CONTINUOUS-TIME IDENTIFICATION OF EXPONENTIAL-AFFINE TERM STRUCTURE MODELS

Arianto Wibowo

FELab
Financial Engineering Laboratory

Graduation Committee:

prof. dr. ir. J. J. W. van der Vegt (chairman), University of Twente, EWI
prof. dr. A. Bagchi (promotor), University of Twente, EWI
prof. dr. S. Aihara, Tokyo University of Science
prof. dr. F. Jamshidian, University of Twente, EWI
prof. dr. W. Albers, University of Twente, EWI
dr. ir. M. H. Vellekoop, University of Twente, EWI
dr. D. Dupont, University of Twente, BBT

Printed by Wöhrmann Printing Service, Zutphen, The Netherlands.

ISBN: 90-365-2442-3

CONTINUOUS-TIME IDENTIFICATION OF EXPONENTIAL-AFFINE TERM STRUCTURE MODELS

DISSERTATION

to obtain
the degree of doctor at the University of Twente,
on the authority of the rector magnificus,
prof. dr. W.H.M. Zijm,
on account of the decision of the graduation committee,
to be publicly defended
on Wednesday, 6 December 2006 at 13:15

by

Arianto Wibowo

born on 21 November 1975
in Bandung, Indonesia

This dissertation has been approved by the promotor

prof. dr. A. Bagchi

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Goal . . . 1
- 1.3 Structure of the Thesis . . . 2

2 An Overview of Term Structure Modeling and Estimation **3**
- 2.1 Term Structure Modeling . . . 3
  - 2.1.1 Bonds and Interest Rates . . . 3
  - 2.1.2 Short Rates Modeling . . . 5
- 2.2 The Exponential-Affine Term Structure Models . . . 8
  - 2.2.1 Basic Assumptions . . . 9
  - 2.2.2 Pricing of European-style Derivatives . . . 10
    - 2.2.2.1 Derivatives with Exponential Pay-off . . . 11
- 2.3 Estimation of Exponential-Affine Term Structure . . . 13
  - 2.3.1 Observed Interest Rate Factors . . . 13
    - 2.3.1.1 Euler Approximation . . . 14
    - 2.3.1.2 Monte-Carlo integration . . . 14
    - 2.3.1.3 Fourier inversion . . . 15
  - 2.3.2 Unobserved Interest Rate Factors . . . 16

3 A Continuous-time Maximum Likelihood Estimation **19**
- 3.1 Introduction . . . 19
- 3.2 Setup and Notation . . . 19
- 3.3 The Likelihood Functional . . . 21
- 3.4 Nonlinear Filtering of Interest Rate Factors . . . 23
  - 3.4.1 Equations for Nonlinear Filtering . . . 23
  - 3.4.2 Numerical Approximation of the Zakai Equation . . . 25
  - 3.4.3 Gaussian Approximation . . . 26
  - 3.4.4 Numerical Approximation of the Gaussian Filtering . . . 29

4 Parameter Estimation of Cox, Ingersol and Ross Model **33**
- 4.1 Introduction . . . 33
- 4.2 Model Setup . . . 33
  - 4.2.1 Cox-Ingersol-Ross Model . . . 33
  - 4.2.2 Measurement Model . . . 34
- 4.3 Maximum Likelihood Estimation . . . 34
  - 4.3.1 Likelihood Functional . . . 34

4.3.2 Nonlinear Filtering 35
4.3.2.1 The Prediction Step 35
4.3.2.2 The Correction Step 37
4.3.3 Gaussian Filtering 37
4.4 Simulated Data 37
4.4.1 Comparison of Filtering Performance 38
4.4.2 Comparison of Likelihood Functionals 42
4.5 Fama-Bliss Data 43
4.6 Conclusions 46

5 A Comparison of Three-factor Models in the European Fixed Income Market 47
5.1 Introduction 47
5.2 Interest Rate Models 47
5.2.1 Balduzzi-Das-Foresi-Sundaram Model 47
5.2.2 Chen Model 48
5.3 Measurement Model 48
5.4 The Euro Swap Rates 49
5.4.1 The Euro Swap Markets 49
5.4.2 Data Description 50
5.4.3 The Implied Term Structure 50
5.4.4 Measurement Equation 52
5.4.5 Estimation Results 53
5.5 The Euribor Futures 55
5.5.1 The Euribor Futures Markets 55
5.5.2 Data Description 56
5.5.3 Measurement Equation 58
5.5.4 Estimation Results 59
5.6 Conclusions 59

6 Conclusions and Future Research 65
6.1 Conclusions 65
6.2 Directions for Further Research 66

A Parameterization of BDFS and Chen Models 69

Bibliography 71

List of Figures

4.1 Numerical solution of the conditional density for $\sigma_0 = 10$ bp. 38
4.2 Comparison of filtering performance. Nonlinear filter vs Gaussian filter for $\sigma_0 = 10$ bp. 39
4.3 Comparison of filtering performance. Nonlinear filter vs Gaussian filter for $\sigma_0 = 0.1$ bp. 40
4.4 RMSE of filtered estimate. Nonlinear filter vs Gaussian filter for $\sigma_0 = 10$ bp. 41
4.5 RMSE of filtered estimate. Nonlinear filter vs Gaussian filter for $\sigma_0 = 0.1$ bp. 41
4.6 Monthly Fama-Bliss data from April 1964 to December 1997. . . . 45

5.1 The Euro swap rates from 2 January 2002 until 30 May 2005 51
5.2 Mean and standard errors between the fitted and observed Euro swap rates from 2 January 2002 until 30 December 2004 (calibration dataset). 54
5.3 Mean and standard errors between the fitted and observed Euro swap rates from 2 January 2005 until 30 May 2005 (validation dataset). 54
5.4 Errors between the fitted and observed Euro swap rates from 2 January 2002 until 30 May 2005. 55
5.5 Euribor futures prices from 2 January 2004 until 30 May 2005. . . . 57
5.6 Fitted Euribor futures prices from 2 January 2004 until 30 May 2005. 60
5.7 Mean and standard errors between the fitted and observed Euribor futures prices from 2 January 2004 until 30 December 2004 (calibration dataset). 61
5.8 Mean and standard errors between the fitted and observed Euribor futures prices from 2 January 2005 until 30 May 2005 (validation dataset). 62
5.9 Errors between the fitted and observed Euribor futures prices from 2 January 2002 until 30 May 2005. 62
5.10 Fitted 3 month Euribor rates from 2 January 2004 until 30 May 2005. 63

List of Tables

4.1 Parameters for the Cox-Ingersol-Ross model. 38
4.2 Monte Carlo results for the Cox-Ingersol-Ross model, $\sigma_0 = 0.1$ bp. . 43
4.3 Monte Carlo results for the Cox-Ingersol-Ross model, $\sigma_0 = 10$ bp. . 44
4.4 Wilcoxon signed rank test for zero median between estimates obtained using the robust likelihood and the Itô likelihood. 44
4.5 Wilcoxon signed rank test for zero median between estimates obtained using the nonlinear and Gaussian filters. 45
4.6 CIR parameter estimated from Fama-Bliss dataset starting April 1964 to December 1997. 46

5.1 Parameter estimates of the Euro swap rates. 53
5.2 Parameter estimates of the Euribor futures. 61

A.1 Parameter constraints for BDFS and Chen models 69
A.2 Drift and volatility parameters of BDFS and Chen models 70

Notation

Symbol	Description
$(\alpha)^*$	Transpose of α
$\mathrm{Tr}(\alpha)$	Trace of α
$\det(\alpha)$	Determinant of α
$\mathrm{diag}(\alpha)$	Diagonal matrix of which the diagonal elements are elements of the vector α
$\langle \alpha, \beta \rangle$	$\mathrm{Tr}((\alpha)^* \beta)$
$\|\alpha\|^2$	$\mathrm{Tr}((\alpha)^* \alpha)$
n_α	Number of elements of the vector α
$\overline{\phi}(X)(t)$	Filtered estimate of $\phi(X(t))$
$\hat{\alpha}$	Estimate of the parameter α
α_k	The k-th element of the vector α
$\alpha_{k,\cdot}$	The k-th row of the matrix α
$\alpha_{\cdot,k}$	The k-th column of the matrix α

1 Introduction

1.1 Motivation

In the early nineties, multi-factor modeling of the term structure of interest rates became an important issue as empirical data suggested that single factor models are inadequate in explaining the complex term structure movement. One highly cited class of models is the class of so-called exponential affine models. Due to the possibility of tracktable pricing formulas for interest rates derivatives, this class of models is quickly becoming popular among researchers and practitioners alike. However, even with its nice properties, estimating parameters of this particular class of models can be quite challenging.

A number of parameter estimation methods have been proposed in the literature. One of the popular estimation methods is the so called state-space filtering method, which is essentially a maximum likelihood estimation method under the assumption that observations are corrupted with additive Gaussian noise. Procedures suggested so far start with time discretization of the underlying model in order to apply known filtering techniques. Moreover, parameter estimates are obtained through maximization of a quasi-likelihood functional. The quasi-likelihood functional involves a stochastic process, the innovation process, that is readily available in the computation of filtered estimates of the interest rate factors. The word 'quasi' is put here because the innovations are assumed to be independent Gaussian random variables, which is not necessarily true.

Theoretical results are available for parameter estimation in continuous-time. With the assumption that observations are corrupted with additive Gaussian noise, the exact likelihood functional can be computed. To our knowledge, continuous-time identification methods have not been explored in the financial context.

1.2 Goal

The goal of this thesis is to formulate and to implement the continuous-time maximum likelihood method for exponential-affine interest rate models. In order to evaluate the performance of the proposed estimation method, we will fit a number of interest rate models to simulated data and to real data from the US and European fixed income markets.

1.3 Structure of the Thesis

This thesis is organized as follows. The second chapter discusses the general mathematical formulation of term structure modeling followed by a more specific exponential-affine class of models. This chapter include an overview of parameter estimation methodologies known in the literature. The continuous-time maximum likelihood method is presented in Chapter 3. Here we present the nonlinear filtering and an approximate Gaussian filter. Algorithms related to the implementation of these filtering algorithms are also presented in the chapter. In chapter 4, we present parameter estimation results of the Cox, Ingersol and Ross model to simulated data and the US term structure. Chapter 5 is dedicated to fitting BDFS and Chen models to the Euro swap rates and Euribor futures prices. Conclusions and directions for further research are given in Chapter 6.

2 An Overview of Term Structure Modeling and Estimation

2.1 Term Structure Modeling

The term structure of interest rates, commonly referred to as the term structure, is defined as a series of zero coupon bond yields of different maturity dates at a particular time. In practice, the term structure can be implied from liquidly traded bonds and other interest rates derivatives. As time evolves, the term structure randomly changes shape. Empirical studies suggest that changes in some part of the term structure also influences the other part of the term structure. By the fact that the term structure is directly related to prices of interest rates derivatives, proper understanding of its future behavior is essential in managing interest rate sensitive portfolios.

2.1.1 Bonds and Interest Rates

Bonds are fixed income instruments that are issued by either governments or corporations. Different kinds of bonds are traded on many exchanges. Some of these are zero coupon bonds, coupon bearing bonds and callable bonds. A *zero coupon bond* is defined as a financial security paying a fixed amount of cash at a future *maturity date* without any intermediate payments in between (without coupons). The fixed amount is called the bond's principal value or the face value. Zero coupon bonds are the most elementary interest rates derivatives. Coupon bearing bonds can be broken down into a portfolio of these zero coupon bonds. For this reason, zero coupon bonds are usually taken as the starting point of mathematical modeling of the term structure of interest rates. While both zero coupon bonds and coupon bearing bonds promise to pay certain cash-flows up to the maturity date of the bond, the callable ones are inherently an option that given certain conditions allow the exchange of the bond for a portion of the equity of its corporate issuer. In general, the issuer of a bond may or may not be able to honor their promise to pay the claims that are written on the contract. Thus, bonds are subjected to default risk of its issuer. However, there are some bonds that are generally regarded as proxies to default-free bonds. For example, the United States treasury bonds and bonds that are issued by economically strong

European countries.

Let us denote the price at time t of a zero coupon bond maturing at time T by $B(t,T)$ and assume that the bond's principal value is normalized to a unit, i.e. $B(T,T)=1$. Note that instead of characterizing bonds with a fixed maturity date T, alternatively, one may parameterize a bond with its *time to maturity* $\tau = T-t$.

The *yield to maturity* $Y(t,T)$ of holding a zero coupon bond in the interval $[t,T]$ is defined as the compounded rate of return of the bond:

$$Y(t,T) = -\frac{1}{T-t}\ln B(t,T), \quad \forall t \in [0,T]. \tag{2.1}$$

The term structure of interest rates, or a yield curve, is the function that relates $Y(t,T)$ to the maturity T. In practice, the term structure may be derived not only from zero coupon bond prices, but also from quoted prices of other actively traded instruments such as coupon bearing bonds, swaps and interest rates futures.

Forward rates $f(t,T_1,T_2)$ are defined as the interest rates that prevail at time t for riskless lending or borrowing over the future time interval $[T_1,T_2]$. It can be realized by a portfolio of two bonds with two different maturities T_1 and T_2. The forward rate is related to the bond price by the following equation:

$$\frac{B(t,T_2)}{B(t,T_1)} = \exp\left(-f(t,T_1,T_2)(T_2-T_1)\right) \tag{2.2}$$

or equivalently

$$f(t,T_1,T_2) = -\frac{\ln B(t,T_2) - \ln B(t,T_1)}{(T_2-T_1)} \tag{2.3}$$

Note that the yield to maturity can be expressed as $Y(t,T) = f(t,t,T)$, that is, the interest rates over the time period $[t,T]$ as seen from time t.

In continuous-time modeling of bond prices, it is useful to consider the limiting case of $f(t,T_1,T_2)$ when $T_2 \longrightarrow T_1$. The resulting expression

$$f(t,T) = \lim_{T_2 \longrightarrow T} f(t,T,T_2)$$

is called the *instantaneous forward rate*. It is interpreted as the instantaneous interest rate which prevails at time T as seen from time t. Unlike bond prices, yield to maturities and forward rates, the instantaneous forward rate is a mathematical idealization rather than a physical quantity observed in the market. Given instantaneous forward rates $f(t,T)$, bond prices are defined by setting

$$B(t,T) = \exp\left(-\int_t^T f(t,s)\ ds,\right) \quad \forall t \in [0,T]. \tag{2.4}$$

Thus if the bond price functional is sufficiently smooth with respect to the maturity T, then instantaneous forward rates $f(t,T)$ can be written as:

$$f(t,T) = -\frac{\partial \ln B(t,T)}{\partial T} \tag{2.5}$$

which follows from (2.3) by taking limits.

The *instantaneous short rate* $r(t)$ is defined as the instantaneous risk-free borrowing and lending interest rate at time t,

$$r(t) = \lim_{T \longrightarrow t} f(t,T).$$

It is the continuous-time version of the interest rate affecting the ordinary savings accounts. Given $\{r(s), 0 \leq s \leq t\}$, the time t value of a savings account subjected to a continuously compounded interest rates is given by:

$$B(t) = \exp\left\{\int_0^t r(s)\ ds\right\} \tag{2.6}$$

where here we have normalized the savings account so that $B(0) = 1$. Equivalently, a savings account solves the differential equation

$$\frac{dB(t)}{B(t)} = r(t)\ dt, \quad B(0) = 1 \tag{2.7}$$

The value of the savings account $B(t)$ represents the amount of cash accumulated up to time t starting from a unit of initial investment at time 0. The connection between short rates and bond prices can easily be made when $r(t)$ is deterministic. In order to avoid arbitrage, the price of a zero coupon bond $B(t,T)$ must be equivalent to a cash amount $\tilde{B} = \exp\left\{-\int_t^T r(s)\ ds\right\}$ invested at time t in a savings account. Therefore,

$$B(t,T) = \exp\left\{-\int_t^T r(s)\ ds\right\}, \quad \forall t \in [0,T]. \tag{2.8}$$

This relationship is less straightforward in a stochastic setup.

2.1.2 Short Rates Modeling

In a stochastic setup, the instantaneous short rate is assumed to be an adapted process $r(t)$ on the filtered probability space $\left(\Omega, \mathbb{F}, \mathbb{P}, (\mathcal{F}(t))_{t\in[0,\tilde{T}]}\right)$ and integrable with respect to the Lebesgue measure over an interval $\left[0, \tilde{T}\right]$ for some $0 < \tilde{T} < \infty$. The filtration $\mathbb{F} = (\mathcal{F}(t))_{t\in\left[0,\tilde{T}\right]}$ is generated by Brownian motions

$W^{\mathbb{P}}(t)$ driving the random process of $r(t)$. Within this stochastic framework, a zero coupon bond is formulated as a derivative instrument on the short rate. Thus, in order to find the fair value $B(t,T)$ of a zero coupon bond, machinery such as the risk neutral valuation method is useful.

Following the risk neutral valuation methodology, we first assume the existence of a savings account satisfying (2.7). A family of bond prices $\{B(t,T),\ 0 \leq t \leq T,\ T \in \left[0, \tilde{T}\right]\}$, is *arbitrage-free* if there exists a probability measure $\mathbb{Q}$ equivalent to $\mathbb{P}$ such that for any $T \in \left[0, \tilde{T}\right]$ the relative bond price process $\left\{Z(t,T) = \frac{B(t,T)}{B(t)},\ \ 0 \leq t \leq T\right\}$ is a martingale under $\mathbb{Q}$ (see e.g. Musiela and Rutkowski (1998) for details). This implies that $Z(t,T) = \mathbf{E}^{\mathbb{Q}}(Z(T,T) \mid \mathcal{F}(t))$ which enable us to express the bond price as:

$$B(t,T) = \mathbf{E}^{\mathbb{Q}}\left(\exp\left\{-\int_t^T r(s)\, ds\right\} \,\middle|\, \mathcal{F}(t)\right), \quad \forall t \in [0,T] \tag{2.9}$$

Given the general expression (2.9), one needs to specify the stochastic process of the underlying short rates $r(t)$. Early approaches to model the instantaneous short rates focused mainly on single-factor models. These models assume that the short rates is driven by a single stochastic factor as it evolves through time e.g. the Vasicek (1977) model which describe the instantaneous short rates as a Gaussian process. A general equilibrium approach to short rates modeling developed by Cox et al. (CIR, 1985) leads to a modification of the mean reverting diffusion model of Vasicek. The model is known as the CIR square root model which under certain technical conditions maintains positivity of the short rates. Further generalizations and modifications can be found in Longstaff and Schwartz (1989) and Hull and White (1990), among many others. Although in most cases one factor models provide tractable bond price equations, where often analytical solutions are known, empirical evidence suggests that bond price movements may depend on more than one driving factor. Moreover, cross-sectional data often suggests that simplistic one factor models fail to produce various shapes of the observed term structure.

Let us now consider the following multi-factor model of the short rates. We denote stochastic processes $\left\{(X_1(t), \ldots, X_{n_X}(t)), t \in [0, \tilde{T}]\right\}$ as stochastic factors of the short rates. These factors satisfy the Itô Stochastic Differential Equations (SDE):

$$dX_i(t) = F_i(t, X(t))\ dt + \sum_{j=1}^{n_W} G_{i,j}(t, X(t))\ dW_j^{\mathbb{P}}(t); \quad i = 1, \ldots, n_X \tag{2.10}$$

or, written in a matrix notation,

$$dX(t) = F(t, X(t))\ dt + G(t, X(t))\ dW^{\mathbb{P}}(t); \quad X(0) = X_0 \tag{2.11}$$

where $X(t) = [X_1(t), \ldots, X_{n_X}(t)]^*$. F: $\left[0, \tilde{T}\right] \times \Re^{n_X} \mapsto \Re^{n_X}$ and G: $\left[0, \tilde{T}\right] \times \Re^{n_X} \mapsto \Re^{n_W \times n_X}$ are assumed to satisfy certain regularity conditions (e.g. Lipschitz and growth conditions) to ensure the existence and uniqueness of the solutions to (2.11). X_0 is a random initial condition satisfying $\mathbf{E}\left[\|X_0\|^2\right] < \infty$ and $\left\{W^{\mathbb{P}}(t),\ t \in [0, \tilde{T}]\right\}$ is a n_W-dimensional Brownian motion. The instantaneous short rate is assumed to be a known function of the interest rate factors,

$$r(t) = \psi(t, X(t)).$$

Theorem 2.1.1 *(Musiela and Rutkowski (1998), Thm B.2.1 pp. 466) Let $W^{\mathbb{P}}$ be a standard n_W-dimensional Brownian motion on a filtered probability space $\left(\Omega, \mathbb{F}, \mathbb{P}, (\mathcal{F}(t))_{t\in[0,\tilde{T}]}\right)$. Suppose that λ is an adapted n_W-valued process such that*

$$\mathbf{E}^{\mathbb{P}}\left[\int_0^{\tilde{T}} \langle \gamma(s), dW(s) \rangle - \frac{1}{2}\int_0^{\tilde{T}} \|\gamma(s)\|^2\, ds\right] = 1 \tag{2.12}$$

Define a probability measure $\mathbb{Q}$ on $(\Omega, \mathcal{F}_{\tilde{T}})$ equivalent to $\mathbb{P}$ by means of the Radon-Nikodym derivative

$$\frac{d\mathbb{Q}}{d\mathbb{P}}\bigg|_{\mathcal{F}_t} = \exp\left\{\int_0^t \langle \gamma(s), dW(s) \rangle - \frac{1}{2}\int_0^t \|\gamma(s)\|^2\, ds\right\}$$

Then the stochastic process $W^{\mathbb{Q}}$, which is given by the formula

$$W^{\mathbb{Q}}(t) = W^{\mathbb{P}}(t) - \int_0^t \gamma(s) ds$$

is a standard n_W-dimensional Brownian motion on the space $(\Omega, \mathbb{F}, \mathbb{Q}, (\mathcal{F}(t))_{t\in[0,\tilde{T}]})$.

In view of the above Girsanov's theorem, let $\gamma(t) = \lambda(t, X(t))$. This implies

$$W^{\mathbb{Q}}(t) = W^{\mathbb{P}}(t) - \int_0^t \lambda(s, X(s))\ ds \tag{2.13}$$

is a n_W−dimensional Brownian motion under the equivalent measure $\mathbb{Q}$. Thus, under the equivalent measure $\mathbb{Q}$, the state process $\left\{X(s),\ t \in [0, \tilde{T}]\right\}$ can be written as:

$$\begin{aligned} dX(t) &= \{F(t, X(t)) + G(t, X(t))\lambda(t, X(t))\} dt + G(t, X(t))\ dW^{\mathbb{Q}}(t), \\ X(0) &= X_0. \end{aligned} \tag{2.14}$$

The vector $\lambda(t, X(t))$ is often interpreted as the *market price of risk*, because the i-th component of $\lambda(t, X(t))$ measures the extent to which risk taken in the i-th factor is compensated by higher expected return. Since equivalent martingale measures are fully characterized by the market price of risk, choosing a martingale measure can be done by imposing certain restrictions on $\lambda(t, X(t))$. For tractability, market price of risk is usually chosen such that the resulting risk neutral dynamics has nice statistical properties.

By the Markov property of the interest rate factors (2.14), the price at time t of a zero coupon bond with maturity T can be written as

$$B(t, T, X(t)) = \mathbf{E}^{\mathbb{Q}}\left(\exp\left\{-\int_t^T \psi(s, X(s))\ ds\right\} \middle| \mathcal{F}(t)\right)$$

It follows from the Feynman-Kac formula (see Theorem 5.7.6 in Karatzas and Shreve (1988)), the bond price equation (2.9) satisfies the following partial differential equation (PDE):

$$\mathcal{P}B(t, T, x) - \psi(t, x) B(t, T, x) = 0 \tag{2.15}$$

where the differential operator $\mathcal{P}$ is given by:

$$\begin{aligned} \mathcal{P}f \quad = \quad & \frac{\partial f}{\partial t} + \sum_{k=1}^{n_X} \frac{\partial f}{\partial x_k} \left(F(t, x) + G(t, x) \lambda(t, x)\right)_k + \\ & \frac{1}{2} \sum_{k=1}^{n_X} \sum_{l=1}^{n_W} (G(t, x) G^*(t, x))_{k,l} \frac{\partial^2 f}{\partial x_k \partial x_l}, \end{aligned} \tag{2.16}$$

with the boundary condition $B(T, T, x) = 1$. Given several interest rates factors, solving the PDE (2.15) can be tedious, and in most cases the solution has to be approximated numerically. Closed form solutions are available for a few specific models. Semi-closed form solutions, in a sense that bond price can be obtained by solving a set of ordinary differential equations (ODE) are available for a small class of models. These classes include the exponential-affine term structure models (Duffie and Kan, 1993) and the exponential-quadratic term structure models (Leippold and Wu, 2002).

2.2 The Exponential-Affine Term Structure Models

The exponential-affine term structure class of models (Duffie and Kan (1996)) is one of the popular classes of models. There are at least three reasons for its popularity. First of all, as the name suggests, bond yields can be expressed as linear functions of the interest rates factors. Secondly, this class of models simplifies the bond pricing PDE into a set of ordinary differential equations. Thirdly, it spans many of the popular one factor and multi-factor models. Examples of interest rate models within the exponential affine class include:

One-factor Vasicek (1977), Cox-Ingersol-Ross (1985) and Hull-White (1990)

Two-factor Longstaff -Schwartz (1992), Chen-Scott (1993)

Three-factor BDFS (1996), Fong-Vasicek (1991), Chen (1996)

2.2.1 Basic Assumptions

Following Duffie and Kan (1996), let $X(t)$ denote a vector containing the short rate factors

$$X(t) = [X_1(t), \ldots, X_{n_X}(t)]^* \in \Re^{n_X \times 1}, \quad t \geq 0. \tag{2.17}$$

where $X(t)$ satisfies the following stochastic differential equation (SDE):

$$dX(t) = (AX(t) + B)\ dt + \Sigma\ \text{diag}\,(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}}\ dW^{\mathbb{Q}}(t) \tag{2.18}$$

where $A,\ \mathcal{A} \in \Re^{n_X \times n_X}$, $B,\ \mathcal{B} \in \Re^{n_X \times 1}$ and $\Sigma \in \Re^{n_X \times n_X}$ are constant matrices. The instantaneous short rates process satisfies the affine relation:

$$r(t) = g_0 + g_1^* X(t) \tag{2.19}$$

where $g_0 \in \Re$ and $g_1 \in \Re^{n_X \times 1}$. $W^{\mathbb{Q}}(t) \in \Re^{n_X \times 1}$ is a vector of $\mathbb{Q}$-Brownian motions.

It is convenient to preserve the structure of the interest rate model under a change of measure in order to take advantage of the features provided by the $\mathbb{Q}$-model (2.18). To derive the appropriate change of measure, we first note that by Girsanov's theorem, $W^{\mathbb{Q}}(t) = W^{\mathbb{P}}(t) - \int_0^t \lambda(s, X(s))\ ds$ is a vector of standard Brownian motions under the equivalent measure $\mathbb{Q}$. This means that in order to preserve the structure of the interest rate factors under the real measure $\mathbb{P}$, $\lambda(t, X(t))$ can be chosen as

$$\lambda(t, X(t)) \quad = \quad \text{diag}\,(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}}\, \Lambda$$

for some constant vector $\Lambda = [\lambda_1, \ldots, \lambda_{n_X}]^*$. With this particular choice of λ, the interest rate factors can now be written as

$$dX(t) = \left(A^{\mathbb{P}} X(t) + B^{\mathbb{P}}\right)\ dt + \Sigma\ \text{diag}\,(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}}\ dW^{\mathbb{P}}(t) \tag{2.20}$$

where

$$\begin{aligned} A^{\mathbb{P}} &= A + \Sigma\ \text{diag}\,(\Lambda)\ \mathcal{A}, \\ B^{\mathbb{P}} &= B + \Sigma\ \text{diag}\,(\Lambda)\ \mathcal{B}. \end{aligned}$$

2.2.2 Pricing of European-style Derivatives

Let $V(t;T,X(t))\,,\ 0\le t\le T$ be the time t value of a derivative instrument that pays off $h(X(T))\in\Re$ at a maturity time T. $V(t;T,X(t))$ satisfies

$$V(t;T,X(t))=\mathbf{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)\,ds\right)h(X(T))\,\middle|\,\mathcal{F}(t)\right] \tag{2.21}$$

where $\mathcal{F}(t)$ is the σ-algebra generated by the Brownian motions driving the interest rate factors. Equivalently, using the Feynman-Kac formula, $V(t;T,x)$ satisfies the following partial differential equation (PDE)

$$\begin{aligned}(g_0+g_1^*x)\,V &= \frac{\partial V}{\partial t}+\sum_{k=1}^{n_X}(Ax+B)_k\frac{\partial V}{\partial x_k}+\\ &\quad \frac{1}{2}\sum_{k=1}^{n_X}\sum_{l=1}^{n_X}(\Sigma\,\mathrm{diag}(\mathcal{A}x+\mathcal{B})\Sigma^*)_{k,\,l}\frac{\partial^2 V}{\partial x_k\partial x_l}\end{aligned} \tag{2.22}$$

where we have substituted $r(t)=g_0+g_1^*X(t)$, subject to the boundary condition $V(T;T,x)=h(x)$.

The solution to (2.22) can be obtained with the help of Arrow-Debreu securities. Let us denote $V_\delta(t;T,X(t),y)$ to be the time t value of an Arrow-Debreu security which pays off

$$h_\delta(X(T);y)=\delta(X(T)-y)$$

at the maturity time T. Thus, $V_\delta(t;T,x,y)$ satisfies the PDE (2.22) subject to the boundary condition $V_\delta(T;T,x,y)=\delta(x-y)$. Consider its Fourier transform in y:

$$\bar{V}_\delta(t;T,x,u)=\int_{\Re^{n_X}}\exp(-i\langle u,y\rangle)\;V_\delta(t;T,x,y)\;dy. \tag{2.23}$$

$\bar{V}_\delta(t;T,x,u)$ satisfies the pricing PDE (2.22) subject to the boundary condition

$$\bar{V}_\delta(T;T,x,u)=\exp(-i\langle u,x\rangle) \tag{2.24}$$

The solution has exponential-affine form (Duffie and Kan (1996))

$$\bar{V}_\delta(t;T,x,u)=e^{C(t;T,\,u)^*x+D(t;T,\,u)}$$

where

$$\frac{dD(t;T,u)}{dt} = g_0-C(t;T,u)^*B-\frac{1}{2}\sum_{k=1}^{n_X}\left[(C(t;T,u)^*\Sigma)_k\right]^2\mathcal{B}_k \tag{2.25}$$

$$\frac{dC(t;T,u)^*}{dt} = g_1^*-C(t;T,u)^*A-\frac{1}{2}\sum_{k=1}^{n_X}\left[(C(t;T,u)^*\Sigma)_k\right]^2\mathcal{A}_{k,\cdot}. \tag{2.26}$$

subject to boundary conditions:

$$C\left(T;T,u\right) = -iu \quad \text{and} \tag{2.27}$$

$$D\left(T;T,u\right) = 0. \tag{2.28}$$

$\mathcal{A}_{k,\cdot}$ denote the k-th row of $\mathcal{A}$. Thus, $V_\delta(t;T,x,y)$ can be obtained via the inverse Fourier transform

$$V_\delta(t;T,x,y) = \frac{1}{(2\pi)^{n_X}} \int_{\Re^{n_X}} \exp\left(i\langle u,y\rangle\right)\ \bar{V}_\delta\left(t;T,x,u\right)\ du$$

Given the price $V_\delta\left(t;T,x,y\right)$ of the Arrow-Debreu security, the solution of the pricing equation (2.22) is given by

$$V\left(t;T,x\right) = \int_{\Re^{n_X}} V_\delta\left(t;T,x,y\right)\ h\left(y\right)\ dy \tag{2.29}$$

2.2.2.1 Derivatives with Exponential Pay-off

For derivatives with an exponential pay-off structure, the current value of the derivatives product can be found without performing a Fourier transformation.

Let $V_a(t;T,X(t),h_0,h_1)$ denote the time t value of a derivative which pays off

$$h_a(X(T)) = \exp(h_0 + h_1^* X(T))$$

at the maturity T. From (2.29), $V_a(t;T,x)$ satisfies:

$$\begin{aligned} V_a\left(t;T,x,h_0,h_1\right) &= \int_{\Re^{n_X}} V^\delta\left(t;T,x,y\right)\ h_a\left(y\right)\ dy \\ &= e^{h_0}\int_{\Re^{n_X}} V^\delta\left(t;T,x,y\right)\ e^{-i\langle(ih_1),y\rangle}\ dy \\ &= e^{h_0}\bar{V}^\delta\left(t;T,x,ih_1\right) \end{aligned}$$

Thus, $V_a\left(t;T,x,h_0,h_1\right)$ can be written as

$$V_a\left(t;T,x,h_0,h_1\right) = \exp\left(C_a\left(t;T,h_0,h_1\right)^* x + D_a\left(t;T,h_0,h_1\right)\right)$$

where

$$\begin{aligned} \frac{dD_a\left(t;T,h_0,h_1\right)}{dt} &= g_0 - C_a\left(t;T,h_0,h_1\right)^* B - \\ &\quad \frac{1}{2}\sum_{k=1}^{n_X}\left[(C_a(t;T,h_0,h_1)^*\Sigma)_k\right]^2 \mathcal{B}_k \end{aligned} \tag{2.30}$$

$$\begin{aligned} \frac{dC_a\left(t;T,h_0,h_1\right)^*}{dt} &= g_1^* - C_a\left(t;T,h_0,h_1\right)^* A - \\ &\quad \frac{1}{2}\sum_{k=1}^{n_X}\left[(C_a(t;T,h_0,h_1)^*\Sigma)_k\right]^2 \mathcal{A}_{k,\cdot} \end{aligned} \tag{2.31}$$

with boundary conditions

$$C_a(T;T) = h_1 \quad \text{and} \tag{2.32}$$
$$D_a(T;T) = h_0. \tag{2.33}$$

Zero Coupon Bonds

A zero coupon bond pays a unit amount of money at the delivery time T. Let $B(t;T,X(t))$ denote the price of a bond at time t. $B(t;T,X(t))$ satisfies:

$$B(t;T,X(t)) = \mathbf{E}^{\mathbb{Q}}\left[\exp\left(-\int_t^T r(s)\,ds\right)h(X(T))\,\middle|\,\mathcal{F}(t)\right]$$

where $h(x)=1$. Thus, $B(t;T,X(t)) = V_a(t,T,X(t),0,\vec{0})$. Sometimes, it is more convenient to express the bond price in terms of the remaining time to maturity $\tau_b(t) = T-t$. The bond price $B(t;t+\tau_b(t),X(t))$ satisfies

$$B(t;t+\tau_b(t),X(t)) = \exp(C_b(\tau_b(t))^* X(t) + D_b(\tau_b(t)))$$

where by (2.31) and (2.30), $C_b(\tau)$ and $D_b(\tau)$ satisfy

$$\frac{dD_b(\tau)}{d\tau} = -g_0 + C_b(\tau)^* B + \frac{1}{2}\sum_{k=1}^{n_X}\left[(C_b(\tau)^*\Sigma)_k\right]^2 \mathcal{B}_k$$
$$\frac{dC_b(\tau)^*}{d\tau} = -g_1^* + C_b(\tau)^* A + \frac{1}{2}\sum_{k=1}^{n_X}\left[(C_b(\tau)^*\Sigma)_k\right]^2 \mathcal{A}_{k,\cdot}$$

with boundary conditions $C_b(0)=\vec{0}$ and $D_b(0)=0$.

Futures on a Zero Coupon Bond

Futures on a zero coupon bond delivers an underlying bond with a constant time to maturity τ_b, at a delivery time T_d. Let $\tau_f(t)$ be the remaining time to delivery of the futures. The futures price $V_f(t;t+\tau_f(t),\tau_b,X(t))$ is given by

$$\begin{aligned} V_f(t;t+\tau_f(t),\tau_b,X(t)) &= \mathbf{E}^{\mathbb{Q}}\left(B(T_d;T_d+\tau_b,X(T_d))\mid\mathcal{F}(t)\right) \\ &= \mathbf{E}^{\mathbb{Q}}\left(e^{C_b(\tau_b)^* X(T_d)+D_b(\tau_b)}\,\middle|\,\mathcal{F}(t)\right) \end{aligned} \tag{2.34}$$

Thus,

$$V_f(t;t+\tau_f(t),\tau_b,X(t)) = e^{C_f(\tau_f(t))X(t)+D_f(\tau_f(t))}$$

where $C_f(\tau)$ and $D_f(\tau)$ satisfy

$$\frac{dD_f(\tau)}{d\tau} = C_f(\tau)^* B + \frac{1}{2}\sum_{k=1}^{n_X}\left[(C_f(\tau)^*\Sigma)_k\right]^2 \mathcal{B}_k \tag{2.35}$$
$$\frac{dC_f(\tau)^*}{d\tau} = C_f(\tau)^* A + \frac{1}{2}\sum_{k=1}^{n_X}\left[(C_f(\tau)^*\Sigma)_k\right]^2 \mathcal{A}_{k,\cdot} \tag{2.36}$$

with boundary conditions $C_f(0) = C_b(\tau_b)$ and $D_f(0) = D_b(\tau_b)$.

2.3 Estimation of Exponential-Affine Term Structure

In implementing a given interest rate model for risk management and pricing purposes, parameters that characterize the model must be estimated from market data. The procedure of estimating model parameters is commonly referred to as *model calibration*. The easiest approach to estimate the parameters is through least squares, by minimizing the sum of squared errors between theoretical prices and their quoted market prices. Derivatives such as bonds, vanilla options on bonds, futures and swaps may be used. The procedure can be easily implemented and repeated whenever new parameter estimates are required. The interest rate factors, that are unobservable, are treated as unknown parameters in the estimation procedure. By doing so, one ignores the statistical characteristics of the underlying stochastic factors. In order to take into account these characteristics, we need to apply more sophisticated statistical techniques such as maximum likelihood estimation.

2.3.1 Observed Interest Rate Factors

Let

$$y(t_k) = H(t_k, X(t_k))$$

be a n_y-dimensional vector observation where $H(t,x) = \vec{C}(t)x + \vec{D}(t)$, $t_k = t_0 + kh$, $k = 1, \ldots, n_T$, h is a fixed step-size, $y(t) \in \Re^{n_y \times 1}$, $\vec{C}(t) \in \Re^{n_y \times n_X}$, $\vec{D}(t) \in \Re^{n_y \times 1}$. The number of observations is assumed to be equal to the number of factors, $n_y = n_X$, and $\vec{C}(t)$ is invertible for all $t \geq 0$.

Maximum likelihood estimation (MLE) methods rely on the computation of the transition density $f_y(y(t_{k+1}) \,|\, y(t_k))$. Let $H^{-1}(t,y) = \vec{C}(t)^{-1}(y - \vec{D}(t))$. The the transition density of y can be determined from the transition density of X:

$$f_{y,\,t_k,\,t_{k+1}}(y(t_{k+1})|\,y(t_k)) = f_{X,\,t_k,\,t_{k+1}}(H^{-1}(y(t_{k+1}))|H^{-1}(y(t_k)))\ \det\left(\frac{\partial H^{-1}(t,y)}{\partial y}\right).$$

Unfortunately, for exponential-affine models, the closed-form density is only available in a few cases; for Gaussian models, or when the interest rate factors are independent square root models. The transition density can always be approximated by numerically solving the forward Kolmogorov equation. However, the curse of dimensionality kicks in when the number of interest rate factor increases. The following methods are relatively simpler to implement and potentially faster than solving the Kolmogorov equation.

2.3.1.1 Euler Approximation

The Euler approximation of the exponential-affine model can be written as

$$X(t_{k+1}) \approx X(t_k) + (A^{\mathbb{P}} X(t_k) + B^{\mathbb{P}})\, h + \Sigma\ \operatorname{diag}(\mathcal{A}X(t_k) + \mathcal{B})^{1/2}\, \sqrt{h}\, \varepsilon_{k+1}$$

where $\{\varepsilon_k,\ k \geq 0\}$ are i.i.d standard Gaussian random variables. Using the above Euler approximation, the conditional density can be approximated as:

$$f_{X,\, t_k,\, t_{k+1}}(X(t_{k+1})|\ X(t_k)) \approx \Phi(\mu_k, \sigma_k)$$

where Φ denotes the multivariate Gaussian density with mean and covariance $\mu_k = X(t_k) + (A^{\mathbb{P}} X(t_k) + B^{\mathbb{P}})\, h$ and $\sigma_k = \Sigma\ \operatorname{diag}(\mathcal{A}X(t_k) + \mathcal{B})\Sigma\ h$ respectively. If the first two moments of the transition density are known exactly, it is better to approximate the transition density with a Gaussian density using these exact moments rather than using moments obtained from the Euler approximation.

2.3.1.2 Monte-Carlo integration

Another way of approximating the transition density $f_{X,\, t_k,\, t_{k+1}}(X(t_{k+1})|\ X(t_k))$ is by a Monte-Carlo integration. This approximation has been used in simulated maximum likelihood estimation methods proposed by Pedersen (1995) and Santa-Clara (2005).

For a small time interval Δ, let $s = t_{k+1} - \Delta > t_k$. We use the Euler approximation to write:

$$X(t_{k+1}) \approx X(s) + (A^{\mathbb{P}} X(s) + B^{\mathbb{P}})\, \Delta + \Sigma\ \operatorname{diag}(\mathcal{A}X(s) + \mathcal{B})^{\frac{1}{2}}\, \sqrt{\Delta}\, \varepsilon_{k+1} \tag{2.37}$$

where $\{\varepsilon_k,\ k \geq 0\}$ are i.i.d standard Gaussian random variables. The transition density can be written as

$$f_{X,\, t_k,\, t_{k+1}}(X(t_{k+1})|\ X(t_k)) = \int_{\Re^{n_X}} f_{X,\, s,\, t_{k+1}}(X(t_{k+1})|\ z)\ f_{X,\, t_k,\, s}(z \mid X(t_k))\, dz$$

With the Euler discretization (2.37), $f_{X,\, s,\, t_{k+1}}(X(t_{k+1})|\ z)$ can be approximated as

$$f_{X,\, s,\, t_{k+1}}(X(t_{k+1})|\ z) \approx \Phi(\mu(z)), \sigma(z))$$

where Φ denotes the multivariate Gaussian density with mean and covariance $\mu(z) = z + (A^{\mathbb{P}} z + B^{\mathbb{P}})\, \Delta$ and $\sigma(z) = \Sigma\ \operatorname{diag}(\mathcal{A}z + \mathcal{B})\Sigma\ \Delta$ respectively. Thus,

$$f_{X,\, t_k,\, t_{k+1}}(X(t_{k+1})|\ X(t_k)) \approx \int_{\Re^{n_X}} \Phi(\mu(z), \sigma(z))\ f_{X,\, t_k,\, s}(z|\ X(t_k))\, dz \tag{2.38}$$

The integral (2.38), can further be approximated using a Monte-Carlo integration. Let $X(s)_{[l]}$, $l = 1, \ldots, N$ be the l-th realization of $X(s)$ which is obtained through a Monte-Carlo simulation starting from $X(t_k)$ at time t_k. Note that simulations of $X(s)$ can be done with Euler or some other higher order discretization of $X(t)$. The integral (2.38) can be approximated as

$$f_{X,\,t_k,\,t_{k+1}}(X(t_{k+1})|\,X(t_k)) \approx \frac{1}{N}\sum_{l=1}^{N}\phi(\mu(X(s)_{[l]}),\sigma(X(s)_{[l]}))$$

2.3.1.3 Fourier inversion

For exponential affine models, the characteristic function can be found by solving a set of ordinary differential equations. By computing the characteristic function at a finite number of points, the conditional density can be approximated by using Fourier transform.

The characteristic function of X can be written as

$$\begin{aligned} F_{X,\,t_k,\,t_{k+1}}(u, X(t_k)) &= \mathbf{E}\left[\exp(i\langle u, X(t_{k+1})\rangle) \mid X(t_k)\right] \\ &= \int_{\Re^{n_X}} \exp\left(i\langle u, z\rangle\right)\; f_{X,\,t_k,\,t_{k+1}}(z|\,X(t_k))\;\; dz \end{aligned}$$

for all $u \in \Re^{n_X \times 1}$ and the imaginary unit i. For exponential-affine models,

$$F_{X,\,t_k,\,t_{k+1}}(u, X(t_k)) = \exp\left(c(h;u)^* X(t_k) + d(h;u)\right)$$

where $c(\tau;u)$ and $d(\tau;u)$ satisfies:

$$\begin{aligned} \frac{d}{d\tau}c(\tau;u)^* &= c(\tau;u)^* A^{\mathbb{P}} + \frac{1}{2}\sum_{k=1}^{n_X}(c(\tau,u)^*\Sigma)_k^2\mathcal{A}_{k,\cdot} \\ \frac{d}{d\tau}d(\tau;u) &= c(\tau;u)^* B^{\mathbb{P}} + \frac{1}{2}\sum_{k=1}^{n_X}(c(\tau,u)^*\Sigma)_k^2\mathcal{B}_k \end{aligned}$$

with boundary conditions $c(0;u) = iu$ and $d(0;u) = 0$. The conditional density can be computed by

$$f_{X,\,t_k,\,t_{k+1}}(X(t_{k+1}))|\,X(t_k)) = \frac{1}{(2\pi)^{n_X}}\int_{\Re^{n_X}} \text{Re}\left(e^{-i\langle u,\,X(t_{k+1})\rangle}\; F_{X,\,t_k,\,t_{k+1}}(u, X(t_k))\right)\; du.$$

This inversion can be implemented using fast Fourier transform (FFT).

2.3.2 Unobserved Interest Rate Factors

Unfortunately, the number of observations are generally larger than the number of factors of the interest rate, $n_y > n_X$. In this case, one may assume that n_X observations are observed without noise while the remaining are corrupted with additive Gaussian noise. However, this is difficult to implement in practice because there is no clear guideline on which part of the term structure can be assumed to be perfectly observed.

Alternatively, one may assume that the whole term structure is observed with noise. The assumption is not unrealistic, considering that term structures are mostly implied from derivative securities for which the computation procedure may involve some sort of interpolation technique. Furthermore, for parameter estimation a unique price has to be determined from bid-ask prices of the derivatives. These will add errors to the observations. Adding errors to all observations means that we will not be able to invert the interest rate factors from observed data. This is where filtering become useful.

Filtering is particularly attractive to estimate parameters of the exponential-affine interest rate models, due to the fact that bond yields can be expressed as linear functions of the latent factors. This method have been used by Pennacchi (1991), Chen and Scott (1993), Duan and Simonato (1999), Lund (1997), De Jong (2000), Geyer and Pichler (1999), Jegadeesh and Pennacchi (1996) and others. In order to illustrate the method, we will outline the approach given in Duan and Simonato. First, the Euler approximation of the interest rate factors can be written as

$$X(t_{k+1}) = X(t_k) + (A^{\mathbb{P}} X(t_k) + B^{\mathbb{P}})\, h + \Sigma\ \mathrm{diag}(\mathcal{A}X(t_k) + \mathcal{B})^{\frac{1}{2}}\ \sqrt{h}\, \varepsilon_{k+1}$$

and the vector of observations is denoted by

$$y(t_{k+1}) = \vec{C}(t_{k+1}) X(t_{k+1}) + \vec{D}(t_{k+1}) + \tilde{\varepsilon}_{k+1}$$

where $\{\varepsilon_k, k \geq 0\}$ and $\{\tilde{\varepsilon}_k, k \geq 0\}$ are mutually independent standard Gaussian random variables. For linear Gaussian models,

$$I(t_k) = y(t_k) - \left\{ \vec{C}(t_k) \mathbf{E}\left[X(t_k) \mid y(t_j), j = 0, \ldots, k-1\right] + \vec{D}(t_k) \right\}$$

are independent Gaussian random variables, which means that obtaining the estimate of the unknown parameter Θ involves the maximization of the log-likelihood

$$L(\Theta) = -\frac{1}{2} \left\{ \sum_{k=0}^{n_T} I(t_k) R(t_k)^{-1} I(t_k) + \log\det\left(R(t_k)\right) \right\} \tag{2.39}$$

The stochastic process $\{I(t_k), k \geq 0\}$ is called *innovation* and $R(t_k)$ denote the covariance of $I(t_k)$, both of which can be readily obtained using the Kalman filtering algorithm.

In the general nonlinear case, the innovations are not necessarily Gaussian. However, in all literature mentioned above, the innovations are assumed to be Gaussian and parameter estimate are obtained by maximizing the (quasi) likelihood (2.39).

A Continuous-time Maximum Likelihood Estimation

3

3.1 Introduction

As we have mentioned in the previous chapter, one characteristic of the estimation methods existing in the literature is that they invariably discretize the observation equation and/or the interest rate (state) equation. The discretization makes it easier to use standard statistical techniques.

In this chapter, we use the continuous-time maximum likelihood method for the parameter estimation of interest rate models. We restrict our analysis to the exponential-affine term structure model in order to exploit the computational advantages offered by this particular class of models. The estimation method proposed in this chapter rests on the assumption that the observation (e.g. bond yields) are contaminated with additive Gaussian noise with a known covariance. This particular assumption placed our method in a similar line of approaches proposed by Pennachi (1991) and others. However, unlike these approaches, we will not discretize either the measurement or the interest rate model.

3.2 Setup and Notation

Let us start by modeling interest rate factors $X_k(t)$, $k = 0, \ldots, n_X$ and $t \geq 0$ as exponential-affine. The interest rate factors satisfy

$$dX(t) = (AX(t) + B)\, dt + \Sigma\, \mathrm{diag}(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}} dW^{\mathbb{Q}}(t) \tag{3.1}$$

where $X(t) = [X_1(t), \ldots, X_{n_X}(t)]^*$. The vector $W^{\mathbb{Q}}(t) = [W_1^{\mathbb{Q}}(t), \ldots, W_{n_X}^{\mathbb{Q}}(t)]^*$ is composed of independent $\mathbb{Q}$-Brownian motions. The superscript '*' denotes matrix transpose. Under the real-world measure $\mathbb{P}$, the interest rate factors satisfy

$$dX(t) = (A^{\mathbb{P}}X(t) + B^{\mathbb{P}})\, dt + \Sigma\, \mathrm{diag}(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}} dW^{\mathbb{P}}(t) \tag{3.2}$$

The matrices $A,\ A^{\mathbb{P}},\ \mathcal{A} \in \Re^{n_X \times n_X}$, $B,\ B^{\mathbb{P}},\ \mathcal{B} \in \Re^{n_X \times 1}$ and $\Sigma \in \Re^{n_X \times n_X}$ are constant matrices. The interest rate is defined as an affine function of the interest rate factors. i.e.

$$r(t) = g_0 + g_1^* X(t), \tag{3.3}$$

where $g_0 \in \Re$ and $g_1 \in \Re^{n_X \times 1}$ are constants.

Given the interest rate model (3.1), the corresponding yield at time t, of a zero coupon bond with maturity $t + \tau(t)$, can be written as an affine function of the interest rate factors $X(t)$. That is,

$$y(t, t + \tau(t)) = -\frac{1}{\tau} \left(C_b(\tau(t))^* X(t) + D_b(\tau(t)) \right) \tag{3.4}$$

The functions $C_b(\tau) \in \Re^{n_X \times 1}$ and $D_b(\tau) \in \Re$ satisfy the following systems of ordinary differential equations:

$$\frac{d}{d\tau} C_b(\tau) = -g_1 + A^* C_b(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_b(\tau)^* \Sigma)_k \right]^2 \mathcal{A}^*_{k,\cdot} \tag{3.5}$$

$$\frac{d}{d\tau} D_b(\tau) = -g_0 + B^* C_b(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_b(\tau)^* \Sigma)_k \right]^2 \mathcal{B}_k \tag{3.6}$$

with boundary conditions $C(0) = \vec{0}$ and $D(0) = 0$. $\mathcal{A}_{k,\cdot}$ and $\mathcal{B}_k$ denote the k-th row of $\mathcal{A}$ and $\mathcal{B}$ respectively.

Let

$$\vec{y}(t) = \begin{bmatrix} y(t, t + \tau_1(t)) \\ \vdots \\ y(t, t + \tau_{n_y}(t)) \end{bmatrix}$$

be the vector of n_y bond yields at time t with various time to maturities $\tau_k(t)$, $k = 1, \ldots, n_y$. By (3.4), we can write

$$\vec{y}(t) = \vec{C}(t) X(t) + \vec{D}(t) \tag{3.7}$$

where

$$\vec{C}(t) = \begin{bmatrix} -\frac{1}{\tau_1(t)} C_b(\tau_1(t))^* \\ \vdots \\ -\frac{1}{\tau_{n_y}(t)} C_b(\tau_{n_y}(t))^* \end{bmatrix}, \ \vec{D}(t) = \begin{bmatrix} -\frac{1}{\tau_1(t)} D_b(\tau_1(t)) \\ \vdots \\ -\frac{1}{\tau_{n_y}(t)} D_b(\tau_{n_y}(t)) \end{bmatrix}.$$

Within the exponential-affine framework, log prices of derivatives such as futures on the interest rate, futures on bond, forwards, and many more, with varying delivery dates can be written as affine functions similar to (3.7). For the sake of clarity, in this chapter we will concentrate only on vector observations involving yields with varying maturities. The estimation method described in this chapter can be easily modified to include other derivatives as long as we are able to transform their prices to an affine function of the interest rate factors.

In continuous-time, we will consider the integrated measurements

$$Y(t) = \int_0^t \vec{y}^{obs}(s)\, ds$$

where $\vec{y}^{obs}(t)$ is the observed value of $\vec{y}(t)$. $Y(t)$ satisfies the following SDE:

$$dY(t) = \left(\vec{C}(t)X(t) + \vec{D}(t)\right) dt + dW_0(t) \tag{3.8}$$

$W_0(t) \in \Re^{n_y \times 1}$ is a vector of Brownian motions, independent to $W^{\mathbb{P}}(t)$, with a known covariance matrix $R \in \Re^{n_y \times n_y}$.

3.3 The Likelihood Functional

Let $\mathcal{Y}(t) = \sigma\{Y(s),\ 0 \le s \le t\}$. Suppose the set of observations $\{Y(s),\ 0 \le s \le T_y\}$ is available for parameter estimation. By the assumption that $Y(t)$ satisfies the measurement equation (3.8) and the observed interest rate $X(t)$ satisfies the state equation (3.2), the log of the likelihood functional is given by (see e.g. Poor (1994))

$$\mathbf{L}(\Theta) = R^{-1} \int_0^{T_y} \left\{ \langle \overline{H}(X)(t),\ dY(t) \rangle - \frac{1}{2} \|\overline{H}(X)(t)\|^2 dt \right\}\ , \tag{3.9}$$

where Θ denotes the vector containing all unknown parameters that describe the dynamics of the state $X(t)$ (both under the risk neutral measure $\mathbb{Q}$ and the real-world measure $\mathbb{P}$) and the measurement $Y(t)$. The bracket term denotes an inner product $< a, b > \triangleq \mathrm{Tr}(a^* b)$ and $\|a\|^2 \triangleq < a, a >$. The time varying variable $\overline{H}(X)(t)$ is defined as the estimate of $H(t, X(t)) = \vec{C}(t)X(t) + \vec{D}(t)$ given $\mathcal{Y}(t)$, i.e.

$$\begin{aligned}
\overline{H}(X)(t) &= \mathbf{E}\left[H(t, X(t)) \mid \mathcal{Y}(t)\right] \\
&= \mathbf{E}\left[\vec{C}(t)\, X(t) + \vec{D}(t) \,\middle|\, \mathcal{Y}(t)\right] \\
&= \vec{C}(t)\, \mathbf{E}\left[X(t) \mid \mathcal{Y}(t)\right] + \vec{D}(t) \\
&= \vec{C}(t)\, \overline{X}(t) + \vec{D}(t)
\end{aligned}$$

where

$$\overline{X}(t) = \mathbf{E}\left[X(t) \mid \mathcal{Y}(t)\right].$$

The estimate of the unknown vector Θ can be found by maximizing the log likelihood functional (3.9). That is,

$$\hat{\Theta} = \operatorname*{argmax}_{\Theta} \mathbf{L}(\Theta).$$

Unfortunately, with the setup given above, the observation process $\{Y(t), t \in [0, T_y]\}$ which is the solution of (3.8) is only an idealization. The real data that one observes in practice cannot be non differentiable everywhere, as the measurement model (3.8) implies. Every measurement instrument acts as a bandpass filter to the original idealized data $\left\{\dot{Y}(t), t \in [0, T_y]\right\}$. Thus the data is always absolutely continuous and in the Wiener space, the set of absolutely continuous functions has measure zero. This means the likelihood functional (3.9) cannot be applied directly to real data.

One way to circumvent the difficulty, due to Balakrishnan (1977), is to try to model the observation process directly with the white noise output. Although modeling white noises directly is intuitively appealing, it brings a host of mathematical complications. This is because, in this theory, one has to work with finitely additive measures on an appropriate Hilbert space and these measures cannot be extended to countably additive measures on that space. Kallianpur and Karandikar (1988) developed a comprehensive theory of nonlinear estimation in this framework. The remarkable advantage of this theory is that, once the mathematical difficulties are resolved, the results obtained are always in the form where real data can be directly used.

Under the white noise framework, Balakrishnan (1977) showed that a correction term must be added to the likelihood functional. The corrected likelihood is given by:

$$\begin{aligned} \mathbb{L}(\Theta) &= R^{-1} \int_0^{T_y} \left\{ \langle \overline{H}(X)(t),\ \vec{y}^{obs}(t) \rangle - \frac{1}{2} \|\overline{H}(X)(t)\|^2 \right\} dt - \\ & R^{-1} \frac{1}{2} \int_0^{T_y} \tilde{P}(t)\, dt \end{aligned} \tag{3.10}$$

where

$$\begin{aligned} \tilde{P}(t) &= \mathbf{E}\left[\|H(X(t)) - \overline{H}(X)(t)\|^2 \,\middle|\, \mathcal{Y}(t) \right]. \\ &= \vec{C}(t) \mathbf{E}\left[\|X(t) - \overline{X}(t)\|^2 \,\middle|\, \mathcal{Y}(t) \right] \vec{C}(t)^*. \end{aligned}$$

Interestingly, the above likelihood functional is the exact limit of the corresponding result of Wong and Zakai for the band-limited processes, as the bandwidth increases without bound. The correction term $-R^{-1} \frac{1}{2} \int_0^T \tilde{P}(t)\, dt$ is sometimes referred to as the *Wong-Zakai correction term*. A review of theoretical results in the white noise framework can be found in (Bagchi (1994)).

For later comparison, we will refer the likelihood functional (3.10) as the robust likelihood, while the following likelihood

$$\mathbf{L}(\Theta) = R^{-1} \int_0^{T_y} \left\{ \langle \overline{H}(X)(t),\ \vec{y}^{obs}(t) \rangle - \frac{1}{2} \|\overline{H}(X)(t)\|^2 \right\} dt \tag{3.11}$$

will be referred to as the Itô likelihood.

3.4 Nonlinear Filtering of Interest Rate Factors

Maximizing the log-likelihood functional is a straightforward exercise, and it can be implemented quickly with many available optimization packages. What is not so clear at this point, is the computation of the conditional expectation $\overline{X}(t)$ and moreover, whether it is possible to compute $\overline{X}(t)$ recursively given an increasing amount of observations through time. Fortunately, the solution to this problem is well-known and recursive solutions have been developed for a large class of models. In control literature, the procedure to estimate an unobserved stochastic variable from a set of observations up to the current time instant is known as *filtering*. In what follows, we are going to highlight some equations for nonlinear filtering that are relevant to our estimation problem. We note that these equations are widely available in standard nonlinear filtering textbooks.

3.4.1 Equations for Nonlinear Filtering

Let us consider the conditional moments of $X(t)$ given the set of observations $\mathcal{Y}(t)$,

$$\overline{\phi}(X)(t) = \mathbf{E}\left[\phi(t, X(t)) \mid \mathcal{Y}(t)\right] \tag{3.12}$$

and denote

$$H(t, x) = \vec{C}(t)x + \vec{D}(t)$$

such that the drift term of the measurement equation (3.8) is $H(t, X(t))$. With the assumption that $W(t)$ and $W_0(t)$ are independent, $\overline{\phi}(X)(t)$ satisfies the *Kushner-Stratonovich* equation (Kushner (1967), Stratonovich (1968)) :

$$\begin{aligned}\overline{\phi}(X)(t) \quad &= \quad \overline{\phi}(X)(0) + \int_0^t \overline{\mathcal{L}^*\phi}(X)(s)\, ds \\ &\quad + \int_0^t \left(\overline{\phi H^*}(X)(s) - \overline{\phi}(X)(s)\, \overline{H^*}(X)(s)\right)\, R^{-1} dI(s)\end{aligned} \tag{3.13}$$

where $\mathcal{L}^*$ is the backward Kolmogorov operator

$$\begin{aligned}\mathcal{L}^* f(t, X) \quad &= \quad -\sum_{k=1}^{n_X} \frac{\partial}{\partial X_k}\left\{(A^{\mathbb{P}}X + B^{\mathbb{P}})_k f(t, X)\right\} + \\ &\quad \frac{1}{2}\sum_{k=1}^{n_X}\sum_{l=1}^{n_X} \frac{\partial^2}{\partial X_k \partial X_l}\left\{\mathrm{diag}(\mathcal{A}X + \mathcal{B})_{k,l} f(t, X)\right\}\end{aligned} \tag{3.14}$$

and $I(t) = Y(t) - \overline{H}(X)(t)$. The stochastic process $I(t)$ is called the *innovation process*.

Algorithm 1 Kalman filtering

For $\mathcal{A} = \mathbf{0}_{n_X \times n_X}$ and $\mathcal{B} = \mathbf{1}_{n_X \times 1}$, the filtered estimate $\hat{X}(t)$ is given by the following Kalman filtering equations:

$$\begin{aligned} K(t) &= P(t)\vec{C}(t)^* & (3.16) \\ d\overline{X}(t) &= (A^{\mathbb{P}}\overline{X}(t) + B^{\mathbb{P}})dt + \\ &\quad K(t)R^{-1}(dY(t) - (\vec{C}(t)\overline{X}(t) + \vec{D}(t))dt) & (3.17) \\ \frac{d}{dt}P(t) &= A^{\mathbb{P}}P(t) + PA^{\mathbb{P}*} - K(t)R^{-1}K(t)^* + \Sigma\Sigma^* & (3.18) \end{aligned}$$

For $\phi(t, X) = X$, The filtering equation (3.13) gives:

$$\begin{aligned} \overline{X}(t) &= \overline{X}(0) + \int_0^t A^{\mathbb{P}}\overline{X}(s) + B^{\mathbb{P}} \, ds \\ &\quad + \int_0^t \left(\overline{XH^*}(X)(s) - \overline{X}(s)\,\overline{H}(X)(s)^*\right) \, R^{-1}dI(s) \end{aligned} \quad (3.15)$$

where $\overline{XH^*}(X)(s)$ will depend on $\overline{XH^*H^*}(X)(s)$. Similarly, $\overline{XH^*H^*}(X)(s)$ will depend on $\overline{XH^*H^*H^*}(X)(s)$ and so forth. Thus, in general we may not be able to compute the filtered estimate $\overline{X}(t)$ using a finite number of equations. One special case where $\overline{X}(t)$ can be computed by a finite number of recursive equations is when $X(t)$ is Gaussian. The recursive filtering procedure for the Gaussian case is given by the celebrated *Kalman filtering* algorithm.

Another way to compute the conditional moments of $X(t)$ is to directly compute the conditional density. Suppose that

$$\overline{\phi}(X)(t) = \int_{\Re^{n_X}} \phi\,(t, X) \; p\,(t, X) \; dX,$$

where $p\,(t, X)$ denotes the conditional density of $X(t)$ given the σ-algebra $\mathcal{Y}(t)$. The filtering equation (3.13) implies that the conditional density $p\,(t, X)$ satisfies the following stochastic integro-differential equation

$$\begin{aligned} dp\,(t, X) &= \mathcal{L}^* p\,(t, X) \; dt + \\ &\quad p(t, X)\left(H\,(t, X) - \int_{\Re^{n_X}} H\,(t, X)\, p\,(t, X) \; dX\right) R^{-1}dI(t). \end{aligned} \quad (3.19)$$

In principle, once the conditional density $p\,(t, X)$ has been found, all conditional moments of $X(t)$ given $\mathcal{Y}(t)$ can be computed. However, solving (3.19) is very difficult due to the integral term appearing on the right hand side of the equation. Very often, it is more convenient to compute the unnormalized conditional density $q(t, X)$ which is defined as

$$p(t, X) = \left(\int_{\Re^{n_X}} q(t, X) \; dX \right)^{-1} q(t, X).$$

The unnormalized conditional density satisfies the Zakai equation (Zakai (1969)):

$$dq(t, X) = \mathcal{L}^* q(t, X) \; dt + \; q(t, X) R^{-1} \langle H(t, X), \; dY(t) \rangle. \tag{3.20}$$

Observe that the integral term in (3.19) is completely absent in the Zakai equation (3.20). This simplifies the computation of $p(t, X)$ considerably. Given $q(t, X)$, the conditional mean and conditional covariance of the interest rate factors are given by:

$$\overline{X}(t) = \mathbf{E}[X(t) \mid \mathcal{Y}(t)] = \frac{\int_{\Re^{n_X}} X \, q(t, X) \; dX}{\int_{\Re^{n_X}} q(t, X) \; dX} \tag{3.21}$$

$$P(t) = \mathbf{E}\left[\|X(t) - \overline{X}(t)\|^2 \mid \mathcal{Y}(t) \right] = \frac{\int_{\Re^{n_X}} \|X - \overline{X}(t)\|^2 \, q(t, X) \; dX}{\int_{\Re^{n_X}} q(t, X) \; dX} \tag{3.22}$$

3.4.2 Numerical Approximation of the Zakai Equation

Although the Zakai equation is relatively easier to solve compared to the Kushner-Stratonovich equation (3.19), in general, a closed-form solution is still difficult to obtain. Thus, we often need to apply numerical methods to approximate the solution of (3.20). For our estimation procedure, we employ the splitting-up method (Le Gland (1989)). This method splits the differential operator (3.20) into two parts, a deterministic differential operator and a stochastic differential operator, and solve them separately.

To illustrate the algorithm, let us introduce the following time discretization on the interval $[0, T_y]$

$$0 = t_0 < t_1 = t_0 + \Delta < \ldots < t_k = t_{k-1} + \Delta < \ldots < t_N = T_y. \tag{3.23}$$

Following LeGland (1989), let $q(t_k, X)$ be approximated by $q^{\Delta}(t_k, X)$. The transition from $q^{\Delta}(t_k, X)$ to $q^{\Delta}(t_k, X)$ can be summarized into the following two steps:

1a. The *prediction step*. Solve

$$\frac{\partial}{\partial t} f(t, X) = \mathcal{L}^* f(t, X) \tag{3.24}$$

between times t_k and t_{k+1} with the initial condition $f(t_k, X) = q^{\Delta}(t_k, X)$ to obtain $f(t_{k+1}, X)$.

2a. The *correction step*. Update $f(t_{k+1}, X)$ from the prediction step by solving:

$$dg(t, X) = g(t, X) R^{-1} \langle H(t, X), dY(t) \rangle \tag{3.25}$$

with the initial condition $g(t_k, X) = f(t_{k+1}, X)$. Furthermore, we assign

$$q^{\Delta}(t_{k+1}, X) = g(t_{k+1}, X). \tag{3.26}$$

The updating equation (3.25) can be explicitly solved, namely:

$$g(t_{k+1}, X) = f(t_{k+1}, X) \; \Phi^{\Delta}(t_{k+1}, X) \tag{3.27}$$

where

$$\log \Phi^{\Delta}(t, X) \;\; = \;\; R^{-1} \int_{t_k}^{t_{k+1}} \left(\langle H(t, X), \; dY(t) \rangle - \frac{1}{2} \left\| H(t, X) \right\|^2 dt \right) \tag{3.28}$$

It was shown by Bensoussan and Glowinski (1989) that when $\Delta \longrightarrow 0$, $q^{\Delta}(t_k, X)$ converges strongly to $q(t_k, X)$.

In practice, the above algorithm is very difficult to implement because the resulting numerical solution often dissipates quickly. A common work-around to this problem is to subsequently normalize the updating equation (3.27) by using the following step in place of step 2a of the above algorithm.

2b. Update $f(t_{k+1}, X)$ from the prediction step by computing

$$q^{\Delta}(t_{k+1}, X) = c_{k+1} \; f(t_{k+1}, X) \;\; \Phi^{\Delta}(t_{k+1}, X) \tag{3.29}$$

where c_{k+1} is a normalizing factor so that $c_{k+1} \int_{\Re^{n_X}} q^{\Delta}(t_{k+1}, X) \; dX = 1$.

In the implementation of the splitting-up algorithm, another approximation is needed in the computation of $\Phi_{k+1}^{\Delta}(t, X)$ as it contains a stochastic integral term involving the measurement $Y(t)$. Here we can use the following approximation for the stochastic integral term

$$\int_{t_k}^{t_{k+1}} \langle H(t, X), \; dY(t) \rangle \approx \int_{t_k}^{t_{k+1}} \langle H(t, X), \; \vec{y}^{obs}(t) \rangle dt$$

(Balakrishnan (1981)).

3.4.3 Gaussian Approximation

The numerical approximation of the conditional density via the splitting-up algorithm can be made to be highly accurate. However, for a multi-factor exponential-affine model, the algorithm is very time consuming. We note that the computation speed may still be acceptable given an identified system. However, for parameter estimation that may require many iterations in the optimization of the likelihood functional, the splitting-up method becomes impractical. In order to speed up the computation, we may need to sacrifice accuracy and use some sort of approximate filter. Here we propose an approximate filter based on Gaussian approximation of the conditional density.

Let us denote

$$\begin{aligned} F(X) &= A^{\mathbb{P}} X + B^{\mathbb{P}} \\ G(X) &= \Sigma \, \mathrm{diag}(\mathcal{A}X + \mathcal{B})^{\frac{1}{2}} \\ H(t, X) &= \vec{C}(t)X + \vec{D}(t) \end{aligned}$$

where $F(X(t))$ and $G(X(t))$ are the drift and volatility term of the interest rate factor respectively, and $H(t, X(t))$ is the drift term of the measurement equation. In general, we can express the nonlinear filtering solution as (see e.g. Poor (1994))

$$\begin{aligned} d\overline{X}(t) &= \overline{F}(X)(t)\, dt + \\ & \quad K(t)R^{-1}\left(dY(t) - \overline{H}(X)(t)\, dt\right) \end{aligned} \tag{3.30}$$

where

$$K(t) = \left(\overline{X\, H^*}(X)(t) - \overline{X}(t)\overline{H^*}(X)(t)\right). \tag{3.31}$$

By substituting $H(X)$ into $K(t)$, we obtain

$$K(t) = P(t)\vec{C}(t)^*$$

where the conditional covariance matrix $P(t) = \mathbf{E}\left[\|X(t) - \overline{X}(t)\|^2 \;\middle|\; \mathcal{Y}(t)\right]$ satisfies

$$\begin{aligned} dP(t) &= \overline{\left(X - \overline{X}(t)\right)\, F^*}(X)(t)\, dt + \overline{F\left(X - \overline{X}(t)\right)^*}(X)(t)\, dt + \\ & \quad \overline{GG^*}(X)(t)dt - K(t)R^{-1}K(t)^* dt + \\ & \quad \left(\overline{\left(X - \overline{X}(t)\right)\left(X - \overline{X}(t)\right)^* H}(X)(t) - P(t)\overline{H}(X)(t)\right)^* \times \\ & \quad R^{-1}\left(dY(t) - \overline{H}(X)(t)\, dt\right). \end{aligned} \tag{3.32}$$

We can further simplify (3.30) and (3.32) to

$$\begin{aligned} d\overline{X}(t) &= (A^{\mathbb{P}}\overline{X}(t) + B^{\mathbb{P}})\, dt + \\ & \quad K(t)R^{-1}\left(dY(t) - (\vec{C}(t)\overline{X}(t) + \vec{D}(t))\, dt\right) \end{aligned} \tag{3.33}$$

and

$$\begin{aligned} dP(t) &= \left(A^{\mathbb{P}}P(t) + P(t)A^{\mathbb{P}*}\right) dt + \\ & \quad \left(\Sigma \, \mathrm{diag}(\mathcal{A}\overline{X}(t) + \mathcal{B})\Sigma^* - K(t)\left(\Sigma_0\Sigma_0^*\right)^{-1} K(t)^*\right) dt + \\ & \quad \left(\left(\overline{\left(X - \overline{X}(t)\right)\left(X - \overline{X}(t)\right)^* H}\right)(X)(t) - P(t)\overline{H}(X)(t)\right)^* \times \\ & \quad R^{-1}\left(dY(t) - (\vec{C}(t)\overline{X}(t) + \vec{D}(t))dt\right) \end{aligned} \tag{3.34}$$

Suppose that at each time instant t we approximate the true conditional density with a Gaussian density with mean $\overline{X}_g(t)$ and covariance $P_g(t)$, i.e.

$$p(t,X) \approx \Phi(t,X;\overline{X}_g(t),P_g(t)). \tag{3.35}$$

Let us denote for any function $\phi : [0,T_y] \times \Re^{n_X} \mapsto \Re$

$$\overline{\phi}_g(X)(t) \triangleq \int_{\Re^{n_X}} \phi\,(t,X)\,\Phi(t,X;\overline{X}_g(t),P_g(t))dX. \tag{3.36}$$

Then,

$$\begin{aligned} d\overline{X}_g(t) &= (A^{\mathbb{P}}\overline{X}_g(t) + B^{\mathbb{P}})\,dt + \\ &\quad K_g(t)R^{-1}\left(dY(t) - (\vec{C}(t)\overline{X}_g(t) + \vec{D}(t))\,dt\right) \end{aligned} \tag{3.37}$$

and

$$\begin{aligned} dP_g(t) &= \left(A^{\mathbb{P}}P_g(t)\; + P_g(t)A^{\mathbb{P}*} - K_g(t)R^{-1}K_g(t)^*\right)dt + \\ &\quad dI_P(t) + \\ &\quad \left(\Sigma\,\mathrm{diag}(\mathcal{A}\overline{X}_g(t) + \mathcal{B})\Sigma^*\right)dt \end{aligned} \tag{3.38}$$

where

$$\begin{aligned} dI_P(t) &= \left(\overline{\left(X - \overline{X}_g(t)\right)\left(X - \overline{X}_g(t)\right)^* H}(X)(t) - P_g(t)\overline{H}_g(X)(t)\right)^* \times \\ &\quad R^{-1}\left(dY(t) - (\vec{C}(t)\overline{X}_g(t) + \vec{D}(t))dt\right) \end{aligned} \tag{3.39}$$

and

$$K_g(t) = P_g(t)\vec{C}^*, \tag{3.40}$$

The above equations are similar to the continuous-time Kalman filtering equation, except for the time-varying term $\left(\Sigma\,\mathrm{diag}(\mathcal{A}\overline{X}_g(t) + \mathcal{B})\Sigma^*\right)$ appearing in (3.38). Due to the Gaussian assumption (3.35) and by substituting $H(x) = \vec{C}(t)x + \vec{D}(t)$ into (3.38), we may follow the lines of proof in (Bensoussan (1992), pp. 104), to simplify the covariance equation (3.38) to

$$\begin{aligned} \frac{d}{dt}P_g(t) &= A^{\mathbb{P}}P_g(t)\; + P_g(t)A^{\mathbb{P}*} - K_g(t)R^{-1}K_g(t)^* + \\ &\quad \Sigma\,\mathrm{diag}(\mathcal{A}\overline{X}_g(t) + \mathcal{B})\Sigma^*. \end{aligned} \tag{3.41}$$

Note that, unlike the Kalman filtering algorithm, equations for the filtered estimate $\overline{X}_g(t)$ and the conditional covariance $P_g(t)$ are coupled.

3.4.4 Numerical Approximation of the Gaussian Filtering

By approximating the stochastic integral (3.37) we obtain systems of ordinary differential equations for the filtered estimate $\overline{X}_g(t)$.

$$\begin{aligned}\frac{d}{dt}\overline{X}_g(t) &= (A^{\mathbb{P}}\overline{X}_g(t) + B^{\mathbb{P}}) + \\ &\quad K_g(t)R^{-1}\left(\vec{y}^{obs}(t) - (\vec{C}(t)\overline{X}_g(t) + \vec{D}(t))\right)\end{aligned} \tag{3.42}$$

and

$$\begin{aligned}\frac{d}{dt}P_g(t) &= A^{\mathbb{P}}P_g(t) + P_g(t)A^{\mathbb{P}*} - K_g(t)R^{-1}K_g(t)^* + \\ &\quad \Sigma\,\mathrm{diag}(\mathcal{A}\overline{X}_g(t) + \mathcal{B})\Sigma^*.\end{aligned} \tag{3.43}$$

In principle the above ordinary differential equations could be approximated accurately using numerical methods such as the Runge-Kutta iteration. However, in the implementation of the filtering equations, we found that it is very difficult to obtain stable numerical solutions when the noise covariance is small with standard numerical integration packages, including the more specialized stiff ODE solvers*. Unfortunately, for financial data, the magnitude of the noise is generally very small. For this reason, we developed a specialized numerical scheme based on backward differences to solve (3.42) and (3.43). In experiments, we found the numerical scheme to be very stable. The algorithm can be described as follows.

Let $t_k = k\Delta$, $k = 0, 1, \ldots$. be discrete time points. Using backward differences, we approximate (3.42) and (3.43) by

$$\begin{aligned}\overline{X}_g^{\Delta}(t_{k+1}) &= \overline{X}_g^{\Delta}(t_k) + (A^{\mathbb{P}}\overline{X}_g^{\Delta}(t_{k+1}) + B^{\mathbb{P}})\,\Delta + \\ &\quad K_g^{\Delta}(t_{k+1})R^{-1}\left(\vec{y}^{obs}(t_{k+1}) - (\vec{C}(t_{k+1})\overline{X}_g^{\Delta}(t_{k+1}) + \vec{D}(t_{k+1}))\right)\Delta\end{aligned} \tag{3.44}$$

and

$$\begin{aligned}P_g^{\Delta}(t_{k+1}) &= P_g^{\Delta}(t_k) + \\ &\quad \Delta\left(A^{\mathbb{P}}P_g^{\Delta}(t_{k+1}) + P_g^{\Delta}(t_{k+1})A^{\mathbb{P}*} - K_g^{\Delta}(t_{k+1})R^{-1}K_g^{\Delta}(t_{k+1})^*\right) + \\ &\quad \Delta\Sigma\,\mathrm{diag}(\mathcal{A}\overline{X}_g^{\Delta}(t_{k+1}) + \mathcal{B})\Sigma^*\end{aligned} \tag{3.45}$$

*This is also true for the simpler Gaussian models. However, for Gaussian models one may still be able to apply steady-state filter as an approximation.

where

$$K_g^{\Delta}(t) = P_g^{\Delta}(t)\vec{C}(t)^*. \tag{3.46}$$

The difference equation (3.44) can be simplified to

$$L_{hs}\overline{X}_g^{\Delta}(t_{k+1}) = R_{hs} \tag{3.47}$$

where

$$L_{hs} = \mathbf{I}_{n_X \times n_X} - A^{\mathbb{P}}\Delta - P_g^{\Delta}(t_{k+1})\vec{C}(t_{k+1})^* R^{-1}\vec{C}(t_{k+1})\Delta,$$

and

$$R_{hs} = \overline{X}_g^{\Delta}(t_k) + B^{\mathbb{P}}\Delta + P_g^{\Delta}(t_{k+1})\vec{C}(t_{k+1})^* R^{-1}\left(\vec{y}^{obs}(t_{k+1}) - \vec{D}(t_{k+1})\right)\Delta$$

The covariance term can be further simplified to

$$\begin{aligned} 0 &= -\left(P_g^{\Delta}(t_k) + \Delta\Sigma\,\mathrm{diag}(\mathcal{A}\overline{X}_g^{\Delta}(t_{k+1}) + \mathcal{B})\Sigma^*\right) + \\ & \quad I_A P_g^{\Delta}(t_{k+1}) + P_g^{\Delta}(t_{k+1})I_A^* + P_g^{\Delta}(t_{k+1})C_{\Delta}^* R^{-1} C_{\Delta} P_g^{\Delta}(t_{k+1}) \end{aligned} \tag{3.48}$$

where

$$I_A = \frac{1}{2}\mathbf{I}_{n_X \times n_X} - A\Delta \text{ and}$$
$$C_{\Delta} = \vec{C}(t_{k+1})\sqrt{\Delta}.$$

Given $\overline{X}_g^{\Delta}(t_k)$ and $P_g^{\Delta}(t_k)$, we will solve for $\overline{X}_g^{\Delta}(t_{k+1})$ and $P_g^{\Delta}(t_{k+1})$ using a fixed-point iteration. We outline the Gaussian filtering algorithm in Algorithm (2).

Algorithm 2 Gaussian filtering

Given $\overline{X}_g^{\Delta}(0)$ and $P_g^{\Delta}(0)$, define discrete time steps $0 = t_0 < t_1 = t_0 + \Delta < \ldots < t_k = t_{k-1} + \Delta < \ldots < t_N = T_y$. The filtered estimate $\overline{X}_g(t_k)$ and the conditional covariance $P_g(t_k)$ are approximated by $\hat{X}_g^{\Delta}(t_k)$ and $P_g^{\Delta}(t_k)$ respectively where for $k = 0, \ldots, N-1$ we execute the following steps:

1. Assign $x_0 := \overline{X}_g^{\Delta}(t_k)$ and $p_0 := P_g^{\Delta}(t_k)$.
2. Compute
$$L_{hs} = \mathbf{I}_{n_X \times n_X} - A^{\mathbb{P}}\Delta - p_0 \vec{C}(t_{k+1})^* R^{-1} \vec{C}(t_{k+1})\Delta$$
and
$$R_{hs} = \overline{X}_g^{\Delta}(t_k) + B^{\mathbb{P}}\Delta + p_0 \vec{C}(t_{k+1})^* R^{-1} \left(\vec{y}^{obs}(t_{k+1}) - \vec{D}(t_{k+1}) \right) \Delta.$$
3. Find x_1 by solving the linear equation $L_{hs} x_1 = R_{hs}$.
4. Find p_1 by solving the continuous time Riccati equation
$$I_A p_1 + p_1 I_A^* + p_1 C_\Delta^* R^{-1} C_\Delta p_1 - \left(\hat{P}_g^{\Delta}(t_k) + \Delta\Sigma \operatorname{diag}(\mathcal{A}x_1 + \mathcal{B})\Sigma^* \right) = 0$$
5. If $\|p_1 - p_0\|^2 + \|x_1 - x_0\|^2 > \epsilon$, assign $p_0 := p_1$ and $x_0 := x_1$ and return to step (2).
6. Assign $\overline{X}_g^{\Delta}(t_{k+1}) = x_1$ and $P_g^{\Delta}(t_{k+1}) = p_1$.

4 Parameter Estimation of Cox, Ingersol and Ross Model

4.1 Introduction

In this chapter we will investigate the performance of the continuous-time maximum likelihood estimation method which is discussed in the previous chapter. We will work exclusively with the Cox-Ingersol-Ross (CIR hereafter) model of the interest rate. Using simulated data, we will first investigate the performance of the Gaussian filter, comparing it to the exact nonlinear filter. Furthermore, using both filtering algorithms, we will investigate the estimation performance of the Itô and Robust likelihood discussed in the previous chapter. We also apply the estimation method to estimate the CIR model on the the US and European term structures.

4.2 Model Setup

4.2.1 Cox-Ingersol-Ross Model

The Cox-Ingersol-Ross interest rate model is a one-factor model given by

$$dr(t) = \kappa(\theta - r(t))\, dt + \sigma\sqrt{r(t)}\, dW(t) \tag{4.1}$$

for which, the corresponding bond yield with time to maturity τ is

$$y(t, t+\tau) = C(\tau)\, r(t) + D(\tau) \tag{4.2}$$

where

$$\begin{aligned} C(\tau) &= \frac{1}{\tau}\frac{2(e^{\gamma\tau}-1)}{2\gamma + (\kappa+\lambda+\gamma)(e^{\gamma\tau}-1)}, & (4.3)\\ D(\tau) &= \frac{2\kappa\theta}{\sigma^2\tau}\log\left(\frac{2\gamma\exp(\tau(\kappa+\lambda+\gamma)/2)}{2\gamma+(\kappa+\lambda+\gamma)(e^{\gamma\tau}-1)}\right), & (4.4)\\ \gamma &= \sqrt{(\kappa+\lambda)^2 + 2\sigma^2}, \text{ and} & (4.5) \end{aligned}$$

and λ is a constant market price of risk parameter.

4.2.2 Measurement Model

In order to apply the continuous maximum likelihood estimation method, we assume that at each time instant t, we have an observation vector $\vec{y}^{obs}(t)$. The integrated observation

$$Y(t) = \int_0^t \vec{y}^{obs}(s)ds \tag{4.6}$$

satisfies the following SDE:

$$dY(t) = (\vec{C}(t)X(t) + \vec{D}(t))dt + dW_0(t). \tag{4.7}$$

where $W_0(t)$ is a vector Brownian motion with a known covariance matrix R which is independent to $W^{\mathbb{P}}(t)$.

Let $y^{obs}(t, t+\tau)$ be the observed value of yields $y(t, t+\tau)$. For parameter estimation, we will use the vector containing observed yields with n_y number of maturities i.e.

$$\vec{y}^{obs}(t) = \begin{bmatrix} y_s^{obs}(t, t+\tau_1) \\ \vdots \\ y_s^{obs}(t, t+\tau_{n_y}) \end{bmatrix} \tag{4.8}$$

Thus, for the CIR model, the matrices $\vec{C}(t)$ and $\vec{D}(t)$ are defined as

$$\vec{C}(t) = \begin{bmatrix} C(\tau_1) \\ \vdots \\ C(\tau_{n_y}) \end{bmatrix}, \ \vec{D}(t) = \begin{bmatrix} D(\tau_1) \\ \vdots \\ D(\tau_{n_y}) \end{bmatrix} \tag{4.9}$$

4.3 Maximum Likelihood Estimation

4.3.1 Likelihood Functional

The unknown parameter Θ driving both the state and observation equations are estimated by maximizing the robust likelihood functional

$$\begin{aligned} \mathbb{L}(\Theta) \ &= R^{-1} \ \int_0^{T_y} \left\{ \langle \vec{C}(t)\, \overline{r}(t) + \vec{D}(t),\ \vec{y}^{obs}(t) \rangle - \frac{1}{2} \| \vec{C}(t)\, \overline{r}(t) + \vec{D}(t) \|^2 \right\} \, dt \\ &\quad - R^{-1} \frac{1}{2} \int_0^{T_y} \vec{C}(t) P(t) \vec{C}(t)^* \, dt \end{aligned} \tag{4.10}$$

Using simulated data, we will compare the performance of the robust likelihood functional with the Itô likelihood

$$\mathbf{L}(\Theta) \ = R^{-1} \ \int_0^{T_y} \left\{ \langle \vec{C}(t)\, \overline{r}(t) + \vec{D}(t),\ \vec{y}^{obs}(t) \rangle - \frac{1}{2} \| \vec{C}(t)\, \overline{r}(t) + \vec{D}(t) \|^2 \right\} \, dt \tag{4.11}$$

4.3.2 Nonlinear Filtering

In order to compute the nonlinear filter, we are going to follow the splitting-up algorithm described in Section 3.4.2. Given the initial condition $q(0,r) = q^{\Delta}(0,r)$, define discrete time steps $0 = t_0 < t_1 = t_0 + \Delta < \ldots < t_k = t_{k-1} + \Delta < \ldots < t_N = T_y$. The unnormalized conditional density $q(t_k, r)$ is approximated by $q^{\Delta}(t_k, r)$ where for $k = 0, \ldots, N-1$.

4.3.2.1 The Prediction Step

At time t_k in the prediction step of the splitting-up method, we will need to solve

$$\begin{aligned} \frac{\partial}{\partial t} f(t,r) &= -\frac{\partial}{\partial r}\left(\kappa(\theta - r)\, f(t,r)\right) - \frac{1}{2}\frac{\partial^2}{\partial r^2}\left(\sigma^2 r\, f(t,r)\right) \\ &= \kappa f(t,r) - (\kappa(\theta - r) - \sigma^2)\frac{\partial}{\partial r} f(t,r) + \frac{1}{2}\sigma^2\, r \frac{\partial^2}{\partial r^2} f(t,r) \end{aligned} \tag{4.12}$$

between times t_k and t_{k+1}, given the initial condition $f(t_k, r) = q^{\Delta}(t_k, r)$. Here we will use the Crank-Nicholson finite difference method to approximate the solution of (4.12).

Let us define a computation grid $\mathbb{T} = \{\tilde{t}_0 = t_k, \tilde{t}_1 = \tilde{t}_0 + \Delta t, \ldots, \tilde{t}_M = t_{k+1}\}$ and $\mathbb{R} = \{r_0 = 0, r_1 = r_0 + \Delta r, \ldots, r_N = r_{N-1} + \Delta r\}$. Let us denote $f_{i,j} = f(\tilde{t}_i, r_j)$. The finite difference approximation of the r.h.s of (4.12) is given by:

$$\mathcal{D}f_{i,j} = \kappa\, f_{i,j} - g_j \frac{q_{i+1,j} - q_{i,j}}{\Delta r} + \frac{1}{2} h_j \frac{q_{i+1,j} - 2q_{i,j} + q_{i,j-1}}{\Delta r^2}, \tag{4.13}$$

$$g_j = (\kappa(\theta - r_j) - \sigma^2), \tag{4.14}$$

$$h_j = \sigma^2 r_j. \tag{4.15}$$

Thus, the Cranks-Nicholson approximation of (4.12) is given by:

$$\frac{f_{i+1,j} - f_{i,j}}{\Delta t} = \frac{1}{2}\left(\mathcal{D}f_{i+1,j} + \mathcal{D}f_{i,j}\right) \tag{4.16}$$

Let us denote:

$$a_j = \frac{1}{2}\frac{h_j}{\Delta r^2} \tag{4.17}$$

$$b_j = (\kappa + \frac{g_j}{\Delta r} - \frac{h_j}{\Delta r^2}) \tag{4.18}$$

$$c_j = (-\frac{g_j}{\Delta r} + \frac{1}{2}\frac{h_j}{\Delta r^2}) \tag{4.19}$$

By rearranging (4.16) we obtain the following:

$$\begin{aligned}
& (-\frac{1}{2}a_j\Delta t)\ f_{i+1,\,j-1} + (1-\frac{1}{2}b_j\Delta t)\ f_{i+1,\,j} + (-\frac{1}{2}c_j\Delta t)\ f_{i+1,\,j-1} \\
= \quad & (\frac{1}{2}a_j\Delta t)\ f_{i,\,j-1} + (1+\frac{1}{2}b_j\Delta t)\ f_{i,\,j} + (\frac{1}{2}c_j\Delta t)\ f_{i,\,j-1}
\end{aligned} \tag{4.20}$$

For large enough r_N, we may approximate the boundary condition by setting $f_{i,\,0} = f_{i,\,N} = 0$ at every time-step $\tilde{t}_i$, $i = 1, \ldots M$. This will allow us to write (4.20) as:

$$A\ \vec{f}_{i+1} = B\ \vec{f}_i \tag{4.21}$$

where

$$A = \Delta t \begin{bmatrix} \frac{1}{\Delta t} - \frac{1}{2}b_1 & -\frac{1}{2}c_1 & 0 & 0 & \cdots & 0 \\ -\frac{1}{2}a_2 & \frac{1}{\Delta t} - \frac{1}{2}b_2 & -\frac{1}{2}c_2 & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & \cdots & 0 & 0 & -\frac{1}{2}a_N & \frac{1}{\Delta t} - \frac{1}{2}b_N \end{bmatrix} \tag{4.22}$$

$$B = \Delta t \begin{bmatrix} \frac{1}{\Delta t} + \frac{1}{2}b_1 & \frac{1}{2}c_1 & 0 & 0 & \cdots & 0 \\ \frac{1}{2}a_2 & \frac{1}{\Delta t} + \frac{1}{2}b_2 & \frac{1}{2}c_2 & 0 & \cdots & 0 \\ \vdots & \vdots & & & & \vdots \\ 0 & \cdots & 0 & 0 & \frac{1}{2}a_N & \frac{1}{\Delta t} + \frac{1}{2}b_N \end{bmatrix} \tag{4.23}$$

and

$$\vec{f}_i = \begin{bmatrix} f_{i,\,0} \\ \vdots \\ f_{i,\,N} \end{bmatrix}, \quad i = 1, \ldots, M.$$

with the initial condition

$$\vec{f}_0 = \begin{bmatrix} q^{\Delta}\left(t_k, r_0\right) \\ \vdots \\ q^{\Delta}\left(t_k, r_N\right) \end{bmatrix}.$$

After iterating (4.21) M times, the solution of (4.12) at time t_{k+1} is given by

$$f\left(t_{k+1}, r_k\right) \approx f_{M,\,k}$$

for $k = 1, \ldots, N$.

4.3.2.2 The Correction Step

At the correction step, for $l = 0, \ldots, M$ we set

$$q^{\Delta}(t_{k+1}, r_l) := c_{k+1}\, f(t_{k+1}, r_l)\ \Phi^{\Delta}(t_{k+1}, r_l) \tag{4.24}$$

where c_{k+1} is a normalizing factor so that $c_{k+1}\sum_{l=0}^{M-1} q^{\Delta}(t_{k+1}, r_l)\,\Delta r = 1$ and

$$\log \Phi^{\Delta}(t, r) \quad = \quad R^{-1}\int_{t_k}^{t_{k+1}} \langle H(t,r)\,,\ \vec{y}^{obs}(t)\rangle - \frac{1}{2}\left\|H(t,r)\right\|^2 dt,$$

$H(t, r) = \vec{C}(t)r + \vec{D}(t)$.

4.3.3 Gaussian Filtering

The Gaussian approximate filter is given by the following systems of ODE:

$$\frac{d}{dt}\overline{r}_g(t) \quad = \quad \kappa(\theta - \overline{r}_g(t))\ + P_g(t)\vec{C}(t)^{*}R^{-1}\left(\vec{y}^{obs}(t) - (\vec{C}(t)\overline{r}_g(t) + \vec{D}(t))\right)$$

and

$$\frac{d}{dt}P_g(t) \quad = \quad -2\kappa P_g(t)\ - P_g(t)^2\vec{C}(t)^{*}R^{-1}\vec{C}(t) + \sigma^2\overline{r}_g(t).$$

4.4 Simulated Data

In this section we will investigate the performance of the algorithms presented in the earlier sections. First of all, we compare the performance of the Gaussian filter to the exact nonlinear filter. Secondly, we will apply the maximum likelihood approach and compare the estimation performance of the Itô likelihood and the robust likelihood using both the exact nonlinear filter and the Gaussian filter. For these studies we model the interest rate according to the one-factor Cox-Ingersol-Ross model.

We will consider two cases where the covariance of the measurement error is $R = \sigma_0^2\mathbf{I}_{9\times 9}$ where $\sigma_0 \in \{10\,\text{bp},\ 0.1\,\text{bp}\}$. We generated 100 paths of the interest rate which resulted in 100×2 data sets, each containing 365 daily noisy bond yields with the following time to maturities: 3 months, 6 months, 1, 2, 3, 5, 7, 10 and 20 years. The parameters for the Cox-Ingersol-Ross model, taken from (de Jong and Santa-Clara (1999)), are listed in Table (4.1). The initial condition is $r(0) = 0.06$. For filtering we assume a Gaussian initial condition with mean $\overline{r}(0) = 0.07$ and variance $P(0) = 0.01$.

κ	θ	σ	λ
0.1862	0.0654	0.0481	-0.0741

Table 4.1: Parameters for the Cox-Ingersol-Ross model.

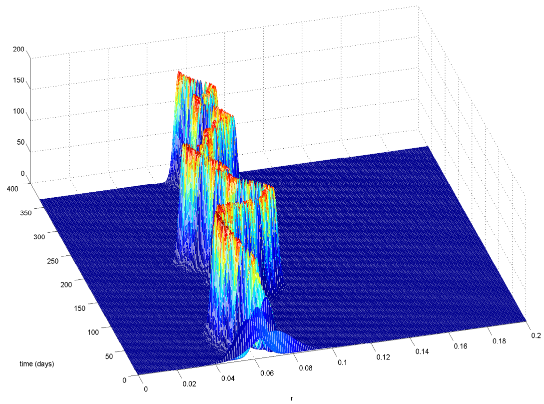

Figure 4.1: Numerical solution of the conditional density for $\sigma_0 = 10$ bp.

4.4.1 Comparison of Filtering Performance

An example of the conditional density for $\sigma_0 = 10$ basis points is shown in Figure (4.1). For a single path, we plot the comparison of the filtered estimates obtained from the nonlinear filter and the approximate Gaussian filter in Figures (4.2) and (4.3). As expected, the filtered estimates given by the Gaussian approximation do not perform as well as the those obtained from the exact nonlinear filter. However, the difference between the two is minimal. This conclusion is further confirmed by the root mean squared errors (RMSE) between exact values and the filtered estimates given in Figures (4.4) and (4.5).

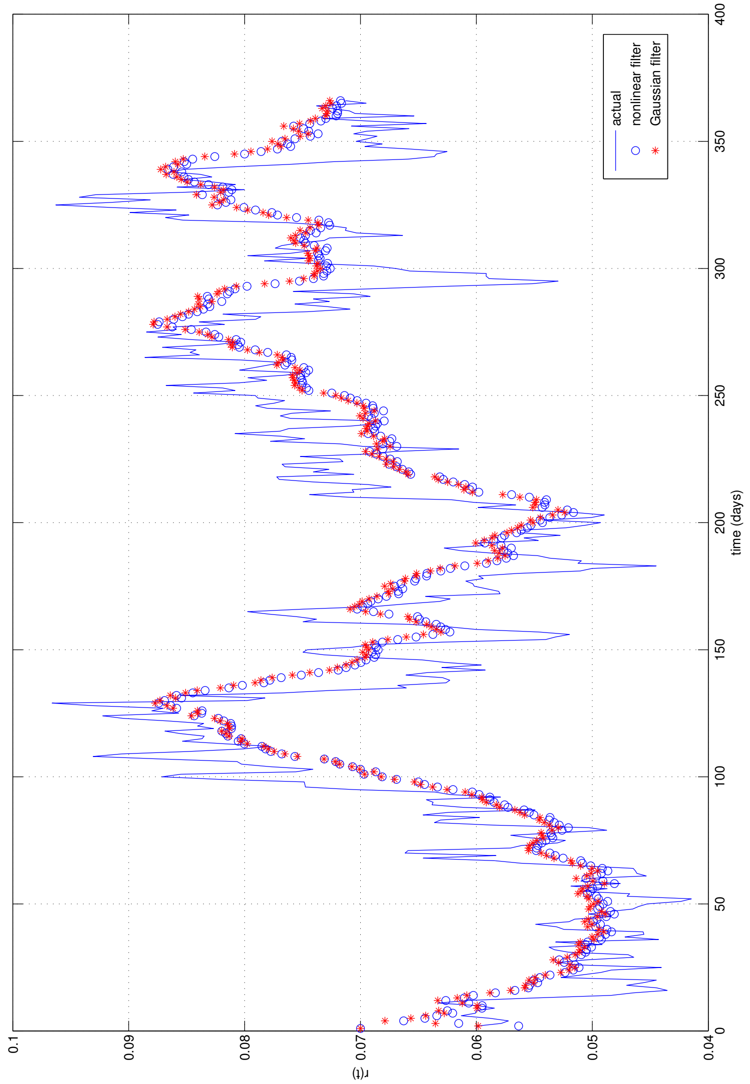

Figure 4.2: Comparison of filtering performance. Nonlinear filter vs Gaussian filter for $\sigma_0 = 10$ bp.

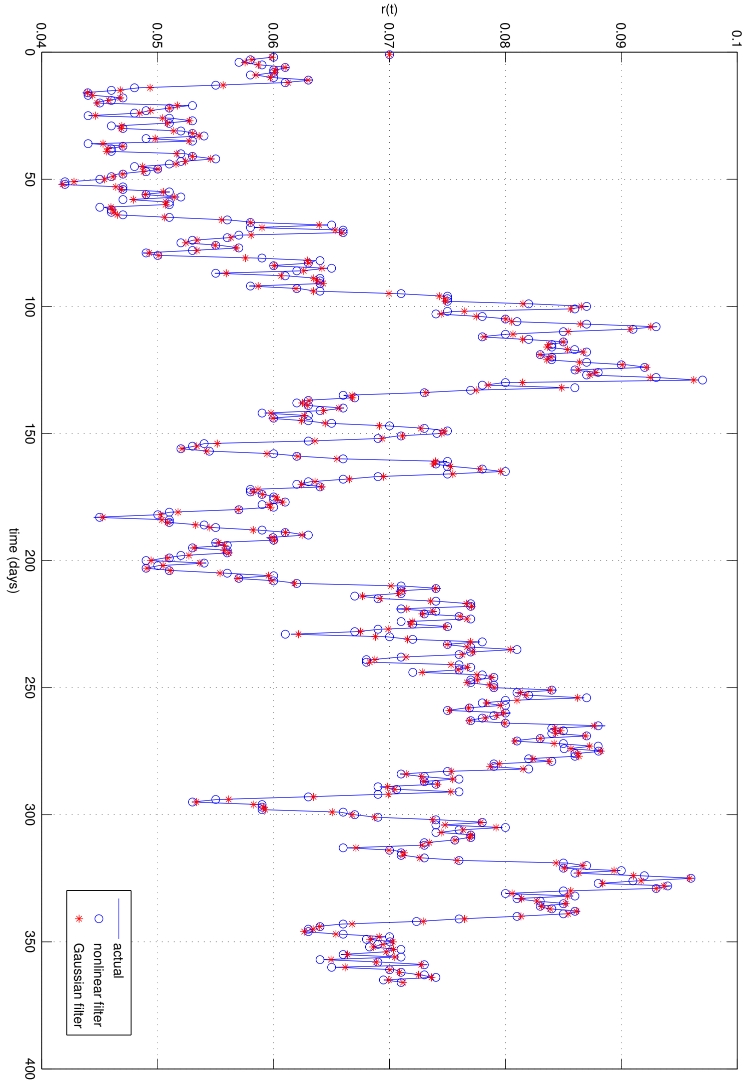

Figure 4.3: Comparison of filtering performance. Nonlinear filter vs Gaussian filter for $\sigma_0 = 0.1$ bp.

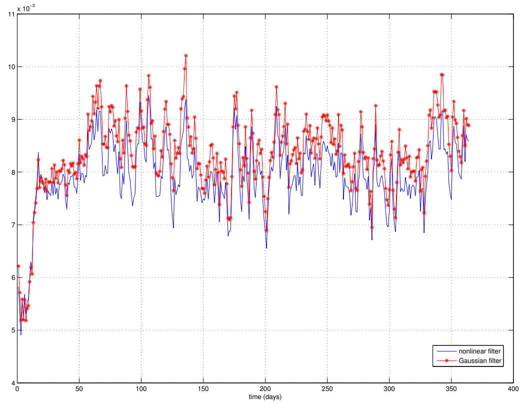

Figure 4.4: RMSE of filtered estimate. Nonlinear filter vs Gaussian filter for $\sigma_0 =$ 10 bp.

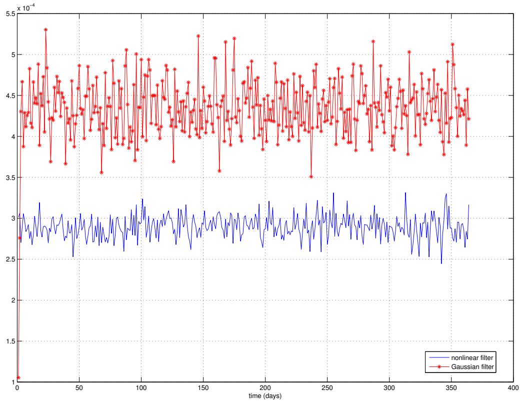

Figure 4.5: RMSE of filtered estimate. Nonlinear filter vs Gaussian filter for $\sigma_0 =$ 0.1 bp.

4.4.2 Comparison of Likelihood Functionals

Although the Gaussian filter performs very well for the simulated data, we would like to know how both filtering algorithms perform in estimating parameters. Furthermore, we ask ourselves whether there are substantial differences between parameter estimates obtained by the Itô likelihood and the robust likelihood. In order to answer these questions, we estimated the parameters of all 200 data sets. For parameter estimation, we have taken true parameter values as starting points for the optimisation procedure. The results are presented in tables (4.2) and (4.3). These tables report the mean and standard deviation of the differences between the estimated parameters and the corresponding true parameter values. We apply Wilcoxon signed rank test to test the hypothesis that the median of the differences are zero. We have chosen this particular statistical test due to questionable normality of the resulting parameter estimates. We also included the results in both tables, indicating whether the null hypothesis (the median of the differences is zero) can be rejected or cannot be rejected at 95% confidence level by 1 and 0 respectively. The results vary across the table. In general, the results suggest that the parameter estimates are biased and that the bias decreases for smaller noise covariance.

In addition, we also apply signed rank test to the parameter estimates obtained from the robust likelihood and the Itô likelihood. We would like to know how significant is the difference between parameter estimates obtained using these two likelihoods for each filtering algorithms. The results obtained from paired signed rank test are given in Table (4.4). Again we indicate whether the null hypothesis (the median of the differences is zero) can be rejected or cannot be rejected at 95% confidence level by 1 and 0 respectively. The results vary across the table. For the smaller noise covariance, using either the nonlinear or Gaussian filter, the difference between parameter estimates obtained using the Itô likelihood and the robust likelihood is not significant. This result is not very surprising because for the smaller noise covariance, the conditional covariance of the filtered estimates that determine the correction factor in the robust likelihood are very small. For the larger noise covariance, using either the nonlinear or Gaussian filter, the difference between the resulting parameter estimates is significant. On the other hand, from Table (4.3), we see parameter estimates obtained by robust likelihood is less biased than those obtained from the Itô likelihood. These led us to conclude the robust likelihood is better than the Itô likelihood.

Furthermore, we also apply signed rank test to the parameter estimates obtained using the nonlinear filter and the Gaussian filter. We would like to know how significant is the difference between parameter estimates obtained using the nonlinear filter and the Gaussian filter. Given the previous conclusion we specifically would like to know if the difference is significant when we use the robust likelihood. The results obtained from paired signed rank test are given in Table (4.5). Again we indicate whether the null hypothesis (the median of the differ-

Parameter	Nonlinear Filter					
	robust likelihood			Itô likelihood		
	Mean	Std. Dev.		Mean	Std. Dev	
κ	0.001644	0.002497	1	0.001797	0.003025	1
θ	-0.000043	0.000302	0	-0.000094	0.000702	1
σ	-0.000256	0.002170	0	-0.000284	0.003254	0
λ	-0.000278	0.000529	1	-0.000438	0.001909	1
Parameter	Gaussian Filter					
	robust likelihood			Itô likelihood		
	Mean	Std.Dev		Mean	Std. Dev.	
κ	0.001706	0.003623	1	0.000225	0.000605	1
θ	-0.000015	0.001224	0	-0.000215	0.000222	1
σ	-0.016366	0.001385	1	0.003447	0.001002	1
λ	0.002297	0.003486	1	-0.001264	0.000726	1

Table 4.2: Monte Carlo results for the Cox-Ingersol-Ross model, $\sigma_0 = 0.1$ bp. This table reports the mean and standard deviation of the errors between the true parameter values and parameter estimates.

ences is zero) can be rejected or cannot be rejected at 95% confidence level by 1 and 0 respectively. For the robust likelihood, the results in the first and third columns suggest that the difference between the mean reversion parameter estimates obtained using the nonlinear filter and the Gaussian filter is not significant while the difference between the volatility parameter estimates using the nonlinear filter and the Gaussian filter is significant.

4.5 Fama-Bliss Data

We apply the continuous-time maximum likelihood method in estimating the CIR model using the Fama-Bliss dataset. The dataset contains monthly bond yields with maturities 1,2,...,5 years from April 1964 to December 1997. Figure (4.6) plots the time series. We assume that the noise covariance is $(0.1 \text{ bp})^2 \mathbf{I}_{5\times 5}$.

Parameter	Nonlinear Filter					
	robust likelihood			Itô likelihood		
	Mean	Std. Dev.		Mean	Std. Dev	
κ	-0.041423	0.034861	1	-0.085835	0.052489	1
θ	0.021074	0.025265	1	0.083220	0.070708	1
σ	0.076237	0.007908	1	0.092651	0.009120	1
λ	0.008500	0.037459	0	0.043658	0.056159	1
Parameter	Gaussian Filter					
	robust likelihood			Itô likelihood		
	Mean	Std. Dev.		Mean	Std. Dev	
κ	-0.017261	0.070081	1	-0.070854	0.039728	1
θ	0.021201	0.043866	1	0.051936	0.048292	1
σ	0.035030	0.005097	1	0.108377	0.009591	1
λ	0.005600	0.067236	0	0.022300	0.041628	1

Table 4.3: Monte Carlo results for the Cox-Ingersol-Ross model, $\sigma_0 = 10$ bp. This table reports the mean and standard deviation of the errors between the true parameter values and parameter estimates.

	$\sigma_0 = 0.1$ bp		$\sigma_0 = 10$ bp	
	Nonlinear filter	Gaussian filter	Nonlinear filter	Gaussian filter
κ	0	0	1	1
θ	0	0	1	1
σ	0	0	1	1
λ	1	1	1	1

Table 4.4: Wilcoxon signed rank test for zero median between estimates obtained using the robust likelihood and the Itô likelihood. The 0 entry indicates that the null hypothesis (median is zero) cannot be rejected at 95% significance level. 1 indicates that the null hypothesis can be rejected at 95% significance level.

	$\sigma_0 = 0.1$ bp		$\sigma_0 = 10$ bp	
	robust likelihood	Itô likelihood	robust likelihood	Itô likelihood
κ	0	1	1	1
θ	0	1	0	1
σ	1	1	1	1
λ	1	1	0	1

Table 4.5: Wilcoxon signed rank test for zero median between estimates obtained using the nonlinear and Gaussian filters. The 0 entry indicates that the null hypothesis (median is zero) cannot be rejected at 95% significance level. 1 indicates that the null hypothesis can be rejected at 95% significance level.

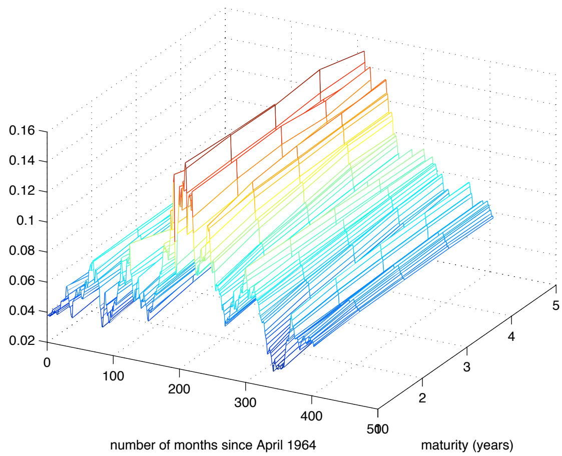

Figure 4.6: Monthly Fama-Bliss data from April 1964 to December 1997.

Parameter estimates are given in Table (4.6), marked CTML. The result obtained by Duan and Simonato (1999) with quasi- maximum likelihood (QML) and De Rossi (2005) with particle filtering (PF) are also given for comparison.

Method	$\hat{\kappa}$	$\hat{\theta}$	$\hat{\sigma}$	$\hat{\lambda}$
CTML	0.125	0.0479	0.0292	-0.0832
QML	0.225	0.0613	0.0700	-0.1112
PF	0.175	0.0592	0.0683	-0.0961

Table 4.6: CIR parameter estimated from Fama-Bliss dataset starting April 1964 to December 1997. Parameter estimates on the second and third rows are those obtained by Duan and Simonato (1999) and De Rossi (2004) respectively.

4.6 Conclusions

Using simulated data, we have investigated the performance between the robust likelihood functional and the Itô likelihood functional in estimating parameters of the one-factor Cox-Ingersol-Ross model. Results suggest that the robust likelihood performs better than the Itô likelihood. The difference is more pronounced when the measurement errors are relatively large. In general, parameter estimates obtained using the continuous-time MLE are reasonably close to the actual value, suggesting that the continuous-time maximum likelihood is a feasible alternative for estimating the exponential-affine term structure models. However, its application is severely limited by extended computational time required in computing the likelihood due to the time consuming nonlinear filtering procedure. For this reason, we proposed an approximate filter by approximating the exact conditional density function by a Gaussian density function. We have performed a comparison between the approximate filter and the exact nonlinear filter. The results suggest that the filtered estimates obtained using the approximate filter are only slightly less accurate than those obtained using the exact nonlinear filter. Furthermore, we did not observe any adverse effect of substituting the exact nonlinear filter for the approximate filter in the estimation procedure.

In the following chapter, we will apply the continuous-time maximum likelihood method in fitting the Euro swap rates and the Euribor futures to 3-factor interest rate models.

A Comparison of Three-factor Models in the European Fixed Income Market

5

5.1 Introduction

In this chapter we will apply the continuous-time maximum likelihood method described in the previous chapter to fit two 3-factor interest rate models to the the Euro swap rates and the Euribor futures prices.

5.2 Interest Rate Models

5.2.1 Balduzzi-Das-Foresi-Sundaram Model

The Balduzzi, Das, Foresi and Sundaram (BDFS, 1996) model belongs to the class of three-factor exponential affine models. It extends the well-known Vasicek model by assuming that both the mean reversion and the volatility are stochastic. The mean reversion level itself is modeled as a mean reverting Gaussian process. To ensure positivity of the instantaneous variance, it is modeled as a square-root process. Under the risk neutral measure, the interest rates factors satisfy the following dynamics:

$$
\begin{aligned}
dr(t) &= \kappa\left(\theta(t) - r(t)\right) dt + \sqrt{v(t)}\, dW_1^{\mathbb{Q}}(t) \\
d\theta(t) &= \nu\left(\bar{\theta} - \theta(t)\right) dt + \zeta\, dW_2^{\mathbb{Q}}(t) \\
dv(t) &= \mu\left(\bar{v} - v(t)\right) dt + \eta\sqrt{v_t}\, dW_3^{\mathbb{Q}}(t)
\end{aligned}
\tag{5.1}
$$

while the equivalent real world dynamics is given by:

$$
\begin{aligned}
dr(t) &= \left(\kappa\left(\theta(t) - r(t)\right) + \lambda_1 v(t)\right)\ dt + \sqrt{v(t)}\, dW_1^{\mathbb{P}}(t) \\
d\theta(t) &= \left(\nu\left(\bar{\theta} - \theta(t)\right) + \lambda_2\zeta\right)\ dt + \zeta\, dW_2^{\mathbb{P}}(t) \\
dv(t) &= \left(\mu\left(\bar{v} - v(t)\right) + \lambda_3\eta v_t\right)\ dt + \eta\sqrt{v_t}\, dW_3^{\mathbb{P}}(t)
\end{aligned}
\tag{5.2}
$$

The mean reverting parameters $\bar{\theta}$ and $\bar{v}$ are both constrained to be positive values. To ensure positivity of the volatility factor $v(t)$ for all $t \geq 0$, the Feller condition $2\mu\bar{v} > \eta^2$ is added. Volatility parameters ζ and η are both constrained to be positive values. For stability, κ, ν, μ and $\mu + \lambda_3\eta$ are assumed to be positive.

5.2.2 Chen Model

Unlike the BDFS, Chen (Chen, 1996) modeled the mean reversion level as a square root model restricting it to positive values. Under the risk neutral measure, the interest rate model satisfies:

$$\begin{aligned} dr(t) &= \kappa\left(\theta(t) - r(t)\right) dt + \sqrt{v(t)}\; dW_1^{\mathbb{Q}}(t) \\ d\theta(t) &= \nu\left(\bar{\theta} - \theta(t)\right) dt + \zeta\sqrt{\theta(t)}\, dW_2^{\mathbb{Q}}(t) \\ dv(t) &= \mu\left(\bar{v} - v(t)\right) dt + \eta\sqrt{v(t)}\, dW_3^{\mathbb{Q}}(t) \end{aligned} \tag{5.3}$$

while the equivalent real world dynamics is given by:

$$\begin{aligned} dr(t) &= \left(\kappa\left(\theta(t) - r(t)\right) + \lambda_1 v(t)\right)\; dt + \sqrt{v(t)}\; dW_1^{\mathbb{P}}(t) \\ d\theta(t) &= \left(\nu\left(\bar{\theta} - \theta(t)\right) + \lambda_2\zeta\theta(t)\right)\; dt + \zeta\sqrt{\theta(t)}\, dW_2^{\mathbb{P}}(t) \\ dv(t) &= \left(\mu\left(\bar{v} - v(t)\right) + \lambda_3\eta v(t)\right)\; dt + \eta\sqrt{v(t)}\, dW_3^{\mathbb{P}}(t) \end{aligned} \tag{5.4}$$

The mean reverting parameters $\bar{\theta}$ and $\bar{v}$ are both constrained to be positive values. To ensure positivity of both the mean reversion $\theta(t)$ and the volatility $v(t)$, the Feller condition $2\nu\bar{\theta} > \zeta^2$ and $2\mu\bar{v} > \eta^2$ are added. Volatility parameters ζ and η are both constrained to be positive values. For stability, κ, ν, μ, $\nu - \lambda_2\zeta$ and $\mu + \lambda_3\eta$ are assumed to be positive.

5.3 Measurement Model

In order to apply the continuous maximum likelihood estimation method, we will first parameterize both BDFS and Chen models into our standard matrix notation of the exponential-affine model:

$$dX(t) = (AX(t) + B)dt + \Sigma\,\mathrm{diag}(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}}\, dW^{\mathbb{Q}}(t) \tag{5.5}$$

where $X(t) = [r(t)\ \theta(t)\ v(t)]^*$. Under the real-world measure, the dynamics of the interest rate factors is given by

$$dX(t) = (A^{\mathbb{P}}X(t) + B^{\mathbb{P}})dt + \Sigma\,\mathrm{diag}(\mathcal{A}X(t) + \mathcal{B})^{\frac{1}{2}}\, dW^{\mathbb{P}}(t) \tag{5.6}$$

The constant matrices A, B, $A^{\mathbb{P}}$, $B^{\mathbb{P}}$, Σ, $\mathcal{A}$ and $\mathcal{B}$ for BDFS and Chen models are given in Appendix A. Furthermore, we assume that at each time instant t, we have an observation vector $\vec{y}^{obs}(t)$. The integrated observation

$$Y(t) = \int_0^t \vec{y}^{obs}(s)ds \tag{5.7}$$

satisfies the following SDE:

$$dY(t) = (\vec{C}(t)X(t) + \vec{D}(t))dt + dW_0(t). \tag{5.8}$$

where $W_0(t)$ is a vector Brownian motion with a known covariance matrix R which is independent of $W^{\mathbb{P}}(t)$. We will later specify the vector of observations $\vec{y}^{obs}(t)$, and time-varying matrices $\vec{C}(t)$ and $\vec{D}(t)$ separately for each case study that we are going to consider in this chapter.

The unknown parameter Θ driving both the state and observation equations is estimated by maximizing the robust log-likelihood functional

$$\begin{aligned}\mathbb{L}(\Theta) \quad = R^{-1} \quad & \int_0^{T_y} \left\{ \langle \vec{C}(t)\, \overline{X}_g(t) + \vec{D}(t),\ \vec{y}^{obs}(t) \rangle - \frac{1}{2} \| \vec{C}(t)\, \overline{X}_g(t) + \vec{D}(t) \|^2 \right\} dt \\ & - \frac{1}{2} \int_0^{T_y} \vec{C}(t) P_g(t) \vec{C}(t)^* \, dt \end{aligned} \tag{5.9}$$

where

$$K_g(t) = P_g(t)^*, \tag{5.10}$$

$$\begin{aligned}\frac{d}{dt}\overline{X}_g(t) \quad &= \quad (A^{\mathbb{P}}\overline{X}_g(t) + B^{\mathbb{P}}) \ + \\ & \quad\ K_g(t) R^{-1} \left(\vec{y}^{obs}(t) - (\vec{C}(t)\overline{X}_g(t) + \vec{D}(t)) \right) \end{aligned} \tag{5.11}$$

and

$$\begin{aligned}\frac{d}{dt}P_g(t) \quad &= \quad A^{\mathbb{P}} P_g(t) \ + P_g(t) A^{\mathbb{P}*} - K_g(t) R^{-1} K_g(t)^* \ + \\ & \quad\ \Sigma \operatorname{diag}(\mathcal{A}\overline{X}_g(t) + \mathcal{B}) \Sigma^*. \end{aligned} \tag{5.12}$$

5.4 The Euro Swap Rates

5.4.1 The Euro Swap Markets

An interest rate swap is a contract between two parties to exchange streams of interest rates payments. Typically, one stream of payments is based on a fixed rate of interest while the other stream is based on a floating rate interest. Only the net cash flows are paid; the notional principal on which the interest rate payments are calculated is not exchanged. In terms of notional principal outstanding, over-the-counter markets for Euro- and US dollar-denominated interest rate derivatives are the largest financial markets in the world. The notional stock of Euro-denominated interest rate swaps and forwards totalled 26.3 trillion at the end June 2002. The US dollar-denominated contracts were slightly smaller at 26.2 trillion.

The pricing of interest rate swaps in general depends on the interest rate used for the floating rate leg of the contract. For Euro swaps, the choice of the floating rate tends to depend on the contract's maturity. For short dated swaps, EONIA (Euro OverNight Index Average) is the most common basis for the floating rate leg. For longer-dated swaps, that we will use in our study, Euribor (European Interbank Offered Rate) remains the key reference rate.

5.4.2 Data Description

We downloaded cross-sectional data consisting of daily Euro swap rates that are available from the website of the central bank of Austria. The Euro swap rates are calculated daily at 11:00 a.m. (Frankfurt time) by an independent institution (International Swap and Derivatives Association, Inc. ISDA) by averaging the interest rates quoted by major European banks. 16 banks indicate the interbank interest rate at which those banks would buy or sell a swap of a given maturity and capital amount. The swap rate quotes are referenced to the six-month Euribor, except for one-year swaps, which are based on the three-month Euribor. In addition, we also downloaded Euribor rates from the same source.

Our dataset contains the 1 month, 2 months, ..., 12 months Euribor rates and 1 year, 2 years,..., 10 years, 12 years, 15 years and 20 years Euro swap rates between 2 January 2002 and 30 May 2005. We divided the dataset into two parts. The first part is the calibration dataset consisting of a subset from 2 January 2002 up to 20 December 2004. This dataset will be used in the parameter estimation of the interest rate models. The remaining data between 2 January 2005 and 30 May 2005, the validation dataset, will be used to asses the performance and the stability of the resulting parameter estimates.

In order to estimate the parameters of BDFS and Chen interest rate models, we will use the continuous-time maximum likelihood method described in the earlier chapter. Unfortunately, under the exponential-affine framework, swap rates cannot be expressed as an affine function of the interest rate factors. Thus, we will need to transform the swap rate into a quantity that can be fitted into the our parameter estimation framework. Here we will use the zero coupon bond yields implied from the Euribor rates and the Euro swap rates.

5.4.3 The Implied Term Structure

We have extracted zero coupon bond yields from the Euribor and swap rates following the standard bootstrap procedure (see e.g. James and Webber (2001)).

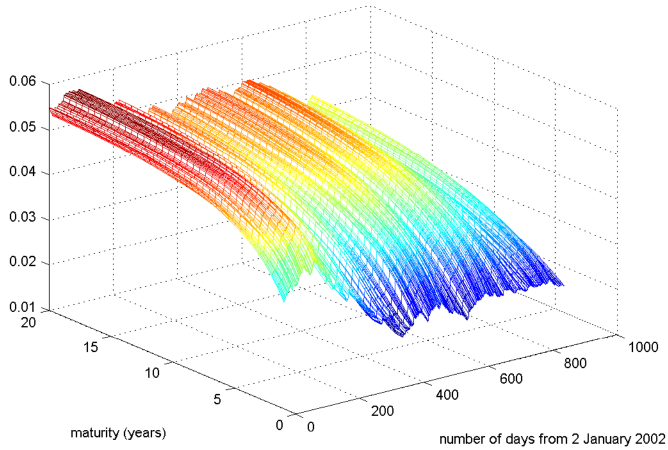

Figure 5.1: The Euro swap rates from 2 January 2002 until 30 May 2005

The procedure can be described as follows. First, we will compute the Euribor-implied zero coupon bond price $B_e(t, t+\tau)$ for $\tau \leq 1$ by using the following equation:

$$B_e\left(t, t+\tau\right) = \frac{1}{1 + r_e\left(t, t+\tau\right)\ \alpha_e\left(t, t+\tau\right)} \tag{5.13}$$

where $r_e\left(t, t+\tau\right)$ denotes the time t value of the 12τ months Euribor rates, and $\alpha_e(t_1, t_2)$ denotes the day-count function in a year unit which correspond to the $actual/360$ day-count convention.

In order to obtain zero coupon bond yields from the swap rates, we will need to approximate the swap rates for $\tau = \frac{1}{2}$ and $\tau = 1$. We need to do this because the day-count convention for calculating the swap rates is different than day-count convention for calculating Euribor rates. Using $B_e(t, t+0.5)$ and $B_e(t, t+1)$ that have been obtained earlier, we define the swap rates as

$$r_s(t, t+0.5) = \frac{1 - B_e(t, t+0.5)}{\alpha_s(t, t+0.5) B_e(t, t+0.5)} \tag{5.14}$$

and

$$r_s(t, t+1) = \frac{1 - B_e(t, t+1)}{\alpha_s(t, t+0.5) B_e(t, t+0.5) + \alpha_s(t+0.5, t+1) B_e(t, t+1)} \tag{5.15}$$

where $\alpha_s(t_1, t_2)$ denotes the day-count function in a year unit which correspond to the $30/360$ day-count convention.

Let $\tau_j = \frac{1}{2}j$, $j = 0, 1, \ldots$. After computing $r_s(t, t+0.5)$ and $r_s(t, t+1)$, we will we need to approximate the swap rates $r_s(t, \tau)$ at $\tau \in \{\tau_3, \tau_4, \ldots \tau_{40}\}$ by interpolating the available swap rates. Given interpolated swap rates, the discount function $B_s(t, t+\tau_j)$, $j = 1, \ldots, 40$ can be recursively computed by:

$$B_s(t, t+\tau_j) = \frac{1 - r_s(t, \tau_j) \sum_{i=1}^{j-1} \alpha_s(t+\tau_{i-1}, t+\tau_i)\ B_s(t, t+\tau_i)}{1 + r_s(t, \tau_j)\ \alpha_s(t+\tau_{j-1}, t+\tau_j)} \tag{5.16}$$

and finally, the swap implied-yield $y_s(t, t+\tau)$ is given by

$$y_s(t, t+\tau) = -\frac{1}{\alpha(t, t+\tau)}\ \ln B_s(t, t+\tau). \tag{5.17}$$

5.4.4 Measurement Equation

Let $y_s^{obs}(t, t+\tau)$ be the observed value of the swap-implied yields $y_s(t, t+\tau)$. For parameter estimation, we will use the vector containing observed swap-implied yields with $n_y = 40$ maturities ranging from 0.5 up to 20 years, i.e.

$$\vec{y}^{obs}(t) = \begin{bmatrix} y_s^{obs}(t, t+\tau_1) \\ \vdots \\ y_s^{obs}(t, t+\tau_{n_y}) \end{bmatrix} \tag{5.18}$$

where $\tau_j = \frac{1}{2}j$, $j = 1, \ldots, 40$. Thus, in this case, the matrices $\vec{C}(t)$ and $\vec{D}(t)$ are defined as

$$\vec{C}(t) = \begin{bmatrix} -\frac{1}{\tau_1(t)} C(\tau_1)^* \\ \vdots \\ -\frac{1}{\tau_{n_y}(t)} C(\tau_{n_y})^* \end{bmatrix},\ \vec{D}(t) = \begin{bmatrix} -\frac{1}{\tau_1(t)} D(\tau_1) \\ \vdots \\ -\frac{1}{\tau_{n_y}(t)} D(\tau_{n_y}) \end{bmatrix} \tag{5.19}$$

where

$$\begin{aligned} \frac{d}{d\tau} C_b(\tau) &= -g_1 + A^* C_b(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_b(\tau)^* \Sigma)_k \right]^2 \mathcal{A}_{k,\cdot}^* \\ \frac{d}{d\tau} D_b(\tau) &= -g_0 + B^* C_b(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_b(\tau)^* \Sigma)_k \right]^2 \mathcal{B}_k \end{aligned} \tag{5.20}$$

with boundary conditions $C(0) = \vec{0}$ and $D(0) = 0$.

Parameter	Model	
	BDFS	Chen
κ	1.07327	0.31727
ν	0.03638	0.78535
$\bar{\theta}$	0.09082	0.06131
ζ	0.01081	0.01046
μ	0.71679	0.00152
$\bar{v}$	0.00014	0.00266
η	0.01423	0.00287
λ_1	0.00002	0.00914
λ_2	0.00186	-0.00230
λ_3	0.00052	0.00292

Table 5.1: Parameter estimates of the Euro swap rates.

5.4.5 Estimation Results

We estimated the parameters of BDFS and Chen models by maximizing the likelihood functional (5.9) given the vector of observations (5.18) and the matrices $\vec{C}(t)$ and $\vec{D}(t)$ in (5.19) . The noise covariance matrix R is assumed to be $(0.1 \text{ bp})^2\mathbf{I}_{40\times 40}$. The parameter estimates are reported in Table (5.1). The fitting performance for the calibration data set and the validation data set are given in Figure (5.2) and (5.3) respectively. The performance of BDFS and Chen models are comparable for the calibration data set with a hint of better fitting provided by the Chen model. With the validation data set, the results are mixed. BDFS seems to capture the short end better than the long end. On the contrary, Chen model seems to capture the long end of the swap rate curve better than the short end. In general, both models fits the euro swap rate curve reasonably well.

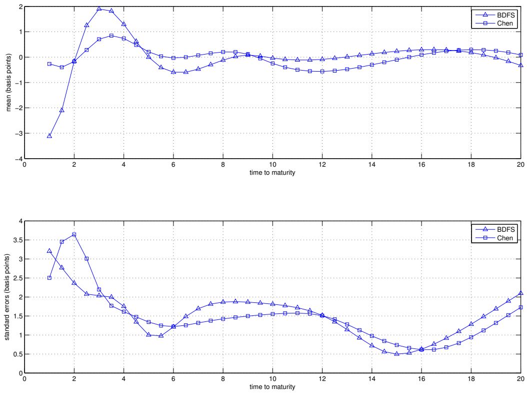

Figure 5.2: Mean and standard errors between the fitted and observed Euro swap rates from 2 January 2002 until 30 December 2004 (calibration dataset).

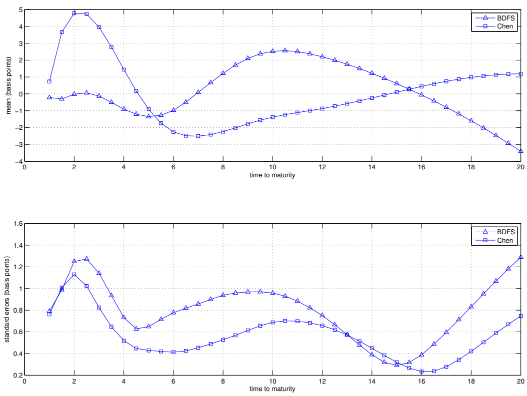

Figure 5.3: Mean and standard errors between the fitted and observed Euro swap rates from 2 January 2005 until 30 May 2005 (validation dataset).

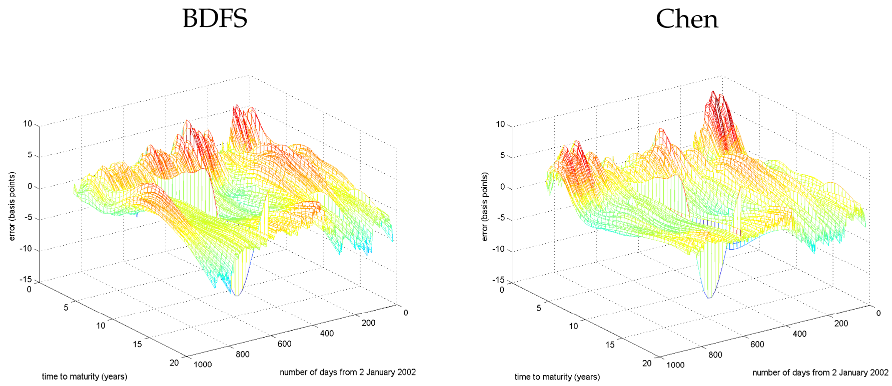

Figure 5.4: Errors between the fitted and observed Euro swap rates from 2 January 2002 until 30 May 2005.

5.5 The Euribor Futures

5.5.1 The Euribor Futures Markets

A futures contract is a binding agreement between two parties to exchange an asset for a fixed price on a specified final settlement date T in the future. The futures price, which is the price at which a given futures contract is entered into, is determined by the usual demand and supply. The futures prices are settled daily through the procedure called marking to market. A futures contract is worth zero when entered into; however, each investor is required to deposit funds into a margin account. The amount that should be deposited when the contract is entered into is known as the initial margin. At the end of each trading day, the balance of the investor's margin account is adjusted in a way that reflects daily movements of the futures prices.

The Euribor futures is a futures contract with a Euribor deposit as the underlying asset. Since 1 January 1999, the Euribor has been used as the European money market reference rate for the unsecured market. Unlike Euro swap contracts, Euribor futures are liquidly traded in major exchanges. The 1-month and 3-month Euribor futures have been traded on the derivatives market since December 1998. The 3-month Euribor future is a contract to engage in a three month loan or deposit a face value of 1.000.000 Euro. The futures price is quoted on a daily basis for the delivery months March, June, September and December, in each case for the 3rd Wednesday of that month. The last trading day of a futures contract is always two exchange trading days prior to the relevant settlement day.

5.5.2 Data Description

Our data consists of quoted Euribor futures prices corresponding to five nearest delivery months. The data set contains Euribor futures prices traded in the Euronext-Liffe between 2 January 2004 to 30 May 2005 downloaded from Econstats. We divided the data into two subsets. The calibration set containing futures prices from 2 January 2004 until 30 December 2004, and the validation set which contains the remaining data up to 30 May 2005. In order to apply the maximum likelihood method described in the earlier chapter, we cannot use the quoted futures prices directly because under the exponential-affine framework the futures price cannot be expressed as an affine function of the interest rate factors. We need a quantity that has an affine relation with the interest rate factors. This will be derived below.

One notable feature of the Euribor futures is that the actual margin adjustment is not based on the *quoted futures price* but based on transformed prices which is termed *actual futures price*. At the delivery of the contract, T, the quoted futures price is defined in terms of the prevailing three-month Euribor rate according to the relation

$$F_q\left(T,T\right) = 1 - r_e(T, T+0.25) \tag{5.21}$$

where $r_e(t, t+\tau)$ is the time t value of the 12τ months Euribor rate. Margin adjustment is given by the actual futures price $F_a(t,T)$ which is defined as:

$$F_a\left(t,T\right) = 1 - 0.25\left(1 - F_q\left(t,T\right)\right). \tag{5.22}$$

This implies, at the delivery time T, the final settlement is

$$F_a\left(T,T\right) = 1 - 0.25\left(1 - \left(1 - r_e(T,T+0.25)\right)\right). \tag{5.23}$$

By substituting

$$r_e(T,T+0.25) = \frac{1}{0.25}\left(\frac{1}{B(T,T+0.25)} - 1\right) \tag{5.24}$$

we obtain

$$F_a\left(T,T\right) = 2 - \frac{1}{B(T,T+0.25)}. \tag{5.25}$$

Thus, the actual futures price at time t is given by

$$F_a\left(t,T\right) = 2 - \mathbf{E}^{\mathbb{Q}}\left[\frac{1}{B(T,T+0.25)}\,\middle|\,\mathcal{F}(t)\right]. \tag{5.26}$$

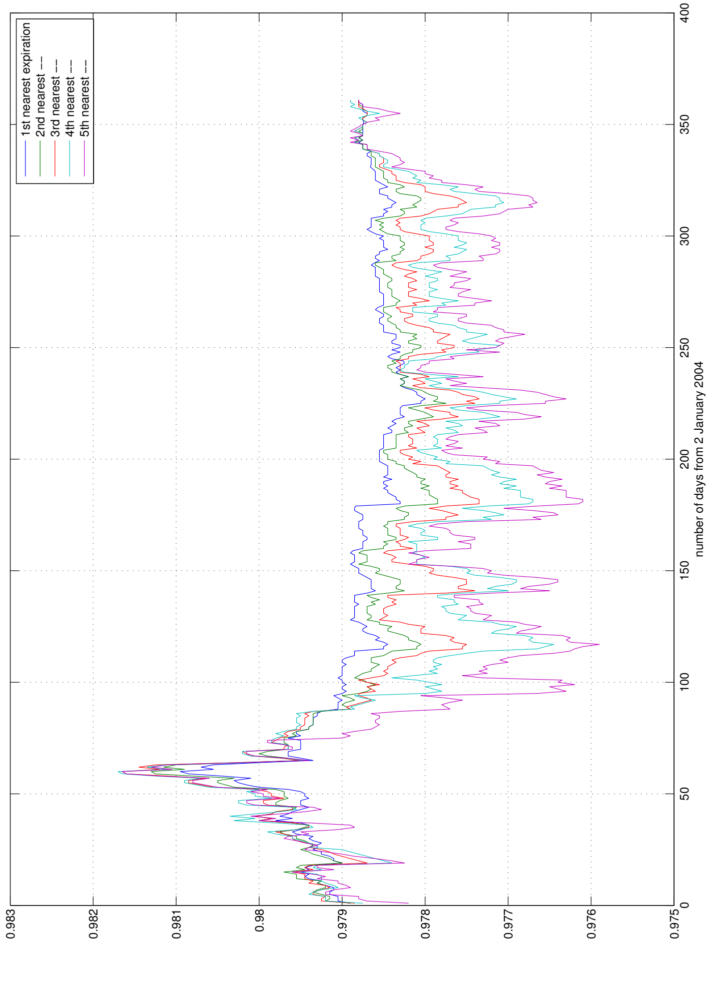

Figure 5.5: Euribor futures prices from 2 January 2004 until 30 May 2005.

where $\mathcal{F}(t)$ is the σ-algebra generated by the interest rate factors $\{X(s), 0 \leq s \leq t\}$. Thus, let us define

$$y_f(t, t+\tau(t)) = \log(2 - F_a(t, t+\tau(t))) = \log \mathbf{E}^{\mathbb{Q}} \left[\frac{1}{B(T, T+0.25)} \,\middle|\, \mathcal{F}(t) \right]. \quad (5.27)$$

where $\tau(t)$ denotes the remaining time to delivery. In the exponential affine framework, $y_f(t, t+\tau(t))$ can be written as affine function of the interest rate factors. Thus, a candidate for the observation in our estimation framework.

5.5.3 Measurement Equation

The quantity $y_f(t, t+\tau(t))$ can be further simplified to

$$y_f(t, t+\tau(t)) = \log \mathbf{E}^{\mathbb{Q}} \left[e^{-C_b(0.25)X(T) - D_b(0.25)} \,\middle|\, \mathcal{F}(t) \right] \quad (5.28)$$

where $C(\tau)$ and $D(\tau)$ satisfy

$$\begin{aligned} \frac{d}{d\tau} C_b(\tau) &= -g_1 + A^* C_b(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_b(\tau)^* \Sigma)_k \right]^2 \mathcal{A}^*_{k,\cdot} \\ \frac{d}{d\tau} D_b(\tau) &= -g_0 + B^* C_b(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_b(\tau)^* \Sigma)_k \right]^2 \mathcal{B}_k \end{aligned} \quad (5.29)$$

with boundary conditions $C(0) = \vec{0}$ and $D(0) = 0$. Thus,

$$y_f(t, t+\tau(t)) = C_f(\tau(t))\, X(t) + D_f(\tau(t)) \quad (5.30)$$

where

$$\begin{aligned} \frac{d}{d\tau} C_f(\tau) &= A^* C_f(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_f(\tau)^* \Sigma)_k \right]^2 \mathcal{A}^*_{k,\cdot} \\ \frac{d}{d\tau} D_f(\tau) &= B^* C_f(\tau) + \frac{1}{2} \sum_{k=1}^{n_X} \left[(C_f(\tau)^* \Sigma)_k \right]^2 \mathcal{B}_k \end{aligned} \quad (5.31)$$

with boundary conditions $C_f(0) = -C_b(0.25)$ and $D_f(0) = -D_b(0.25)$.

Given a set of delivery times $\{T_1, \ldots, T_{n_y}\}$, $n_y = 5$, let $\tau_j(t) = T_j - t$, $j = 1, \ldots, n_y$ be the remaining times to delivery of the futures. At each time instant t we construct an observation vector containing futures prices and the yield implied from 3 months Euribor. Let $y_e(t, t+0.25)$ be the yield implied from 3-months Euribor,

$$y_e\,(t, t+0.25) = -\frac{1}{\alpha_e(t, t+0.25)} \log \frac{1}{1 + r_e\,(t, t+0.25)\;\alpha_e\,(t, t+0.25)}.$$

The vector of observations is given by

$$\vec{y}^{obs}(t) = \begin{bmatrix} y_f^{obs}(t, t+\tau_1(t)) \\ \vdots \\ y_f^{obs}(t, t+\tau_{n_y}(t)) \\ y_e^{obs}(t, t+\tau_e) \end{bmatrix} \tag{5.32}$$

where $y_f^{obs}(t, t+\tau)$ and $y_e^{obs}(t, t+0.25)$ denote the observed values of $y_f(t, t+\tau)$ and $y_e(t, t+0.25)$ respectively. Thus, the matrices $\vec{C}(t)$ and $\vec{D}(t)$ are defined as

$$\vec{C}(t) = \begin{bmatrix} C_f(\tau_1(t))^* \\ \vdots \\ C_f(\tau_{n_y}(t))^* \\ C_b(0.25) \end{bmatrix}, \ \vec{D}(t) = \begin{bmatrix} D_f(\tau_1(t)) \\ \vdots \\ D_f(\tau_{n_y}(t)) \\ D_b(0.25) \end{bmatrix}. \tag{5.33}$$

5.5.4 Estimation Results

We estimated the parameters of BDFS and Chen models by maximizing the likelihood functional (5.9) given the vector of observations (5.32) and the matrices $\vec{C}(t)$ and $\vec{D}(t)$ in (5.33) . The noise covariance matrix R is assumed to be $(0.1 \text{ bp})^2 \mathbf{I}_{6\times 6}$. The parameter estimates are reported in Table (5.2). Fitting performance for the calibration data set and the validation data set are given in Figures (5.7) and (5.8) respectively. As seen from the figure, the fitting performance of both models are comparable. In Figure (5.6), both models are able to capture the Euribor futures prices with excellent accuracy. Furthermore, the models are able to track the underlying 3-month Euribor reasonably well as depicted in Figure (5.10).

5.6 Conclusions

In this chapter we have estimated 3-factor exponential affine models to the Euro swap rates and the Euribor futures data. The results are encouraging, showing

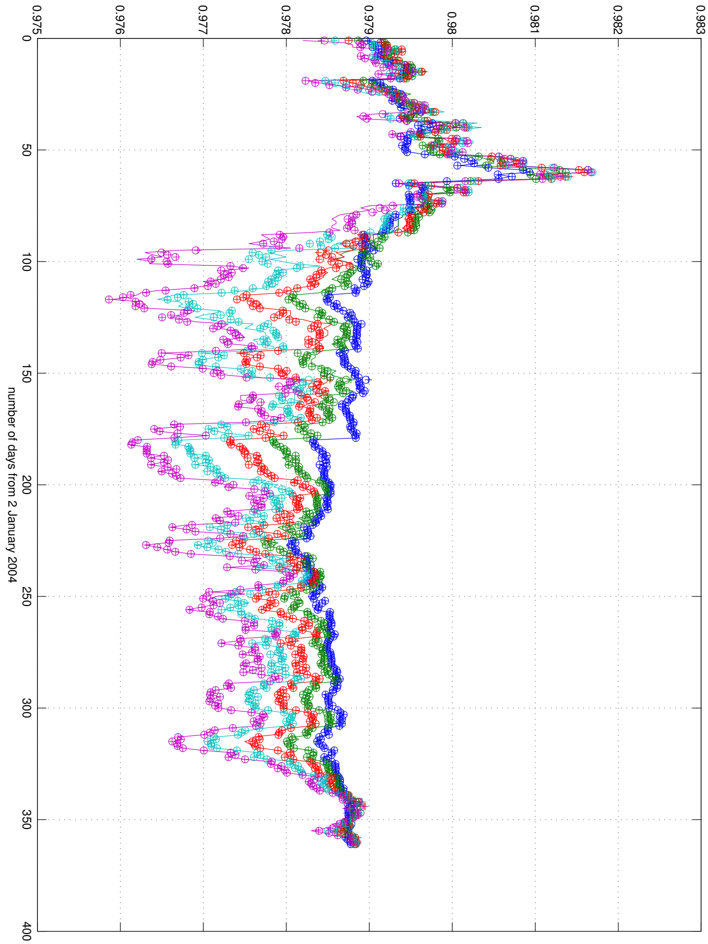

Figure 5.6: Fitted Euribor futures prices from 2 January 2004 until 30 May 2005. The solid line represents the observed futures prices, while filtered estimates obtained from the BDFS and Chen model are represented by the circles and the crosses respectively.

Parameter	Model	
	BDFS	Chen
κ	1.83898	2.19415
ν	0.00480	0.00567
$\bar{\theta}$	0.03463	0.42005
ζ	0.00171	0.03858
μ	1.75782	1.87003
$\bar{v}$	0.00004	0.02297
η	0.01222	0.01093
λ_1	0.00175	-0.08109
λ_2	-0.00461	-0.10111
λ_3	0.00240	-0.03922

Table 5.2: Parameter estimates of the Euribor futures.

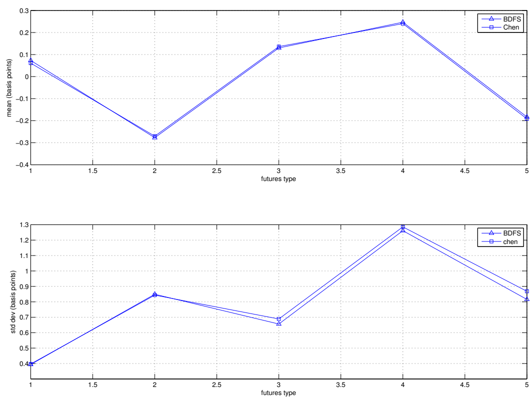

Figure 5.7: Mean and standard errors between the fitted and observed Euribor futures prices from 2 January 2004 until 30 December 2004 (calibration dataset).

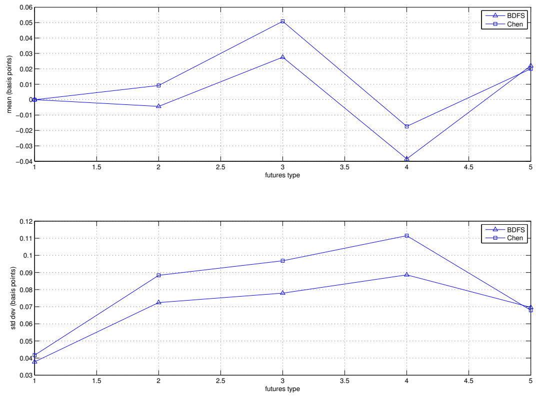

Figure 5.8: Mean and standard errors between the fitted and observed Euribor futures prices from 2 January 2005 until 30 May 2005 (validation dataset).

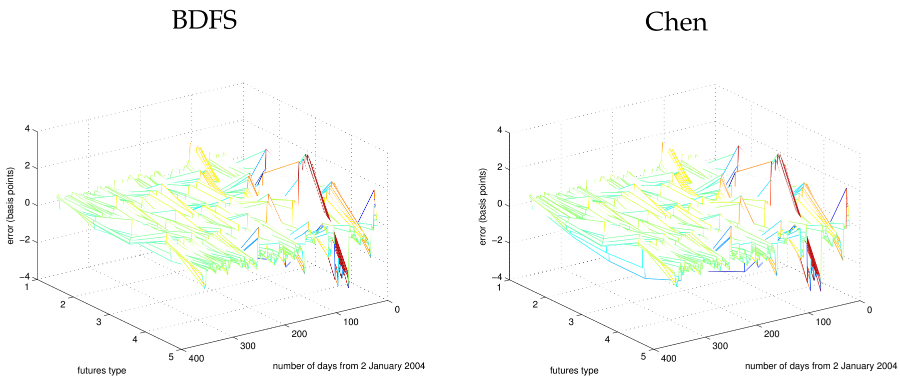

Figure 5.9: Errors between the fitted and observed Euribor futures prices from 2 January 2002 until 30 May 2005.

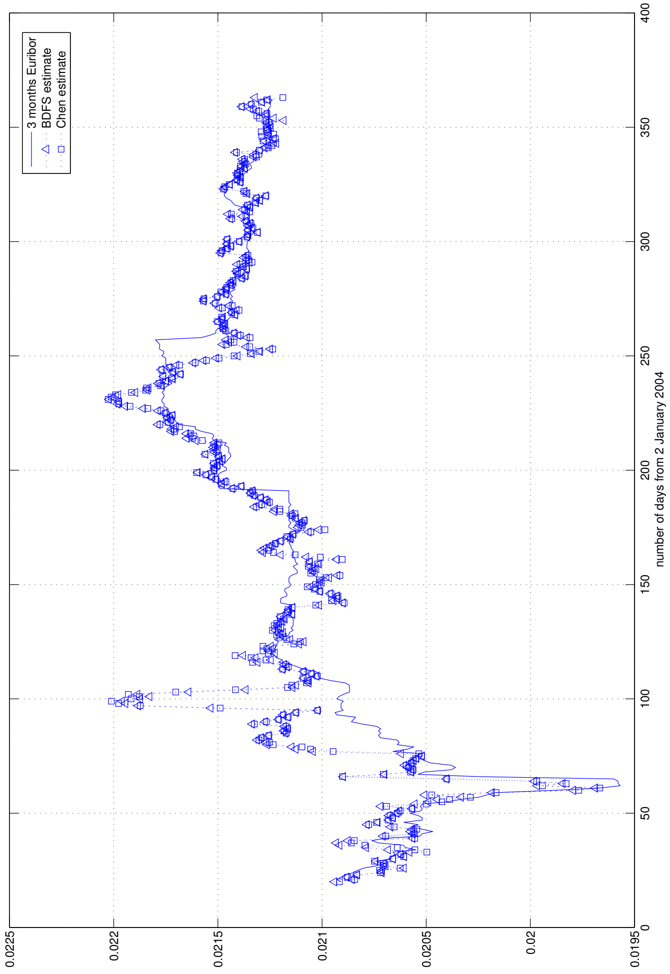

Figure 5.10: Fitted 3 month Euribor rates from 2 January 2004 until 30 May 2005.

that both models are capable in explaining the movement of the term structure of the Euro swap rates and the Euribor futures. For the Euro swap data, considering the larger cross-sectional data that was fitted, we found the models adequately explain the movement of the swap rates. Although both models encounter difficulty in fitting simultaneously the short and long end of the term structure. For the Euribor futures, the three factor models fit the data exceptionally well and able to capture simultaneously the movement of the underlying 3 month Euribor rates.

Conclusions and Future Research 6

6.1 Conclusions

In this thesis we have presented a continuous-time maximum likelihood method to estimate the exponential-affine term structure models. The method rests on the assumption that observations are affine functions of interest rate factors, and observed with additive Gaussian noise with known covariance matrix. This place our estimation method in the state-space approach of term structure estimation. Unlike methods found in literature, we do not discretize either the interest rate model nor the observation model. Within the framework of exponential-affine model that we presented here, maximum likelihood estimates can be obtained with an exact likelihood functional and not by an approximate quasi-likelihood functional.

The most common approach in formulating observations in continuous-time is to write the integrated observations as an stochastic differential equation. Thus, the additive noise is essentially additive Brownian motion to the integrated observations. Furthermore, under this assumption the likelihood functional will involve stochastic integration of the filtered estimate w.r.t. the integrated observation. Due to the stochastic integral term, computation of the likelihood functional is difficult. Another problem arise because real data cannot be non differentiable everywhere. One way to circumvent the difficulty, due to Balakrishnan (1977), is to formulate the measurement in the white noise framework. The resulting robust likelihood functional can be immediately applied to real data.

The computation of the likelihood functional will involve computation of filtered estimates of the interest rate factors. We found the filtering procedure to be the most challenging part of this research. In principle, filtered estimates of the interest rate factors and other higher conditional moments can be found by numerically solving the Zakai equation. For nonlinear model, the CIR model, it is feasible to compute the filtering densities through Zakai equation. However, it does come with a few quirks, mainly when the noise covariance is assumed to be small. When the noise covariance is small, the splitting-up method which is outlined in Chapter 3 becomes very inefficient. First of all, for the prediction step, one will need to use small space discretization because the conditional densities are very thin. Secondly, for the update step, one will need to use small time discretization due to limitation of the machine precision to handle large exponential values. These problems essentially render the use of exact nonlinear filtering impractical for multi-factor models.

Due to the large computation time imposed by the use of nonlinear filtering, we propose a Gaussian approximation to the nonlinear filter. Using Gaussian approximation, the filtering procedure amounts to solving a system of ordinary differential equations. The ODE are similar to the (continuous-time) Kalman filtering except that the equations between the conditional means and conditional covariance are coupled. In experiments, we found that a small noise covariance presents quite a challenge to standard ODE solvers. Using these standard solvers, it is very difficult, if not impossible, to obtain a stable numerical solution to the filtering ODE. In Chapter 3, we included a backward differentiation scheme to solve the filtering ODE. Unlike off-the-shelves ODE solver that are not designed specifically for the Gaussian filtering ODE, in our numerical scheme we have taken advantage of the continuous-time Riccati equation appearing in the finite difference equation. In experiments, we have found the numerical scheme to be very stable.

In Chapter 4, we have conducted Monte-Carlo analysis of each of the components of the continuous-time maximum likelihood method. First is the comparison between the exact nonlinear filtering and the approximate Gaussian filtering. We found the Gaussian filtering to be quite accurate. The filtered estimates resulted from the approximate filter is very close to the filtered estimates from the exact nonlinear filter. Comparison of the likelihoods confirms that the robust likelihood is indeed better than the Itô likelihood. Furthermore, the difference between the estimates obtained using the nonlinear filter and those obtained using the Gaussian filter is generally small, although, Wilcoxon test have indicated some mixed results.

In Chapter 5, we applied the continuous-time estimation method to the euro swap rates and the Euribor futures. For the euro swap rates, the parameter estimation is done using the yield implied from the swap rates. We compared two 3-factor models of interest rate, BDFS and Chen models. The result indicated that both model adequately captures the movement of the swap rates, although none can simultaneously capture the short and long end of the swap rate curve. The result for Euribor futures is quite encouraging. Both models fit the futures prices very well as evident by their small pricing errors. Both models are also able to track the underlying 3-month Euribor rates.

6.2 Directions for Further Research

We first note that the application of the method presented here is not limited to the term structure of interest rates. There are a number of asset pricing models that can be categorized as exponential-affine, e.g. the Heston stochastic volatility model. By applying the Heston model in the commodity, foreign exchange or energy markets, estimation of parameters using the method described in this thesis is possible using observations from forward or futures prices.

The assumption that the noise covariance is known is the main limitation of the method presented in this thesis. Unfortunately for the general exponential-affine model, there are no exact guidelines in finding the noise covariance. However, for the Gaussian models, it is possible to include the noise covariance as part of the unknown parameters. The details of the procedure and proof of convergence can be found in (Bagchi (1975)). One may conduct Monte-Carlo studies in order to provide indications that the method will work for the more general exponential-affine. We did not pursue this possibility in the thesis due the difficulty in finding parameter estimates that (globally) maximize the likelihood functional. Adding more unknowns to the optimization procedure will only worsen the problem.

We found the optimization of the likelihood functional to be quite challenging. In optimizing the likelihood functional, simulated annealing, grid search and multi-starting points with local optimization procedure have been used in the literature. In this thesis, we have adopted the last approach. Parameter estimates of real-data reported in chapter 4 and 5 are parameter estimates that lead to the highest likelihood among 100 locally optimizing parameter estimates. Experiments indicated that it is difficult to estimate the parameter that has little influence to the observation. For example, in the case of yield observations in the CIR model, the parameter κ does not really affect the value of yield. The parameter $\kappa + \lambda$ is. As a result, there are many local minima's that lead to almost similar value of $\hat{\kappa} + \hat{\lambda}$. Identifying an equivalent model where the risk neutral drift is $\tilde{\kappa}$ while the drift under the real measure is $\tilde{\kappa} - \lambda$ helps. Optimization of the likelihood is increasingly difficult when the noise covariance decreases. Assuming that we know the true noise covariance, if one is able to measure the effect of assigning a larger noise covariance to the resulting parameter estimates, it may be possible to use parameter estimates obtained using larger covariance as the starting point to the actual estimation with the true (smaller) noise covariance.

A Parameterization of BDFS and Chen Models

Model	Additional	Stationarity	Boundary
BDFS	$\bar{\theta} > 0, \bar{v} > 0$	$\kappa > 0, \nu > 0, \mu > 0$	$2\mu\bar{v} - \eta^2 > 0$
	$\zeta > 0, \eta > 0$	$\mu - \lambda_3\eta > 0$	
Chen	$\bar{\theta} > 0, \bar{v} > 0$	$\kappa > 0, \nu > 0, \mu > 0$	$2\nu\bar{\theta} - \zeta^2 > 0$
	$\zeta > 0, \eta > 0$	$\nu - \lambda_2\zeta > 0, \mu - \lambda_3\eta > 0$	$2\mu\bar{v} - \eta^2 > 0$

Table A.1: Parameter constraints for BDFS and Chen models

Model	$A^{\mathbb{P}}$	$B^{\mathbb{P}}$	A	B
BDFS	$\begin{bmatrix} -\kappa & \kappa & \lambda_1 \\ 0 & -\nu & 0 \\ 0 & 0 & -\mu+\lambda_3\eta \end{bmatrix}$	$\begin{bmatrix} 0 \\ \nu\bar{\theta}+\lambda_2\zeta \\ \mu\bar{v} \end{bmatrix}$	$\begin{bmatrix} -\kappa & \kappa & 0 \\ 0 & -\nu & 0 \\ 0 & 0 & -\mu \end{bmatrix}$	$\begin{bmatrix} 0 \\ \nu\bar{\theta} \\ \mu\bar{v} \end{bmatrix}$
Chen	$\begin{bmatrix} -\kappa & \kappa & \lambda_1 \\ 0 & -\nu+\lambda_2\zeta & 0 \\ 0 & 0 & -\mu+\lambda_3\eta \end{bmatrix}$	$\begin{bmatrix} 0 \\ \nu\bar{\theta} \\ \mu\bar{v} \end{bmatrix}$	$\begin{bmatrix} -\kappa & \kappa & 0 \\ 0 & -\nu & 0 \\ 0 & 0 & -\mu \end{bmatrix}$	$\begin{bmatrix} 0 \\ \nu\bar{\theta} \\ \mu\bar{v} \end{bmatrix}$

Model	Σ	$\mathcal{A}_\xi$	$\mathcal{B}_\xi$
BDFS	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & \eta^2 \end{bmatrix}$	$\begin{bmatrix} 0 \\ \zeta^2 \\ 0 \end{bmatrix}$
Chen	$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$	$\begin{bmatrix} 0 & 0 & 1 \\ 0 & \zeta^2 & 0 \\ 0 & 0 & \eta^2 \end{bmatrix}$	$\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Table A.2: Drift and volatility parameters of BDFS and Chen models

Bibliography

[A. Bensoussan and Rascanu, 1989] A. Bensoussan, R. G. and Rascanu, A. (1989). Approximation of zakai equation by the splitting-up method. In Zabczyk, J., editor, *Stochastic Systems and Optimization*. Springer-Verlag.

[Ait-Sahalia, 1996] Ait-Sahalia, Y. (1996). Testing continuous-time models of the spot interest rate. *Review of Financial Studies*, (9):427–470.

[Ait-Sahalia and Kimmel, 1996] Ait-Sahalia, Y. and Kimmel, R. (1996). Estimating affine multifactor term structure models using closed-form likelihood expansions. Working Paper.

[Babbs and Nowman, 1999] Babbs, S. H. and Nowman, K. B. (1999). Kalman filtering of generalized vasicek term structure models. *Journal of Financial and Quantitative Analysis*, (34):115–130.

[Bagchi, 1975] Bagchi, A. (1975). Continuous time systems identifications with unknown noise covariance. *Automatica*, 11:533–536.

[Bagchi, 1994] Bagchi, A. (1994). Some recent results in finitely additive white noise theory. *Acta Applicandae Mathematicae*, 35:27–47.

[Balakrishnan, 1977] Balakrishnan, A. V. (1977). Likelihood ratios for signals in additive white noise. *Applied Mathematics and Optimization*, 3(4):341–356.

[Benes, 1981] Benes, V. E. (1981). Exact finite-dimensional filters for certain diffusions with nonlinear drift. *Stochastics*, 5:65–92.

[Bensoussan, 1992] Bensoussan, A. (1992). *Stochastic Control of Partially Observable Systems*. Cambridge University Press.

[Bernoth and von Hagen, 2003] Bernoth, K. and von Hagen, J. (2003). The performance of the euribor futures market: Efficiency and the impact of ecb policy announcements. Working Paper, University of Bonn.

[Black et al., 1990] Black, F., Derman, E., and Toy, W. (1990). A one-factor model of interest rates and its application to treasury bond options. *Financial Analyst Journal*, pages 33–39.

[Black and Karasinski, 1991] Black, F. and Karasinski, P. (1991). Bond and option pricing when shortrates are lognormal. *Financial Analyst Journl*, pages 52–59.

[Black and Scholes, 1973] Black, F. and Scholes, M. (1973). The pricing of options and corporate liabilities. *Journal of Political Economy*, 81:637–654.

[Brandt and Santa-Clara, 2001] Brandt, M. and Santa-Clara, P. (2001). Simulated likelihood estimation of diffusions with and application to exchange rate dynamics in incomplete markets. Working paper, Wharton School.

[Chen, 1995] Chen, L. (1995). *Interest Rate Dynamics, Deivatives Pricing and Risk Management*. Lecture Notes in Economics and Mathematical Systems. Springer-Verlag, 435 edition.

[Chen, 1996] Chen, L. (1996). *Stochastic Mean and Stochastic Volatility – A Three-Factor Model of the Term Structure of Interest Rates and Its Applications to the Pricing of Interest Rate Derivatives*. Blackwell Publishers, Oxford, U.K.

[Chen and Scott, 1993] Chen, R. and Scott, L. (1993). Maximum likelihood estimation for a multifactor equiblirium model of the term structure of interest rates. *Journal of Fixed Income*, 3:14–31.

[Chen and Scott, 1995] Chen, R. and Scott, L. (1995). Interest rate options in multifactor cox-ingersoll-ross models of the term structure. *Journal of Fixed Income*, pages 53–72.

[Constantinides, 1992] Constantinides, G. (1992). A theory of the nominal term structure of interest rates. *Review of Financial Studies*, 5:531–552.

[Cox et al., 1985] Cox, J., Ingersoll, J., and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, 53:385–408.

[Dai and Singleton, 2000] Dai, Q. and Singleton, K. (2000). Specification anlysis of affine term structure models. *Journal of Finance LV*, pages 1943–1978.

[Duan and Simonato, 1999] Duan, J. C. and Simonato, J. G. (1999). Estimating and testing exponential-affine term structure models by kalman filter. *Review of Quantitative Finance and Accounting*, 13.

[Duffie and Kan, 1996] Duffie, D. and Kan, R. (1996). A yield-factor model of interest rates. *Mathematical Finance*, 6(4):379–406.

[Fackler, 2000] Fackler, P. L. (2000). Multi-factor option pricing. Working Paper, North Carolina State University.

[Florchinger, 1991] Florchinger, P. (1991). Time discretization of the zakai equation for diffuaion processes observed in correlated noise. *Stochastics and Stochastic Reports*, 35:233–256.

[Geyer and Pichler, 1999] Geyer, A. L. J. and Pichler, S. (1999). A state-space approach to estimate and test multifactor cox-ingersol-ross models of the term structure. *Journal of Financial Research*, (23):107–130.

[Hull, 2003] Hull, J. C. (2003). *Options, Futures, and Other Derivatives*. Prentice Hall Finance Series. Prentice Hall, 5 edition.

[Ito, 2000] Ito, K. (2000). Gaussian filter for nonlinear filtering problems. In *Proceedings of the 39th IEEE Conference on Decision and Control*, pages 1218–1223.

[Ito and Rozovskii, 2000] Ito, K. and Rozovskii, B. (2000). Approximation of the kushner equation for nonlinear filtering. *SIAM Journal on Control and Optimization*, 38(3):893–915.

[J. C. Cox and Ross, 1985] J. C. Cox, J. I. and Ross, S. (1985). A theory of the term structure of interest rates. *Econometrica*, (53):385–407.

[James and Webber, 2002] James, J. and Webber, N. (2002). *Interest Rate Modelling*. John Wiley and Sons.

[Jegadeesh and Pennacchi, 1996] Jegadeesh, N. and Pennacchi, G. G. (1996). The behavior of interest rates implied by the term structure of eudollar futures. *Journal of Money, Credit and Banking*, (28):426–446.

[Jong and Santa-Clara, 1999] Jong, F. D. and Santa-Clara, P. (1999). Time-series and cross-section information in the affine term structure models. *Journal of Financial and Quantitative Analysis*, (34):131–157.

[Kailath, 1998] Kailath, T. (1998). Detection of stochastic processes. *IEEE Transactions on Information Theory*, 44(6):2230–2258.

[Kallianpur and Karandikar, 1988] Kallianpur, G. and Karandikar, R. L. (1988). *White Noise Theory of Prediction, Filtering and Smoothing*. Gordon and Breach, London.

[Kushner, 1967] Kushner, H. J. (1967). Dynamical equations for optimal nonlinear filtering. *Journal of Differential Equations*, 3:179–190.

[Kushner and Budhiraja, 2000] Kushner, H. J. and Budhiraja, A. S. (2000). A nonlinear filtering algorithm based on an approximation of the conditional distribution. *IEEE Transactions Automatic Control 45*, 3:580–585.

[Lund,] Lund, J. Non-linear kalman filtering techniques for term structure models. Working Paper, Aarhus School of Business.

[Mehra, 1971] Mehra, R. K. (1971). On-line identification of linear dynamic systems with applications to kalman filtering. *IEEE Transactions on Automatic Control*, AC-16(1):12–20.

[Mitter and Moro, 1983] Mitter, S. K. and Moro, A., editors (1983). *Nonlinear Filtering and Stochastic Control*. Number 972 in Lecture Notes in Mathematics. Springer-Verlag.

[Musiela and Rutkowski, 1998] Musiela, M. and Rutkowski, M. (1998). *Martingale Methods in Financial Engineering*, volume 36 of *Applications of Mathematics*. Springer-Verlag.

[Øksendal, 2000] Øksendal, B. (2000). *Stochastic Differential Equations: An Introduction with Applications*. Universitext. Springer-Verlag, 5-th edition.

[P. Balduzzi and Sundaram, 1996] P. Balduzzi, S. R. F. and Sundaram, R. (1996). A simple approach to three factor affine term structure models. *Journal of Fixed Income*, (6):43–53.

[Pennacchi, 1991] Pennacchi, G. G. (1991). Identifying the dynamics of real interest rates and inflation: Evidence using survey data. *Review of Financial Studies*, (4):53–86.

[Piazzesi, 2003] Piazzesi, M. (2003). Affine term structure models. http://gsbwww.uchicago.edu/fac/monika.piazzesi/research/.

[Poor, 1994] Poor, H. V. (1994). *An introduction to signal detection and estimation*. Springer texts in electrical engineering. Springer, second edition.

[Protter, 1990] Protter, P. (1990). *Stochastic Integration and Differential Equations: A New Approach*, volume 21 of *Applications of Mathematics*. Springer-Verlag.

[Remonola and Wooldridge, 2003] Remonola, E. M. and Wooldridge, P. D. (2003). The euro interest rate swap market. BIS Quarterly.

[Stratonovich, 1968] Stratonovich, R. L. (1968). *Conditional Markov Processes and Their Applications to the Theory of Optimal Control*, pages 113–142. Elsevier.

[Vasicek, 1977] Vasicek, O. A. (1977). An equilibrium characterization of the term structure. *Journal of Financial Economics*, (5):177–188.

Summary

This thesis addresses the problem of parameter estimation of the exponential-affine class of models, which is a class of multi-factor models for the short rate. We propose a continuous-time maximum likelihood estimation method to estimate the parameters of a short rate model, given set of observations that are linear with respect to the interest rate factors. We assume that observations are corrupted by Gaussian noise with a known covariance, which lead to a maximum likelihood estimation method for partially observed systems. Unlike other approaches in the literature, we do not discretize either the interest rate model or the observation model.

By assuming that the observations are noisy, parameter estimation involves a filtering step to estimate the unknown interest rate factors. In order to filter/estimate the unknown factors one needs to solve the Zakai partial differential equation (PDE) for the conditional density. It is possible to approximate the solution of the Zakai PDE with finite differences accurately. We have tested this approach to the CIR model.

Unfortunately, by numerically solving the Zakai PDE, the estimation method becomes impractical due to the heavy computational burden. In order to circumvent this difficulty, we propose a Gaussian approximation that resulted in coupled ordinary differential equations (ODE). For a reasonably large noise covariance, one can employ standard ODE solvers. However, it is increasingly difficult to obtain stable numerical solution when the noise covariance decreases. We develop a numerical scheme that addresses this problem. From experiments, we have found that the proposed numerical method is very stable. With the Gaussian filtering, we estimate the parameters of Chen and Balduzzi-Das-Sundaram-Foresi (BDFS) models using the Euro swap rates and Euribor future prices. Theoretical swap rates and future prices computed from estimated models are in good agreements with the quoted swap rates and futures prices.

Samenvatting

Dit proefschrift pakt het probleem aan van parameterschatting van exponentieel affiene modellen, die een klasse vormen van multifactor-modellen voor het korte rentetarief. Gegeven een reeks observaties die met betrekking tot de rentetarieffactoren lineair zijn, stellen wij een schattingsmethode voor van maximumwaar schijnlijkheid in de continue tijd om de parameters van een kort rentetariefmodel te schatten. Wij veronderstellen dat de observaties door Gaussische ruis met een bekende covariantie worden beinvloed, hetgeen tot een methode van maximumwaarschijnlijkheid schatting voor gedeeltelijk waargenomen systemen leidt. In tegenstelling tot andere benaderingen in de literatuur discretiseren wij het rentetariefmodel en het observatiemodel niet.

Door te veronderstellen dat de observaties ruis bevatten, impliceert de parameterschatting een filterende stap om zo de onbekende rentetarieffactoren te schatten. Om de onbekende factoren te filteren/schatten, moet men de Zakai partiële differentiaalvergelijking (PDV) voor de voorwaardelijke kansdichtheid oplossen. Het is mogelijk om de oplossing van de Zakai PDV nauwkeurig te benaderen. Wij hebben deze benadering op het CIR model getest.

Door de Zakai PDV numeriek op te lossen, wordt de schattingsmethode onpraktisch door de grote rekenintensiviteit. Om deze moeilijkheid te omzeilen, stellen wij een Gaussische benadering voor, die in gekoppelde gewone differentiaalvergelijkingen (GDV) resulteert. Voor een redelijk grote ruiscovariantie kan men standaard solvers op de GDV toepassen. Het wordt echter steeds moeilijker om stabiele numerieke oplossingen te verkrijgen naarmate de ruiscovariantie vermindert. We hebben een numerieke regeling ontwikkeld, die dit probleem aanpakt. We hebben vervolgens experimenteel geconstateerd dat deze numerieke methode zeer stabiel is. Met Gaussisch filteren schatten we de parameters van de Chen en Balduzzi-Das-Sundaram-Foresi (BDFS) modellen, waar we de Euro ruilmiddeltarieven en de Euribor future prijzen hebben gebruikt. De ruilmiddeltarieven en de future prijzen die door de geschatte modellen worden voorspeld, zijn in goede overeenkomst met de opgegeven ruilmiddeltarieven en de future prijzen.

Acknowledgments

First, I would like to thank Prof. Arun Bagchi for accepting me as a PhD student in his group. I also would like to thank him for his guidance and for providing the necessary enthusiasm and encouragement during the period that I have spent in Twente.

I thank the committee members; Prof. S. Aihara, Prof. F. Jamshidian, Prof. W. Albers, Dr. M. Vellekoop and Dr. D. Dupont; for their time and effort in reviewing this dissertation and for their valuable comments and remarks.

There are many people that I would like to thank for making my time in Enschede so enjoyable. Thanks go to the (former) graduate students at the Systems and Control group; Agung, Agoes, Vishi, Hendra, Ram, Javi, Jaroslav, Saikat; and to all staff members in the group. I owe a special note of gratitude to Emad Imreizeeq for his friendship. I thank the members of Twente BS group and ITC-CF support group.

Finally, I would like to thank my family and Irene for their continuing support and encouragement in difficult times. This dissertation is dedicated to my parents with love and thanks for everything they have done and for their love for me.