{pmatrix} dt + \begin{pmatrix}
\sqrt{\sigma_0^2 + \sigma_1^2 r(t)} & 0 \\
0 & \sqrt{s_0^2 + s_1^2 \theta(t)}
\end{pmatrix} \begin{pmatrix}
dW_r(t)^* \\
dW_\theta(t)^*
\end{pmatrix}.
\]

In the following we show how a linear combination of two yields can be used to approximate the central tendency, which is obviously an unobservable factor. From (2.2) we know that yields can be expressed by
\[
R(t, \tau_i) = -\frac{A(t, \tau_i)}{\tau_i} - \frac{B_1(t, \tau_i) r(t)}{\tau_i} - \frac{B_2(t, \tau_i) \theta(t)}{\tau_i} \quad i = \{1, 2\}.
\]

\(^3\) $\Lambda(t)$ corresponds to $\lambda(t)'$ of equation (1.46).

\(^4\) i.e. $\sigma(X, t)\lambda(t)'$ in the notation of formula (1.46).
Solving for $r$ from the first yield, substituting into the second and rearranging leads to

$$
\tau_1 B_1(t, \tau_2) R(t, \tau_1) - \tau_2 B_1(t, \tau_1) R(t, \tau_2) = B_1(t, \tau_2) [A(t, \tau_1) + B_2(t, \tau_1) \mu(t)] - B_1(t, \tau_1) [A(t, \tau_2) + B_2(t, \tau_2) \mu(t)].
$$

Note that this quantity does not depend on $r$. Therefore $\mu$ can be written in the form

$$
\mu(t) = \frac{B_1(t, \tau_2) [\tau_1 R(t, \tau_1) - A(t, \tau_1)] - B_1(t, \tau_1) [\tau_2 R(t, \tau_2) - A(t, \tau_2)]}{B_1(t, \tau_2) B_2(t, \tau_1) - B_1(t, \tau_1) B_2(t, \tau_2)}.
$$

This equation justifies the following approximation $\hat{\mu}$ for $\mu$

$$
\mu(t) = a_0 + a_1 [B_1(t, \tau_2) \tau_1 R(t, \tau_1) - B_1(t, \tau_1) \tau_2 R(t, \tau_2)].
$$

Using this approximation the model can be estimated by applying the method of Maximum Likelihood.

The BDFS Three-Factor Model

An extension of the two-factor model described in the last section is the Balduzzi, Das, Foresi and Sundaram model, which belongs to the three-factor affine family (see equations (2.26) (2.27) and (2.28)). BDFS term structures are highly flexible, giving rise to hump- and spoon-shaped yield curves. However the lack of explicit formulas makes it awkward to calibrate the parameters to market data, which is a major disadvantage. The Fong and Vasicek model, where $\mu$ is constant, is a special case of the BDFS model.

3.3 Economic Models

Understanding the links between output, prices, interest rates and money supply has been a major objective of economists for decades. The standard macroeconomic model is the IS-LM\textsuperscript{5} model that presumes an equilibrium between output and income as well as an equilibrium between money supply and money demand. In contrast to this approach financial models of interest rates emphasize no-arbitrage at the expense of an economic context. Thus, this section is concerned with understanding how the economic concepts can be applied to financial models of interest.

The economic model that we regard is directly based on the IS-LM framework and was formulated by Tice and Webber [12]. As already mentioned the

\textsuperscript{5}LM stands for “Liquidity and Money”, while IS signifies “Investment and Savings”.
LM relationship comes from an assumption of equilibrium in the money market. In particular, it is supposed that the demand for real money increases in income and decreases in the short-term rate \( r(t) \). The expenditure, meaning the IS-Line also increases in income and decreases in the short rate, but besides that it depends on the public spending that is modeled as constant. In our case these relationships are specified by the following equations:

\[
m^d(t) = ky(t) - ur(t) \quad \text{(LM)},
\]

\[
e(t) = a - br(t) + cy(t) \quad \text{(IS)},
\]

where \( m^d \) denotes the (log-) demand for real money and \( e(t) \) the expenditure. Both, money demand and expenditure are affine functions of (log-) income \( y(t) \) and the short-term rate \( r(t) \). \( k, u, a, b, c \) are constants. It is supposed that the equilibrium dynamics are given by

\[
dm^d(t) = \alpha(m_s - m^d(t))dt + \sigma_m dW_m(t),
\]

\[
dy(t) = \beta(e(t) - y(t))dt + \sigma_y dW_y(t),
\]

so that money demand reverts to the level of money supply \( m_s \), which is assumed to be constant, and income to the level of expenditure. \( W_m(t) \) and \( W_y(t) \) are independent standard Brownian motions under \( \mathbb{P} \).

### 3.3.1 The General Framework

In a general framework we suppose to have \( n \) state variables \( X_1(t), \ldots, X_n(t) \) and \( n \) economic variables \( m_i(t) \) and \( \overline{m}_i(t) \), where

\[
m_i(t) = m_i(X_1(t), \ldots, X_n(t)) \quad i = 1, \ldots, n,
\]

\[
\overline{m}_i(t) = \overline{m}_i(X_1(t), \ldots, X_n(t)) \quad i = 1, \ldots, n.
\]

We suppose that the variable \( m_i(t) \) reverts to the level \( \overline{m}_i(t) \), so that \( \overline{m}_i(t) \) is interpreted as the equilibrium level of \( m_i(t) \). In particular, the dynamics of the equilibrium relationship are assumed to evolve according to the SDE

\[
dm_i(t) = \alpha_i(\overline{m}_i(t) - m_i(t))dt + S_i dW_t \quad i = 1, \ldots, n,
\]

where \( \alpha = (\alpha_1, \ldots, \alpha_n)' \) is a constant vector, \( S \) is a constant \( n \times n \) matrix (\( S_i \) denotes the \( i^{th} \) row) and \( W_t \) is a \( n \)-dimensional standard Brownian motion under \( \mathbb{P} \). If the functions \( M(t) = (m_1(t), \ldots, m_n(t))' \) and \( \overline{M}(t) = (\overline{m}_1(t), \ldots, \overline{m}_n(t))' \) are sufficiently regular we can invert (3.29) to

\footnote{Instead of \( r(t) \), the difference between the long-term interest rate and the long run future inflation is often used to model this relationship.}
find the process followed by the state variables $X(t) = (X_1(t), \ldots, X_n(t))'$. Under the assumption that the dynamics of $X(t)$ are given by (1.27), thus

$$dX(t) = \mu(X(t))dt + \sigma(X(t))dW_t,$$

we obtain by means of Itô’s formula

$$dM(t) = (M_x\mu + \frac{1}{2}h)dt + M_x\sigma dW_t,$$

where $M_x = \{\frac{\partial m_i}{\partial X_j}\}_{i,j=1,\ldots,n}$ and $h = \{h_i\}_{i=1,\ldots,n}$ is defined as

$$h_i = \sum_{k,l,j=1}^n \frac{\partial^2 m_i}{\partial X_k \partial X_l} \sigma_{kj} \sigma_{lj}.$$ 

From (3.29) we immediately get

$$\sigma(X(t)) = M_x^{-1}S, \quad \mu(X(t)) = M_x^{-1}\text{diag}(\alpha)(M(t) - M(t)) - \frac{1}{2}M_x^{-1}h. \quad (3.30) \quad (3.31)$$

### 3.3.2 IS - LM Framework

We can now apply the results of the previous section to our IS-LM relationship defined by (3.24)-(3.26). The economic variables are $m^d(t)$ and $e(t)$, whereas $r(t)$ and $y(t)$ are the underlying state variables, on which the values of $m^d(t)$ and $e(t)$ depend. Hence, we invert (3.26) to obtain

$$dr(t) = \alpha_r(\beta_r + \gamma_r y(t) - r(t))dt + \sigma_r dW_r(t), \quad (3.32)$$

$$dy(t) = \alpha_y(\beta_y - \gamma_y r(t) - y(t))dt + \sigma_y dW_y(t), \quad (3.33)$$

where

$$\alpha_r = \frac{\alpha u + \beta bk}{u},$$

$$\beta_r = \frac{\beta ak - \alpha m_s}{\alpha u + \beta bk},$$

$$\gamma_r = \frac{k(\alpha - \beta(1-c))}{\alpha u + \beta bk},$$

$$\alpha_y = \beta(1-c),$$

$$\beta_y = \frac{a}{1-c},$$

$$\gamma_y = \frac{b}{1-c}.$$
Furthermore $\sigma_r = \sqrt{\frac{k^2 \sigma_y^2 + \sigma_m^2}{\mu}}$, so that $W_r(t)$, defined by

$$dW_r(t) = -\frac{\sigma_m}{u \sigma_r} dW_m(t) + \frac{k \sigma_y}{u \sigma_r} dW_y(t),$$

is a (with $W_y$ correlated) Brownian motion. Obviously this is a two-factor affine model under $\mathbb{P}$ that can easily be transformed into a two-factor affine model under $\mathbb{P}^*$ by assuming an appropriate market price of risk (see equation (2.30)). We can re-parameterize the system in terms of $r(t)$ and $x(t) = \beta_r + \gamma_r y(t)$ to get the following equations,

$$dr(t) = \alpha_r (x(t) - r(t)) dt + \sigma_r dW_r(t), \quad (3.34)$$
$$dx(t) = \alpha_x (\beta_x r(t) - (1 - \beta_x) \mu - x(t)) dt + \sigma_x dW_x(t), \quad (3.35)$$

for appropriate constants $\alpha_x$, $\beta_x$, $\mu$, $\sigma_x$ and $W_x(t) = W_y(t)$.

**Remarks.** This model is a generalized Vasicek model, where $r(t)$ reverts to $x(t)$, and $x(t)$ to a weighted sum of $r(t)$ and $\mu$. In a way, it gives an economic justification for interest rate models with a second stochastic factor identified with the drift function. We have already encountered this idea in section (3.2.2), where we described the BDF model which also has a stochastic drift, but whose dynamics are specified by two generalized CIR processes.

$\mu$ can be interpreted as a variable that is controlled by the monetary authorities via $m^s$ and the public spending $a$. Hence $\mu$ could reasonably be time-dependent, which provides some economic justification for time-dependent variables that in general need to be treated with caution.

### 3.4 Non-Affine Models - Consol Models

We present here a non-affine model with the aim to put the advantages of affine models across. We will see in chapter 5 that affine yield factor models can be constructed which allow to relate yields of long maturities to the state variables. Another (earlier) approach to include a long term rate into an interest rate model in order to have a reasonable economic interpretation are consol models. They are two factor models whose state variables are the consol rate and the short rate. The consol rate can be defined as the yield on a bond that has a continuous coupon paid at a constant rate $c$ and infinite maturity, i.e. we have to consider an economy with an infinite horizon date, $T^* = \infty$. The price of the consol at time $t$ is then given by

$$C(t) = \mathbb{E}_{\mathbb{P}^*} \left( \int_t^\infty c \exp \left( - \int_t^s r(u) du \right) ds \bigg| \mathcal{F}_t \right) = \int_t^\infty c P(t, s) ds. \quad (3.36)$$
We suppose without loss of generality that $c = 1$ and substitute $r(u)$ by the $\mathcal{F}_t$-measurable random variable $l(t)$, which gives

$$C(t) = \int_t^\infty \exp\left(-\int_t^s l(t)du\right)ds = \int_t^\infty \exp(-l(t)(s-t))ds = \frac{1}{l(t)} \quad (3.37)$$

So the consol rate is simply defined as the reciprocal of its price $l(t) = C(t)^{-1}$ and can be seen as an approximation of a long-term rate of interest. Brennan and Schwartz used it to extend the short rate model to a two-factor model, in which the short rate $r$ and the long-term rate $l$ are intertwined. Since $C(t) = l(t)^{-1}$, we may work directly with the price of the consol. Then the two-dimensional process $(r(t), C(t))$ evolves according to the following pair of SDEs

$$dr(t) = \mu_r(r, C)dt + \sigma_r(r, C)dW^*_t(t), \quad (3.38)$$
$$dC(t) = \mu_C(r, C)dt + \sigma_C(r, C)dW^*_l(t) \quad (3.39)$$

under the martingale measure $\mathbb{P}^*$. The Application of Itô’s formula to (3.36) with $c = 1$ implies that

$$\mu_C = r(t)C(t) - 1.$$ 

For any choice of the short-term rate coefficients $\mu_r$ and $\sigma_r$, $\sigma_C$ can be chosen consistently, which was conjectured by Black and confirmed by Duffie, Ma and Yong [7]. They could also show that under some technical regularity conditions $C(t)$ is necessarily of the form $g(r(t))$ for a function $g$, that is the unique solution of the following ordinary differential equation,

$$g'(r)\mu_r(r, g(r)) - rg(r) + \frac{1}{2}\sigma_r^2(r, g(r)) + 1 = 0.$$ 

Thus the consol model can be reduced to a model with only one state variable $r$, whose dynamics are given by (3.38) and with $C(t) = g(r(t))$ or $l(t) = 1/g(r_t)$. This is rather surprising as the aim of this model was to provide two state variables for the term structure, the short rate $r(t)$ and the long rate $l(t)$.

**Remark.** Although the technical regularity conditions that are imposed to have this result may rule out some interesting cases, modeling $C(t)$ or $l(t)$ is always fraught with difficulty, as the diffusion term of the consol has to be chosen consistently with the solution of a non-trivial fix point problem involving the drift and the diffusion of the short rate.
3.5 Criteria for Model Selection

Generally speaking, affine models are those among the wide range of possible interest rate models which fulfill most of the criteria that are crucial for the model selection. Their theoretical as well as their practical properties are decisive for their frequent utilization.

We have now seen a number of examples of affine interest rate models (and one non-affine model) that all could be used in practice. So, the first question that arises is “Which model is the best one and should consequently be chosen.” Although there is no general answer to this question, it is certainly possible to state important features which good interest rate models should have and which determine the model selection. So when choosing a particular model the following criteria should be considered:

- **Tractability:** Interest rate models should have explicit or easy numerical solutions for bond prices and other instruments such as caplets or swaptions (see 4.1.2). As the solutions of the ricatti equations (2.13-2.14) can be quickly computed numerically in cases where explicit solutions are not available, the tractability of affine models is beyond doubt, which is one important reason for their popularity.

- **Fitting the yield curve:** The idea behind fitting a model to the current market yield curve is to fit it to given points and judge the goodness of fit regarding the resulting residuals. To a large extend this is purely a function of how many parameters the model contains. In other words, a model must have few enough parameters that a good fit is significant, and enough to ensure that a good fit is possible. One-factor affine models, for example the Vasicek model, do not have a large range of shapes and will provide a poor fit to some initial yield curves. Multi-factor affine models, however, offer large flexibility due to a larger number of parameters.

- **Good dynamics:** Beside fitting the current market prices, a model should also be able to fit, to some extend, the way that prices change over time. The dynamical features that may be desirable to match could be the dynamics of the short rate as well as the dynamics of the whole yield curve. By allowing time-dependent coefficients, which entails the problem of the right starting time, the dynamics of the market can sometimes be better reproduced. However, good results, for instance, of the short rate dynamics can also be attained by historical estimation of its mean reversion level, its mean reversion rate and its volatility. The historical estimation of these parameters that
are elements of many commonly used affine models implicates the estimation of the market price of risk in order to be able to move from $\mathbb{P}$ to $\mathbb{P}^*$.

- **Economic interpretation:** The chosen factors, i.e. state variables should have an economic interpretation. In most of the examples that exist in literature the first factor is identified with the short rate, whereas the second factor (in two-factor models) can be related to
  
  - the *volatility of the short-term rate* (Longstaff & Schwartz: section 3.2.1, Vasicek & Fong),
  - the *mean level of the short-term rate* (Balduzzi, Das & Foresi: section 3.2.2),
  - the *inflation* (Heston, Pearson & Sun),
  - the *spread between long and short-term rates* (Schaefer & Schwartz: consol rate model) and
  - the *long-term rate* (Brennan & Schwartz: consol rate model, Duffie & Kan [8]: affine yield-factor model, chapter 5)

Both, the Brennan and Schwartz and the Schaefer and Schwartz model belong to the class of consol models, whose use could raise a number of difficulties as we examined in section 3.4. Some of the other mentioned affine models include factors which have indeed an economic relevance, however which cannot be observed directly (e.g. the volatility of the short-term rate). So the affine yield factor model developed by Duffie and Kan, which we examine in detail in chapter 5, seems to be a promising approach, as the state variables are identified with the yields which are of course observable.

- **Two-factor model:** The choice of a two factor model seems to be a good compromise between flexibility and complexity. The calibration and solution of three (or more) factor models which provide high flexibility are already rather complicated, whereas single factor models are hardly adequate to describe the behavior of long rates, which play an important role, for example when pricing interest rate dependent products with long maturities.
Chapter 4

Calibration and Estimation

The process of fitting an interest rate model to market data is generally known as estimation. Depending on the kind of data, there are two notions that must be distinguished: calibration and historical estimation. Calibration means fitting the parameters of the model to current market data, whereas historical estimation of the parameters is based on statistical methods which are used to filter information out of historical time series. As the valuation of financial products should be market consistent, the calibration to current market prices, is in most cases preferable. Nevertheless it is still important to look at historical data. If the parameter estimates from historical and current data are systematically out of line, then one may be inclined to investigate the cause.

The next question that arise is, which aspects of the complete set of available market data should be fitted. For example, one could try to fit

- the current yield curve (which is usually the first objective),
- current bond option (cap, swaption) prices or
- the current volatility structure of bond options.

In practice it is usually only possible to fit a few aspects as the models are not adequate to match all market prices of all available interest rate products. In effect, interest rate models can be regarded as methods of interpolation and extrapolation. On the one hand, the parameters of the model have to be chosen so that at least one aspect of the market data is well reproduced by the model. On the other hand, using the found parameters other instruments may now be priced consistently with the market data.

In the following sections we describe different methods for both, calibration and historical estimation. In order to test and compare these approaches,
we apply them to the Vasicek model using real market data. We have chosen the Vasicek model because of its simplicity, with the aim to illustrate each method clearly, however being aware of the fact that it is not an appropriate model to fit the data well. So we do not focus on the absolute results, since there are many other models which are more adequate.

Before starting to explain various ways of calibration and estimation we comment on different types of data and their preparation.

4.1 Obtaining a Data Set

4.1.1 Market Data for the Current Yield Curve

Calibrating a model to the current yield curve means trying to reproduce the curve of the today’s continuously-compounded spot interest rate \( T \rightarrow R_M^{M}(0, T) \) (see (1.4)) for different maturities \( T \). However, these market yields \( R_M^{M}(0, T) \) are not available in this form and have to be calculated from EURIBOR- and swap rates. So we use the 1-12 month EURIBOR rates to calculate \( R_M^{M}(0, \frac{n}{12}), n = 1, \ldots, 12 \) and for the longer maturities 1, 2, \ldots, 30 years the corresponding swap rates. From equation (1.18) we know that the swap rates \( S(0, T_n), T_n = 1, 2, \ldots, 30 \) are given by

\[
S(0, T_n) = 1 - P(0, T_n) \sum_{i=1}^{n} \tau_i P(0, T_i).
\]  

(4.1)

We can rearrange (4.1) to express \( P(0, T_n) \) in terms of bond prices with shorter maturity times \( T_i, i < n \) and the swap rate \( S(0, T_n) \),

\[
P(0, T_n) = \frac{1 - S(0, T_n) \sum_{i=1}^{n-1} \tau_i P(0, T_i)}{1 + S(0, T_n) \tau_n}.
\]  

(4.2)

On the basis of this equation the zero bond prices can be calculated by the bootstrapping method. The yields are then easily deduced from the bond prices.

**Remarks.** When converting the EURIBOR rates to yields the day-count convention must be taken into consideration.

It is worth remarking that \( R_M^{M}(0, 1) \) can be derived from the EURIBOR rate as well as from the swap rate. However, \( L(0, 1) \) and \( S(0, 1) \) which should be equal (having already considered the day-count convention) do not coincide. This difference must therefore be incorporated into the EURIBOR rates.

\(^1\)We write \( R_M^{M} \) for the market yields.

\(^2\)They are denoted by \( L \) and \( S \) respectively.
4.1.2 Market Data for Bond Options

We introduce now the two main derivative products of the interest rate market, namely *caps* and *swaptions*, to which interest rate models can be calibrated.

**Interest Rate Caps**

A *cap* is a contract that can be viewed as a payer interest rate swap (compare equation (1.15)), where each exchange payment is executed only if it has positive value. The cap discounted payoff at time $t$ is therefore given by

$$\sum_{i=1}^{n} P(t, T_i) \tau_i (L(T_{i-1}, T_i) - K)^+, \quad (4.3)$$

where $T_0$ is the first reset date, $T_i$, $i = 1, \ldots, n$ are the payment times, $\tau_i = T_i - T_{i-1}$ and $K$ is a fixed rate, termed the *strike* of the cap contract. Each summand of the above sum defines a contract that is called *caplet*. Let $c_i$ be the value at time $t$ of the $i$th caplet. It is market practice to price a caplet $c_i$ with Black’s formula, i.e.

$$c_i = P(t, T_i) \tau_i \text{Bl}(K, F(t, T_{i-1}, T_i), \sigma_i \sqrt{T_{i-1}}), \quad (4.4)$$

where, denoting by $\phi$ the standard gaussian distribution function,

$$\text{Bl}(K, F, \sigma_i \sqrt{T_{i-1}}) = F \phi(d_1(K, F, \sigma_i \sqrt{T_{i-1}})) - K \phi(d_2(K, F, \sigma_i \sqrt{T_{i-1}})),$$

$$d_1(K, F, \sigma_i \sqrt{T_{i-1}}) = \ln(F/K) + \sigma_i^2 T_{i-1}/2 \sigma_i \sqrt{T_{i-1}},$$

$$d_2(K, F, \sigma_i \sqrt{T_{i-1}}) = \ln(F/K) - \sigma_i^2 T_{i-1}/2 \sigma_i \sqrt{T_{i-1}}.$$

$F$ is the simply-compounded forward rate and $\sigma_i$ the Black’s volatility of the $i$th caplet. As in the classical Black and Scholes option pricing setup, Black’s formula is based on the assumption that the underlying of the option, in our case the forward rate $F$, is log-normal and evolves under $\mathbb{P}^*$ according to

$$dF(t) = \sigma_i F(t) dW_t^{*}. \quad (4.5)$$

So $\sigma_i$ coincides with the actual forward rate volatility if the forward rate is the solution of the above SDE. The cap price $c$ is now $\sum_{i=1}^{n} c_i(\sigma_i)$ and the Black’cap volatility is $\sigma$ such that $c = \sum_{i=1}^{n} c_i(\sigma)$. In some sense $\sigma$ represents an average volatility of the set of individual caplets. Caps are quoted in the market in terms of Black’s volatility $\sigma$. 


Definition 4.1. ATM Cap. A cap with payment times \( T_i, i = 1, \ldots, n, \) 
\( \tau_i = T_i - T_{i-1} \) and strike \( K \) is said to be at-the-money (ATM) if and only if
\[
K = K_{ATM} = S(0, T_n) = \frac{1 - P(0, T_n)}{\sum_{i=1}^{n} \tau_i P(0, T_i)}.
\]
The cap is instead said to be in-the-money (ITM) if \( K < K_{ATM} \) and out-of-the-money (OTM) if \( K > K_{ATM} \).

The pricing of caps: We now show that a cap is actually equivalent to a portfolio of European zero coupon put options. This equivalence is used to derive explicit formulas for cap prices under analytically tractable short-rate models like the Vasicek model. The arbitrage free price at time \( t \) of the \( i^{th} \) caplet is according to the general pricing formula (1.22)
\[
c_i = \mathbb{E}_\mathbb{P}_s \left( \exp \left( - \int_t^{T_i} r(s) \, ds \right) \tau_i (L(T_{i-1}, T_i) - K) \, \right| \mathcal{F}_t \right)
\]
\[
= \mathbb{E}_\mathbb{P}_s \left[ \mathbb{E}_\mathbb{P}_s \left( \exp \left( - \int_t^{T_i} r(s) \, ds \right) \tau_i (L(T_{i-1}, T_i) - K) \, \right| \mathcal{F}_{T_i-1} \right] \, \mathcal{F}_t \right)
\]
\[
= \mathbb{E}_\mathbb{P}_s \left[ \exp \left( - \int_t^{T_i} r(s) \, ds \right) \, \mathcal{F}_{T_i-1} \right] \mathcal{F}_t \right)
\]
\[
= \mathbb{E}_\mathbb{P}_s \left[ \exp \left( - \int_t^{T_i} r(s) \, ds \right) P(T_i-1, T_i) \tau_i (L(T_{i-1}, T_i) - K) \, \right| \mathcal{F}_t \right]
\]
(4.7)

Using the definition of \( L(T_{i-1}, T_i) \), we obtain
\[
c_i(t) = \mathbb{E}_\mathbb{P}_s \left[ \exp \left( - \int_t^{T_i-1} r(s) \, ds \right) (1 - (1 + K \tau_i) P(T_i-1, T_i)) \, \right| \mathcal{F}_t \right]
\]
\[
= (1 + K \tau_i) \mathbb{E}_\mathbb{P}_s \left[ \exp \left( - \int_t^{T_i-1} r(s) \, ds \right) \frac{1}{1 + K \tau_i} - P(T_i-1, T_i) \, \right| \mathcal{F}_t \right],
\]
(4.8)

which is actually an European put option price with strike \( \frac{1}{1 + K \tau_i} \) and nominal value \( (1 + K \tau_i) \). Finally, cap prices are simply obtained by summing up the prices of the underlying caplets.

So, this formula is used to get explicit expressions for the model cap prices which we need, for instance, when fitting an interest rate model to market
ATM cap volatilities which is one possibility of calibration. However, this kind of calibration can be fraught with difficulty, since the price produced by a short-rate model is not compatible with Black’s market formula, which we investigate in detail in section 4.2.2.

Swaptions

Swaptions are options on a forward start interest rate swap (IRS). There are two main types, a payer version and a receiver version. A European payer swaption is an option giving the right (and no obligation) to enter a payer forward start IRS at a given future time, the swaption maturity. Usually the swaption maturity coincides with the first reset date \( T_0 \) of the underlying IRS. So the payoff at time \( T_0 \) of a payer swaption on a forward start swap with strike \( K \) is

\[
\left( \sum_{i=1}^{n} P(T_0, T_i) \tau_i (S(t, T_0, T_n) - K) \right)^+,
\]

where \( S(t, T_0, T_n) \) is the forward start swap rate (1.19). It is market practice to value swaptions with a Black-like formula. Precisely, the price of the above payer swaption (at time zero) is

\[
s = \text{Bl}(K, S(0, T_0, T_n), \sqrt{T_0}) \sum_{i=1}^{n} \tau_i P(0, T_i),
\]

where \( \sigma \) is now a volatility parameter that is quoted in the market that is different from the corresponding \( \sigma \) in the caps case. Again, Black’s formula would be “correct” if the forward start swap rates \( S(t, T_0, T_n) \) were log-normal with volatility \( \sigma \). Similar to a cap a swaption is said to be ATM if and only if its strike is equal to the forward start swap rate corresponding to the underlying swap.

Concerning the calibration of interest rate models to swaption volatilities one has to face the same problem as in the case of the calibration to cap volatilities, since the distribution of the forward start swap rate derived by the model is not compatible with Black’s formula. Moreover one also has to decide which maturity the underlying swap should have because the whole market volatility surface, as function of the swaption maturity and the maturity of the underlying swap, can never be fitted.
4.1.3 Which Market Rate should be used for the Short-Term Rate?

The short rate \( r(t) \) is the key interest rate in all models, even though it cannot be directly observed. As the short rate is defined to have an instantaneous holding period, one could assume that the overnight rate would be the best approximation. However, this assumption is rebutted by the high volatility and the low correlation with other yields. So the one- or three month spot rate, \( R^M(0, 1/12) \) or \( R^M(0, 1/4) \) respectively, is often taken to be the best approximation. One reason for this choice is their liquidity. For our calibration examples we always take the one month rate as surrogate for the short-term rate.

4.2 Calibration to Current Market Data

In this section we exemplify the calibration to current market data by means of the Vasicek and alternatively the Hull and White model. At first we calibrate the Vasicek model to the current term structure and then to cap volatilities.

4.2.1 Calibrating the Vasicek Model to the Current Term Structure

As already explained in section 4.1.1 we calibrate the Vasicek model to market yields \( R^M(0, T_n) \), where \( T_n \) are the different maturities

\[
1/12, 2/12, \ldots, 11/12, 1, 2, \ldots, 30.
\]

The Vasicek model yields are given by

\[
R(0, T_n) = -\frac{A(T_n)}{T_n} + \frac{B(T_n)}{T_n} r(0),
\]

(4.11)

where \( A \) and \( B \) are the functions of (3.10). We denote by \( R^M(0, T) \) the vector of market yields and by \( R(0, T) \) the vector of model yields. For the today’s short rate \( r(0) \) we use the one-month rate \( R^M(0, 1/12) \). We now have to choose the parameters \( \alpha, \beta, \sigma \) of the Vasicek model (3.4), so that \( R(0, T) \) best matches \( R^M(0, T) \). Our approach is to minimize

\[
Res_R(\alpha, \beta, \sigma) = (R^M(0, T) - R(0, T))^T (R^M(0, T) - R(0, T)),
\]

(4.12)
the sum of squared deviations. Consequently, our calibration problem is in fact a nonlinear least-squares problem. In order to solve this we use the Matlab function “lsqnonlin” choosing the large scale optimization algorithm that is based on a subspace trust region method. We describe this algorithm in detail in appendix A.1.

For the least-squares minimization it is necessary to specify the initial parameters $\alpha_0, \beta_0$ and $\sigma_0$, where the Matlab routine starts searching the minimum of (4.12). Since the final result is highly sensitive to these start values we generate various $\alpha_0, \beta_0$ and $\sigma_0$ randomly and calculate some optimization steps. The values with minimal residual are taken as start parameters for the final least-squares optimization which then supplies the final parameters. We use this kind of algorithm as well for all other calibration issues that we describe in the following sections. It is worth mentioning that we also define a set of lower and upper bounds between which the parameters must range (see appendix A.1.1).

So when fitting the Vasicek model to market yields from 31 March 2006 using the above described algorithm we found the following optimal parameters.

<table>
<thead>
<tr>
<th>Date</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
<th>Res$_R$</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 March 2006</td>
<td>0.02468</td>
<td>0.48995</td>
<td>0.06042</td>
<td>1.95828e-005</td>
</tr>
</tbody>
</table>

Table 4.1: Parameters for the Vasicek Model, Calibration to Market Yields, 31 March 2006

Figure 4.1 illustrates the deviations between market yields and the yields produced by the Vasicek model with the above parameters. Although the differences are rather small (the maximal deviation is 0.00126 for maturity 7 years) it is obvious that the Vasicek model does not fit the market data in an optimal way because it lacks a large range of possible shapes.

**Stability of the Parameters over Time**

As all parameters of the Vasicek model are supposed to be constants, they should not vary much over time. In order to see if the parameters really behave like this, we investigate their development between January 2005 and March 2006 by calibrating the model every month to the corresponding market data using the same method as described above. Figure 4.2 shows the

---

3 $Res$ stands for residuals.
4 For the large scale algorithm, the number of equations (the number of elements of equation (4.12)) must be at least as many as the parameters that we want to find.
5 If the model was written in form $dr(t) = \beta(\mu - r(t))dt + \sigma dW^*_t$, $\mu$ would be 0.05038.
parameter values for $\alpha$, $\beta$ and $\sigma$ since January 2005. While $\alpha, \sigma$ are relatively constant, $\beta$, the mean reversion rate, varies extremely.

### 4.2.2 Calibrating the Vasicek Model to Cap Volatilities

For illustration purposes we now try to calibrate the Vasicek model to market ATM cap volatilities. As we will see the Vasicek model is not appropriate for this kind of calibration since the current term structure can not be fitted simultaneously when calibrating to cap volatilities. In the next section we will get to know the Hull and White model, which allows a calibration to both kinds of data, however with the disadvantage of having a time-dependent parameters.

Fitting to cap volatilities means computing the market cap price by inserting the market cap volatility into Black’s formula at first and then adapting the parameters of the model, so that the model cap prices best match the market cap prices obtained by Black’s formula. Table 4.2 contains the market quoted ATM cap volatilities for different maturity times $T_n$ to which we try to fit the model. The underlying caplets of each cap have maturities $T_{ni}$, which are up to one year equally three-month spaced and after one year
Using the same algorithm as described in section 4.2.1, we minimize now the sum of the square differences between market and model cap prices,

$$\min Re_{cap} = \min \left( \sum_{T_n} (Cap^M(T_n) - Cap(T_n))^2 \right),$$  \hspace{1cm} (4.13)

where $Cap^M$ and $Cap$ denote the market and the model cap price respectively. The model cap price is obtained by means of formula (4.8). With this approach we find the parameters of table 4.3, which do not really match the parameters that we obtained when calibrating to the current yields.

Figure 4.3 opposes the market cap prices against the model cap prices, as well as the market cap volatilities against the model cap implied volatilities. The model cap implied volatility is the volatility parameter that must be plugged into Black’s formula (4.4) in order to obtain the model price.

Although the model implied volatility does not differ much from the market volatility in this case, there are some problems with this calibration method. Firstly, the parameters which are supposed to be constant vary again a lot over time, which we have already experienced when calibrating the model to market yields. Secondly, as already hinted at the beginning,
CHAPTER 4. CALIBRATION AND ESTIMATION

<table>
<thead>
<tr>
<th>Maturity $T_n$ (years)</th>
<th>ATM volatilities $\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>2</td>
<td>14.09%</td>
</tr>
<tr>
<td>3</td>
<td>15.94%</td>
</tr>
<tr>
<td>4</td>
<td>16.50%</td>
</tr>
<tr>
<td>5</td>
<td>16.74%</td>
</tr>
<tr>
<td>6</td>
<td>16.73%</td>
</tr>
<tr>
<td>7</td>
<td>16.64%</td>
</tr>
<tr>
<td>8</td>
<td>16.50%</td>
</tr>
<tr>
<td>9</td>
<td>16.33%</td>
</tr>
<tr>
<td>10</td>
<td>16.17%</td>
</tr>
<tr>
<td>12</td>
<td>15.87%</td>
</tr>
<tr>
<td>15</td>
<td>15.31%</td>
</tr>
<tr>
<td>20</td>
<td>14.56%</td>
</tr>
<tr>
<td>25</td>
<td>14.06%</td>
</tr>
<tr>
<td>30</td>
<td>13.67%</td>
</tr>
</tbody>
</table>

Table 4.2: ATM Cap Volatilities, EUR, 31 March 2006

<table>
<thead>
<tr>
<th>Date</th>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>31 March 2006</td>
<td>0.0063</td>
<td>0.1274</td>
<td>0.0115</td>
</tr>
</tbody>
</table>

Table 4.3: Parameters for the Vasicek Model - Calibration to Cap Volatilities, 31 March 2006

Figure 4.3: Market- versus Vasicek Model Cap Prices and Volatilities, 31 March 2006
the yields produced by the Vasicek model with $\alpha, \beta$ and $\sigma$ from Table 4.3 are not in accordance with the market yields.

So the Vasicek model is not an appropriate model for fitting cap volatilities, something that aims at all affine models in alleviated form. Generally speaking, the basic problem with this approach is the fact that the cap price calculated by an affine model is not compatible with Black's market formula, since the model forward rate dynamics are never of form (4.5). More specifically, for no choice of the model parameters does the distribution of the forward rate produced by an affine model coincide with the distribution of the "Black"-like forward rate following (4.5). For this reason the so-called "log-normal forward LIBOR models" were developed, where the dynamics of $F$ are specified by (4.5).

Nevertheless we investigate once again the calibration to cap volatilities by means of another short rate model, namely the Hull and White model, since it enables an exact fit of the term structure of interest rates while being calibrated to cap volatilities at the same time. Nevertheless the above stated objections are still true.

### 4.2.3 Calibrating the Hull-White Extended Vasicek Model

In this section we consider the following extension of the Vasicek model which was analyzed by Hull and White

$$dr(t) = (\alpha(t) - \beta r(t))dt + \sigma dW^*,$$

(4.14)

where $\alpha(t)$ is time-dependent and chosen so as to exactly fit the current term structure of interest rates. It can be shown that $\alpha$ must be

$$\alpha(t) = \frac{\partial f^M(0,t)}{\partial t} + \beta f^M(0,t) + \frac{\sigma^2}{2\beta}(1 - e^{-2\beta t}),$$

(4.15)

where $f^M$ denotes the market instantaneous forward rate. The bond price and $B(t,T)$ have the same form as in the classical Vasicek model (3.10), however $A(t,T)$ is now given by

$$A(t,T) = \ln \left(\frac{P^M(0,T)}{P^M(0,t)}\right) + B(t,T) f^M(0,t) - \frac{\sigma^2}{4\beta}(1 - e^{-2\beta t})B(t,T)^2,$$

(4.16)

where $P^M$ is the market bond price. It is therefore obvious that current (i.e. $t = 0$) market yields can always be exactly reproduced by the model.

We now try to find the other parameters $\beta$ and $\sigma$ by calibrating the model to the market cap volatilities taking the same approach as in the previous
Table 4.4: Parameters for the Hull-White Model - Calibration to Cap Volatilities

<table>
<thead>
<tr>
<th>Param./ Date</th>
<th>10/05</th>
<th>11/05</th>
<th>12/05</th>
<th>01/06</th>
<th>02/06</th>
<th>03/06</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\beta$</td>
<td>0.0221</td>
<td>0.0219</td>
<td>0.0214</td>
<td>0.0236</td>
<td>0.0201</td>
<td>0.0255</td>
</tr>
<tr>
<td>$\sigma$</td>
<td>0.0072</td>
<td>0.0072</td>
<td>0.0071</td>
<td>0.0069</td>
<td>0.0066</td>
<td>0.0068</td>
</tr>
</tbody>
</table>

Figure 4.4: Hull-White Model Implied Cap Volatilities, 31 March 2006 and Stability of the Parameters over 6 month (October 2005 - March 2006)

section. The cap volatility curve implied by the Hull and White model is shown in figure 4.4. Notice that it generates a decreasing cap volatility curve instead of reproducing the humped market curve. Furthermore the calibrated value of the mean-reversion parameter $\beta$ shown in table 4.4 is extremely small, which is a common situation and can be traced back to the dominance of the increasing part of the market volatility curve. In spite of these disadvantages the parameters remain relatively stable as illustrated in figure 4.4.

It is worth mentioning that the Hull and White extension of the Vasicek model is - despite some disadvantages that we have already stated and the fact that it can produce negative interest rates - one of the historically most important interest rate models. Nevertheless, the time-dependency of $\alpha$ is a disadvantage, since it creates the necessity to justify the right starting time.

4.3 Historical Estimation

Although it usually preferable to estimate parameter values from prices by implied calibration methods as described in the last section, it is still nec-
essary to compare these results to historical parameter estimates. Here we exemplify the two main methods, namely the *General Method of Moments* (GMM) and the *Maximum Likelihood Method* (ML) again with the Vasicek model.

### 4.3.1 Maximum Likelihood Method

The idea behind the Maximum Likelihood Method is to find parameter values for which the actual outcome has the maximum probability. Suppose we have observed a time series \( r(t_i), i = 1, \ldots, n \) whose transition density, i.e. the likelihood that \( r(t_i) \) will move from state \((t_i, r(t_i))\) at time \( t_i \) to state \((t_{i+1}, r(t_{i+1}))\) at time \( t_{i+1} \),

\[
p(t_{i+1}, r(t_i); t_i, r(t_i)) | \theta)
\]

is known. The density depends on the parameter set \( \theta \). In the general case the transition density at time \( t \) is conditional on \( F_t \), however when the process is Markov it is only conditional on values at time \( t_i \). As our interest rate models are based on Markov processes we consider only this case. The joint density of our observations is

\[
p(r(t_1), \ldots, r(t_n) | \theta) = p_0(r(t_1) | \theta) \prod_{i=1}^{n} p(t_{i+1}, r(t_i); t_i, r(t_i)) | \theta),
\]

where \( p_0 \) is some former density for \( r(t_1) \). The likelihood function is then given by

\[
L(\theta) = \prod_{i=1}^{n} p(t_{i+1}, r(t_i); t_i, r(t_i)) | \theta).
\]

Finally, \( \hat{\theta} = \arg \max_{\theta} L(\theta) \) is the maximum likelihood estimate of \( \theta \). Maximizing \( L \) places the observed time series at the maximum of the joint density function. Since \( \ln L \) is monotonically increasing, it is often more convenient to maximize \( \ln L \) instead of \( L \).

#### Maximum Likelihood Estimation for the Vasicek Model

For the Vasicek model with \( \theta = (\alpha, \beta, \sigma) \), the transition density function for the short rate \( r(t) \) process is because of the gaussian distribution

\[
p(t_{i+1}, r(t_{i+1}); t_i, r(t_i)) | \theta) = \frac{1}{\sqrt{2\pi \text{var} t_i}} \exp \left( -\frac{1}{2} v^2(r(t_i), r(t_{i+1}), \Delta t_i) \right),
\]

(4.20)
where $\Delta t_i = t_{i+1} - t_i$,

$$
\text{var}_{t_i} = \frac{\sigma^2}{2\beta} (1 - e^{-2\beta \Delta t_i}),
$$

(4.21)

$$
v(r(t_i), r(t_{i+1}), \Delta t_i) = \frac{r(t_{i+1}) - (\frac{\alpha}{\beta} + (r(t_i) - \frac{\alpha}{\beta}) e^{-\beta \Delta t_i})}{\sqrt{\text{var}_{t_i}}}. \tag{4.22}
$$

Hence the log-likelihood which we will maximize is

$$
\ln L = -\frac{n-1}{2} \ln 2\pi - \frac{1}{2} \sum_{i=1}^{n} \left[ \ln \left( \frac{\sigma^2}{2\beta} (1 - e^{-2\beta \Delta t_i}) \right) + v(r(t_i), r(t_{i+1}), \Delta t_i) \right].
$$

(4.23)

Using now daily data\(^6\) of the one-month rate $R(0, \frac{1}{12})$ between March 1991 and March 2006 as surrogate for the short-term rate we find the following parameters, which are not consistent with the values that we obtained when calibrating to current market data (compare with Table 4.1).

<table>
<thead>
<tr>
<th>$\alpha$</th>
<th>$\beta$</th>
<th>$\sigma$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.002472</td>
<td>0.1546</td>
<td>0.0056</td>
</tr>
</tbody>
</table>

Table 4.5: Maximum Likelihood Parameters for the Vasicek Model

Alternatively not regarding $R(0, \frac{1}{12})$ as substitute for the short-term rate, the “real” transition density function is\(^7\)

$$
p(t_{i+1}, R_{t_{i+1}}; t_i, R_t)|\theta) = \frac{1}{\sqrt{2\pi \text{var}_{t_i}}} \exp \left( -\frac{1}{2} \tilde{v}^2 (R_t, R_{t_{i+1}}, \Delta t_i) \right), \tag{4.24}
$$

where

$$
\tilde{v}(R_t, R_{t_{i+1}}, \Delta t_i) = \frac{R_{t_{i+1}} - R_t e^{-\beta \Delta t_i}}{\sqrt{\text{var}_{t_i}}} \left( -\frac{A(1/12)}{1/12} + \frac{B(1/12) \alpha}{1/12 \beta} \right) (1 - e^{-\beta \Delta t_i}).
$$

(4.26)

and where $A$ and $B$ are the functions from (3.10). Although this approach is more precise it leads to the same results.

\(^6\)i.e. $\Delta t_i = \frac{1}{256}$, considering public holidays

\(^7\)We now write $R$ instead of $r$ in order to emphasize that we regard the one-month spot rate and not the short-term rate, although using the same data.
One reason why the parameter values are not in accordance with the values from our calibration to current yields, is the fact that we only regard \( R(0, \frac{1}{12}) \) for the historical estimation, whereas for the calibration to current market data all yields with different maturities are taken into account. Furthermore, we have already seen that the parameters obtained by implied calibration methods vary a lot over time, so it would be rather surprising if the historical estimates exactly match the current calibration.

### 4.3.2 General Method of Moments

Basically, the idea of this method is to compare certain functions of a sample, called moments, with their theoretical values. The values of the parameters are then chosen so that the values of the theoretical moments are close to their sample values.

We suppose to have a process \( X(t) \) that depends on a parameter vector \( \theta \). Given a function \( f \), one may in principle compute the theoretical expectation \( g(\theta) \) of \( f(X) \),

\[
g(\theta) = \mathbb{E}(f(X)|\theta). \tag{4.27}
\]

Using a set of observations \( \hat{X}(t_i), i = 1, \ldots, n \), the sample moment \( \hat{f} \) is calculated by

\[
\hat{f} = \frac{1}{n} \sum_{i=1}^{n} f(\hat{X}(t_i)). \tag{4.28}
\]

The aim is to choose \( \theta \) so that \( \hat{f} = g(\theta) \). In fact, when \( f \) is vector valued, \( \theta \) is given by

\[
\hat{\theta} = \arg \min_{\theta} J(\theta) = (\hat{f} - g)'W(\theta)(\hat{f} - g), \tag{4.29}
\]

where \( W \) is a weighting matrix that can be chosen in an optimal way, which was investigated by Hansen [9].

Of course the estimated value \( \hat{\theta} \) of \( \theta \) depends crucially on the function \( f \). Thus, this fact immediately arises two questions. Firstly, how should \( f \) be chosen and secondly, how can \( g(\theta) \) be calculated? In general, computing the theoretical moments \( g(\theta) \) might be a serious practical problem. So, the solution is to generate \( f \), so that by construction \( g(\theta) \) has known values, for instance \( g(\theta) = 0 \). Hansen and Scheinkman [10] give construction possibilities in order to generate moments with this property.

---

\(^8\)This set of observations could be the short-term rate but also other yields, for instance.
Finding Moments for the Vasicek Model

In order to obtain some moment conditions in the case of the Vasicek Model, one could use the Euler discretization of (3.4), i.e.

\[ r(t+1) = \alpha \Delta t + (1 - \beta \Delta t) r(t) + \sigma \sqrt{\Delta t} z(t+1), \]  

(4.30)

where \( z \) is a standard gaussian variable. We set \( \varepsilon(t+1) = \sigma \sqrt{\Delta t} z(t+1) \). If the model is correctly specified, then \( \varepsilon(t+1) \sim N(0, \sigma^2 \Delta t) \) is iid normally distributed and serially uncorrelated, which gives us the following moment conditions:

\[ \cdot \varepsilon(t+1) \sim N(0, \sigma^2 \Delta t): \]

\[ \mathbb{E}[\varepsilon(t+1)] = 0, \]  
\[ \mathbb{E}[\varepsilon(t+1)^2 - \sigma^2 \Delta t] = 0, \]  
\[ \mathbb{E}[\varepsilon(t+1)^3] = 0, \]  
\[ \cdot \varepsilon(t+1) \) is serially uncorrelated: \]

\[ \mathbb{E}[\varepsilon(t+1) \varepsilon(t)] = 0, \]  
\[ \mathbb{E}[\varepsilon(t+1) \varepsilon(t-1)] = 0, \]  

Pros and Cons of GMM

A big advantage of GMM is that its use generally does not require a knowledge of the distribution of \( \varepsilon(t) \), just its moments \( f \). The ignorance of the transition density function in cases where it is available is disadvantageous, since GMM only uses information about the moments \( f \) and does not make use of other possible information. Beside that, there is also the already mentioned problem of the choice of which moments to use.

\[ ^9 \text{For instance, in the case of equation (4.31), } f(r) = \varepsilon(t+1) = r(t+1) - (\alpha \Delta t + (1 - \beta \Delta t) r(t)). \]
Chapter 5
Affine Yield-Factor Models

A yield factor model as analyzed by Duffie and Kan [8] is characterized by the fact that the $i^{th}$ component of the state variable $X(t)$ in an affine model (see section 2.1) can be viewed as the yield at time $t$ on a zero-coupon bond of maturity $\tau_i$, for fixed maturities $\tau_1, \ldots, \tau_n$. This is very convenient feature since the abstract state variables now get a reasonable economic interpretation.

5.1 General Affine Yield-Factor Model

In order to have an affine factor model with $P(t, t + \tau) = \exp(A(\tau) + B(\tau)'X(t))$, where $X_i$ is the yield of maturity $\tau_i$,

$$X_i(t) = -\frac{\ln P(t, t + \tau_i)}{\tau_i}, \quad i \in \{1, \ldots, n\}, \quad (5.1)$$

not only the initial conditions $A(0) = 0$ and $B(0) = 0$ (2.15) and the parameter restrictions from theorem 2.3 must hold, but also the following constraints resulting from (5.1)

$$A(\tau_i) = B_j(\tau_i) = 0, \quad j \neq i, \quad B_i(\tau_i) = -\tau_i \quad (5.2)$$

must be fulfilled for all $i$.

There are two possible ways to construct an affine yield-factor model. One is to suppose from the beginning that the state variables are yields and to ensure that the coefficients for the process (2.10) are chosen so that (5.2) is satisfied. The other, indirect approach is to allow $X$ to be any general state process for an arbitrary affine model and to attempt a change of variables from the original state vector $X(t)$ to a new yield state vector $R(t) = (R(t, t + \tau_1), \ldots, R(t, t + \tau_n))'$.  

58
Proposition 5.1. If the model is affine in a general process $X(t)$, meaning that it is of form (2.10) and that $P(t, t + \tau)$ can be expressed by $\exp(A(\tau) + B(\tau)'X(t))$, then it can be reformulated in a way which is affine in $R(t)$, i.e. $P(t, t + \tau) = \exp(\tilde{A}(\tau) + \tilde{B}(\tau)'R(t))$.

Proof. From (2.2) we know that $R(t, t + \tau_i)$ is given by

$$R(t, t + \tau_i) = \frac{-A(\tau_i)}{\tau_i} X(t), \text{ for } i = 1, \ldots, n.$$  

In vector notation this is

$$R(t) = (R(t, t + \tau_1), \ldots, R(t, t + \tau_n))' = k + KX(t)$$

for a constant vector $k$ and a constant matrix $K$ (given $\tau_1, \ldots, \tau_n$). Assuming that $K$ is invertible this implies that

$$X(t) = K^{-1}((R(t) - k)).$$

Hence

$$P(t, t + \tau) = \exp(A(\tau) + B(\tau)'K^{-1}(R(t) - k)) = \exp(\tilde{A}(\tau) + \tilde{B}(\tau)'R(t)),$$

where $\tilde{A}(\tau) = A(\tau) - B(\tau)'K^{-1}k$ and $\tilde{B}(\tau)' = B(\tau)'K^{-1}$.

So, provided that the matrix $K$ is non-singular, the change of variables from a general state vector $X(t)$ given by (2.10) to $R(t)$ is possible. In this case, we can write

$$dR(t) = (\tilde{\alpha} + \tilde{\beta}R(t))dt + \tilde{S}D(R(t))dW_t,$$  

(5.3)

where

$$\tilde{\alpha} = K\alpha - K\beta K^{-1}k,$$  

(5.4)

$$\tilde{\beta} = K\beta K^{-1},$$  

(5.5)

$$\tilde{S} = KS,$$  

(5.6)

$$\tilde{D}(R(t)) = \begin{pmatrix} \sqrt{\tilde{\gamma}_1'R(t) + \tilde{\delta}_1} & 0 \\ \vdots & \ddots \\ 0 & \sqrt{\tilde{\gamma}_n'R(t) + \tilde{\delta}_n} \end{pmatrix},$$  

(5.7)

$$\tilde{\gamma}_i' = \gamma_i'K^{-1},$$  

(5.8)

$$\tilde{\delta}_i = \delta_i - \gamma_i'K^{-1}k.$$  

(5.9)
Remarks. The volatility function \((\tilde{S}\tilde{D}(\cdot))\) is rather difficult to calibrate to observed volatilities from current, since the elements of the matrix \(K\) are of the form \((-B_j(\tau_i)/\tau_i)\), which depend via the solution of the Ricatti equations (2.13) and (2.14) on the original parameters \(\alpha, \beta, \gamma, \delta\) and \(S\).

So, from a practical point of view it is advisable to start with an affine factor model for which the state vector \(X(t)\) is already treated as a vector of yields for fixed maturities \(\tau_1, \ldots, \tau_n\). Then the parameters \(S\) and \(D\) could be chosen directly from calibration under consideration that the boundary conditions \(A(0) = 0\) and \(B(0) = 0\) as well as (5.2) must be fulfilled. Furthermore, theorem 2.3, which guarantees a solution to the differential equation (2.10), must also be respected.

Although there are no theoretical results describing how certain coefficients can be fixed in advance, while others can be adjusted afterward in order to achieve consistency with all the above mentioned conditions, in practice it is always possible to fix \(S\) and \(D\) at first and then adjust the drift, i.e. the parameters \(\alpha\) and \(\beta\). In the next sections this method is explained in detail for a two-factor version of the model.

5.2 A Two-Factor Affine Model of the “Long”- and the Short-Term Rate

We will now consider a two-factor affine yield model, in which one of the factors is the short-term rate itself. For the second factor we take the yield with maturity \(\tau_2 = 10\) years\(^1\), which could be interpreted as a long-term rate. Furthermore we simplify the matrix \(D(X(t))\), so that all diagonal elements are of the same form \(\gamma'X(t) + \delta\). In this special case (2.13) and (2.14) can be written

\[
\frac{\partial A(\tau)}{\partial \tau} = B(\tau)'\alpha + \frac{\delta}{2} q(\tau), \quad (5.10)
\]

\[
\frac{\partial B_1(\tau)'}{\partial \tau} = B(\tau)'\beta_1 + \frac{\gamma_1}{2} q(\tau) - 1, \quad (5.11)
\]

\[
\frac{\partial B_2(\tau)'}{\partial \tau} = B(\tau)'\beta_2 + \frac{\gamma_2}{2} q(\tau), \quad (5.12)
\]

where \(\beta_i\) is the \(i^{th}\) column of the matrix \(\beta\), \(\gamma_i\) the \(i^{th}\) component of the vector \(\gamma\) and

\[
q(\tau) = B(\tau)'SS'B(\tau) = \sum_{i=1}^{2} \sum_{j=1}^{2} B_i(\tau)B_j(\tau)s_is'_j
\]

\(^1\)One could also choose any other maturity.
5.2.1 Deterministic Volatility

In the case of deterministic volatility, meaning that the volatility is independent of \( X(t) \), thus defined by \( \gamma = 0 \), the equations (5.11) and (5.12) form a linear system and have the standard solution

\[
B_i(\tau) = \sum_{j=1}^{2} \psi_{ij} \exp(\lambda_j \tau) + \phi_i, \quad i \in \{1, 2\},
\]

(5.13)

where \( \psi_{ij} \) and \( \phi_i \) are constants and \( \lambda_j \) are the roots of the characteristic equation,

\[
\det \begin{pmatrix} \beta_{11} - \lambda & \beta_{12} \\ \beta_{21} & \beta_{22} - \lambda \end{pmatrix} = 0.
\]

(5.14)

Thus, we have

\[
\lambda_{1,2} = \frac{1}{2}(\beta_{11} + \beta_{22} \pm \sqrt{\Delta}),
\]

(5.15)

where \( \Delta = \beta_{11}^2 + \beta_{22}^2 + 4\beta_{12}\beta_{21} - 2\beta_{11}\beta_{22} \). Considering the initial condition \( B(0) = 0 \), the solution for \( B \) is then

\[
B_1(\tau) = \frac{1}{4\sqrt{\Delta}(\beta_{12}\beta_{21} - \beta_{11}\beta_{22})}[(\beta_{11} - \beta_{22} + \sqrt{\Delta})(\beta_{11} + \beta_{22} - \sqrt{\Delta})e^{\lambda_1 \tau} - (\beta_{11} - \beta_{22} - \sqrt{\Delta})(\beta_{11} + \beta_{22} + \sqrt{\Delta})e^{\lambda_2 \tau} - 4\beta_{22}\sqrt{\Delta},
\]

\[
B_2(\tau) = \frac{\beta_{12}}{2\sqrt{\Delta}(\beta_{12}\beta_{21} - \beta_{11}\beta_{22})}[(\beta_{11} + \beta_{22} - \sqrt{\Delta})e^{\lambda_1 \tau} - (\beta_{11} + \beta_{22} + \sqrt{\Delta})e^{\lambda_2 \tau} + 2\sqrt{\Delta}.
\]

(5.16)

The constraints (5.2) for \( B \), in our case for \( \tau_2 = 10 \) can then be written explicitly by setting

\[
B_1(10) = 0 \quad \text{and} \quad B_2(10) = -10.
\]

(5.17)

As the first factor is the short-term rate itself, (5.2) only implies these two conditions, since \( B(0) = 0 \) must hold anyway. By putting \( B \) into (5.10), \( A \) can be obtained by integration.

5.2.2 Calibrating the Deterministic Volatility Model to the Current Term Structure

As already indicated the calibration of a general affine yield model causes some difficulty since all the mentioned constraints must be obeyed. In this
special case things simplify significantly. Firstly, we have explicit solutions for the Ricatti equations (5.10)- (5.12) and secondly, the constrains of theorem 2.3 are automatically fulfilled because $\gamma = 0$ and $\delta_1 = \delta_2$. Hence we only have to consider (5.17).

We choose the same approach as for the calibration of the Vasicek model to the current term structure (see section 4.2.1), i.e. minimizing the sum of square deviations between model and current market yields. However, it is not as straightforward as in the case of the Vasicek model since we have to consider which and how many parameters - depending on the other parameter values that are “free” for calibration - have to be adjusted in order to satisfy (5.17). To overcome this problem we apply the following approach:

For fixed $\beta_{11}$ we determine the parameters $\beta_{12}, \beta_{21}$ and $\beta_{22}$ by solving the two non-linear equations (5.17). Having then fixed $\beta, \alpha_1, S$ and $\delta$ we define $\alpha_2$ by setting $A(10) = 0$, which is the necessary condition of (5.2) for $A$. The exact optimization algorithm that we used is the following:

I. Generate start values:

- For $j = 1, 2, \ldots, j_{\text{max}}$
  - Generate $\beta_{11}^j$ randomly.

- For $k = 1, 2, \ldots, k_{\text{max}}$
  - Solve equations (5.17) for determining $\beta_{12}^j, \beta_{21}^j, \beta_{22}^j$.

- For $i = 1, 2, \ldots, i_{\text{max}}$
  - Generate $\alpha_1^i, (s_1s_1')^i, (s_1s_2')^i, (s_2s_2')^i, \delta^i$ randomly as start values in order to calculate some steps of the following optimization.\(^3\)
  - Minimize
  \[
  \text{Res}_{j,k,i} = \sum_{T_n} (R^B(0,T_n) - R(0,T_n))^2 \tag{5.18}
  \]
  with respect to $\alpha_1, s_1s_1', s_1s_2', s_2s_2'$ and $\delta$ leaving $\beta_{jk}$ unchanged.\(^4\) The model yields are computed by means of the explicit solution (5.16) of $B$ and $A$, adjusting $\alpha_2$ so that $A(10) = 0$ is fulfilled.

\(^2\)As the solution of (5.17) for $\beta_{12}, \beta_{21}$ and $\beta_{22}$ is not unique the $k$- loop is necessary to find the most appropriate parameters.

\(^3\)We only generate $s_1s_1', s_1s_2', s_2s_2'$ where $s_i$ is the $i^{th}$ row of $S$ as the values of $S$ only appear in these combinations.

\(^4\)We use again the Matlab function “lsqnonlin” for this least-square problem.
II. Final optimization:

- Do a last optimization of (5.18) with respect to $\alpha_1, s_1s_1', s_1s_2', s_2s_2'$ and $\delta$ using the values with minimal residual found by the above loops as start parameters.

The reason why we do this kind of optimization is the fact that $\beta$ is subject to non-linear constraints which can usually not be implemented in standard optimization routines for the non-linear least-squares problem. In fact, $\beta_{12}, \beta_{21}, \beta_{22}$ and $\alpha_2$ are adjusted to match the constraints, whereas all the other parameters are chosen to fit the market yields at best.

Since we want to find parameters that remain relatively stable over time or at best are the same for each point in time, we minimize (5.18) for different dates simultaneously. In fact we try to minimize

$$\sum_{t_i} \sum_{T_n} (R^M(t_i, T_n) - R(t_i, T_n))^2,$$

where $t_i$ corresponds to the end-of-month dates from October 2005 to March 2006. $T_n$ are again the different maturities $1/12, 2/12, \ldots, 11/12, 1, 2, \ldots, 30$.

For these data we found the following parameters.

<table>
<thead>
<tr>
<th>$\beta_{11}$</th>
<th>$\beta_{12}$</th>
<th>$\beta_{21}$</th>
<th>$\beta_{22}$</th>
<th>$\delta$</th>
</tr>
</thead>
<tbody>
<tr>
<td>-0.8671</td>
<td>0.8561</td>
<td>-0.1467</td>
<td>0.1452</td>
<td>0.0146</td>
</tr>
<tr>
<td>$\alpha_1$</td>
<td>$\alpha_2$</td>
<td>$s_1s_1'$</td>
<td>$s_1s_2'$</td>
<td>$s_2s_2'$</td>
</tr>
<tr>
<td>0.0046</td>
<td>0.0010</td>
<td>0.8835</td>
<td>0.1011</td>
<td>0.0133</td>
</tr>
</tbody>
</table>

Table 5.1: Parameters for the Two-Factor Deterministic Volatility Model between October 2005 and March 2006

Table 5.2 summarizes the residuals obtained by using the above parameters for every month from October to March.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1.88e-05</td>
<td>1.84e-05</td>
<td>0.21e-05</td>
<td>0.20e-05</td>
<td>0.59e-05</td>
<td>1.61e-05</td>
<td>6.34e-05</td>
</tr>
</tbody>
</table>

Table 5.2: Residuals for the Two-Factor Deterministic Volatility Model, October 2005 - March 2006

The results of December, January and February are already satisfactory, whereas the other ones could be improved by slightly adapting the parameters.
which we investigate later on. Figure 5.1 also confirms this findings. When applying the above parameters to data from April we also achieve a promising result since the residual is only 1.069e-05, although we did not use data from April in our initial optimization. The graphical result of April is also shown in Figure 5.1.

As already indicated we now adjust our parameters for each month in order to achieve better results, especially for October, November, March and April. The values for $\beta$ remain unchanged, whereas $\alpha_1, s_1 s'_1, s_1 s'_2, s_2 s'_1$ and $\delta$ are slightly modified to fit the data in a better way. Indeed, we always use the optimal parameters of the previous month as start values for the next optimization rather than generating new initial values for each month as described in the above algorithm. This procedure allows to keep the values relatively constant over time. Table 5.3 contains the so obtained values for $\alpha_1, s_1, s'_1, s'_2$ and $\delta$ as well as the corresponding residuals for each month, which have improved in all cases.

<table>
<thead>
<tr>
<th>Month</th>
<th>$\alpha_1$</th>
<th>$\alpha_2$</th>
<th>$s_1 s'_1$</th>
<th>$s_1 s'_2$</th>
<th>$s_2 s'_1$</th>
<th>$\delta$</th>
<th>Res</th>
</tr>
</thead>
<tbody>
<tr>
<td>Oct.</td>
<td>0.0032</td>
<td>0.0011</td>
<td>0.8887</td>
<td>0.0958</td>
<td>0.0136</td>
<td>0.0147</td>
<td>0.53e-05</td>
</tr>
<tr>
<td>Nov.</td>
<td>0.0032</td>
<td>0.0011</td>
<td>0.8874</td>
<td>0.0963</td>
<td>0.0136</td>
<td>0.0146</td>
<td>0.38e-05</td>
</tr>
<tr>
<td>Dec.</td>
<td>0.0037</td>
<td>0.0009</td>
<td>0.8701</td>
<td>0.0976</td>
<td>0.0129</td>
<td>0.0132</td>
<td>0.20e-05</td>
</tr>
<tr>
<td>Jan.</td>
<td>0.0040</td>
<td>0.0009</td>
<td>0.8714</td>
<td>0.098</td>
<td>0.0127</td>
<td>0.013</td>
<td>0.12e-05</td>
</tr>
<tr>
<td>Feb.</td>
<td>0.0039</td>
<td>0.0009</td>
<td>0.8727</td>
<td>0.0991</td>
<td>0.0131</td>
<td>0.0128</td>
<td>0.33e-05</td>
</tr>
<tr>
<td>Mar.</td>
<td>0.0049</td>
<td>0.0008</td>
<td>0.8708</td>
<td>0.1009</td>
<td>0.0125</td>
<td>0.0133</td>
<td>0.21e-05</td>
</tr>
<tr>
<td>Apr.</td>
<td>0.0045</td>
<td>0.0008</td>
<td>0.8693</td>
<td>0.1019</td>
<td>0.0132</td>
<td>0.013</td>
<td>0.57e-05</td>
</tr>
</tbody>
</table>

Table 5.3: Monthly Parameters for the Two-Factor Deterministic Volatility Model, October 2005 - April 2006

The behavior of the parameters $\alpha_1, \alpha_2, s_2 s'_2, \delta$ over time is then graphically illustrated by figure 5.2, which points out that the variation of the parameters is very small. Considering the results of March as an example, figure 5.3 shows that the market- and the model yield curves are now nearly identical.

### 5.2.3 Stochastic Volatility

If $\gamma \neq 0$, we have to assure the non-negativity of $v(X(t)) := \gamma' X(t) + \delta$ (see (2.12)) by restricting the coefficients. For this reason we consider the “hyperplane”

$$H = \{(x_1, x_2) : \gamma_1 x_1 + \gamma_2 x_2 + \delta = 0\}, \quad (5.20)$$
Figure 5.1: Market- versus Model Yields, Two-Factor Deterministic Volatility Model, October 2005 - April 2006
where $v = 0$. Without loss of generality we take $\gamma_2 = 1$, if $\gamma_1 \neq 0$, so that $x_2 = -(\gamma_1 x_1 + \delta)$ on $H$. There, the drift function of $v(X(t))$ is consequently

$$
\gamma'(\alpha + \beta x) = \gamma_1[\alpha_1 + \beta_{11}x_1 + \beta_{12}x_2] + \gamma_2[\alpha_2 + \beta_{21}x_1 + \beta_{22}x_2]
$$

$$
= \gamma_1[\alpha_1 + \beta_{11}x_1 - \beta_{12}(\gamma_1 x_1 + \delta)]
$$

$$
+ [\alpha_2 + \beta_{21}x_1 - \beta_{22}(\gamma_1 x_1 + \delta)],
$$

$$
= k_1 + k_2 x_1
$$

(5.21)

where

$$
k_1 = \gamma_1(\alpha_1 - \beta_{12}\delta) + \alpha_2 - \beta_{22}\delta.
$$

(5.22)

$$
k_2 = \gamma_1\beta_{11} - \gamma_1^2\beta_{12} + \beta_{21} - \gamma_1\beta_{22}.
$$

(5.23)

Thus, in order to guarantee $v(X(t))$ to be non-negative, we will have to assume that it has a sufficiently strong positive drift on $H$. Therefore the model has to satisfy another condition

$$
k_1 > 0, \quad \text{and} \quad k_2 = 0
$$

(5.24)

besides the general requirements (5.10)-(5.12) and (5.2).
5.2.4 Calibrating the Stochastic Volatility Model to the Current Term Structure

As we do not have explicit solutions for the differential equations (5.10) in this case, we need numerical methods for solving them. In effect, we have to solve a two-point boundary value problem (BVP) for the differential equations (5.10), where the boundary conditions are

\[ A(0) = 0 \quad A(10) = 0, \]
\[ B_1(0) = 0 \quad B_1(10) = 0, \]
\[ B_2(10) = -10. \]

Having fixed the parameters \( \beta_{12}, \gamma_1, S_5, \alpha_1, \delta \), setting \( \gamma_2 = 1 \) and deriving \( \beta_{21} \) from \( k_2 = 0 \) (5.24), it is then possible to find the 3 parameters \( \beta_{11}, \beta_{22} \) and \( \alpha_2 \) by solving the above BVP. More precisely, we determine all our parameters in the following way:

I. Generate start values:

---

5We fix again \( s_1s_1', s_2s_2', s_1s_2' \).
• For $j = 1, 2, \ldots, j_{\text{max}}$

1. Generate $\beta_{11}, \beta_{22}$ and $\alpha_2$ randomly as initial guess for the BVP algorithm.

2. Generate $\beta_{12}, \gamma_1, S^j, \alpha_1^j, \delta^j$ randomly as start values in order to calculate some steps of the following minimization. Set $\gamma_2 = 1$ and calculate $\beta_{21}$ by means of (5.23).

3. Minimize

$$\text{Res}^j = \sum_{T_n} (R^M(0, T_n) - R(0, T_n))^2$$

(5.25)

with respect to $\beta_{12}^j, \gamma_1^j, S^j, \alpha_1^j, \delta^j$. In each optimization step the model yields are calculated by solving the above BVP problem including the determination of the parameters $\beta_{11}, \beta_{22}$ and $\alpha_2$. Having then found all the parameters, the solution of (5.10) is calculated by a Runge-Kutta method for all maturities (until 30), since the solution of the BVP problem is only obtained in the interval $[0, 10]$.

4. Save the parameters if $(\text{Res}^j < \text{Res}^{j-1})$.

II. Final optimization:

• Do a last optimization as described in item 3, using the parameters with minimal residual found by the above loop as start parameters.

Finally one has to check if the parameters are chosen so as to satisfy the conditions of of theorem 2.3. For solving our BVP problem we use the Matlab function “bvp4c” which is a finite difference code that implements the three stage Lobatto Illa formula.

Analogously to the deterministic case we minimize (5.19) for the same data as before, which allows to obtain constant parameters for different points in time. With this method we found the parameters listed in table 5.4, which also satisfy the conditions of theorem 2.3. The sum of the residuals obtained

\begin{table}[h]
\centering
\begin{tabular}{cccccc}
\hline
$\alpha_1$ & $\alpha_2$ & $\beta_{11}$ & $\beta_{12}$ & $\beta_{21}$ & $\beta_{22}$ \\
\hline
-0.0101 & 0.0045 & -0.6592 & 1.0675 & -0.1609 & 0.1733 \\
\hline
$s_1s'_1$ & $s_2s'_2$ & $s_1s'_2$ & $\gamma_1$ & $\delta$ \\
\hline
0.2945 & 0.0586 & 0.0204 & -0.4263 & 0.0246 \\
\hline
\end{tabular}
\caption{Parameters for the Two-Factor Stochastic Volatility Model, October 2005 - March 2006}
\end{table}
by using these parameters for every month from October to March is smaller
than in the deterministic volatility case, which can be read off table 5.5.
Similar to the deterministic volatility case these results can be improved by
optimizing (5.25) separately for each month using the parameters from table
5.4 as initial values. Table 5.6 presents the parameters for each month which
fulfill the conditions of theorem 2.3 as well.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>0.46e-05</td>
<td>0.75e-05</td>
<td>1.52e-05</td>
<td>0.38e-05</td>
<td>0.43e-05</td>
<td>1.18e-05</td>
<td>4.73e-5</td>
</tr>
</tbody>
</table>

Table 5.5: Residuals for the Two-Factor Stochastic Volatility Model, October 2005 - March 2006

From the following figures one can gather that the parameters remain
stable while the result are really satisfactory since they could be improved
with respect to the results presented in table (5.5), where we used in contrast
to the current approach exactly the same parameters for each month. Figure
5.5 shows the comparison between model and market yields from March and
October as example. It is obvious that the fit is very good for both month
although the parameters do not change much.
CHAPTER 5. AFFINE YIELD-FACTOR MODELS

Figure 5.4: Parameters over Time, Two-Factor Stochastic Volatility Model, October 2005 - March 2006

Figure 5.5: Market- versus Model Yields, Two-Factor Stochastic Volatility Model, October 2005, March 2006
Using the parameters from Table 5.4 as initial values for the calibration to data from April\textsuperscript{6} we also found satisfying results with parameters that do not change significantly which can also be seen in Table 5.6.

5.2.5 Conclusion

On account of our calibration results we can conclude that both models, the deterministic and the stochastic volatility model are appropriate means to fit the term structure of interest rates. Both models enable a good fit to market yields from different dates with unchanging parameters. This can be seen as advantage over models with time dependent parameters, as the Hull-White model described in Section 4.2.3, where one parameter is a function of the market instantaneous forward rate. The fitting results of the stochastic volatility model are slightly better than those of the deterministic volatility model, whereas the numeric calibration procedure is more stable in the latter case, since there are explicit solutions to the Ricatti equations. In the stochastic volatility case however, we have to solve them in each optimization step numerically respecting the boundary constraints, which makes the calibration very time-intensive and less stable.

\textsuperscript{6}April data was not taken into consideration when optimizing (5.19).
Appendix A

Numerical Methods for Calibration

We explain here some methods that we used when calibrating our models to market data. For most of these procedures Matlab already provides good algorithms which we incorporated into our own programs. In the following sections we dwell on the methods that form a basis of our nonlinear least-squares data fitting problems.

A.1 Trust-Region Methods for Nonlinear Minimization

Our calibration issues described in chapter 4 are principally a matter of nonlinear least-squares problems. However, before referring to this special kind of minimization problem, we consider the general case, i.e. the general unconstrained minimization problem

$$\min_{x \in \mathbb{R}^n} f(x),$$  \hspace{1cm} (A.1)

where $f: \mathbb{R}^n \to \mathbb{R}$ is a twice continuously differentiable function. The basic idea is to approximate $f$ with a simpler function $\psi$, which reasonably reflects the behavior of $f$ in a neighborhood around the current point $x_k$, the so-called trust region. In the standard trust region method [13] for problem (A.1), $\psi$ is the quadratic approximation for $f$ defined by the first two terms of the Taylor approximation to $f$ at $x_k$. The trust-region subproblem is typically formulated by

$$\min_{s \in \mathbb{R}^n} \left\{ \psi_k(s) = g_k s + \frac{1}{2} s' H_k s : ||D_k s|| \leq \Delta_k \right\}, \hspace{1cm} (A.2)$$
where \( g_k = \nabla f(x_k) \), the gradient of \( f \) at the current point \( x_k \), and \( H_k \) is the Hessian matrix \( \nabla^2 f(x_k) \). \( D_k \) is a non-singular scaling matrix and \( \Delta_k \) is a positive scalar representing the trust region size. \( \| \cdot \| \) denotes the 2-norm.

In Newton’s method with a trust region strategy, each iterate \( x_k \) has therefore a neighborhood regulated by \( \Delta_k \) and \( D_k \) so that

\[
f(x_k + s) \approx f(x_k) + \psi_k(s), \| D_k s \| \leq \Delta_k.
\]

(A.3)

In other words \( \psi_k \) is a model of the reduction in \( f \) within a neighborhood of the current point \( x_k \). So, by finding \( s_k \) that minimizes \( \psi_k \) we also minimize \( f(x_k + s) \). If the step is satisfactory, in the sense that \( f(x_k + s_k) < f(x_k) \), \( x_{k+1} = x_k + s_k \), otherwise the \( x_k \) remains unchanged and the trust region is shrunk by adapting \( \Delta_k \).

Thus, the essential step is the solution of the trust-region subproblem. The Matlab method follows the Branch, Coleman and Li [2] approach, which restricts the trust region subproblem to a two-dimensional subspace \( S = \langle s_1, s_2 \rangle \). This is necessary, because accurate solutions of (A.2) require too much time in the case of large-scale problems. So the first direction of our subspace \( s_1 \) is the gradient whereas the second direction \( s_2 \) is determined with the aid of a preconditioned conjugate gradient process (PCG). PCG is a popular way to solve general large symmetric positive definite systems of linear equations

\[
Hs = -g.
\]

(A.4)

In our context we obtain this equation by differentiating \( \psi \) with respect to \( s \) and setting the derivative 0. So the solution \( s \) would be the arg min of \( \psi \). The Hessian matrix can assumed to be symmetric, however it is guaranteed to be positive definite only in the neighborhood of a strong minimizer. The PCG algorithm exists when a direction of negative curvature, i.e. \( s^T H s \leq 0 \), is encountered. So the output direction \( s_2 \) is either a direction of negative curvature or an approximate solution of the Newton system \( Hs_2 = -g \). Solving (A.2) becomes much easier and faster than in the unrestricted case since in the subspace the problem is only two-dimensional.

Summarizing the mentioned ideas the whole minimization algorithm consists basically in 4 steps:

- Formulation of the 2-dimensional trust region subproblem using PCG to determine the subspace directions.

- Resolution of the now two-dimensional problem (A.2) to determine \( s_k \).

- If \( f(x_k + s_k) < f(x_k) \) then \( x_{k+1} = x_k + s_k \).

\[\text{provided that } H \text{ is positive definite.}\]
• Adjustment of $\Delta_k$

These four steps are repeated until convergence. For some special cases of $f$ the procedure can be simplified, for instance for our nonlinear least-squares, as we see in section (A.1.2).

### A.1.1 Box Constraints

The box constrained problem is of the form

$$\min_{l \leq x \leq u} f(x),$$

(A.5)

where $l$ is a vector of lower bounds and $u$ is vector of upper bounds, which can also be equal to $-\infty$ and $\infty$. Two techniques are used to generate a sequence of strictly feasible points. Firstly, instead of the unconstrained Newton step, i.e. solving (A.4) to define the two dimensional subspace $S$, a scaled modified Newton step is introduced and secondly reflections are used to increase the stepsize (see Coleman and Li [5] for details).

### A.1.2 Nonlinear Least-Squares

An important special case for $f(x)$ which is relevant for our purposes is the nonlinear least-square problem

$$f(x) = \frac{1}{2} \sum_{i=1}^{n} f_i^2(x) = \frac{1}{2} ||F(x)||^2$$

(A.6)

$F(x)$ is a vector-valued function whose $i^{th}$-component is $f_i(x)^2$. The structure of this function is exploited to enhance efficiency. Instead of trying to find $s_2$ of the two-dimensional subspace by solving (A.4), the normal equations, i.e.

$$J'Js_2 = J'F,$$

(A.7)

where $J$ is the Jacobian of $F(x)$ are solved for determining $s_2$. These equations are derived by differentiating $\tilde{\psi} = ||Js + F||^2$ with respect to $s$ and setting the derivative 0. In this particular case $s'g + s'Hs$ can be approximated by $\tilde{\psi}$ which allows to avoid the calculation of the second derivatives.

---

\(^2\)In our calibration problems $F$ is always the difference between the market- and the model yields.
Bibliography


