12-1-2014

Empirical Studies on Interest Rate Derivatives

Xudong Sun
University of Nevada, Las Vegas

EMPIRICAL STUDIES ON INTEREST RATE DERIVATIVES

by

Xudong Sun

Bachelor of Science in Mathematics
Dalian University of Technology, China
2006

Master of Science – Mathematical Sciences
Dalian University of Technology, China
2008

A dissertation submitted in partial fulfillment of
the requirements for the

Doctor of Philosophy – Mathematical Sciences

Department of Mathematical Sciences
College of Sciences
The Graduate College

University of Nevada, Las Vegas, USA
December 2014

UNLV | GRADUATE COLLEGE

We recommend the dissertation prepared under our supervision by

Xudong Sun

entitled

Empirical Studies on Internet Rate Derivatives

is approved in partial fulfillment of the requirements for the degree of

Doctor of Philosophy - Mathematical Sciences

Department of Mathematical Sciences

Hongtao Yang, Ph.D., Committee Chair

Jichun Li, Ph.D., Committee Member

Michael Marcozzi, Ph.D., Committee Member

Monika Neda, Ph.D., Committee Member

Pengtao Sun, Ph.D., Committee Member

Jianzhong Zhang, Ph.D., Graduate College Representative

Kathryn Hausbeck Korgan, Ph.D., Interim Dean of the Graduate College

December 2014

ABSTRACT

EMPIRICAL STUDIES ON INTEREST RATE DERIVATIVES

by

Xudong Sun

Dr. Hongtao Yang, Examination Committee Chair
Associate Professor of Mathematics
University of Nevada, Las Vegas, USA

Interest rate models are the building blocks of financial market and the interest rate derivatives market is the largest derivatives market in the world. In this dissertation, we shall focus on numerical pricing of interest rate derivatives, estimating model parameters by Kalman filter, and studying various models empirically.

We shall propose a front-fixing finite element method to price the American put option under the quadratic term structure framework and compare it with a trinomial tree method and common finite element method. Numerical test results show the superiority of our front-fixing finite element method in the aspects of computing the option and free boundary simultaneously with high accuracy. We shall also employ the Kalman filter and its variant techniques to estimate parameters of the affine term structure models as well as quadratic term structure models. Various comparisons of different Kalman filter performance and both the in-sample fit and out-sample fit for Monte Carlo simulations as well as real treasury yield data are presented. In addition, we shall propose a general one-factor interest rate model and apply a homotopy perturbation method to valuate bond prices. One of the attractive qualities of the approximated solution of homotopy perturbation method is its fast speed of achieving the same accuracy compared to the tree method.

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members and support from my family.

I would like to express my deepest gratitude to my advisor, Dr. Yang for his guidance, support, and expertise. I would like to thank my committee members Dr. Jichun Li, Dr. Michael Marcozzi, Dr. Monika Neda, Dr. Pengtao Sun, and Dr. Jianzhong Zhang for their time, expertise and support. I would also thank my parents and my wife Qianzhi for their constant support and encouragement over the years.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

An interest rate derivative is a derivative whose underlying asset is the right to pay or receive a notional amount of money at a given interest rate, for example, bonds and their options, swaps, captions, and floors. The interest rate derivatives are popular among investors with customized cash flow needs or specific views on the interest rate movements. The interest rate derivatives market is the largest derivatives market in the world. According to the quarterly report of the Office of the Comptroller of the currency (OCC), derivative contracts remain concentrated in interest rate products, which comprise 81% of total derivative notional amounts $236.8 trillion of the US market in the second quarter of 2014. It is clearly important to model interest rate in order to price and to understand these financial derivatives. In this dissertation, we shall study various one-factor models of interest rates and their calibration by using the observed market data.

This chapter is organized as follows. In §1.1, we introduce basic elements about the interest rate derivatives. §1.2 presents the overview of two methodologies of pricing interest rate contingent claims under a single factor framework and some popular and extensively used single factor interest rate models among the researchers and market

practitioners. At the end, a outline of this dissertation is summarized in §1.4.

1.1 Bonds, Yields and Forward Rates

A bond is a contract, paid up front, which guarantees the holder a known amount on a known date in the future. The known amount is referred as the principal value or face value and the known date as the maturity date. The bond may also pay a known cash dividend (called the coupon) at fixed times during the life of the contract. If there is no dividend payment the bond is known as a zero-coupon bond or a pure discount bond. The main purpose of issuing a bond is the raising of capital, and the up-front premium that can be thought of as a loan to the government or to the company that has issued the bond. The Securities and Financial Markets Association (SIFMA) classifies the bond market into five sectors: government and agency, corporate, municipal, mortgage backed, asset backed and collateralized debt obligation, and funding.

The U.S. bond market is the largest and most active bond market in the world. As of June 2014, the size of the outstanding U.S. bond market debt was $39.9 trillion. The most important part of the bond market is the government bond market due to its size and liquidity. Because of the inverse relation between bond prices and interest rates, the bond market is often used to indicate change in interest.

Let $B(t,T)$ denote the price of a zero-coupon bond at time t that pays one dollar at time T, i.e., $B(T,T) = 1$ for any T. At time t, the yield to maturity $y(t,T)$ of zero-coupon bond $B(t,T)$ is the continuously compounded rate of return that increases

the bond price to one at time T. Then we have

$$B(t,T) = e^{-(T-t)y(t,T)}. \tag{1.1}$$

The yield rates as a function of maturity T is also called the yield curve. We denote $r(t)$ the instantaneous risk-free interest rate that is also called short rate. The short rate can be treated as a yield rate that has an extremely short maturity, i.e.,

$$r(t) = \lim_{T \to t} y(t,T).$$

Let $f(t,T_1,T_2)$ denote the forward rate that is agreed upon at time t for a risk-free loan starting at time T_1 and ending at time T_2. The forward rate can be expressed in terms of two bond prices, i.e.,

$$f(t,T_1,T_2) = \frac{\ln B(t,T_1) - \ln B(t,T_2)}{T_2 - T_1}. \tag{1.2}$$

If $T_1 = T_2$, we have the instantaneous forward rate at time t for a loan starting at time T for an infinitesimal period of time. We have

$$f(t,T) = f(t,T,T).$$

Letting $T_2 = T$ and $T_1 \to T^+$ in (1.2), we obtain

$$f(t,T) = -\frac{\partial \ln B(t,T)}{\partial T} = -\frac{1}{B(t,T)} \frac{\partial B(t,T)}{\partial T},$$

where the bond price is assumed to be differentiable. Equivalently, the bond price can be expressed in terms of forward rates as

$$B(t,T) = e^{-\int_t^T f(t,\tau) d\tau}.$$

Thus by (1.1), we have

$$y(t,T) = \frac{1}{T-t}\int_t^T f(t,\tau)d\tau,$$

which implies that

$$r(t) = f(t,t).$$

In summary, we have presented the definitions of zero-coupon bond, yield rates, forward rates, and the short rates as well as the relations between any two of them. Actually, the above relations show us the perfectly equivalent ways of expressing the same information. For instance, if a complete term structure of forward rates is known, we can compute the zero-coupon bond price and spot rate.

1.2 One-Factor Models of Interest Rates

This section is devoted to outline the approaches of single factor models of interest rates and to review most popular and widely used models. All these models use one single specific factor as the sole state variable to summarize all the information about the term structure at any time. As a result, the price of any interest rate contingent claim will be affected by only short term rate and the time to maturity. For instance, at time t, the price of a zero-coupon bond maturing at time T $(T > t)$ has the form of

$$B(t,T) = B(t,T,r(t)).$$

There are two basic methodologies of pricing interest rate contingent claims under a single factor framework being extensively used by researchers and market practitioners, namely the partial differential equation approach and the martingale approach.

The equivalency of these two approaches can be shown by the Feynman-Kac Theorem. A brief review of these two approaches will be given in the following two subsections.

1.2.1 PDE approach of pricing interest rate derivatives

Let us assume that the dynamics of the short interest rates is governed by the diffusion process

$$dr(t) = \mu(t, r(t))dt + \sigma(t, r(t))dW(t), \tag{1.3}$$

where $W(t)$ is a one dimensional standard Brownian motion under the real-world measure $\mathbb{P}$, $\mu(\cdot, \cdot)$ and $\sigma(\cdot, \cdot)$ are given real valued functions that totally determine the behavior of the short rate. Let $V(t)$ denote the value of an interest rate contingent at time t. Resulting from the one factor assumption, $V(t)$ depends on the short rate $r(t)$. We write

$$V(t) = V(t, T, r(t)). \tag{1.4}$$

Applying Ito's lemma to $V(t)$, we have

$$dV(t) = \frac{\partial V}{\partial r}dr + \frac{\partial V}{\partial t}dt + \frac{1}{2}\frac{\partial^2 V}{\partial r^2}(dr)^2.$$

Substituting (1.3) into the above equation gives us

$$dV(t) = \left[\frac{\partial V}{\partial t} + \mu(t, r(t))\frac{\partial V}{\partial r} + \frac{\sigma(t, r(t))}{2}\frac{\partial^2 V}{\partial r^2}\right] + \left[\frac{\partial V}{\partial r}\sigma(t, r(t))\right] dW(t).$$

If we divide both sides by $V(t)$, then we have the instantaneous relative return on the contingent claim:

$$\frac{dV(t)}{V(t)} = \mu_V dt + \sigma_V dW(t), \tag{1.5}$$

where

$$\mu_V = \frac{1}{V(t)}\left[\frac{\partial V}{\partial t} + \mu(t, r(t))\frac{\partial V}{\partial r} + \frac{\sigma(t, r(t))}{2}\frac{\partial^2 V}{\partial r^2}\right], \tag{1.6}$$

$$\sigma_V = \frac{1}{V(t)}\left[\frac{\partial V}{\partial r}\sigma(t, r(t))\right]. \tag{1.7}$$

Now, consider a portfolio consisted of x_1 units of the interest rate derivative $V_1(t) = V(t, T_1, r(t))$ and x_2 units of the interest rate derivative $V_2(t) = V(t, T_2, r(t))$. The portfolio value satisfies the following process:

$$P(t) = x_1 V_1(t) + x_2 V_2(t).$$

As V_1 and V_2 are interest rate contingent claims, their prices have forms of (1.5), i.e.,

$$\frac{dV_1(t)}{V_1(t)} = \mu_{V_1} dt + \sigma_{V_1} dW(t),$$
$$\frac{dV_2(t)}{V_2(t)} = \mu_{V_2} dt + \sigma_{V_2} dW(t).$$

Therefore, the variations of the portfolio value are given by

$$\begin{aligned} dP(t) &= x_1 dV_1(t) + x_2 dV_2(t) \\ &= (x_1 \mu_{V_1} V_1 + x_2 \mu_{V_2} V_2)\, dt + (x_1 \sigma_{V_1} V_1 + x_2 \sigma_{V_2} V_2)\, dW(t). \end{aligned}$$

In order to get a risk-neutral position, we need to choose x_1 and x_2 to reduce the volatility of $dP(t)$ to 0. According to arbitrage-free theory, the return rate of the portfolio has to be the same as the risk-free rate. Thus, the following system must be satisfied:

$$x_1 \sigma_{V_1} V_1 + x_2 \sigma_{V_2} V_2 = 0,$$

$$x_1 \mu_{V_1} V_1 + x_2 \mu_{V_2} V_2 = (x_1 V_1 + x_2 V_2) r.$$

The system has a non trivial solution if and only if

$$\frac{\mu_{V_1} - r(t)}{\sigma_{V_1}} = \frac{\mu_{V_2} - r(t)}{\sigma_{V_2}},$$

which must hold for any T_1 and T_2. So each side of the above equation must be independent of the contingent claim and we denote it by

$$\lambda(t, r(t)) = \frac{\mu_V - r(t)}{\sigma_V}.$$

Here $\lambda(t, r(t))$ is called the market risk premium. After substituting μ_V and σ_V by their definitions in (1.6) and (1.7), we end up with a second order parabolic partial differential equation

$$\frac{\partial V}{\partial t} + (\mu(t, r(t)) - \lambda(t, r(t))\sigma(t, r(t)))\frac{\partial V}{\partial r} + \frac{\sigma(t, r(t))^2}{2}\frac{\partial^2 V}{\partial r^2} - r(t)V = 0. \qquad (1.8)$$

We can conclude that all interest rate contingent claim in a no-arbitrage one factor model must satisfy the fundamental equation (1.8). Different models have different choices of input functions $\lambda(t, r(t))$, $\mu(t, r(t))$ and $\sigma(t, r(t))$, while different interest rate contingent claims will satisfy the same type of partial differential equation with different boundary conditions. For instances, if we consider V as a zero-coupon bond $B(t, T)$ with maturity T, then we have

$$\frac{\partial B}{\partial t} + (\mu(t, r(t)) - \lambda(t, r(t))\sigma(t, r(t)))\frac{\partial B}{\partial r} + \frac{\sigma(t, r(t))^2}{2}\frac{\partial^2 B}{\partial r^2} - r(t)B = 0 \qquad (1.9)$$

with the final condition

$$B(T, T) = 1. \qquad (1.10)$$

If we consider V as a call option on a zero-coupon bond $B(t, T)$ with maturity date $T_C < T$, then we have

$$\frac{\partial C}{\partial t} + (\mu(t, r(t)) - \lambda(t, r(t))\sigma(t, r(t)))\frac{\partial C}{\partial r} + \frac{\sigma(t, r(t))^2}{2}\frac{\partial^2 C}{\partial r^2} - r(t)C = 0 \qquad (1.11)$$

with the final condition

$$C(T_C) = (B(t, T_C) - K)^+$$

where K is the strike price.

Theoretically, we can price zero-coupon bond by solving the final value problem (1.9)–(1.10) if the real valued functions $\mu(t, r(t))$, $\sigma(t, r(t))$ and $\lambda(t, r(t))$ are given. Functions $\mu(t, r(t))$ and $\sigma(t, r(t))$ can be specified by examining long-term statistical properties of the short rate. However, specifying $\lambda(t, r(t))$ is harder due to its non-observable property.

If we apply Feynman-Kac Theorem to equation (1.9), then we are able to express the price of zero-coupon bond as an expectation:

$$B(t, T) = \mathrm{E}^{\mathbb{P}}\left[e^{-\int_t^T r(s)ds - \frac{1}{2}\int_t^T \lambda^2(s,r(s))ds - \int_t^T \lambda(s,r(s))dW(s)} | \mathbb{F}_t\right], \qquad (1.12)$$

where $\mathbb{F}_t$ is the sigma-algebra generated by the past information of process $W(t)$ up to time t under the $\mathbb{P}$ measure.

1.2.2 Martingale approach of pricing interest rate derivatives

The key concept of martingale approach is the equivalent martingale measure. Let $\mathbb{Q}$ be the risk-neutral measure, under which all discounted security prices are martingales. Let us assume that the short rate $r(t)$ under the risk-neutral measure $\mathbb{Q}$ is given by

$$dr(t) = \alpha(t, r(t))dt + \beta(t, r(t))dW^{\mathbb{Q}(t)}, \qquad (1.13)$$

where $W^{\mathbb{Q}}(t)$ is a standard Brownian motion under $\mathbb{Q}$ measure. Then at time t, the price of a traded security $V(t)$ is given by

$$V(t) = D(t)\mathrm{E}_t^{\mathbb{Q}}\left[\frac{V(T)}{D(T)}\right], \tag{1.14}$$

where $D(t) = e^{\int_0^t r(s)ds}$. For instance, the zero-coupon bond price is given by

$$B(t,T) = \mathrm{E}_t^{\mathbb{Q}}\left[e^{-\int_t^T r(s)ds}\right].$$

The equivalency between PDE approach and martingale approach can be established by the Feynman-Kac representation and Girsanov's Theorem. Without loss of generality, let us consider the case of non-dividend paying security price $V(t)$ at time t, which satisfies equation (1.14). According to the Feynman-Kac formula, $V(t)$ satisfies the following PDE:

$$\frac{\partial V}{\partial t} + \frac{1}{2}\beta(t,r(t))\frac{\partial^2 V}{\partial r^2} + \alpha(t,r(t))\frac{\partial V}{\partial r} - r(t)V = 0. \tag{1.15}$$

Since $V(t)$ satisfies both (1.8) and (1.15), we must have

$$\sigma(t,r(t)) = \beta(t,r(t)),$$
$$\lambda(t) = \frac{\mu(t,r(t)) - \alpha(t,r(t))}{\sigma(t,r(t))}.$$

We also have

$$dW(t) = dW^{\mathbb{Q}}(t) - \lambda(t)dt$$

by comparing (1.3) and (1.13). Further, by Girsanov's Teorem, $\mathbb{P}$-measure and $\mathbb{Q}$-measure are related by the formula

$$\frac{d\mathbb{Q}}{d\mathbb{P}} = e^{-\int_0^t \lambda(s)dW(t) - \frac{1}{2}\int_0^t \lambda^2(s)ds}.$$

1.3 Review of Some Popular Models

In this section we will review in detail most well known and widely used one factor interest rate models.

Merton (1973)

Merton was the pioneer to propose a general stochastic process as a model for short rates. He assumed that the short rate process follows the stochastic differential equation:

$$dr(t) = \mu dt + \sigma dW(t), \tag{1.16}$$

where μ and σ are constant. The explicit solution of SDE (1.16) is

$$r(t) = r(s) + \mu t + \sigma \int_s^t dW(s)$$

for $t \geq s \geq 0$. Thus the short rate $r(t)$ is normally distributed:

$$r(t)|\mathbb{F}_s \sim \mathbb{N}\left(r(s) + (t-s)\mu, (t-s)\sigma^2\right).$$

According to equation (1.9), the zero-coupon bond price satisfies

$$\frac{\partial B}{\partial t} + (\mu + \lambda\sigma)\frac{\partial B}{\partial r} + \frac{\sigma^2}{2}\frac{\partial^2 B}{\partial r^2} - r(t)B = 0$$

with the final condition $B(T,T) = 1$, where the market risk premium λ is assumed constant. This partial differential equation can be solved explicitly and the solution is

$$B(t,T) = e^{-\tau r(t) - \frac{\tau^2(\mu - \lambda\sigma)}{2} + \frac{\tau^3\sigma^2}{6}}$$

where $\tau = T - t$ is the time to maturity. Apparently, the bond price is an increasing function of the time to maturity which is unrealistic because the bond price will blow out when it has an infinite maturity date. From the bond price, one can easily obtain the yield rate

$$y(t,T) = -\frac{\ln B(t,T)}{T-t} = r(t) + \frac{(T-t)(\mu - \lambda\sigma)}{2} - \frac{(T-t)^3\sigma^2}{6}$$

Vasicek (1977)

Vasicek assumed that the short rate process follows the Ornstein-Uhlenbeck process:

$$dr(t) = \kappa(\theta - r(t))dt + \sigma dW(t) \tag{1.17}$$

where κ, θ and σ are positive constants and $W(t)$ is a standard Brownian motion. In this setting, when the short rate $r(t)$ goes apart from its long term level θ, $r(t)$ tends to come back to θ at a mean-reverting speed κ. Vasicek also makes an assumption that the market risk premium λ is constant. The solution of SDE (1.17) is

$$r(t) = \theta + (r(s) - \theta)e^{-\kappa(t-s)} + \sigma \int_s^t e^{-\kappa(t-s)} dW(u)$$

for $t \geq s \geq 0$, which implies that the short rate is normally distributed

$$r(t)|\mathbb{F}_s \sim \mathbb{N}\left(\theta + (r(s) - \theta)e^{-\kappa(t-s)}, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)})\right).$$

As a result of the normal distribution, Vasicek model suffers from the fact that the interest rate can become negative, which is not only impractical but also incompatible with no arbitrage theory.

The price of zero-coupon bond can be formulated in the following partial differ-

ential equation

$$\frac{\partial B}{\partial t} + (\kappa(\theta - r(t)) - \lambda\sigma)\frac{\partial B}{\partial r} + \frac{\sigma^2}{2}\frac{\partial^2 B}{\partial r^2} = 0$$

with the final condition $B(T,T) = 1$. The above partial differential equation can be solved explicitly with the solution

$$B(t,T) = e^{a(\tau)r(t)+b(\tau)}$$

where

$$\begin{aligned}
&a(\tau) = \frac{1}{\kappa}(e^{-\tau\kappa} - 1),\\
&b(\tau) = \frac{\sigma^2}{4\kappa^3}(1 - e^{-2\tau\kappa}) + \frac{1}{\kappa}(\theta - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{\kappa^2})(1 - e^{-\tau\kappa}) - (\theta - \frac{\lambda\sigma}{\kappa} - \frac{\sigma^2}{\kappa^2})\tau,\\
&\tau = T - t.
\end{aligned}$$

The yield rate is given by

$$y(t,T) = -\frac{1}{\tau}(a(\tau)r(t) + b(\tau)).$$

Cox, Ingersoll, Ross (1985)

Cox, Ingersoll, and Ross (CIR) choose $\mu(t, r(t))$ the same as the one in Vasicek model but they let $\sigma(t, r(t)) = \sigma\sqrt{r(t)}$ and the market risk premium $\lambda(t, r(t)) = \lambda\sqrt{r(t)}$. Thus, the short rate process satisfies

$$dr(t) = \kappa(\theta - r(t))dt + \sigma\sqrt{r(t)}dW(t) \tag{1.18}$$

where κ, θ and σ are positive constants and $W(t)$ is the standard Brownian motion. The drift factor ensures mean reversion of the short rate towards the long run term

value θ with speed of adjustment κ. The volatility factor, $\sigma\sqrt{r(t)}$ avoids the possibility of negative interest rates and an interest rate of zero is also precluded if the condition

$$2\kappa\theta \geq \sigma^2$$

is met. Since the volatility is proportional to the interest rate, the volatility becomes small when the interest rate is at a low level. As a consequence, when the interest rate gets close to zero, its evolution is dominated by the drift factor, which drags the interest rate upwards. The SDE (1.18) has a unique positive solution

$$r(t) = \theta + (r(s) - \theta)e^{-\kappa(t-s)} + \sigma e^{-\kappa(t-s)} \int_s^t e^{\kappa(u-s)} \sqrt{r(u)} dW(u)$$

for $t \geq s \geq 0$. It has been shown that the short rate $r(t)$ has a distribution of non-central chi-square, i.e.,

$$r(t)|\mathbb{F}_s \sim \chi(a, b, c)$$

with b degrees of freedom and non-central parameter c, where

$$\begin{aligned} a &= \frac{4\kappa r(t)}{\sigma^2(1 - e^{-\kappa(t-s)})}, \\ b &= \frac{2\kappa\theta}{\sigma^2}, \\ c &= \frac{4\kappa e^{-\kappa(t-s)}}{\sigma^2(1 - e^{-\kappa(t-s)})}. \end{aligned}$$

Consequently, the mean and variance of $r(t)$ given $r(s)$ are given by

$$\mathrm{E}(r(t)|r(s)) = \theta + (r(s) - \theta)e^{-\kappa(t-s)}$$

and

$$\mathrm{V}(r(t)|r(s)) = r(s)\frac{\sigma^2}{\kappa}(e^{-\kappa(t-s)} - e^{-2\kappa(t-s)}) + \theta\frac{\sigma^2}{2\kappa}\left(1 - e^{-\kappa(t-s)}\right)^2.$$

The zero-coupon bond price can be solved through PDE

$$\frac{\partial B}{\partial t} + (\kappa(\theta - r(t)) - \lambda\sigma)\frac{\partial B}{\partial r} + \frac{\sigma^2}{2}\frac{\partial^2 B}{\partial r^2} - r(t)B = 0. \tag{1.19}$$

with the boundary condition $B(T,T) = 1$. The solution of the above PDE is in form of

$$B(t,T) = e^{a(\tau) - b(\tau)r}$$

where

$$\begin{aligned}
a(\tau) &= \frac{2\kappa\theta}{\sigma^2}\ln\left(\frac{2\gamma e^{\frac{(\gamma+\kappa+\lambda)\tau}{2}}}{(\gamma+\kappa+\lambda)(e^{\gamma\tau}-1)+2\gamma}\right),\\
b(\tau) &= \frac{2(e^{\gamma\tau}-1)}{(\gamma+\kappa+\lambda)(e^{\gamma\tau}-1)+2\gamma},\\
\gamma &= \sqrt{(\kappa+\lambda)^2+2\sigma^2},\\
\tau &= T - t.
\end{aligned}$$

The yield rate can be easily expressed in terms of $a(\tau)$ and $b(\tau)$

$$y(t,T) = -a(\tau) + b(\tau)r.$$

Hull and White (1990)

One major shortcoming of the above time-invariant models is that these models cannot be calibrated to effective yield curves. To overcome these imperfections, Hull and White (1990) introduce a class of models that allow both $\mu(t, r(t))$ or/and $\sigma(t, r(t))$ be to time-dependent. The most general Hull-White model follows

$$dr(t) = (\theta(t) - \kappa(t)r(t))dt + \sigma(t)r^{\beta}(t)dW(t)$$

with risk premium

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

where λ and γ are non-negative parameters. The time varying functions $\theta(t)$, $\kappa(t)$ and $\sigma(t)$ can be calibrate exactly to current market prices. The short rate, zero-coupon bond price, and yield rate are not analytically obtainable anymore.

Under the Hull and White framework, if β is set to zero and κ and σ are positive constants, then we have the extended Vasicek model

$$dr(t) = \kappa \left(\frac{\theta(t)}{\kappa} - r(t) \right) dt + \sigma dW(t).$$

The parameter κ is considered as the adjustment speed and $\frac{\theta(t)}{\kappa}$ can be considered as the time dependent reversion level. The parameter $\theta(t)$ can be calibrated using the initial term structure as

$$\theta(t) = \frac{\partial}{\partial t} f(0, t) + \kappa y(0, t) + \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa t}).$$

The zero-coupon bond price can be explicitly expresses as

$$B(t, T) = e^{-a(t,T) - b(t,T)r}$$

with

$$\begin{aligned} b(\tau) &= \frac{1}{\kappa}(1 - e^{\kappa(\tau)}), \\ a(t, T) &= -\ln \left(\frac{B(0, T)}{B(0, t)} \right) - b(T - t) f(0, t) + \frac{\sigma^2}{4\kappa} b(T - t)^2 (1 - e^{-2\kappa t}). \end{aligned}$$

Here we have used the bond prices $B(0, \cdot)$ and the forward rates $f(0, \cdot)$ current term structure which can be computed from the market data.

Hull and White have also developed trinomial tree method to calibrate the model to market data ([33]). Nonetheless, the extended Vasicek model still suffers from allowing for negative interest rates.

Black and Karasinski (1991)

Instead of modeling the short rates $r(t)$, Black and Karasinski propose that $\ln(r(t))$ follows the extended Hull-White model ([11]):

$$d\ln(r(t)) = (\theta(t) - \kappa(t)\ln(r(t)))dt + \sigma(t)dW(t)$$

Black and Karasinski also suggested a binomial tree approach to calibrate the parameters fitting to the yield curve, the volatility curve.

The Quadratic Model

The Quadratic model is firstly studied by Beaglehole and Tammey [9]. It has been studied in the context of both theoretical analysis and empirical test [1], [40], [41]. The one factor quadratic model is based on an Ornstein-Uhlenbeck state process $X(t)$ that follows

$$dX(t) = (\alpha(t) - \beta(t)X(t))dt + \sigma dW(t)$$

where $W(t)$ is a standard Brownian motion under the risk-neutral measure. Then the short rate process $r(t)$ is specified by

$$r(t) = \frac{1}{2}X(t)^2.$$

One property that the quadratic short rate model, Cox-Ingersoll-Ross model and Black-Karasinski model share is the short rates are never negative. The quadratic

short rate model not only exhibits a nice analytic tractability but also is able to capture the non-linearity of the time series and more flexible for model design. The zero-coupon bond that can be expressed as a function of $x(t)$ instead of $r(t)$ satisfies the following partial differential equation

$$\frac{\partial B}{t} + (\alpha(t) - \beta(t)x)\frac{\partial B}{\partial x} + \frac{1}{2}\sigma(t)^2\frac{\partial^2 B}{\partial x^2} - \frac{1}{2}x^2 P = 0$$

with the final condition $B(T,T) = 1$ for any $-\infty < x < \infty$. It has been shown that the bond price $B(t,T)$ has the following form

$$B(t,T,x) = e^{-A(t,T)-B(t,T)x-\frac{1}{2}C(t,T)x^2},$$

where $A(t,T)$, $B(t,T)$, and $C(t,T)$ are the solutions of the final value problem (4.10)–(4.13). We shall study this model in detail in Chapter 2 and §4.2.

1.4 Dissertation Outline

The remainder of this dissertation is organized as follows. In Chapter 2 we shall discuss the details of the quadratic interest rate model and develop three numerical algorithms to evaluate American put option on zero-coupon bond. Chapter 3 is devoted to parameter estimation of the models with constant parameters such as the Vasicek model, the CIR model, and the quadratic model by using Kalman filter, the extended Kalman filter, and the unscented Kalman filter techniques. We shall also propose a special Kalman technique to efficiently estimate the parameters of the quadratic model. In Chapter 4, we shall calibrate the time-dependent models such as the Hull-White model, the extended CIR model, and the quadratic model to the market data and compare these models' performance of capturing yield rates'

movements. In Chapter 5, we shall propose a family of one factor models to restrict the short rate in the range of $(0, 1)$, which makes more practical sense. We further apply the homotopy perturbation method (HPM) to approximate the zero-coupon bond prices and compare the performance of HPM and the trinomial tree method. The conclusion remarks and future work are given in Chapter 6.

CHAPTER 2

THE QUADRATIC MODEL

2.1 Introduction

Let $X(t)$ be a mean-reverting process which is the solution of the following stochastic differential equation:

$$dX(t) = (\alpha(t) - \beta(t)X(t))dt + \sigma(t)dW(t), \tag{2.1}$$

where W is a standard Brownian motion under the risk-neutral measure and $\alpha(t)$, $\beta(t)$ and $\sigma(t)$ are deterministic functions of t. Then the quadratic model of the short interest rate process $r(t)$ is given by

$$r(t) = \frac{1}{2}X(t)^2. \tag{2.2}$$

As for the Black-Karasinski model ([11]) and Cox-Ingersoll-Ross Model ([16]), the short interest rates are never negative under the quadratic model. Besides, it has more flexibility in term structure and can outperform affine models in explaining historical bond price behavior in the United States ([1, 41]).

By Itô's formula, we have

$$dr(t) = \left(\frac{1}{2}\sigma(t)^2 + \alpha(t)X(t) - \beta(t)X(t)^2\right)dt + \sigma(t)X(t)dW(t). \tag{2.3}$$

Since process $X(t)$ can take any real number as its value and function $\zeta(x) = \frac{1}{2}x^2$ is not invertible on $(-\infty, \infty)$, we can not obtain a SDE for $r(t)$ by replacing $X(t)$ in terms of $r(t)$ in (2.3). It means that $r(t)$ can not be regarded as an independent variable for the quadratic model, i.e., the prices of interest rate derivatives are functions of $X(t)$ instead of $r(t)$. However, we can formally obtain the following SDE for $X(t) = \sqrt{2r(t)}$:

$$dr(t) = \left(\frac{1}{2}\sigma(t)^2 + \sqrt{2}\alpha(t)\sqrt{r(t)} - 2\beta(t)r(t)\right)dt + \sqrt{2}\sigma(t)\sqrt{r(t)}dW(t). \tag{2.4}$$

When parameters are constant, this SDE becomes the double square root model proposed by Longstaff ([43]):

$$dr(t) = \left(\frac{\nu^2}{4} - \kappa\sqrt{r(t)} - 2\beta r(t)\right)dt + \nu\sqrt{r(t)}dW(t), \tag{2.5}$$

where $\kappa = -\sqrt{2}\alpha$ and $\nu = \sqrt{2}\sigma$. Here ν and β are corresponding to σ and λ in [43]. It should be pointed out that $\alpha < 0$ is assumed in [43].

It clearly follows from the above discussion that the interest rate process of the quadratic model defined by (2.2) is not the solution of SDE (2.4). Indeed, SDE (2.4) ((2.5)) does not have a positive solution. Otherwise, process $X(t) = \sqrt{2r(t)}$ will be a positive solution of SDE (2.1), which is impossible since the solution of SDE (2.1) is a Gaussian process. Beaglehole and Tenney pointed out in [9] that SDE (2.5) should go with the additional condition: process $r(t)$ is reflected whenever it reaches zero. They have also shown numerically that the bond prices given by the formula in [43] do not match the ones obtained by Monte Carlo simulations for the modified model. However, as indicated in [9], it is not difficult to verify that the bond price formula

in [43] with the parameters specified as above is the one under the quadratic model in [35] after replacing $\sqrt{r}$ by $x/\sqrt{2}$ and r by $x^2/2$.

A remedy for rescuing SDE (2.5) is to replace $\frac{\nu^2}{4}$ in (2.5) by δ:

$$dr(t) = \left(\delta - \kappa\sqrt{r(t)} - 2\beta r(t)\right) dt + \nu\sqrt{r(t)}dW(t). \tag{2.6}$$

Notice that it becomes the CIR model when $\kappa = 0$. Therefore, as for the CIR model, this SDE has positive solutions when

$$\delta > \frac{1}{4}\nu^2.$$

This claim can be justified by examining the SED for $X(t) = 2\sqrt{r(t)}$:

$$dX(t) = \left(\frac{\eta}{X(t)} - \kappa - \beta X(t)\right) dt + \nu dW(t),$$

where

$$\eta = \delta - \frac{1}{4}\nu^2.$$

When $\eta > 0$, the drift term goes to positive infinity as $X(t) \to 0^+$, which makes that zero is not accessible. Unfortunately, no analytic formula is available for zero-coupon bond prices under the dynamics (2.6) of interest rates. We shall study this model in our future work.

The remaining of this chapter is organized as follows. In §2.2, a front-fxing finite element method is considered to solve the free boundary value problem for American put options on zero-coupon bond. In §2.3, we propose a new trinomial method for both European and American options by transferring the SDE (2.1) into a SDE without drift term. It should be pointed out that we does not need to change the

paths of the tree to ensure positive probabilities. Lattice methods including binomial and trinomial methods are more attractive to practitioners since they can be easily implemented and are more flexible to compute option prices and hedge ratios at any given point. In §2.4, we consider a finite element method to solve the variational inequality problem for American put options. In the last section, §2.5, we give two examples to to examine the convergence of the proposed methods in the previous sections and compare the quadratic model with other the Hull-White model and the CIR model.

2.2 A Front Fixing Finite Element Methods for American Put Options

In this section, we shall apply the front fixing finite element method to solve the free boundary problem for American put options.

Consider the American put option on a T^*-maturity zero-coupon bond. The option expiration date is T ($< T^*$), its exercise price is K, and its payoff function is

$$g(x,t) = \max(K - P(x,t;T^*), 0)$$

where $P(x,t;T^*)$ is the T^* bond price at (x,t). The put price will be denoted by $p(x,t)$. Notice that the bond price $P(x,t;T^*)$ goes to 0 as $x \to \pm\infty$, which means that the option payoff $g(x,t)$ approaches the exercise price is K as $x \to \pm\infty$. Since the option can be exercised at any time up to its expiration date, there should be two critical values $\varphi_1(t)$ and $\varphi_2(t)$ at any time t such that the put should be exercised when $x \le \varphi_1(t)$ or $x \ge \varphi_2(t)$. As usual, we can show that $p(x,t)$ and $\varphi_i(t)$ solve the

following free boundary problem:

$$p_t + \frac{1}{2}\sigma(t)^2 p_{xx} + (\alpha(t) - \beta(t)x)p_x - \frac{1}{2}x^2 p = 0, \quad \varphi_1(t) < x < \varphi_2(t) \tag{2.7}$$

$$p(\varphi_i(t), t) = g(\varphi_i(t), t), \quad 0 < t \leq T, \quad i = 1, 2 \tag{2.8}$$

$$p_x(\varphi_i(t), t) = g_x(\varphi_i(t), t), \quad 0 < t \leq T, \quad i = 1, 2. \tag{2.9}$$

$$p(x, T) = g(x, T), \quad \varphi_1(T) < x < \varphi_2(T). \tag{2.10}$$

Consider the variable transforms

$$\begin{aligned} &\tau = T - t, \\ &y = \frac{x - \varphi_1(t)}{\varphi_2(t) - \varphi_1(t)}, \\ &\psi_i(\tau) = \varphi_i(T - t), \quad i = 1, 2., \\ &u(y, \tau) = p(x, t). \end{aligned}$$

Then the spatial domain is changed from $\varphi_1(t) < x < \varphi_2(t)$ to $0 < y < 1$. Problem (2.7)–(2.10) becomes

$$u_t + \mathcal{L}u = 0, \quad 0 < y < 1, \quad 0 < \tau \leq T, \tag{2.11}$$

$$u(0, \tau) = f(0, \tau; \psi_1, \psi_2), \quad 0 < \tau \leq T, \tag{2.12}$$

$$u(1, \tau) = f(0, \tau; \psi_1, \psi_2), \quad 0 < \tau \leq T, \tag{2.13}$$

$$u_y(0, \tau) = f_y(0, \tau; \psi_1, \psi_2), \quad 0 < \tau \leq T, \tag{2.14}$$

$$u_y(1, \tau) = f_y(1, \tau; \psi_1, \psi_2), \quad 0 < \tau \leq T, \tag{2.15}$$

$$u(y, 0) = u_0(y), \quad 0 \leq y \leq 1, \tag{2.16}$$

where

$$\mathcal{L} = -a(\psi_1, \psi_2)u_{yy} + b(y, \tau; \psi_1, \psi_2, \psi_1', \psi_2')u_y + c(y, \tau; \psi_1, \psi_2)u,$$

$$\Delta\psi = \psi_2 - \psi_1,$$

$$a(\psi_1, \psi_2) = \frac{\sigma^2}{2\Delta\psi(\tau)^2},$$

$$b(y, \tau; \psi_1, \psi_2, \psi_1', \psi_2') = \beta y + \frac{\beta\psi_1 - \alpha - y\Delta\psi' - \psi_1'}{\Delta\psi},$$

$$c(y, \tau; \psi_1, \psi_2) = \frac{1}{2}\left((y\Delta\psi)^2 + 2y\Delta\psi + \psi_1^2\right),$$

$$f(y, \tau; \psi_1, \psi_2) = g(y\Delta\psi + \psi_1, T - \tau),$$

$$u_0(y) = f(y, 0; \psi_1(0), \psi_2(0)).$$

To apply the finite element method, we shall integrate the natural boundary conditions (2.14) and (2.15) into the variational problem and treat the essential boundary conditions (2.12) and (2.13) as two nonlinear equations from which ψ_1 and ψ_2 can be solved. Define the bilinear form $\mathcal{B}$ as follows:

$$\mathcal{B}(v, w; \tau, \psi_1, \psi_2) = a(\psi_1, \psi_2)(v_y, w_y) + (b(y, \tau; \psi_i, \psi_i')v_y, w) + (c(y, \tau; \psi_1, \psi_2)v, w),$$

where $(\cdot, \cdot)$ denotes the inner product of $L^2(\Omega)$, the space of square integrable functions on $\Omega = (0, 1)$. Let $H^1(\Omega)$ be the usual Sobolev space, and let $H^{-1}(\Omega)$ be its dual space . Define

$$V = \left\{\omega : \omega \in L^2(0, T; H^1(\Omega)), \quad \omega_\tau \in L^2(0, T, L^2(\Omega)), \quad \omega_y(y, \tau) = f_y(y, \tau) \text{ on } \partial\Omega\right\}.$$

The variational form for problem (2.11)–(2.16) is: Find $u \in V$ and $\psi_i \in C([0, T]) \cap$

$C^1((0,T])$ for $i = 1, 2$ such that $u(0) = u_0$ and for $0 < \tau \leq T$

$$(u_\tau, \omega) + \mathcal{B}(v, w; \tau, \psi_1, \psi_2) = \mathcal{F}(w; \tau, \psi_1, \psi_2), \quad \forall \omega \in V, \tag{2.17}$$

$$u(0, \tau) = f(0, \tau; \psi_1, \psi_2), \tag{2.18}$$

$$u(1, \tau) = f(0, \tau; \psi_1, \psi_2) \tag{2.19}$$

where

$$\mathcal{F}(w; \tau, \psi_1, \psi_2) = G(1, \tau; \psi_1, \psi_2)\omega(1, \tau) - G(0, \tau; \psi_1, \psi_2)\omega(0, \tau),$$

$$G(y, \tau, \psi_1, \psi_2) = a(\psi_1, \psi_2) f_y(y, \tau; \psi_1, \psi 2).$$

Let $\Pi_y : 0 = y_0 < y_1 < \ldots < y_N = 1$ and $\Pi_\tau : 0 = \tau_0 < \tau_1 < \ldots < \tau_M = T$ be the partitions of $[0, 1]$ and $[0, T]$, where M and N are positive integers. Then $h_j = y_j - y_{j-1}$ and $k_i = \tau_i - \tau_{i-1}$ are the lengths of the sub-intervals $[y_{j-1}, y_j]$ and $[\tau_{i-1}, \tau_i]$, respectively. Let V_h be the piece-wise linear element subspace of V with respect to partition Π_y, where $h = \max_{1 \leq j \leq N}(h_j)$. Denote the basis function of V_h by $\omega_1, \omega_2, \ldots, \omega_N$ such that $\omega_j(y_i) = \delta_{i,j}$ for $j = 0, 1, \ldots, N$ and $i = 0, 1, \ldots, N$, where δ_{ij} is the Kronecker delta.

Values $\psi_1(0)$ and $\psi_2(0)$ are determined according to the the optimal exercise, i.e., the following equation

$$P(x, T; T^*) = K.$$

Then $u_0(y) = f(y, 0; \psi_1(0), \psi_2(0))$ is known. Let

$$\psi_1^0 = \psi_1(0), \quad \psi_2^0 = \psi_2(0), \quad u_h^0 = \sum_{j=0}^{N} u_0(y_j)\omega_j(y).$$

The finite element approximation the the variational problem (2.17)–(2.19) by Crank-Nicolson scheme in time is: For $m = 1, 2, \ldots, N$, find $u_h^m \in V_h$ and ψ_1^m and ψ_2^m such

that for $m = 1, \ldots, M$

$$(\delta_\tau u_h^m, \omega) + \mathcal{B}_m\left(u_h^{m-\frac{1}{2}}, \omega\right) = \mathcal{F}\left(\omega, \tau_m; \psi_1^{m-\frac{1}{2}}, \psi_1^{m-\frac{1}{2}}\right), \quad \forall \omega \in V_h, \tag{2.20}$$

$$u_h^m(0) = f(0, \tau_m; \psi_1^m, \psi_2^m), \tag{2.21}$$

$$u_h^m(1) = f(1, \tau_m; \psi_1^m, \psi_2^m), \tag{2.22}$$

where

$$\begin{aligned}
&\mathcal{B}_m(u, \omega) = \mathcal{B}(u, \omega; \tau_{m-\frac{1}{2}}, \psi_1^m, \psi_2^m, \delta_\tau \psi_1^m, \delta_\tau \psi_2^m), \\
&u_h^{m-\frac{1}{2}} = \frac{u_h^m + u_h^{m-1}}{2}, \quad \tau_{m-\frac{1}{2}} = \frac{\tau_m + \tau_{m-1}}{2}, \quad \psi_i^{m-\frac{1}{2}} = \frac{\psi_i^m + \psi_2^{m-1}}{2}, \\
&k_m = \tau_m - \tau_{m-1}, \quad \delta_\tau u_h^m = \frac{u_h^m - u_h^{m-1}}{k_m}, \quad \delta_\tau \psi_i^m = \frac{\psi_i^m - \psi_i^{m-1}}{k_m}.
\end{aligned}$$

We can rewrite (2.20)–(2.22) in the matrix form:

$$\left(A + \frac{1}{2} k_m B_m\right) U^m = \left(A - \frac{1}{2} k_m B_m\right) U^{m-1} + k_m F_m \tag{2.23}$$

$$u_1^m = f(0, \tau_m; \psi_1^m, \psi_2^m) \tag{2.24}$$

$$u_N^m = f(1, \tau_m; \psi_1^m, \psi_2^m) \tag{2.25}$$

where

$$A = (\omega_j, \omega_i)_{N \times N}, \quad B_m = (\mathcal{B}_m(\omega_j, \omega_i))_{N \times N}, \quad U^m = (u_1^m, \ldots, u_N^m),$$

$$F_m = (-G_m(0), 0, \ldots, 0, G_m(1)), \quad G_m(y) = G\left(y, \tau_m; \psi_1^{m-\frac{1}{2}}, \psi_1^{m-\frac{1}{2}}\right).$$

To save computational time, we shall express B_m as the linear combination of the matrices independent of m. Indeed, we have by simple calculation

$$B_m = c_m^{(1)} B^{(1)} + c_m^{(2)} B^{(2)} + c_m^{(3)} B^{(3)} + c_m^{(4)} B^{(4)} + c_m^{(5)} B^{(5)} + c_m^{(5)} A,$$

where

$$c_m^{(1)} = a(\psi_1^m, \psi_2^m), \quad c_m^{(2)} = \beta - \frac{\Delta\delta_\tau\psi^m}{\Delta\psi_m}, \quad c_m^{(3)} = \frac{\beta\psi_1^m - \alpha - \delta_\tau\psi_1^m}{\Delta\psi_m},$$
$$c_m^{(4)} = \frac{1}{2}\left(\Delta\psi^m\right)^2, \quad c_m^{(5)} = \Delta\psi^m\psi_1^m, \quad c_m^{(6)} = \frac{1}{2}\left(\psi_1^m\right)^2,$$
$$B^{(1)} = ((\omega_y)_j, (\omega_y)_i)_{N\times N}, \quad B^{(2)} = (y_j(\omega_y)_j, \omega_i)_{N\times N},$$
$$B^{(3)} = ((\omega_y)_j, \omega_i)_{N\times N}, \quad B^{(4)} = (y_j^2\omega_j, \omega_i)_{N\times N}, \quad B^{(5)} = (y_j\omega_j, \omega_i)_{N\times N}.$$

Notice that matrices $A, B^{(1)}, \ldots, B^{(5)}$ are independent of m. Once these matrices are computed and stored, we can assemble B_m fast with least amount of work at each time step. We regard U^m as an implicit function of ψ_1^m and ψ_2^m determined by equation (2.23). Then we can treat (2.24) and (2.25) as a system of two nonlinear equations for ψ_1^m and ψ_2^m. We can rewrite them as:

$$\begin{cases} L_1(\psi_1^m) = u_1^m - K + P(\psi_1^m, T - \tau_m; T^*) = 0, \\ L_1(\psi_2^m) = u_N^m - K + P(\psi_2^m, T - \tau_m; T^*) = 0 \end{cases} \tag{2.26}$$

which can be solved by Broyden method quickly.

To sum up, for a given tolerance ε, our front-fixing finite element method can be implemented as follows:

Algorithm 2.1. A front-fixing finite element method for American Puts

---

1. Compute matrices A, $B^{(1)}$, $B^{(2)}$, $B^{(3)}$, $B^{(4)}$, and $B^{(5)}$.
2. For $m = 1, 2, \ldots, M$, do
   - Let $\psi_i^m = \psi_i^{m-1}, \quad i = 1, 2.$
   - For $j = 1, 2, \ldots,$ do
     - Build system (2.23).

* Build system (2.23).

* Solve system (2.23) by Thomas Algorithm.

* Solve non-linear system (2.26) for the new approximations of ψ_1^m and ψ_2^m.

* If the norm of the difference of the old and new approximations of ψ_1^m and ψ_2^m is less than or equal to ε, then terminate the loop.

End do

– Solve system (2.23) for a better approximation of U^m.

End do

2.3 A Trinomial Method

In this section, we shall develop a trinomial method to compute the bond option price at a given point (x_0, t_0). Consider an option on a T^*-maturity zero-coupon bond. The option expiration date is T $(< T^*)$ and its exercise price is K. The option payoff is

$$g(x,t) = \begin{cases} \max(P(x,t;T^*) - K, 0) & \text{for a call,} \\ \\ \max(K - P(x,t;T^*), 0) & \text{for a put,} \end{cases}$$

where $P(x,t;T^*)$ is the bond price.

Let

$$a(t) = \exp\left(\int_{t_0}^{t} \beta(s)ds\right), \quad b(t) = \int_{t_0}^{t} a(s)\alpha(s)ds + x_0. \tag{2.27}$$

Then by Ito's formula, the new process $Y(t) = a(t)X(t) - b(t)$ follows

$$dY(t) = \gamma(t)dW(t),$$

where $\gamma(t) = a(t)\sigma(t)$. It should be pointed out that $X(t_0) = x_0$ is corresponding to $Y(t_0) = 0$. For a given positive integer M, let $k = (T - t_0)/M$ be the step size in

time. Denote by h the mesh size in spatial variable y. Let

$$y_m = mh, \quad t_m = t_0 + mk, \quad m = 0, 1, 2, \ldots, M.$$

Then we setup our tree for process $Y(t)$ by using nodes (y_j, t_m) for $j = 0, \pm 1, \pm 2, \ldots, \pm m$, $m = 0, 1, 2, \ldots, M$. Let p_m^-, p_m^0, and p_m^+ be the probabilities by which the tree branches from node (y_j, t_{m-1}) to nodes (y_{j-1}, t_m), (y_j, t_m), and (y_{j+1}, t_m), respectively. Notice that

$$Y(t_m) - Y(t_{m-1}) \approx \gamma(t_{m-1})\sqrt{k}Z$$

for some standard norm random variable Z. Matching the mean and variance, we can obtain the system for p_m^-, p_m^0, and p_m^+:

$$\begin{aligned}
&p_m^- + p_m^0 + p_m^+ = 1, \\
&p_m^- y_{j-1} + p_m^0 y_j + p_m^+ y_{j+1} = y_j, \\
&p_m^- y_{j-1}^2 + p_m^0 y_j^2 + p_m^+ y_{j+1}^2 = y_j^2 + k\gamma^2(t_{m-1}).
\end{aligned}$$

Solving this system, we get

$$p_m^- = p_m^+ = \frac{k\gamma^2(t_{m-1})}{2h^2}, \quad p_m^0 = 1 - \frac{k\gamma^2(t_{m-1})}{h^2}.$$

In order to ensure that p_m^0 is nonnegative for given k, we need the following restriction on h:

$$h \geq \gamma_{\max}\sqrt{k},$$

where $\gamma_{\max}$ is the maximum value of $\gamma(t)$ over $[t_0, T]$.

Let

$$\xi(y, t) = \frac{y + b(t)}{a(t)}.$$

Then at time level $t = t_m$, the x-coordinates for the nodes of the corresponding tree for process $X(t)$ are $x_j^m = \xi(y_j, t_m)$ for $j = 0, \pm 1, \pm 2, \ldots, \pm m$. Denote by V_j^m the approximation of the put value at node (x_j^m, t_m). Let

$$V_j^M = g\left(x_j^M, t_M\right), \quad j = 0, \pm 1, \pm 2, \ldots, \pm M.$$

We have the following trinomial algorithms to compute V_0^0, the approximation of the American and European prices at point (x_0, t_0):

Algorithm 2.2. A trinomial method for American options

For $m = M - 1, \ldots, 1, 0$, do

- $q_m = \dfrac{k\gamma(t_m)}{h^2}$;
- $p_m^0 = 1 - q_m$;
- $p_m^- = p_m^+ = \dfrac{1}{2} q_m$;
- For $j = -m, \ldots, m$, do
  - $r_j^m = \dfrac{1}{2}\left(x_j^m\right)^2$;
  - $v_j^m = \dfrac{1}{1 + k r_j^m}\left(p_m^- V_{j-1}^{m+1} + p_m^0 V_j^{m+1} + p_m^+ V_{j+1}^{m+1}\right)$;
  - $V_j^m = \max\left(v_j^m, g(x_j^m, t_m)\right)$;

  End do

End do

Algorithm 2.3. A trinomial method for European options

For $m = M-1, \ldots, 1, 0$, do

- $q_m = \dfrac{k\gamma(t_m)}{h^2}$;
- $p_m^0 = 1 - q_m$;
- $p_m^- = p_m^+ = \dfrac{1}{2} q_m$;
- For $j = -m, \ldots, m$, do
  - $r_j^m = \dfrac{1}{2}\left(x_j^m\right)^2$;
  - $V_j^m = \dfrac{1}{1+kr_j^m}\left(p_m^- V_{j-1}^{m+1} + p_m^0 V_j^{m+1} + p_m^+ V_{j+1}^{m+1}\right)$;

  End do

End do

It is easy to see that this trinomial method is numerically stable. In fact, we have the following stability estimates: for the American option,

$$\max_{-(M-m)\le j\le M-m} \left|V_j^m\right| \le \max\left(\max_{-M\le j\le M}\left|V_j^0\right|, \max_{m\le \ell\le M}\max_{-(M-\ell)\le j\le M-\ell} g\left(x_j^\ell, t_\ell\right)\right)$$

for $m = 1, 2, \ldots, M$, and for the European option

$$\max_{-(M-m)\le j\le M-m} \left|V_j^m\right| \le \max_{-M\le j\le M}\left|V_j^0\right|$$

for $m = 1, 2, \ldots, M$.

Remark 2.1. It is well-known that American and European calls on zero-coupon bonds have the same values when the interest rate process is always positive (see [55]).

2.4 A Finite Element Method for American Put Options

Let $p(y,t)$ be the American put option price when the new process $Y(t)$ takes value y at time t. Here we have let $t_0 = 0$ and $x_0 = 0$ for the definitions of $a(t)$ and $b(t)$ in (2.27). Then $p(y,t)$ is the solution of the following variational inequality problem:

$$
\begin{aligned}
&p_t + \frac{1}{2}\gamma(t)p_{yy} - \frac{1}{2}\xi^2(y,t)p \geq 0, \quad y \in R, \quad 0 \leq t < T, \\
&p(y,t) \geq g(\xi(y,t),t), \quad y \in R, \quad 0 \leq t < T, \\
&\left(p_t + \frac{1}{2}\gamma(t)p_{yy} - \frac{1}{2}\xi^2(y,t)p\right)(p(y,t) - g(\xi(y,t),t)) = 0, \quad y \in R, \quad 0 \leq t < T, \\
&p(y,T) = g(\xi(y,T),T), \quad y \in R.
\end{aligned}
$$

We should mention that the above variational inequality problem can also be derived by applying the variable substitution $x = \frac{y-b(t)}{a(t)}$ to problem (2.7)–(2.10). As known in §2.2, the option price is equal to its payoff when $|x|$ is sufficiently large. Thus we can pick a large negative number Y_1 and a large positive number Y_2 such that

$$
p(y,t) = g(\xi(y,t),t), \quad \forall y \leq Y_1 \text{ or } y \geq Y_2, \quad 0 \leq t \leq T.
$$

Let

$$
\begin{aligned}
u(y,t) &= p(y,T-t), \quad \rho(t) = \frac{1}{2}\gamma(T-t), \\
f(y,t) &= g(\xi(y,T-t),T-t), \quad R(y,t) = \frac{1}{2}\xi^2(y,T-t).
\end{aligned}
$$

We can rewrite the above variational inequality problem over the bounded interval

$\Omega = (Y_1, Y_2)$:

$$u_t - \rho(t)u_{yy} + R(y,t)u \geq 0, \quad y \in \Omega, \quad 0 < t \leq T, \tag{2.28}$$

$$u(y,t) \geq f(y,t), \quad y \in \Omega, \quad 0 < t \leq T, \tag{2.29}$$

$$(u_t - \rho(t)u_{yy} + R(y,t)u)\,(u(y,t) - f(y,t)) = 0, \quad y \in \Omega, \quad 0 < t \leq T, \tag{2.30}$$

$$u(Y_1,t) = f(Y_1,t), \quad 0 < t \leq T, \tag{2.31}$$

$$u(Y_2,t) = f(Y_2,t), \quad 0 < t \leq T, \tag{2.32}$$

$$u(y,0) = f(y,0), \quad y \in \Omega. \tag{2.33}$$

Let $L^2(\Omega)$ be the space of square integrable functions on Ω and $(\cdot,\cdot)$ be its inner product. We denote $H^{-1}(\Omega)$ the dual space of usual Sobolev space $H^1(\Omega)$. Let

$$\mathcal{U} = \big\{v : v \in L^2(0,T;H^1(\Omega)), v_t \in L^2(0,T;L^2(\Omega)),$$
$$\text{and } v = f \text{ a.e. on } \partial Q_T, v \geq G \text{ a.e. in } Q_T\big\},$$

where $Q_T = \Omega \times (0,T)$ and $\partial Q_T = \partial\Omega \times (0,T) \cap [Y_1, Y_2] \times 0$. Define the bilinear form

$$a(t,u,v) = \int_{Y_1}^{Y_2} \big(u_y(y,t)v_y(y,t) + R(y,t)u(y,t)v(y,t)\big)dy$$

Then the variational problem for (2.28)-(2.33) is: Find $u \in \mathcal{U}$ such that

$$(u_t, v-u) + a(t,u,v-u) \geq 0, \quad \text{a.e. } t \in (0,T], \quad \forall v \in \mathcal{U} \tag{2.34}$$

Now we consider the finite element approximations to parabolic variational inequality (2.4). Let $\Pi_h : Y_1 = y_0 < y_1 < \ldots < y_N = Y_2$ be a partition of $[Y_1, Y_2]$ for a given positive integer N and $h = \max_{1\leq j\leq N}(y_j - y_{j-1}) < 1$. Denote the linear element space under the partition Π_h by V_h. For another given positive integer M,

let $k = T/M$ be the step size in time and $t_m = mk$ for $m = 0, 1, \ldots, M$. Let

$$\mathcal{U}^m = \big\{ v \in V_h : v(y_0) = f(y_0, t_m), v(y_N) = f(y_N, t_m),$$

$$v(y_j) \geq f(y_j, t_m), j = 1, 2, \ldots, N-1 \big\}$$

Denote w_h^m the approximation of $w(t_m, y)$ in $\mathcal{U}^m$ and define the following quantities:

$$t^{m-\frac{1}{2}} = t_m - \frac{1}{2}\tau, u_h^{m-\frac{1}{2}} = \frac{1}{2}(u_h^{m-1} + u_h^m), \delta_\tau u_h^m = \frac{u_h^m - u_h^{m-1}}{\tau}.$$

Let u_h^0 be the interpolant of $f(y, 0)$ in $\mathcal{U}^m$. The finite element approximation of (2.4) is: Find $u_h^m \in \mathcal{U}^m$ for $m = 1, 2, \ldots, M$ such that

$$(\delta_\tau u_h^m, v - u_h^m) + a\left(t_{m-\frac{1}{2}}, u_h^{m-\frac{1}{2}}, v - u_h^m\right) \geq 0, \quad \forall v \in \mathcal{U}^m. \tag{2.35}$$

Here the Crank-Nicholson scheme was used in time.

It is not difficulty to rewrite (2.35) into a linear complementarity problem as follows:

$$A_m U^m \geq B_m, \quad U^m \geq F^m, \quad (A_m U^m - B_m)(U^m - F^m) = 0 \tag{2.36}$$

for $m = 1, \ldots, M$. We can easily verify that A_m is a tridiagonal M-matrix when k is sufficiently small. Thus, the above linear complimentary problem can solved efficiently by the algorithm developed in([17]). Our numerical tests show that problem (2.36) can be solved by this algorithm with less than 10 iterations when $k = h$.

2.5 Numerical Examples

In this section, we shall give several numerical examples to examine the convergence of the proposed methods in the previous sections and compare the quadratic model with other the Hull-White model and the CIR model.

Example 2.1. In this example, we want to test our numerical algorithms by assuming that $f(T)$ and $g(T)$ are determined by the quadratic model with constant parameters. Recall that $A(0,T)$, $B(0,T)$ and $C(0,T)$ can be computed by (4.14)–(4.16). We set $\sigma = 0.06$, $\beta = 0.4$, and $\alpha = 0.16$. Then the long-term expected value of $x(t)$ is $\alpha/\beta = 0.4$. For $x(0) = 0.4$, we have the initial interest rate $r(0) = x(0)^2/2 = 0.08$. One-year American put options written on 5-year and 30-year bonds with face value \$100 will be considered. The option exercise prices are chosen to be 87% and 83% of the current forward bond prices respectively.

In Figures 2.1–2.2, we display the L^2-norm and H^1-norm of errors of American option prices in finite element method and front-fixing finite element method. The L^2-errors and H^1-errors are computed between two successive finite element approximations of option prices against step size h in time. We also display the L^2 norm of $\psi_i^m - \psi_i^{2m}$ with respect to the number of time steps $m = 0, 1, \ldots, M$ for the early exercise interest rate in Figures 2.3–2.4. We can observe that the both finite element method and front fixing finite element method with Crank-Nicolson scheme converges quadratically and linearly in the L^2-norm and H^1-norm as expected. For the early exercise interest rate, the rate of convergence of finite element method is one, while the rate of convergence of front fixing finite element method is greater than one in L^2-norm.

Next, we want to compare the L^2-error of the front-fixing element method (FFEM) with the usual finite element method (FEM). Tables 2.1–2.3 display the L^2-errors for today's option prices and L^2-errors for early exercise interest rate as compared with the "exact values" computed by FFEM and FEM with 16000 time steps for the top

and bottom numbers in each cell. We can see that the front fixing finite element method provides more accurate and more stable results and converges more quickly than the usual finite element method, especially for the early exercise interest rate.

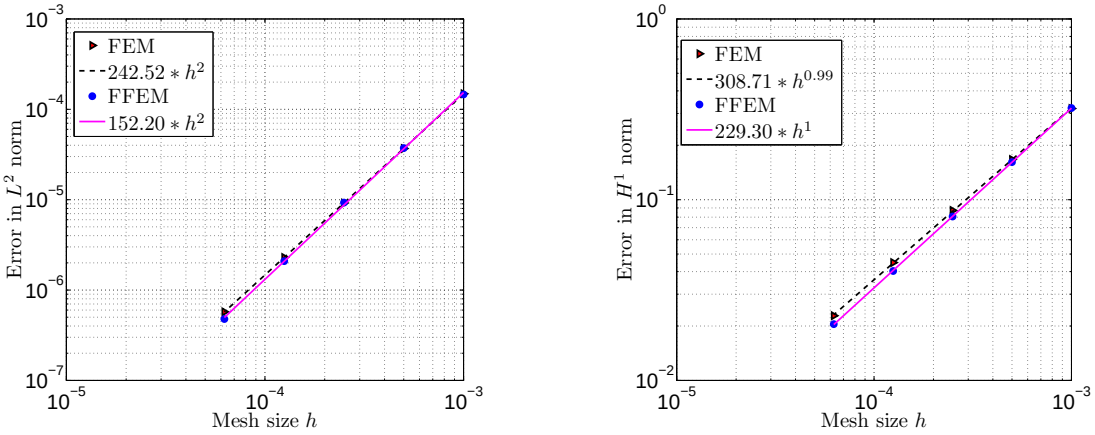

Figure 2.1: Convergence of American Option Price: $T^* = 5$, $K = 75$

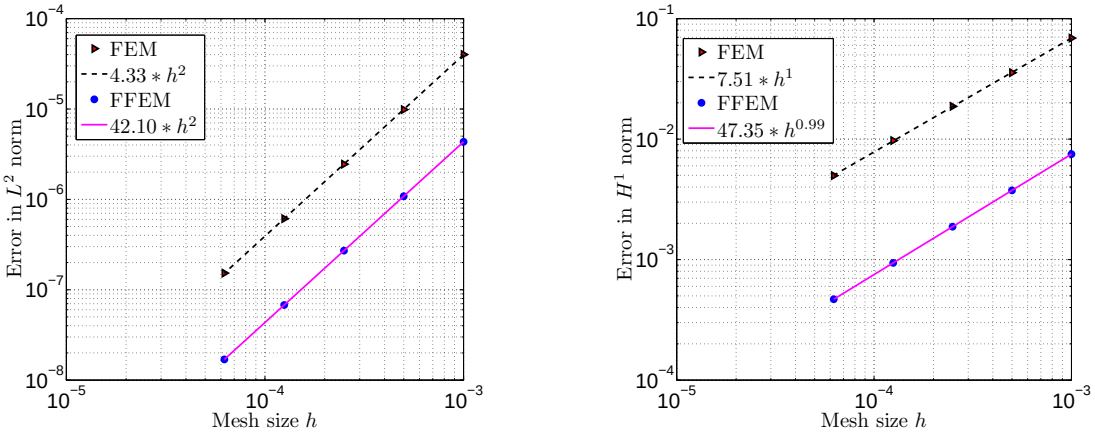

Figure 2.2: Convergence of American Option Price: $T^* = 30$, $K = 10$

Example 2.2. In this example, we assume that the interest rate process $r(t)$ evolves

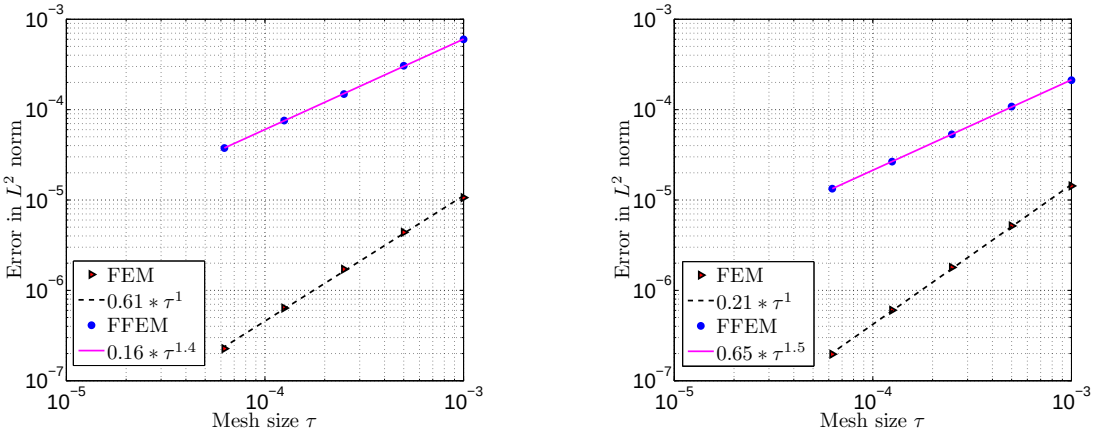

Figure 2.3: Convergence of Early Exercise Interest Rate: $T^* = 5$, $K = \$75$

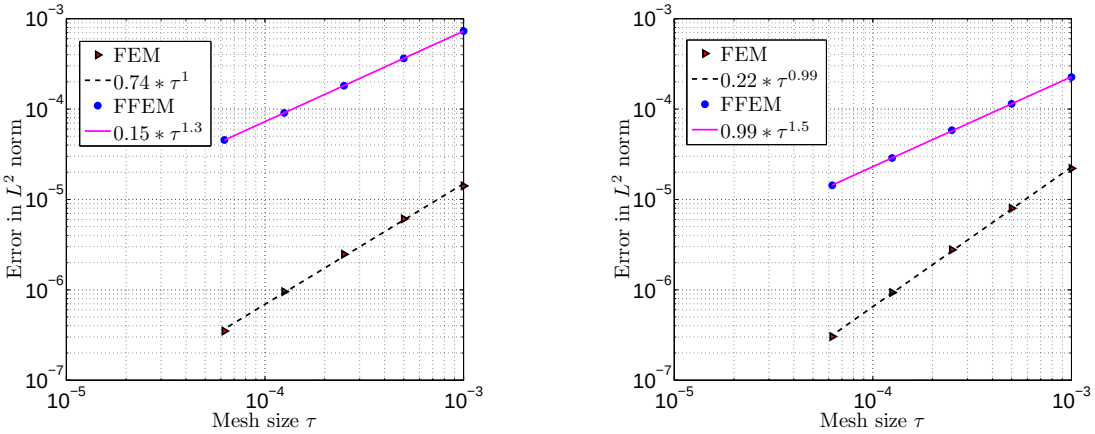

Figure 2.4: Convergence of Early Exercise Interest Rate: $T^* = 30$, $K = 10$

according to the two-factor CIR model as in [32, 69] (see [39]):

$$r(t) = x_1(t) + x_2(t),$$

$$dx_i = \kappa_i(\theta_i - x_i)dt + \sigma_i\sqrt{x_i}dW_i(t), \quad i = 1, 2,$$

where κ_i is the speed of mean reversion, θ_i is the long term interest rate, σ_i is a positive constant, $W_1(t)$ and $W_2(t)$ are two independent standard Brownian motion. The zero-coupon bond price under this two-factor model is given by

$$P(r, t, T) = P_1(x_1, t, T)P_2(x_2, t, T),$$

Table 2.1: L^2 errors for today's option prices

$\sigma = 0.06$, $\alpha = 0.16$, $\beta = 0.4$				
	$T^* = 5$, $K = \$75$		$T^* = 30$, $K = \$10$	
M	FEM	FFEM	FEM	FFEM
500	2.24×10^{-7}	4.29×10^{-9}	2.35×10^{-7}	5.96×10^{-8}
	2.24×10^{-7}	4.98×10^{-9}	2.34×10^{-7}	6.32×10^{-8}
1000	5.58×10^{-8}	8.32×10^{-10}	5.75×10^{-8}	1.08×10^{-8}
	5.61×10^{-8}	1.53×10^{-9}	5.75×10^{-8}	1.37×10^{-8}
2000	1.43×10^{-8}	1.73×10^{-10}	1.50×10^{-8}	1.45×10^{-9}
	1.40×10^{-8}	4.04×10^{-10}	1.49×10^{-8}	2.69×10^{-9}
4000	4.05×10^{-9}	2.02×10^{-10}	4.17×10^{-9}	1.81×10^{-10}
	3.54×10^{-9}	8.74×10^{-11}	3.78×10^{-9}	4.57×10^{-10}
8000	1.34×10^{-9}	3.04×10^{-10}	1.35×10^{-9}	2.92×10^{-10}
	9.12×10^{-10}	1.14×10^{-11}	9.26×10^{-10}	4.95×10^{-11}

Table 2.2: L^2 errors for early exercise interest rates

$\sigma = 0.06$, $\alpha = 0.16$, $\beta = 0.4, T^* = 5, K = \$75$				
	left free boundary		right free boundary	
M	FEM	FFEM	FEM	FFEM
500	2.19×10^{-7}	4.01×10^{-9}	2.34×10^{-7}	3.12×10^{-8}
	2.20×10^{-7}	5.18×10^{-9}	2.34×10^{-7}	3.47×10^{-8}
1000	5.74×10^{-8}	6.48×10^{-10}	6.01×10^{-8}	5.42×10^{-9}
	5.75×10^{-8}	1.45×10^{-9}	6.02×10^{-8}	7.47×10^{-9}
2000	1.45×10^{-8}	1.91×10^{-10}	1.48×10^{-8}	6.82×10^{-10}
	1.39×10^{-8}	3.54×10^{-10}	1.47×10^{-8}	1.47×10^{-9}
4000	4.17×10^{-9}	2.46×10^{-10}	4.08×10^{-9}	1.82×10^{-10}
	3.59×10^{-9}	7.16×10^{-11}	3.66×10^{-9}	2.48×10^{-10}
8000	1.36×10^{-9}	3.36×10^{-10}	1.35×10^{-9}	3.19×10^{-10}
	8.97×10^{-10}	8.88×10^{-12}	9.56×10^{-10}	2.69×10^{-11}

where

$$P_i(x_i, t, T) = A_i(T-t)e^{-B_i(T-t)x_i}$$

Table 2.3: L^2 errors for early exercise interest rates

$\sigma = 0.06$, $\alpha = 0.16$, $\beta = 0.4, T^* = 30, K = \$10$				
	left free boundary		right free boundary	
M	FEM	FFEM	FEM	FFEM
500	2.24×10^{-7}	4.29×10^{-9}	2.35×10^{-4}	$5.96.27 \times 10^{-5}$
	5.32×10^{-4}	1.72×10^{-5}	5.47×10^{-4}	8.41×10^{-5}
1000	2.66×10^{-4}	1.52×10^{-5}	2.72×10^{-4}	3.00×10^{-5}
	2.66×10^{-4}	7.41×10^{-6}	2.73×10^{-4}	3.03×10^{-5}
2000	1.34×10^{-4}	1.36×10^{-5}	1.38×10^{-4}	1.54×10^{-5}
	1.33×10^{-4}	2.96×10^{-6}	1.37×10^{-4}	1.04×10^{-5}
4000	6.86×10^{-5}	1.25×10^{-5}	7.06×10^{-5}	1.29×10^{-5}
	6.62×10^{-5}	1.07×10^{-6}	6.80×10^{-5}	3.28×10^{-6}
8000	4.00×10^{-5}	1.37×10^{-5}	4.09×10^{-5}	1.45×10^{-5}
	3.32×10^{-5}	3.03×10^{-7}	3.40×10^{-5}	8.34×10^{-7}

is the bond price when the short-term rate process is $x_i(t)$ and

$$A_i(\tau) = \left(\frac{a_i e^{b_i \tau}}{b_i(e^{a_i \tau} - 1) + a_i} \right)^{c_i}, \quad B_i(\tau) = \frac{e^{a_i \tau} - 1}{b_i(e^{a_i \tau} - 1) + a_i},$$
$$a_i = (\kappa_i^2 + 2\sigma_i^2)^{\frac{1}{2}}, \quad b_i = (\kappa_i + a_i)/2, \quad c_i = 2\kappa_i \theta_i / \sigma_i^2.$$

The quadratic model is fitted to the two-factor CIR model as follows:

$$\sigma = \sqrt{(\sigma_1^2 x_1(0) + \sigma_2^2 x_2(0))/(2r(0))},$$
$$f(T) = \sqrt{\sigma_1^2 x_1(0) B_1(T)^2 + \sigma_2^2 x_2(0) B_2(T)^2}/\sigma,$$
$$g(T) = \log(P(x_1(0), x_2(0), 0; T),$$

where $x(0) = \sqrt{2(x_1(0) + x_2(0))}$.

The bond face value is \$100. The parameters for the TCIR model are given in Table 2.4. For the first group of parameters, processes $x_1(t)$ and $x_2(t)$ are always positive since $\frac{\kappa_1 \theta_1}{\sigma_1^2} > 1/2$ and $\frac{\kappa_2 \theta_2}{\sigma_2^2} > 1/2$, and for the second group of parameters, since $\frac{\kappa_1 \theta_1}{\sigma_1^2} > 1/2$ and $\frac{\kappa_2 \theta_2}{\sigma_2^2} < 1/2$, process $x_1(t)$ is always positive, but processes $x_2(t)$ can be zero with a positive probability ([16]).

Now we consider one-year American put options written on bonds with expiration dates 5-year, 10-year, 15-year and 20-year. Their exercise prices are given as the percentage of the current forward bond price: 87%, 88%, 89%, and 90%. We assume that the initial term structure is determined by the tow-factor CIR model with constant parameters, and then we calibrate quadratic model, extended CIR model, and Hull-White model to the initial term structure. The comparison of the option prices under all these models is presented in Figures 2.5–2.6. Both figures demonstrate that all three models can accurately capture the put option price's movements. When the volatility and the mean reverting speed are small in Case I, the put option prices under quadratic model, extended CIR model, and Hull-White model are very close to the one under the tow-factor CIR model, especially for the ones with short expiration dates. When the tow-factor CIR model has bigger volatility and mean reverting speed in Case II, both the extended CIR model and Hull-White model tend to underestimate the put option prices, while the quadratic model tend to overestimate the put option prices. Nonetheless, all three models are able to capture the put option price curve's shapes.

Table 2.4: Parameters for the TCIR model

Group	σ_1	κ_1	θ_1	$x_1(0)$	σ_2	κ_2	θ_2	$x_2(0)$
I	0.03	0.10	0.05	0.05	0.03	0.01	0.05	0.05
II	0.15	0.16	0.04	0.04	0.15	0.10	0.05	0.05

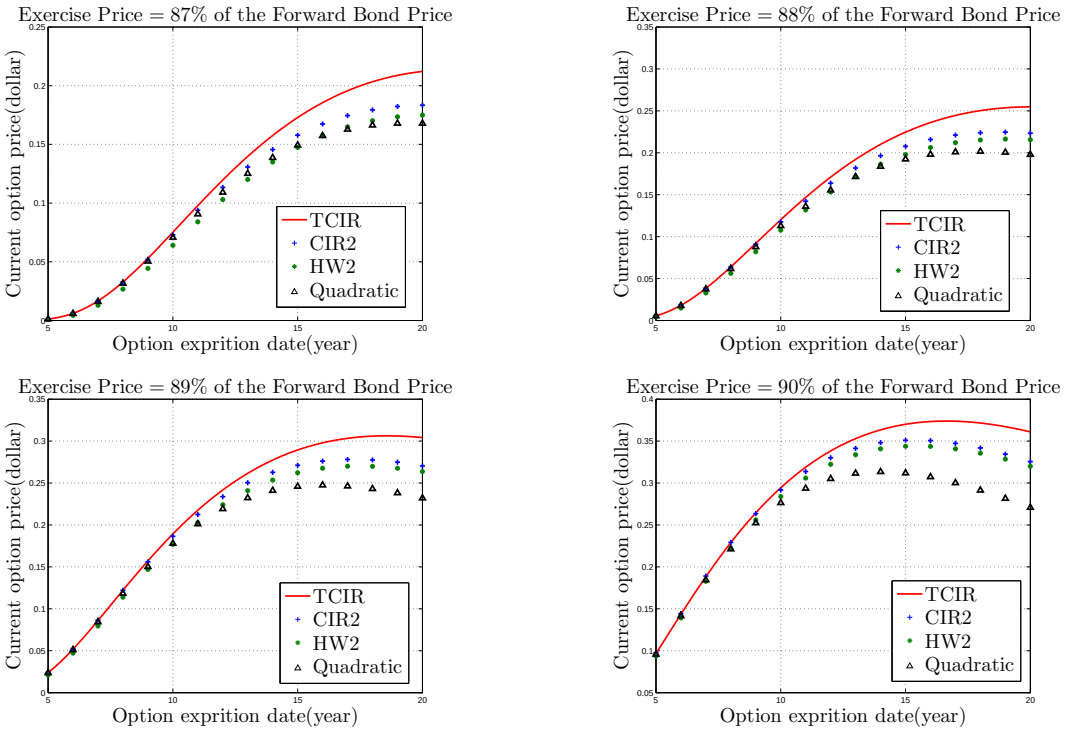

Figure 2.5: American Put Options: Case I

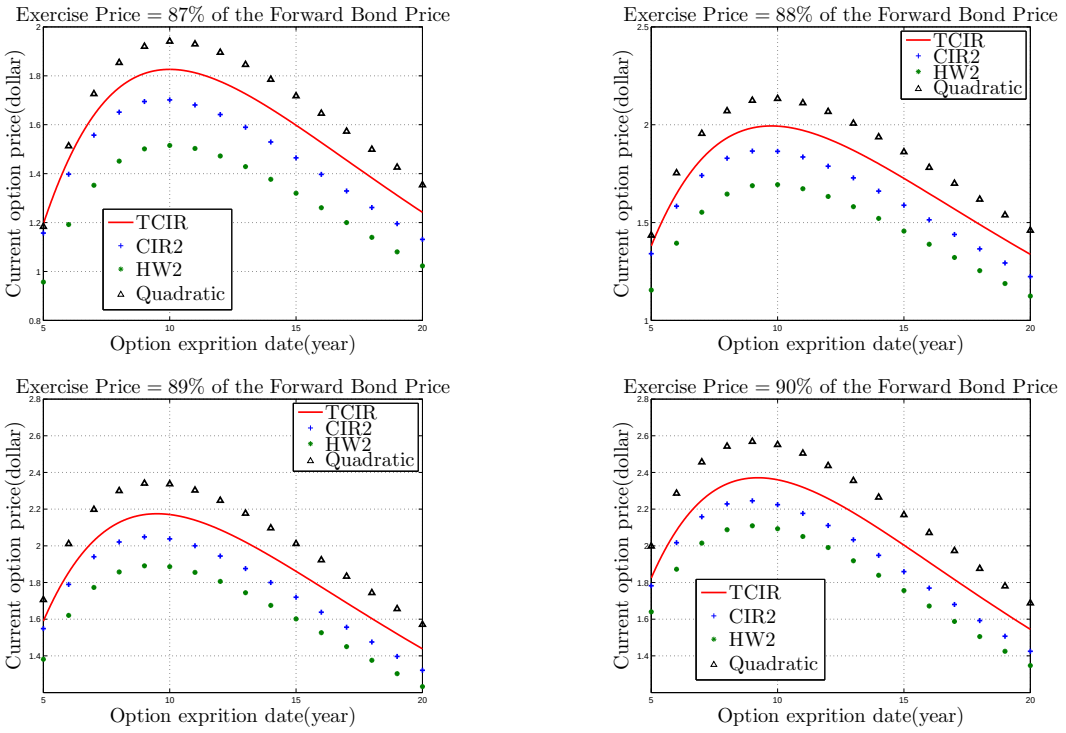

Figure 2.6: American Put Options: Case II

CHAPTER 3

PARAMETER ESTIMATION BY KALMAN FILTER

3.1 Literature Review

Besides modeling the the term structure of interest rates, an extensive study has been focusing on estimating the model parameters. Several prominent methods have been developed to estimate the model parameters over the decades. The maximum likelihood method is naturally employed with the time-series approach, the cross-section approach or their combination (see [61, 14, 7, 2] and the references cited therein).

The maximum likelihood estimation requires the state variable is observable and the likelihood function is analytically known. The simulation-based maximum likelihood approach is developed by using the approximation of the likelihood function when the state is observable ([60, 63]). Without the knowledge of the likelihood function, we may use the available moments of the distribution, which results so call the the quasi-maximum likelihood estimation ([12, 62, 23]). Moreover, Hansen [31] proposes the Nobel prize winning method – generalized method of moments (GMM) that reduces the reliance on distribution assumptions by matching the empirical mo-

ments with the theoretical ones. Chan et al. apply GMM to a variety of continuous models of the short-term riskless rate models [13]. Later, Ingram and Lee[34] and Dai and Singleton [18] proposed the simulated method of moments (SMM) to time-series estimation. As the name suggests, SMM minimizes the reliance on distribution assumptions by matching the empirical moments with the simulated ones.

Another popular estimation technique is the efficient method of moments(EMM) developed by Gallant and Tauchen[26]. The EMM is widely used to estimation term structure models when maximum likelihood is unfeasible. For instance, by using EMM, Andersen and Lund estimate the stochastic volatility models[3], Dai and Singleton estimate the affine models[19], and Ahn et al. estimate the quadratic term structure models[1]. Since the short interest rates are unobservable, the methods based on the Kalman filter have been developed (see [24, 54, 27, 20, 4, 6] and references cited therein). The observed variables are the interest rate derivatives such as yield rates, bond prices, caps, and so on.

It would be beneficial to practitioners to know which estimation method produces the most accurate results. Zhou [70] compares the finite sample properties of EMM, GMM, QMLE, and MLE for a square-root interest rate diffusion model by Monte Carlo simulation and concludes that MLE achieves the most efficient estimation method, QMLE is less efficient than MLE but it provides the best inference, and EMM provide better inference than GMM and MLE in a high volatility scenario. Duffee and Stanton [25] also study the finite-sample properties of MLE, EMM, and Kalman filter methods for term structure models. They conclude that MLE works well for simple models and produces strongly biased parameter estimates when the

model includes some flexible specification terms. EMM performs poorly even in the simplest term structure settings and the linearized Kalman filter is a tractable and reasonably accurate estimation technique. According to the previous study results and Duffee and Stanton's recommendation, this dissertation adopts Kalman filter and its variations to estimate term structure models.

3.2 The Kalman Filter For the Affine Models

3.2.1 The affine models

In affine models, the instantaneous interest rate $r(t)$ is assumed to follow the stochastic differential equation:

$$dr(t) = \kappa(\theta - r(t))dt + \sigma r^{\nu}(t)dW(t) \tag{3.1}$$

where $W(t)$ is a Wiener process under the risk neutral measure and κ, θ, and σ are positive parameters. We have the the Vasicek model and the CIR model for $\nu = 0$ and $\nu = \frac{1}{2}$, respectively. Here parameter κ can be interpreted as the speed of reversion that characterizes the velocity at which $r(t)$ evolves around its long term mean level θ.

Let $P(r, t; T)$ be the T-bond price when the interest rate is r at time t. Then it is the solution of the fundamental partial differential equation

$$P_t + \frac{1}{2}\sigma^2 P_{rr} + \kappa\,(\theta - r)\,P_r - rP = 0, \quad -\infty < x < \infty,\ 0 \le t < T$$

subject to the final condition

$$P(r, T; T) = 1.$$

The favorable feature of the above model is that the yield is an affine function of the short interest rate $r(t)$. In fact, the zero coupon bond under affine models takes the following form:

$$P(t,T;\Psi) = e^{-a(T-t;\Psi)-b(T-t;\Psi)r(t)}. \tag{3.2}$$

where

$$\Psi = (\theta,\kappa,\sigma), \quad \gamma = \sqrt{\kappa^2+2\sigma^2}, \tag{3.3}$$

$$b(\tau;\Psi) = \begin{cases} \dfrac{1-e^{-\kappa\tau}}{\kappa} \text{ for } \nu = 0, \\ \dfrac{2(e^{\gamma\tau}-1)}{2\gamma+(\kappa+\gamma)(e^{\gamma\tau}-1)} \text{ for } \nu = \dfrac{1}{2}, \end{cases} \tag{3.4}$$

$$a(\tau;\Psi) = \begin{cases} \left(\theta - \dfrac{\sigma^2}{2\kappa^2}\right)(\tau - b(\tau;\Psi)) + \dfrac{\sigma^2}{4\kappa}b^2(\tau;\Psi) \text{ for } \nu = 0, \\ -\dfrac{2\kappa\theta}{\sigma^2}\ln\left(\dfrac{2\gamma e^{(\gamma+\kappa)\tau/2}}{2\gamma+(\gamma+\kappa)(e^{\gamma\tau}-1)}\right) \text{ for } \nu = \dfrac{1}{2}, \end{cases} \tag{3.5}$$

Let $z(t,T)$ denote the time continuously compounded yield on a zero coupon bond of maturity T. Then we have

$$z(t,T) = -\frac{\ln P(t,T;\Psi)}{T-t} = \frac{a(T-t;\Psi)+b(T-t;\Psi)r(t)}{T-t}. \tag{3.6}$$

3.2.2 The state-space formulation

To deal with the estimation problem, it is reasonable to assume that the yields for different maturities are observed with measurement errors. Based on (3.6), n yields can be represent as the following system for n bonds with different maturities $\{T_1, T_2, \cdots, T_n\}$ at time t:

$$\underbrace{\begin{bmatrix} z(t,T_1) \\ z(t,T_2) \\ \vdots \\ z(t,T_n) \end{bmatrix}}_{z(t)} = \underbrace{\begin{bmatrix} \frac{a(t,T_1)}{T_1-t} \\ \frac{a(t,T_2)}{T_2-t}) \\ \vdots \\ \frac{a(t,T_n)}{T_n-t} \end{bmatrix}}_{\mathrm{A(t)}} + \underbrace{\begin{bmatrix} \frac{b(t,T_1)}{T_1-t} \\ \frac{b(t,T_2)}{T_2-t}) \\ \vdots \\ \frac{b(t,T_n)}{T_n-t} \end{bmatrix}}_{\mathrm{B(t)}} r(t) + \underbrace{\begin{bmatrix} \upsilon_1(t) \\ \upsilon_2(t) \\ \vdots \\ \upsilon_N(t) \end{bmatrix}}_{\upsilon_t} \tag{3.7}$$

i.e.,

$$z(t) = A + Br(t) + \upsilon(t) \tag{3.8}$$

where $\upsilon \sim \mathcal{N}(0, R)$ with the covariance matrix $R = \text{diag}(\rho_1^2, \ldots, \rho_n^2)$.

The transition equations are slightly more complex. For the Vasicek Model, we can solve the stochastic differential equation (3.1) to obtain

$$r(t) = \theta(1 - e^{-\kappa(t-s)}) + e^{-\kappa(t-s)}r(s) + \int_s^t e^{-\kappa(t-u)}\sigma dW(u)$$

for all $0 \leq s \leq t$. Thus the transition equation for Vasicek Model is as follow:

$$r(t_i) = \theta(1 - e^{-\kappa(t_i - t_{i-1})}) + e^{-\kappa(t_i - t_{i-1})}r(t_{i-1}) + \epsilon(t_i), \tag{3.9}$$

where

$$\epsilon(t_i) = \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i - s)}\sigma dW_s.$$

It follows from the properties of Ito integral that

$$\epsilon_i \mid \mathcal{F}_{t_{i-1}} \sim \mathcal{N}(0, q_i),$$

where

$$q_i = \text{Var}\left[\epsilon(\text{t}_\text{i})|\mathcal{F}_{\text{t}_{\text{i}-1}}\right] = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t_i - t_{i-1})}\right).$$

Hence

$$r(t_i) \mid \mathcal{F}_{t_{i-1}} \sim \mathcal{N}\left(\mu_i, q_i\right), \tag{3.10}$$

where

$$\mu_i = \theta(1 - e^{-\kappa(t_i - t_{i-1})}) + e^{-\kappa(t_i - t_{i-1})}r\left(t_{i-1}\right).$$

However, the transition density in CIR model follows a non-central χ-squared distribution, which is rather difficult to handle. For the purpose of simulation, we use the first two moments of the non-central χ-squared distribution:

$$\mathrm{E}\left[r(t_i) \mid \mathcal{F}_{t_{i-1}}\right] = \theta + (r(t_{i-1}) - \theta)e^{-\kappa(t_i - t_{i-1})}, \tag{3.11}$$

$$\mathrm{Var}\left[\mathrm{r(t_i)} \mid \mathcal{F}_{\mathrm{t_{i-1}}}\right] = \frac{\sigma^2}{2\kappa}\left(1 - e^{-\kappa(t_i - t_{i-1})}\right)\left(\theta + (2r(t_{i-1}) - \theta)e^{-\kappa(t_i - t_{i-1})}\right). \tag{3.12}$$

3.2.3 Kalman filter

In 1960, Rudolph Kalman published his famous paper [38] proposing a powerful linear filtering technique named after him. The Kalman filter is a means that provides an efficient recursive algorithm to estimate the state of a process and unobservable parameters. The Kalman filter is an optimal estimator in a sense of minimizing the mean of the squared error. The reason that Kalman filter is widely used in a lot of areas such as tracking objects, economics and finance, navigation, computer vision, and so on is that it supports estimations of past, present and future. A good Kalman filter tutorial is presented by Terejanu in [65]. We only outline the general Kalman filter algorithm here and skip all the details that can be found in [65]. The Model implementations will be discussed in detail in the following sections.

Consider a stochastic linear system

$$x_k = Ax_{k-1} + Bu_{k-1} + w_{k-1}, \tag{3.13}$$

$$z_k = Hx_k + v_k, \tag{3.14}$$

where the control input u_k is a known nonrandom variable, variable w_k captures the uncertainties of the model, and v_k denotes the noise measurement. We assume that

$w_k \sim \mathcal{N}(0, Q_k)$ and $v_k \sim \mathcal{N}(0, P_k)$ and that w_k and w_j, v_k and v_j, and w_k and v_k are are uncorrelated.

The Kalman filter algorithm includes the following three steps:

Initialization:

$$x_{0|0} = \mu_0 \text{ with error covariance } P_0.$$

Model Forecast step:

$$x_{k|k-1} = Ax_{k-1|k-1} + Bu_{k-1},$$

$$P_{k|k-1} = AP_{k-1|k-1}A^T + Q_{k-1}.$$

Data Assimilation Step:

$$x_{k|k} = x_{k|k-1} + K_k\left(z_k - Hx_{k|k-1}\right),$$

$$K_k = P_{k|k-1}H^T\left(HP_{k|k-1}H + R\right)^{-1},$$

$$P_{k|k} = \left(I - K_kH\right)P_{k|k-1},$$

where P_j is the covariance of the state and K_k is the Kalman gain.

To estimate the parameters in the model, Kalman filter is usually combined with the MLE when the state vector dynamics is Gaussian with the normally distributed noise or the QMLE when the state vector dynamics is not Gaussian or the noise is not normally distributed.

The Kalman filter algorithm for the Vasicek and CIR models is described in detail as follows.

Step 1: Initializing the state vector. First, we need to find the appropriate starting values for recursion. The unconditional mean and variance of transition sys-

tems are good choices. The unconditional mean, for both the CIR and Vasicek models, has the following form:

$$r_{0|0} = \mathrm{E}\left[r_0\right] = \theta \tag{3.15}$$

The unconditional variance is

$$P^r_{0|0} = \mathrm{Var}\left[\mathrm{r_0}\right] = \begin{cases} \dfrac{\sigma^2}{2\kappa} & \text{for the Vasicek model,} \\ \dfrac{\sigma^2\theta}{2\kappa} & \text{for the CIR model.} \end{cases} \tag{3.16}$$

Step 2: Forecasting the measurement equation. Assume that we have an optimal estimate $r_{i-1|i-1} = \mathrm{E}\left[r_{i-1} \mid \mathcal{F}_{i-1}\right]$ with $P^r_{i-1|i-1} = \mathrm{Var}\left[\mathrm{r_{i-1}} \mid \mathcal{F}_{\mathrm{i-1}}\right]$ at time $i-1$. The conditional forecast of the measurement equation has the following form:

$$z_{i|i-1} = \mathrm{E}\left[z_i \mid \mathcal{F}_{i-1}\right] = A + B\mathrm{E}\left[r_i \mid \mathcal{F}_{i-1}\right] = A + Br_{i-1|i-1}. \tag{3.17}$$

The associated conditional variance is,

$$P^z_{i|i-1} = \mathrm{Var}\left[\mathrm{z_i} \mid \mathcal{F}_{\mathrm{i-1}}\right] = B\mathrm{Var}\left[\mathrm{r_i} \mid \mathcal{F}_{\mathrm{i-1}}\right]B^T + R = BP^r_{i|i-1}B^T + R. \tag{3.18}$$

Step 3: Updating the inference about the state vector. The observed true value of the measurement system, z_i gives us a sense of the error in the conditional prediction, which can be denoted as

$$\zeta_i = z_i - z_{i|i-1}. \tag{3.19}$$

At the current point in the Kalman filter algorithm, this prediction error is used to update our inference about the unobserved transition system. This updating takes the form of revising our conditional expectation with the underlying

expression

$$r_{i|i} = \mathrm{E}\left[r_i \mid \mathcal{F}_i\right] = \mathrm{E}\left[r_i \mid \mathcal{F}_{i-1}\right] + K_i\zeta_i = r_{i|i-1} + K_i\zeta_i, \tag{3.20}$$

where

$$K_i = \mathrm{Var}\left[\mathrm{r_i} \mid \mathcal{F}_{\mathrm{i-1}}\right]B^T\mathrm{Var}\left[\mathrm{z_i} \mid \mathcal{F}_{\mathrm{i-1}}\right]^{-1} = P^x_{i|i-1}B^T{P^z_{i|i-1}}^{-1}. \tag{3.21}$$

is called the Kalman gain matrix. The gain matrix determines the weight given to the new observation in the updated state system forecast. The conditional variance of the stated system is also updated as follow:

$$P^r_{i|i} = \mathrm{Var}\left[\mathrm{r_i} \mid \mathcal{F}_{\mathrm{i}}\right] = (I - K_iB)P^r_{i|i-1}. \tag{3.22}$$

Step 4: Constructing the likelihood function. The previous four steps have to be repeated for each discrete time step in the data sample. In our test, we use weekly US treasury yield data over a period of ten years. To actually implement this algorithm to estimate the parameters, we initialize the state vector using equation (3.15) and (3.16) and then iterate on equations (3.17) to (3.22). At each step, we generate a measurement-system prediction error ζ_i and a prediction error variance $\mathrm{Var}\left[\mathrm{r_i} \mid \mathcal{F}_{\mathrm{i-1}}\right]$. With the first two moments, we can construct the log-likelihood function for Vasicek model and log-quasi-likelihood function for the CIR model. It will have the following form

$$\mathcal{L}(\theta) = -\frac{nN\ln(2\pi)}{2} - \frac{1}{2}\sum_{i=1}^{N}\left(\ln(\det(P^r_{i|i-1}) + \zeta_i^T{P^r_{i|i-1}}^{-1}\zeta_i\right). \tag{3.23}$$

In other words, Step 1 through Step 4 are used for the construction of a log-likelihood function. To find the optimal parameters, we need to treat the above

log-likelihood function as our objective function and use nonlinear numerical optimization toolbox in MATLAB to find the maximum (or minimum of the opposite of the likelihood function).

3.3 The Kalman Filter For the Quadratic Model

Recall that the instantaneous interest rate $r(t)$ for the quadratic model is assumed to be given by (Chapter 2)

$$r(t) = \frac{1}{2}x(t)^2,$$

where the state variable $x(t)$ that follows the stochastic differential equation:

$$dx(t) = \kappa(t)(\theta(t) - x(t))dt + \sigma(t)dW_t. \tag{3.24}$$

This SDE is identical to the one in Vasiceck Model, which means the state transition equation should be the same too. We also have state transition equations as (3.9):

$$x(t_i) = \underbrace{\theta(1 - e^{-\kappa(t_i - t_{i-1})})}_{\text{E}} + \underbrace{e^{-\kappa(t_i - t_{i-1})}}_{\text{F}} x(t_{i-1}) + \epsilon(t_i) \tag{3.25}$$

or

$$x_i = E + Fx_{i-1} + \epsilon_i \tag{3.26}$$

with the same definition of $\epsilon(t_i)$. Unlike the Vasicek Model and CIR Model, the yield-to-maturity of the quadratic model is a quadratic function of the state variable instead of linear functions.

$$z(t,T) = -\frac{\ln\left(P(x(t);t,T)\right)}{T-t} = \frac{\frac{1}{2}c(t,T)x(t)^2 + b(t,T)x(t)d + a(t,T)}{T-t} \tag{3.27}$$

where $a(t,T)$, $b(t,T)$, and $c(t,T)$ are defined in (4.14)–(4.16).

To estimation the parameters κ, θ and σ in the above model, we need to employ the same assumption as we did for the Vasiceck Model and CIR Model that the yields for the different maturities have measurement errors. Analog to equation (3.7), n yields can be represent as the following system for n bonds with different maturities $[T_1, T_2, \cdot, T_n]$ at time t:

$$\underbrace{\begin{bmatrix} z(t,T_1) \\ z(t,T_1) \\ \vdots \\ z(t,T_n) \end{bmatrix}}_{z(t)} = \underbrace{\begin{bmatrix} \frac{a(t,T_1)}{T_1-t} \\ \frac{a(t,T_2)}{T_2-t} \\ \vdots \\ \frac{a(t,T_n)}{T_n-t} \end{bmatrix}}_{A^Q} + \underbrace{\begin{bmatrix} \frac{b(t,T_1)}{T_1-t} \\ \frac{b(t,T_2)}{T_2-t} \\ \vdots \\ \frac{b(t,T_n)}{T_n-t} \end{bmatrix}}_{B^Q} x(t) + \frac{1}{2} \underbrace{\begin{bmatrix} \frac{c(t,T_1)}{T_1-t} \\ \frac{c(t,T_2)}{T_2-t} \\ \vdots \\ \frac{c(t,T_n)}{T_n-t} \end{bmatrix}}_{C^Q} x(t)^2 + \underbrace{\begin{bmatrix} \upsilon_1(t) \\ \upsilon_2(t) \\ \vdots \\ \upsilon_N(t) \end{bmatrix}}_{\upsilon_t}, \tag{3.28}$$

i.e.,

$$z(t) = A^Q + B^Q x_t + C^Q x_t^2 + \upsilon_t \tag{3.29}$$

where $\upsilon \sim \mathcal{N}(0, R)$ with the covariance matrix $R = \text{diag}(\rho_1^2, \ldots, \rho_n^2)$. To deal with the nonlinear measurement, we implement and compare three Kalman filter variations, namely, the extended Kalman filter, unscented Kalman filter and Kalman filter for the quadratic measurement function (3.29).

3.3.1 The extended Kalman filter

A natural way to apply Kalman filter to the nonlinear measurement function is to linearize it at $x^f(t)$ obtained from the forecast step, i.e.,

$$\begin{aligned} z(t) &= A^Q + B^Q x(t) + C^Q x(t)^2 + \upsilon_t \\ &= A^Q + B^Q x^f(t) + C^Q \left(x^f(t)\right)^2 + B^Q \left(x(t) - x^f(t)\right) \\ &+ 2C^Q x^f(t) \left(x(t) - x^f(t)\right) + C^Q \left(x(t) - x^f(t)\right)^2 + \upsilon_t \end{aligned}$$

$$
\begin{aligned}
&\approx A^Q + B^Q x^f(t) + C^Q \left(x^f(t)\right)^2 + B^Q \left(x(t) - x^f(t)\right) \\
&+ 2C^Q x^f(t) \left(x(t) - x^f(t)\right) + \upsilon_t \\
&= A^Q - C^Q \left(x^f(t)\right)^2 + \left(B^Q + 2C^Q x^f(t)\right) x(t) + \upsilon_t \\
&= A + Bx(t) + \upsilon_t,
\end{aligned}
$$

where

$$
A = A^Q - C^Q \left(x^f(t)\right)^2, \quad B = B^Q + 2C^Q x^f(t).
$$

The above approximation together with (3.26) forms the following stochastic linear system:

$$
x_i = E + Fx_{i-1} + \epsilon_i, \tag{3.30}
$$

$$
z_i = A + Bx_i + \upsilon_i. \tag{3.31}
$$

We can apply the standard Kalman filter in §3.2.3 to estimate the parameters of the system, which leads to the following extended Kalman filter algorithm:

Step 1: Initializing the state vector. The unconditional mean has the following form:

$$
x_{0|0} = \mathrm{E}\left[x_0\right] = \theta.
$$

The unconditional variance is

$$
P^x_{0|0} = \mathrm{Var}\left[\mathrm{x}_0\right] = \frac{\sigma^2}{2\kappa}.
$$

Step 2: Forecasting the measurement equation. With an optimal estimate $x_{i-1|i-1} = \mathrm{E}[x_{i-1}| \mid \mathcal{F}_{i-1}]$ with $P^x_{i-1|i-1} = \mathrm{Var}[\mathrm{x_{i-1}}| \mid \mathcal{F}_{\mathrm{i-1}}]$ available at time $i-1$, the conditional forecast of the measurement equation has the following form:

$$z_{i|i-1} = \mathrm{E}[z_i \mid \mathcal{F}_{i-1}] \approx A + B\mathrm{E}[x_i \mid \mathcal{F}_{i-1}] = A + Bx_{i-1|i-1}. \tag{3.32}$$

The associated conditional variance is,

$$P^z_{i|i-1} = \mathrm{Var}[\mathrm{z_i} \mid \mathcal{F}_{\mathrm{i-1}}] \approx B\mathrm{Var}[\mathrm{x_i} \mid \mathcal{F}_{\mathrm{i-1}}]B^T + R = BP^x_{i|i-1}B^T + R.$$

Step 3: Updating the inference about the state vector. The error in the conditional prediction is

$$\zeta_i = z_i - z_{i|i-1}.$$

and the updating conditional expectation is defined as

$$x_{i|i} = \mathrm{E}[x_i \mid \mathcal{F}_i] = \mathrm{E}[x_i \mid \mathcal{F}_{i-1}] + K_i\zeta_i = x_{i|i-1} + K_i\zeta_i,$$

where the Kalman gain is

$$K_i = \mathrm{Var}[\mathrm{x_i} \mid \mathcal{F}_{\mathrm{i-1}}]B^T\mathrm{Var}[\mathrm{z_i} \mid \mathcal{F}_{\mathrm{i-1}}]^{-1} = P^x_{i|i-1}B^TP^z_{i|i-1}.$$

The conditional variance of the stated system is updated as follow:

$$P^x_{i|i} = \mathrm{Var}[\mathrm{x_i} \mid \mathcal{F}_{\mathrm{i}}] = (I - K_iB)P^x_{i|i-1}.$$

Step 4: Constructing the likelihood function. The log-likelihood function has the following form

$$\mathcal{L}(\theta) = -\frac{nN\ln(2\pi)}{2} - \frac{1}{2}\sum_{i=1}^{N}\left(\ln(\det(P^x_{i|i-1})) + \zeta_i^T {P^x_{i|i-1}}^{-1}\zeta_i\right).$$

We should point out that the main advantage of the extended Kalman filter is that the estimation procedure is almost the same as the standard Kalman filter recursion and is very easy to use, to understand and computationally efficient. However, the extended Kalman filter has some limitations such as not working in considerable nonlinearities, the differentiable requirement on state transition equation and measurement equation and unstablity of computing Jacobian matrices for high dimension problems. Fortunately, both the state transition equation and the measurement equation are first order and second order polynomials in the quadratic term structure and we do not have to worry about their nonlinearity, differentiability and computation of the Jacobian matrices.

3.3.2 The unscented Kalman filter

The unscented Kalman filter (UKF) was proposed by Julier and Uhlman in their sensational paper[37]. Unlike the extended Kalman filter, the unscented Kalman filter does not linearize the measurement equation. Rather than approximation the nonlinear measurement function $z(t)$ in equation (3.29), the unscented Kalman filter approximates the conditional distribution of the state variable $x(t)$ using the unscented transformation [52]. The unscented transformation is a method of using the statistical linearization technique that is used to linearize a nonlinear function of a random variable through a linear regression between n points drawn from the prior distribution of the random variable. The UKF is a derivative-free alternative to the extended Kalman filter, thus it overcomes the extended Kalman filter's limitations such as the posterior mean and covariance could be corrupted due to the state distri-

bution propagated analytically through the first-order linearization of the nonlinear equations. Therefore, UKF is more accurate than Taylor series linearization [67]. In [68], Wan and Van der Merwe employ UKF for nonlinear estimation and conclude that UKF consistently achieves a better level of accuracy than EKF within a number of application domains including parameter estimation. Even though the UKF is extensively used in the engineering literature, it has not been widely used in the empirical asset pricing literature [15]. Among fewer papers that use UKF to estimate the term structure models, Leippold and Wu estimate a series of multi-currency quadratic models by using UKF [53], Nyholm and Rositsa also apply UKF to estimate multi-factor quadratic models. Besides, unlike most other papers using standard Kalman filter to estimate the affine term structure models, Christoffersen et al. conduct extensive study on showing that UKF outperforms the standard and extended Kalman filter in parameter estimation and forecasting swap rates and caps for affine term structure models with nonlinear instruments. In this section, we shall rely on the unscented Kalman filter to estimate the quadratic interest rate models similar to Leippold and Wu and Nyholm and Vido-Koleva.

To understand the unscented Kalman filter, we need to be clear about the unscented transform that is used to statistically linearize the nonlinear functions. Consider an $n-$ dimension random variable x with mean $\bar{x}$ and covariance P_x is propagated through a nonlinear function $y = f(x)$. To calculate the mean and covariance of y, we construct a matrix $\mathcal{X}$ of $2n+1$ sigma points χ_i with the corresponding weights

w_i as follows:

$$\begin{aligned}
&\chi_0 = \bar{x},\\
&\chi_i = \bar{x} + (\sqrt{(n+\lambda)P_x})_i \quad i = 1, 2, \ldots, n,\\
&\chi_i = \bar{x} - (\sqrt{(n+\lambda)P_x})_i \quad i = n+1, n+2, \ldots, 2n,\\
&w_0^m = \frac{\lambda}{n+\lambda},\\
&w_0^c = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta),\\
&w_i^m = w_i^c = \frac{1}{2(n+\lambda)} \quad i = 1, 2, \ldots, 2n,
\end{aligned}$$

where $\lambda = \alpha^2(n+\kappa) - n$ is a scaling parameter, α controls the spread of the sigma points around $\bar{x}$ and is set to a small positive value, κ is a secondary scaling parameter that is usually set to 0, and β incorporates prior knowledge of the distribution of x. When the distribution of x_t is Gaussian, we have $\beta = 2$, $\kappa = 3 - n$ or 0, and $\alpha = 1$ for low dimensional problems. Here $(\sqrt{(n+\lambda)P_x})_i$ denotes the ith row of the square root of the matrix. It is easy to see that for any α, κ, β, we have:

$$\sum_{i=0}^{2n} w_i^m \chi_i = \bar{x}, \tag{3.33}$$

$$\sum_{i=0}^{2n} w_i^m (\chi_i - \bar{x})(\chi_i - \bar{x})' = \sum_{i=0}^{2n} w_i^c (\chi_i - \bar{x})(\chi_i - \bar{x})' = P_x. \tag{3.34}$$

These sigma points are propagated through the nonlinear function

$$y_i = f(\chi_i) \quad i = 0, 1, \ldots, 2n$$

and the mean and covariance of y are approximated using a weight sample mean and

covariance of the posterior sigma points:

$$\bar{y} \approx \sum_{i=0}^{2n} w_i^m y_i, \tag{3.35}$$

$$P_y \approx \sum_{i=0}^{2n} (y_i - \bar{y})(y_i - \bar{y})'. \tag{3.36}$$

As discussed in [51], the approximations of the unscented transform are accurate to the third order for Gaussian inputs for all nonlinearities and at least the second order for non-Gaussian inputs. A simple comparison of EKF and UKF is show in Figure 3.1 for a 2−dimensional system. The true mean and covariance propagation using Monte-Carlo sampling are shown in the left; the results of EKF are shown in the center and the right plots show the performance of the unscented transform with 5 sigma points. It is clear that the unscented transform does a better job than EKF to approximate the mean and covariance of the nonlinear function y. The Unscented Kalman Filter(UKF) is a straightforward extension of the Kalman filter incorporating the unscented transform of (3.33), (3.34), (3.35) and (3.36). Notice that the transition equation (3.26) in quadratic interest model is linear, we do not need to calculate $\bar{x}$ and P_x through equations (3.33) and (3.34). In another word, we can use the standard Kalman filter for $\bar{x}$ and P_x.

The algorithm of UKF for parameter estimation in system (3.26) and (3.31) is formulated in the following algorithm:

Step 0: Calculating the weights. Since the weights are independent to the rest steps, we should calculate the weight at the very beginning.

$$w_0^m = \frac{\lambda}{n + \lambda},$$

$$w_0^c = \frac{\lambda}{n+\lambda} + (1 - \alpha^2 + \beta),$$
$$w_i^m = w_i^c = \frac{1}{2(n+\lambda)} \quad i = 1, 2, \ldots, 2n.$$

Step 1: Initializing the state vector. First, we also need to assign the unconditional mean and variance of transition systems starting values for recursion as we did in the standard Kalman filter for Vasicek models since they share the same transition equation.

$$x_{0|0} = \mathrm{E}\left[x_1\right] = \mathrm{E}\left[x_1 \mid \mathcal{F}_0\right] = \theta.$$

The unconditional variance for the Quadratic state variable is

$$P_{0|0}^x = \mathrm{Var}\left[\mathrm{x}_1\right] = \mathrm{Var}\left[\mathrm{x}_1 \mid \mathcal{F}_0\right] = \frac{\sigma^2}{2\kappa}.$$

Step 2: Predicting state variable. After having an estimate $x_{i-1|i-1} = \mathrm{E}\left[x_{i-1}| \mid \mathcal{F}_{i-1}\right]$ with $P_{i-1|i-1}^x = \mathrm{Var}\left[\mathrm{x_{i-1}}| \mid \mathcal{F}_{\mathrm{i-1}}\right]$ at time $i-1$, we can predict the state variable and its covariance at time t by

$$x_{i|i-1} = C + F x_{i-1|i-1},$$
$$P_{i|i-1}^x = F P_{i-1|i-1}^x F' + Q.$$

Step 3: Computing the $2n+1$ sigma points. With the state prediction, the sigma

points can be calculated as

$$\chi_0 = x_{i|i-1}.$$

$$\chi_i = x_{i|i-1} + (\sqrt{(n+\lambda)P^x_{i|i-1}})_i, \quad i = 1, 2, \ldots, n,$$

$$\chi_i = x_{i|i-1} - (\sqrt{(n+\lambda)P^x_{i|i-1}})_i, \quad i = n+1, n+2, \ldots, 2n.$$

Step 4: Forecasting the measurement equation. The measurement forecasting is updated by propagating the sigma points though the nonlinear measurement function $z(t)$ in equation (3.31). The conditional forecast of the measurement equation has the following form:

$$z_{i|i-1} = \sum_{i=0}^{2n} w_i z(\chi_i).$$

The associated conditional variance is

$$P^z_{i|i-1} = \sum_{i=0}^{2n} w_i^c [z(\chi_i) - z_{i|i-1}][z(\chi_i) - z_{i|i-1}]' + R.$$

Step 5: Updating the inference about the state vector. The observed true value of the measurement system, z_i gives us a sense of the error in the conditional prediction, which can be denoted as

$$\zeta_i = z_i - z_{i|i-1}.$$

The Kalman gain matrix for UFK is defined as

$$K_i = \sum_{i=0}^{2n} w_i^c [\chi_i - x_{i|i-1}][z(\chi_i) - z_{i|i-1}]' {P^z_{i|i-1}}^{-1}$$

and the state variable is updated by

$$x_{i|i} = x_{i|i-1} + K_i \zeta_i$$

with covariance

$$P^x_{i|i} = P^x_{i|i-1} - K_i P^z_{i|i-1} K_i'.$$

Step 6: Constructing the likelihood function. Like the log-quasi-likelihood function in the CIR model, the log-quasi-likelihood function of the quadratic model has the following form

$$\mathcal{L}(\theta) = -\frac{nN\ln(2\pi)}{2} - \frac{1}{2}\sum_{i=1}^{N}\left(\ln(\det(P^x_{i|i-1})) + \zeta_i^T {P^x_{i|i-1}}^{-1}\zeta_i\right). \tag{3.37}$$

Again, to find the optimal parameters, we need to treat the above log-quasi-likelihood function as our objective function and use nonlinear numerical optimization toolbox in MATLAB to find the maximum (or minimum of the opposite of the likelihood function).

3.3.3 The quadratic Kalman filter

Even though the unscented Kalman filter does a fine job estimating the parameters in the quadratic model, we still hope to find a way to apply the standard Kalman filter directly to the quadratic model since the nonlinear measurement equation is in a relatively simple form of quadratic function. Monfort, Renne and Roussellet propose the quadratic Kalman filter to the linear transition equation and quadratic measurement equation in any dimension [57]; meanwhile we independently discover the same technique in 1−dimension case. Here we shall demonstrate quadratic Kalman filter

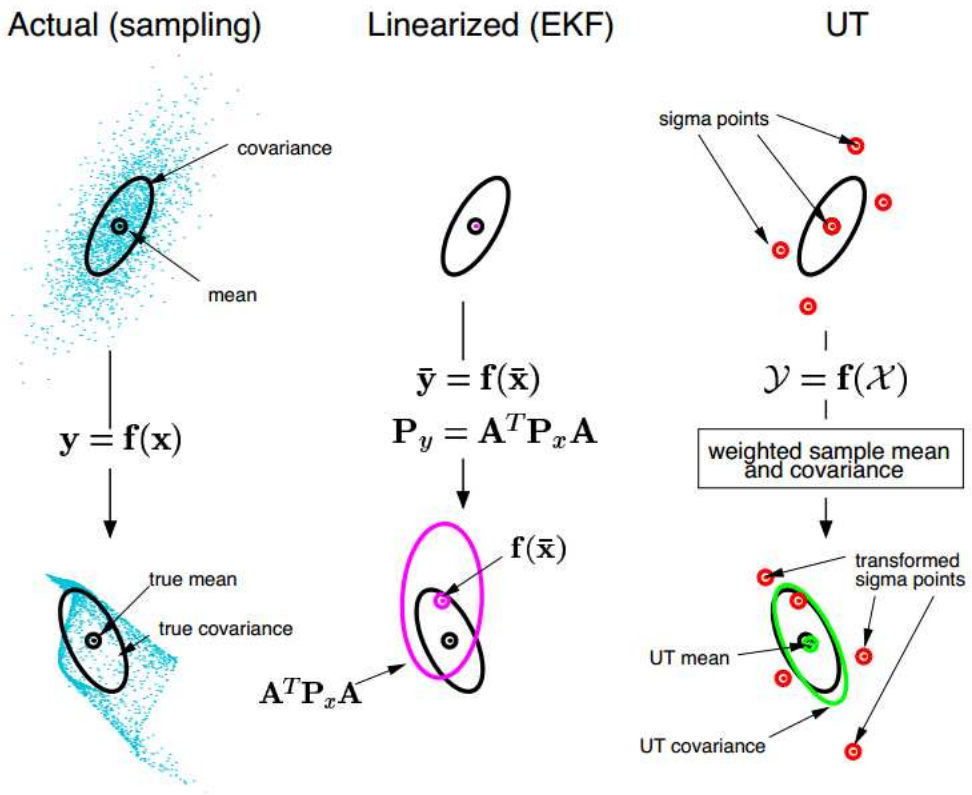

Figure 3.1: Comparison of the UT for mean and covariance propagation.

estimator for 1−dimension problem that is formulated in (3.30) and (3.31). The basic idea behind the quadratic Kalman filter is increasing the dimension of the transition and measurement equations to reduce the nonlinearality in (3.31).

Equations (3.30) and (3.31) form the so-called linear-quadratic state-space model:

$$x_i = E + Fx_{i-1} + \epsilon_i,$$

$$z_i = A + Bx_i + Cx_i^2 + \upsilon_i.$$

In order to reduce the above nonlinear system to a linear system, we regard x_i^2 as a new state variable and let

$$X_i = \begin{bmatrix} x_i \\ x_i^2 \end{bmatrix}.$$

Then the measurement equation can be expressed as a linear equation of X_i:

$$z_i = A + DX_i + \upsilon_i \tag{3.38}$$

where $D = \begin{bmatrix} B & C \end{bmatrix}$. If we can find a linear transition equation for the new Variable X_i, then we can estimate the parameters by using the standard Kalman filter. It is easy to get

$$x_i^2 = (E + Fx_{i-1} + \epsilon_i)^2 = E^2 + 2(E + \epsilon_i)Fx_{i-1} + F^2x_{i-1}^2 + 2E\epsilon_i + \epsilon_i^2 \tag{3.39}$$

Thus the transition equation of X_i satisfies

$$X_i = \alpha + \beta_i X_{i-1} + \eta_i \tag{3.40}$$

where

$$\alpha = \begin{bmatrix} E \\ E^2 \end{bmatrix}, \quad \beta_i = \begin{bmatrix} F & 0 \\ 2(E + \epsilon_i)F & F^2 \end{bmatrix}, \quad \eta_i = \begin{bmatrix} \epsilon_i \\ 2E\epsilon_i + \epsilon_i^2 \end{bmatrix}.$$

Now equations (3.40) and (3.38) form a linear state-space model, on which the standard Kalman filter could be applied. The following theorem shows the first two conditional moments of X_i.

Theorem 3.1. *The first two conditional moments of X_i in the transition equation* (3.40) *are given by*

$$\mathrm{E}\left[X_i \mid X_{i-1}\right] = \alpha + \beta X_{i-1} \tag{3.41}$$

$$\mathrm{Var}\left[\mathrm{X_i} \mid \mathrm{X_{i-1}}\right] = P^x_{i|i-1}(X_{i-1}) \tag{3.42}$$

where

$$\beta = \begin{bmatrix} F & 0 \\ 2EF & F^2 \end{bmatrix}, \quad P^x_{i|i-1}(X_{i-1}) = \begin{bmatrix} Q & 2Q(E+Fx_{i-1}) \\ 2Q(E+Fx_{i-1}) & Q(E+Fx_{i-1})^2 + 2Q^2 \end{bmatrix}.$$

Proof. It follows from transition equation (3.40) that

$$\begin{aligned} \mathrm{E}\left[X_i \mid X_{i-1}\right] &= \mathrm{E}_{i-1}[\alpha + \beta_i X_{i-1} + \epsilon_i] \\ &= \alpha + \mathrm{E}_{i-1}[\beta_i] X_{i-1} \\ &= \alpha + \beta X_{i-1}. \end{aligned}$$

Since $\mathrm{E}_{i-1}[\epsilon_i^k] = 0$ when i is odd, the variance can be calculated according to equation ((3.39)).

$$\begin{aligned} \mathrm{Var}_{i-1}[x_i^2] &= \mathrm{Var}_{i-1}[(E^2 + 2Ex_{i-1} + F^2 x_{i-1}^2) + (2E\epsilon_i + 2Fx_{i-1}\epsilon_i + \epsilon_i^2)] \\ &= \mathrm{Var}_{i-1}[(2E + 2Fx_{i-1})\epsilon_i \epsilon_i + \epsilon_i^2] \\ &= 4(E + F^x_{i-1})^2 \mathrm{Var}_{i-1}[\epsilon_i] + \mathrm{Var}_{i-1}[\epsilon_i^2] \\ &= 4(E + F^x_{i-1})^2 Q + 2Q^2. \end{aligned}$$

To find $\mathrm{Var}_{i-1}[X_i]$, we only need to compute $\mathrm{Cov}_{i-1}[x_i, x_i^2]$ that equals

$$\begin{aligned}
\mathrm{Cov}_{i-1}[x_i, x_i^2] &= \mathrm{Cov}_{i-1}[\epsilon_i, 2(E + Fx_{i-1})\epsilon_i + \epsilon_i^2] \\
&= 2(E + Fx_{i-1})\mathrm{Var}_{i-1}[\epsilon_i, \epsilon_i] + \mathrm{Cov}_{i-1}[\epsilon_i, \epsilon_i^2] \\
&= 2(E + Fx_{i-1})Q + \mathrm{E}_{i-1}[\epsilon_i^3] - \mathrm{E}_{i-1}[\epsilon_i]\mathrm{E}_{i-1}[\epsilon_i^2] \\
&= 2(E + Fx_{i-1})Q.
\end{aligned}$$

Finally, the conditional covariance matrix

$$\mathrm{Var}_{i-1}[X_i] = \begin{bmatrix} Q & 2Q(E + Fx_{i-1}) \\ 2Q(E + Fx_{i-1}) & Q(E + Fx_{i-1})^2 + 2Q^2 \end{bmatrix}.$$

□

Next, we need to derive the first two unconditional moments of X_i to initialize Kalman filter.

Theorem 3.2. *The unconditional expectation μ and covariance matrix Σ of X_i are given as follows:*

$$\mu = (I - \beta)^{-1}\alpha \tag{3.43}$$

$$\mathrm{Vec}[\Sigma] = (I_4 - \beta \otimes \beta)^{-1}\mathrm{Vec}[P^x(\mu)] \tag{3.44}$$

. where $\mathrm{Vec}[\cdot]$ *is the vectorization of a matrix, i.e., a linear transformation which converts the matrix into a column vector and*

$$\begin{aligned}
P^x(\mu) &= \begin{bmatrix} Q & 2Q(E + V) \\ 2Q(E + V) & Q(E^2 + W) + 2Q^2 \end{bmatrix} \\
V &= [F, 0]\mu \\
W &= [2EF, F^2]\mu.
\end{aligned}$$

Proof. By applying the law of total expectation and Theorem 3.1, we have the unconditional mean μ

$$\begin{aligned}
\mu &= \lim_{i\to\infty} \mathrm{E}[X_i \mid X_0] \\
&= \lim_{i\to\infty} \mathrm{E}_0\left[\mathrm{E}_1\left[\mathrm{E}_2\left[\cdots \mathrm{E}_{i-1}\left[X_i\right]\right]\right]\right] \\
&\vdots \\
&= \lim_{i\to\infty} \mathrm{E}_0\left[\mathrm{E}_1\left[\mathrm{E}_2\left[\cdots \mathrm{E}_{i-2}\left[\alpha + \beta X_{i-1}\right]\right]\right]\right] \\
&= \lim_{i\to\infty} (I + \beta + \beta^2 + \cdots + \beta^{i-1})\alpha + \beta^i \alpha X_0 \\
&= (I - \beta)^{-1}\alpha.
\end{aligned}$$

Next, the unconditional variance Σ can be derived by using we use law of total variance as well as Theorem 3.1.

$$\begin{aligned}
\mathrm{Var}[X_i] =& \mathrm{E}\left[\mathrm{Var}_{i-1}[X_i]\right] + \mathrm{Var}\left[\mathrm{E}_{i-1}[X_i]\right] \\
=& \mathrm{E}\left[P^x_{i|i-1}(X_{i-1})\right] + \mathrm{Var}\left[\alpha + \beta X_{i-1}\right] \\
=& P^x_{i|i-1}\left(\mathrm{E}[X_{i-1}]\right) + \beta \mathrm{Var}[X_{i-1}]\beta^T \\
=& P^x(\mu) + \beta \mathrm{Var}[X_{i-1}]\beta^T
\end{aligned}$$

Let $i \to \infty$, we have the following matrix equation

$$\Sigma = P^x(\mu) + \beta\Sigma\beta^T$$

whose solution can be expressed as follows

$$\mathrm{Vec}[\Sigma] = (I_4 - \beta \otimes \beta)^{-1}\mathrm{Vec}[P^x(\mu)]$$

□

Now with the above two theorems, we can use the standard Kalman filter to estimate parameters in the quadratic model. The corresponding algorithm is as follows.

Step 1: Initializing the state vector. We initialize the state vector by the unconditional mean and variance:

$$X_{0|0} = \mathrm{E}\left[X_0\right] = \mu$$

$$P^x_{0|0} = \mathrm{Var}\left[\mathrm{X_0}\right] = \Sigma$$

where μ and Σ are defined in Theorem 3.2.

Step 2: Forecasting the measurement equation. Suppose we have optimal estimate $X_{i-1|i-1} = \mathrm{E}\left[X_{i-1}| \mid \mathcal{F}_{i-1}\right]$ with $P^x_{i-1|i-1} = \mathrm{Var}\left[\mathrm{X_{i-1}}| \mid \mathcal{F}_{\mathrm{i-1}}\right]$ available at time $i-1$, then the conditional forecast of the measurement equation has the following form:

$$z_{i|i-1} = \mathrm{E}\left[z_i \mid \mathcal{F}_{i-1}\right] = A + B\mathrm{E}\left[X_i \mid \mathcal{F}_{i-1}\right] = A + BX_{i-1|i-1}.$$

The associated conditional variance is given by,

$$P^z_{i|i-1} = \mathrm{Var}\left[\mathrm{z_i} \mid \mathcal{F}_{\mathrm{i-1}}\right] = B\mathrm{Var}\left[\mathrm{X_i} \mid \mathcal{F}_{\mathrm{i-1}}\right]B^T + R = BP^x_{i|i-1}B^T + R.$$

Step 3: Updating the inference about the state vector. The error between the observed data and the conditional prediction is

$$\zeta_i = z_i - z_{i|i-1}$$

and the updating conditional expectation is defined as

$$\mathrm{E}\left[X_i \mid \mathcal{F}_i\right] = \mathrm{E}\left[X_i \mid \mathcal{F}_{i-1}\right] + K_i\zeta_i = X_{i|i-1} + K_i\zeta_i,$$

where the Kalman gain is

$$K_i = \text{Var}\left[\text{X}_\text{i} \mid \mathcal{F}_{\text{i}-1}\right] B^T \text{Var}\left[\text{z}_\text{i} \mid \mathcal{F}_{\text{i}-1}\right]^{-1} = P^x_{i|i-1} B^T {P^z_{i|i-1}}^{-1}.$$

The conditional variance of the stated system is updated as follows:

$$P^x_{i|i} = \text{Var}\left[\text{X}_\text{i} \mid \mathcal{F}_\text{i}\right] = (I - K_i B) P^x_{i|i-1}.$$

Step 4: Constructing the likelihood function. The quasi-log-likelihood function has the following form

$$\mathcal{L}(\theta) = -\frac{nN\ln(2\pi)}{2} - \frac{1}{2}\sum_{i=1}^{N}\left(\ln(det(P^x_{i|i-1})) + \zeta_i^T {P^x_{i|i-1}}^{-1}\zeta_i\right).$$

As seen above, the quadratic Kalman filter algorithm is identical to the standard Kalman filter algorithm once we convert the linear-quadratic state-space model to linear state-space model. Since the X_i is not normally distributed, we construct the quasi-log-likelihood function using the first two moments rather than building the log-likelihood function.

3.4 Numerical Results

3.4.1 Simulation results

This subsection is devoted to apply the above theoretical discussion of Kalman filter to the Vasicek and CIR models, and the extended Kalman filter, unscented Kalman filter and quadratic Kalman filter to quadratic model. We first place the one-factor Vasicek, CIR and quadratic models into state-space form, then simulate various term structure outcomes using known parameters, and further proceed to estimate the

model parameters. This simulation procedure is intended to show how effective these estimation techniques work in terms of identifying parameters. In particular, a sequence of steps are followed in the simulation:

1. The first step is to simulate the underlying state variable paths. The state variables are simulated from the discretized solution to their attendant stochastic differential equations. In the estimation, we start with an arbitrary set of parameters for each term structure sample path and construct monthly observations over a 10-year time horizon with maturities ranged from 1-month to 10-years. Actually, to achieve a better approximation of the underlying stochastic processes that govern the state variables, we simulate the state variables daily over the 10-year period, but we only use the monthly observations in the data. Finally, to be consistent to our discussion in above sections, we assume that the zero-coupon rates are observed with a normally distributed independent error term.

2. We proceed to employ the multi-start optimization solver of MATLAB to find the optimal parameter set. The actual optimization problem in our models does not have constraints on the parameter values, but we impose the lower and upper bound to make sure that the optimal parameter set found is reasonable; for instance, we constraint $\theta \in [0, 1]$ in Vasicek model since θ is the long term interest rate and it does not make any sense it is negative or over 1.

3. We simulate a sample path for the term structure of interest rates and apply the estimation algorithm 500 times. This may not be a sufficient number of

simulations, but the procedure is rather time consuming and the results do demonstrate the accuracy of the approach.

The following tables summarize the estimation results of application the Kalman filter to the Vasicek and CIR models and the extended Kalman filter, unscented Kalman filter and quadratic Kalman filter to the quadratic model. We compare the true values (TV) and the mean estimate (ME) over the 500 simulations and the standard deviation (SD) of the estimates are also shown in these tables.

Table 3.1: Parameter Estimation for the Vasicek and CIR Models

Model	Vasicek			CIR		
Parameters	TV	ME	SD	TV	ME	SD
θ	0.06	0.06002	8.74×10^{-4}	0.10	0.10	1.33×10^{-5}
κ	0.05	0.05	0	0.10	0.10	2.08×10^{-5}
σ	0.02	0.02001	7.60×10^{-4}	0.025	0.02504	1.97×10^{-4}

Table 3.2: Parameter Estimation for the Quadratic Model

Method		EKF		UKF		QKF	
Parameters	TV	ME	SD	ME	SD	ME	SD
θ	0.22	0.217	0.031	0.220	3.64×10^{-3}	0.219	9.07×10^{-3}
κ	0.2	0.196	0.028	0.200	3.45×10^{-3}	0.200	7.03×10^{-3}
σ	0.1	0.098	0.022	0.100	3.79×10^{-3}	0.101	9.19×10^{-3}

Table 3.1 summarizes the estimation results of the Vasicek and CIR models by Kalman filter. In both instances, 500 estimations were performed using 1-month, 3-month, 6-month, 2-year, 5-year and 10-year zero rates. Table 3.1 shows that Kalman

filter works well for the Vasicek and CIR models. In particular, it estimates the parameter set closely to the mean-reversion, long term interest rate and volatility parameters. Moreover, the standard errors are quite small.

Table 3.2 displays the results of three Kalman filter variations - EKF, UKF, and QKF for the quadratic model. It shows that all three methods work well on estimating parameters. However, EKF's mean is less closed to the "true value" and it has a larger standard errors compared to UKF and QKF in both relative and absolute terms. Among these three estimation techniques, UKF's performance is encouraging, and the mean estimates of all parameters are extremely closed to the "true" values with small standard errors. The special estimator–QKF also works very well. The mean estimates of QKF is close to the "true value". QKF outperforms EKF in identifying all three parameters, although its performance is slightly inferior to UKF's. Overall, we may conclude that both UKF and QKF are reasonable and encouraging estimation techniques for the quadratic interest rate model.

3.4.2 Actual results

In this section, we apply the proposed estimation techniques to U.S. Treasury zero-coupon yield curve data ranging from January 1970 to December 2000. The sample consists of monthly yield observations with maturities of 1,3, 6, 9, 12, 15, 18, 21, 24, 30, 36, 48, 60, 72, 84, 96, 108, and 120 months. These data are constructed by Diebold and Li [22] based on end-of-month Center for Research Security Prices(CRSP) government bond files. Figure 3.2 displays the evolution of the actual U.S. Treasury zero-coupon yield curve and Table 3.3 gives the statistics of the data set. The zero-

coupon rates incorporated into our estimation include six observations with 1-month, 6-month, 1-year, 2-year, 5-year, and 10-year terms to maturity. We shall assume these zero-coupon rate data are observed with independent normally distributed errors.

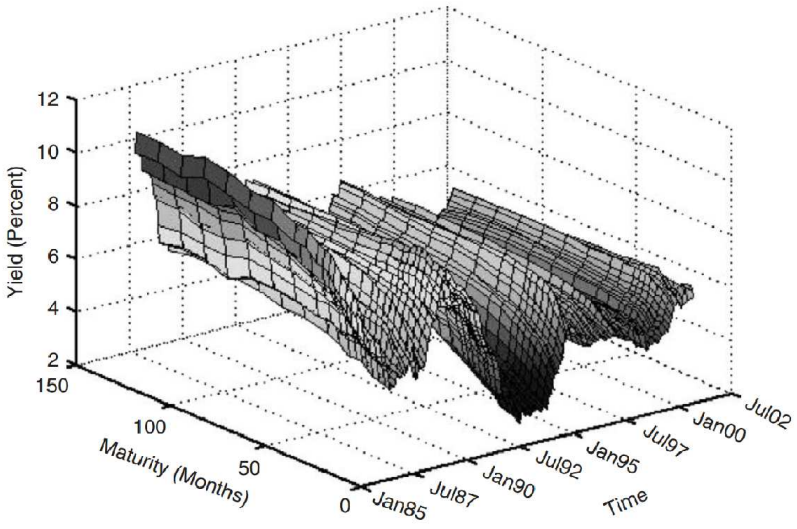

Figure 3.2: U.S. Treasury Zero-coupon Yield Rate Curve

We follow the similar steps to find the optimal estimation of the parameter set as ones in the simulation. However, instead of computing the standard deviations for the estimates, we compute the standard errors from the Fisher information matrix. In particular, if ψ is used to denote the vector of standard errors for 3 parameters, then we calculate each individual standard error as follows:

$$\psi_i = \sqrt{H_{ii}^{-1}}$$

for $i = 1, 2, 3$ where H is the Hessian matrix. Table 3.4 summarizes the estimation results for the Vasicek and CIR models. The results of identifying parameters of quadratic interest rate model estimated by EKF, UKF, and QKF are shown in Table 3.5. The results of estimation from the actual data are consistent to the results of estimation from simulated data.

Table 3.3: The Descriptive Statistics of U.S. Treasury Zero-coupon Yield Rates from January 1980 to December 1989

Maturity	Mean	Std. dev.	Minimum	Maximum
1-month	0.0856	0.7903	0.1616	0.392
6-month	0.0927	0.7985	0.1648	0.0530
1-year	0.0952	0.6934	0.1582	0.0544
2-year	0.0982	0.6036	0.1565	0.0583
5-year	0.1021	0.4826	0.1582	0.0664
10-year	0.1040	0.4105	0.1494	0.0709

Table 3.4: The Estimated Parameters for the Vasicek and CIR Models

Parameters	Vasicek		CIR	
	Estimate	Std. Error	Estimate	Std. Error
θ	0.153	0.0074	0.153	0.0087
κ	0.115	0.0110	0.010	0.0129
σ	0.039	0.0033	0.104	0.0081

Table 3.5: The Estimated Parameters for the Quadratic Model

Parameters	EKF		UKF		QKF	
	Estimate	Std. Error	Estimate	Std. Error	Estimate	Std. Error
θ	0.511	0.0073	0.510	0.0068	0.511	0.0156
κ	0.119	0.0123	0.122	0.0120	0.123	0.0080
σ	0.079	0.0062	0.076	0.0057	0.255	0.0050

CHAPTER 4

EMPIRICAL TESTS

In this chapter, we use the empirical estimation results in §3.4.2 to compare a variety of term structure models. We first compare the in-sample fit and out-sample forecasting properties of the Vasicek, CIR, and quadratic models with constant parameters, which has been extensively studies in literature. However, to our knowledge, there is no empirical result of out-sample forecasting between constant parameter term structure models and time-dependent parameter models and among the time-dependent parameter models, i.e. the extended models. We review and compare three models: the extended Vasicek model, the extended CIR model, and the quadratic model with time-dependent parameters.

4.1 Calibration of the Affine Models

In §3.2 , we have considered the affine models with constant parameters. It is known that the model with constant parameters can not capture the whole yield curve. To overcome this drawback, one needs to allow one or more parameters to be time-dependent. Here we only consider the simplest case that θ is a function of time t, i.e.,

the following model

$$dr(t) = \kappa(\theta(t) - r(t))dt + \sigma r^{\nu}(t)dW_t \tag{4.1}$$

under the risk-neutral measure. It is called the Hull-White model or the extended CIR model when $\nu = 0$ or $\nu = \frac{1}{2}$. As usual, the zero-coupon bond is given by

$$P(t,T) = P(r_t, t, T) = e^{-a(t,T)-b(t,T)r_t}, \tag{4.2}$$

where $a(t,T)$ and $b(t,T)$ are the solutions of differential equations

$$a_t(t,T) + \kappa\theta(t)(t)b(t,T) - \frac{1-2\nu}{2}\sigma^2 b(t,T)^2 = 0, \quad 0 \le t < T, \tag{4.3}$$

$$b_t(t,T) - \kappa b(t,T) - \nu\sigma^2 b(t,T)^2 + 1 = 0, \quad 0 \le t < T, \tag{4.4}$$

with the final conditions

$$a(T,T) = 0, \tag{4.5}$$

$$b(T,T) = 0. \tag{4.6}$$

Since κ and σ are constant, we can obtain by solving equation (4.4) with the final condition (4.6)

$$b(t,T) = \begin{cases} \frac{1}{\kappa}\left(1 - e^{-\kappa(T-t)}\right), & \text{for the HW model,} \\ \frac{2\left(e^{\gamma(T-t)}-1\right)}{(\gamma+\kappa)\left(e^{\gamma(T-t)}-1\right)+2\gamma}, & \text{for the ECIR model,} \end{cases}$$

where $\gamma = \sqrt{\kappa^2 + 2\sigma^2}$.

Now assume that we observe the short rate r_0 and the yield curve $y(T)$ at current time ($t = 0$) for all maturities $T \in [0,\ T^*]$, where T^* is a given positive number. By (4.2), we have

$$e^{-a(0,T)-b(0,T)r_0} = e^{-Ty(T)},$$

which leads to

$$a(0,T) = Ty(T) - b(0,T)r_0.$$

Next, we need the inverse problem to determine $\theta(t)$ and $a(t,T)$ by equation (4.3), the final condition (4.6) and the known function $a(0,T)$.

The Hull-White model: The inverse problem can solved analytically to get ([32])

$$\begin{aligned}
\theta(t) &= \frac{\partial a(0,t)}{\partial t} + \frac{1}{\kappa}\left(\frac{\partial^2 a(0,t)}{\partial t^2} + g(t)\right), \\
a(t,T) &= a(0,T) - a(0,t) - b(t,T)\frac{\partial a(0,t)}{\partial t} + \frac{1}{2}b^2(t,T)g(t), \\
g(t) &= \left(\frac{\partial b(0,t)}{\partial t}\right)^2 \int_0^t \left(\frac{\sigma}{\frac{\partial b(0,s)}{\partial s}}\right)^2 ds.
\end{aligned}$$

The ECIR model: We have from (4.3)

$$\int_t^T \theta(s)b(s,T)ds = -\frac{a(t,T)}{\kappa}, \quad 0 \le t \le T.$$

Letting $t = 0$, we obtain the following integral equation for θ ([16])

$$\int_0^T \theta(s)b(s,T)ds = -\frac{a(0,T)}{\kappa}, \quad 0 \le T \le T^*.$$

This is a Volterra integral equation of the first kind. Differentiating the above equation twice, we get the following Volterra integral equation of the second kind:

$$\theta(T) + \int_0^T \theta(s)\frac{\partial^2 b(s,T)}{\partial T^2}ds = -\frac{1}{\kappa}\frac{\partial^2 a(0,T)}{\partial T^2}, \quad 0 \le T \le T^*. \tag{4.7}$$

By Theorem 3.1 of [42], equation (4.7) has a unique solution of $\theta(t)$ if $\frac{\partial^2 b(s,T)}{\partial T^2}$ and $\frac{\partial^2 a(0,T)}{\partial T^2}$ are continuous. As suggested in [69], the block-by-block method in section 7.6 of [42] can be used to solve this integral equation accurately and efficiently.

To simplify the notation, we make the following substitutions before we outline the block-by-block method. Let

$$K(s,T) = \frac{\partial^2 b(s,T)}{\partial T^2}, \quad f(T) = -\frac{1}{\kappa}\frac{\partial^2 a(0,T)}{\partial T^2}.$$

Then equation (4.7) can be rewritten as

$$\theta(T) + \int_0^T \theta(s)K(s,T)ds = f(T), \quad 0 \le T \le T^*. \tag{4.8}$$

Let the step size be $h = T/M$ and time partition be $t_m = mh$ for $m = 0, 1, \cdots, M$ with the half points $t_{m+h/2} = t_m + h/2$ for $m = 0, 1, \cdots, M-1$. Denote θ_m the approximation of $\theta(t_m)$. Then the block-by-block method can be formulated as follows: for $m = 0, 1, \cdots, M/2-1$, compute θ_{2m+1} and θ_{2m+2} by

$$\begin{cases} a_m\theta_{2m+1} + b_m\theta_{2m+2} = p_m \\ \\ c_m\theta_{2m+1} + d_m\theta_{2m+2} = q_m \end{cases} \tag{4.9}$$

where

$$\begin{aligned}
a_m &= 1 + \frac{h}{2}K(t_{2m+1}, t_{2m+1/2}) + \frac{h}{6}K(t_{2m+1}, t_{2m+1}), \\
b_m &= -\frac{h}{12}K(t_{2m+1}, t_{2m+1/2}), \\
c_m &= \frac{4h}{3}K(t_{2m+2}, t_{2m+1}), \\
d_m &= 1 + \frac{h}{2}K(t_{2m+2}, t_{2m+2}), \\
p_m &= h(t_{2m+1}) + \frac{h}{6}K(t_{2m+1}, t_{2m})\theta_{2m} - \theta h 4K(t_{2m+1}, t_0)\theta_0 - \frac{h}{3}\sum_{i=0}^{2m} w_{m,i}K(T_{2m+1}, t_i), \\
q_m &= h(t_{2m+2}) - \frac{h}{3}\sum_{i=0}^{2m} w_{m,i}K(T_{2m+2}, t_i), \\
\{w_{m,0}, w_{m,1}, \cdots, w_{m,m-1}, w_{m,m}\} &= \{1, 4, 2, \cdots, 2, 4, 1\}.
\end{aligned}$$

The Simpson's rule is applied to solve the numerical integration in the above block-by-block method. System (4.9) has a unique solution when h is small enough. Fur-

thermore, the error estimate has been shown in [42] that

$$\max_{1\le m\le M} |\theta(t_m) - \theta_m| \le Ch^4$$

for sufficiently smooth functions $K(t,s)$ and $h(t)$ and some constant C independent of h.

Remark 4.1. Recall that the bond price (4.2) can also be given in terms of the forward rate f_t^T:

$$P(t,T) = e^{-\int_t^T f_t^s ds}.$$

Then we can determine the yield $y(T)$ by the forward rate $f(T) = f_0^T$:

$$y(T) = \frac{1}{T}\int_t^T f(s)ds.$$

4.2 Calibration of the Quadratic Model

The calibration of the quadratic model (2.2) is to determine one or more time-dependent parameters by the observed market data such as the current yield curve (or forward rate curve), forward rate volatility curve, etc. Here we only consider the simplest case: determine $\alpha(t)$ when σ and β are known constants.

Since $x(t)$ is the real independent variable, the T-maturity zero-coupon bond price is a function of $x = x(t)$ instead of $r(t)$, denoted by $P(x(t),t;T)$. As usual, we know that $P(x,t;T)$ is the solution of the fundamental partial differential equation

$$P_t + \frac{1}{2}\sigma(t)^2 P_{xx} + (\alpha(t) - \beta(t)x)\, P_x - \frac{1}{2}x^2 P = 0, \quad -\infty < x < \infty,\ 0 \le t < T$$

subject to the final condition

$$P(x,T;T) = 1.$$

It is known that the bond price $P(x,t;T)$ takes the following form:

$$P(x,t;T) = e^{-a(t,T)-b(t,T)x-\frac{1}{2}c(t,T)x^2}.$$

Here $a(t,T)$, $b(t,T)$, and $c(t,T)$ are the solution of the following final value problem of a system of ordinary differential equations:

$$a_t(t,T) - \gamma(t)b^2(t,T) + \alpha(t)b(t,T) + \gamma(t)c(t,T) = 0, \tag{4.10}$$

$$b_t(t,T) - (2\gamma(t)c(t,T) + \beta(t))\, b(t,T) + \alpha(t)c(t,T) = 0, \tag{4.11}$$

$$c_t(t,T) - 2\gamma(t)c^2(t,T) - 2\beta(t)c(t,T) + 1 = 0, \tag{4.12}$$

$$a(T,T) = b(T,T) = c(T,T) = 0, \tag{4.13}$$

for $0 \le t \le T$, where $\gamma(t) = \frac{1}{2}\sigma(t)^2$. When α, β and σ are constant, we can solve the above system to get ([35])

$$a(t,T) = c_1\tau + c_2\left(h(\tau)\left(2\beta e^{\mu\tau} + c_3\right) - c_4\right) - \frac{1}{2}\log(2\mu h(\tau)), \tag{4.14}$$

$$b(t,T) = \frac{\alpha}{\mu}\left(e^{\mu\tau} - 1\right)^2 h(\tau), \tag{4.15}$$

$$c(t,T) = \left(e^{2\mu\tau} - 1\right) h(\tau), \tag{4.16}$$

where

$$h(\tau) = \left((\mu+\beta)e^{2\mu\tau} + \mu - \beta\right)^{-1}, \quad \tau = T - t, \quad \mu = \sqrt{\sigma^2 + \beta^2},$$
$$c_1 = \frac{1}{2}\left(\frac{\alpha^2}{\mu^2} - \frac{\sigma^2}{\mu - \beta}\right), \quad c_2 = \frac{\alpha^2}{\mu^3}, \quad c_3 = \frac{\sigma^2 - \beta^2}{\mu + \beta}, \quad c_4 = \frac{2\beta + \mu}{2(\beta + \mu)}.$$

Let

$$b(0,T) + c(0,T)x_0 = f(T), \tag{4.17}$$

$$a(0,T) + b(0,T)x_0 + \frac{1}{2}c(0,T)x_0^2 = g(T), \tag{4.18}$$

where $f(T)$ can be determined by the current term structure of sport or forward rate volatilities and $g(T)$ can be determined by the current term structure of interest rates or forward rates.

As pointed out in [35], constant σ can be determined by $\sigma(0) = f_T(0,0)/x(0)$, where $x(0)$ can be determined by the current interest rate $r(0)$ and $f(0,0)$ the current forward-rate volatility. In order to determine constant parameters β, we need to know more information from the current term structure. Suppose $f(T)$ in (4.17) and $g_{TT}(0,0)$ in (4.18) are known. It follows from (4.10)–(4.13) that

$$\beta(0) = -\frac{c_{TT}(0,0)}{2}, \quad \frac{1}{2}\sigma^2(0) = a_{TT}(0,0).$$

Thus, β can be solved by

$$\beta(0) = \frac{2g_{TT}(0) - 2f_{TT}(0)x_0 - \sigma^2(0)}{2x_0}$$

Once constants $\beta = \beta(0)$ and $\sigma = \sigma(0)$ are determined, $c(t,T)$ can be analytically solved from (4.12) and (4.13):

$$c(t,T) = \left(e^{2\mu\tau} - 1\right) h(\tau),$$

where

$$\tau = T - t, \quad \mu = \sqrt{\sigma^2 + \beta^2}, \quad h(\tau) = \left((\mu + \beta)e^{2\mu\tau} + \mu - \beta\right)^{-1}.$$

With $c(t,T)$, $b_T(0,T)$ and $b_{TT}(0,T)$ can be solved from (4.17). By using formula (8.12) in [35], $\alpha(t)$ can be recovered as follows:

$$\alpha(t) = (c_T(0,t))^{1.5}(c_T(0,t)b_{TT}(0,t) - c_{TT}(0,t)b_T(0,t))$$

By Theorem 8.1 in [35], when the term structure of the forward rate $f(t,T)$ is known at time t with maturity T, we have the following analytic formulas for $a(t,T)$ and $b(t,T)$:

$$b(t,T) = c(t,T)\sqrt{c_T(0,t)}\left(\frac{b(0,T)-b(0,t)}{c(0,T)-c(0,t)} - \frac{b_T(0,t)}{c_T(0,t)}\right), \tag{4.19}$$

$$a(t,T) = a(0,T) - a(0,t) - \tilde{a}(c(t,T), b(t,T), \frac{b_T(0,t)}{\sqrt{c_T(0,t)}}, \sigma^2 c(t,T)), \tag{4.20}$$

where

$$\begin{aligned}
b_T(t,T) &= -c_T(0,T)x(0) + \sqrt{c_T(t,T)(2f(t,T) - \sigma^2 c(t,T))}, \\
a_T(t,T) &= \frac{1}{2}\left(\frac{b_T(t,T)^2}{c_T(t,T)} + \sigma^2 c(t,T)\right), \\
\tilde{a}(x,y,z,w) &= \frac{1}{2}\ln(1+xw) + \frac{xz^2 + 2yz - wy^2}{2(1+wx)}.
\end{aligned}$$

4.3 Forward Curve Representation

In last section, we calibrate the Hull-White, extended CIR, and quadratic models to the current term structure. To determine $\theta(t)$ in the Hull-White and extended CIR models and $\alpha(t)$ in the quadratic model, we need to know the volatility σ, the speed of adjustment κ in the Hull-White and extended CIR models and β in the quadratic model. Constant parameters σ, κ or β can be determined by Kalman filter and its variations. Besides the constant parameters, we need to provide the current instantaneous interest rate and forward interest rate $f(t)$ observed from the current term structure. However, not only the instantaneous interest rate cannot be observed from the market, but also the treasury yield data described in Section 3.4 is not a smooth function but a set of discrete points at any time t. So we need to construct a smooth function of time t to model the forward curve $f(t)$ and set the instantaneous

interest rate $r(t) = f(t,t)$. The Nelson-Siegel-Svensson model is popular among practitioners for modeling the yield curve. The model is first proposed by Nelson and Siegel ([59]) and extended by Svensson ([64]). The Nelson-Siegel-Svensson model is widely used by the central banks ([30]) and market practitioners ([29]) and extensively studied in academia ([28, 21, 22]). We shall look into the original formulation of Nelson and Siegel and the extension of Svensson.

Let $f_t(\tau)$ be the forward rate at time t for maturity τ. The Nelson and Siegel model for the forward rate curve is given by:

$$f_t(\tau) = \beta_1 + \beta_2 \exp(-\tau/\lambda) + \beta_3 \lambda \exp(-\tau/\lambda). \tag{4.21}$$

The Nelson-Siegel forward rate curve can be viewed as a constant plus a polynomial times an exponential decay term. The relationship between the yield curve $y(\tau)$ and the forward rate at time t satisfies

$$y(\tau) = \frac{1}{\tau} \int_0^\tau f_t(s) ds.$$

Then the corresponding yield curve is

$$y(\tau) = \beta_1 + \beta_2 \left[\frac{1 - \exp(-\tau/\lambda)}{\tau/\lambda} \right] + \beta_3 \left[\frac{1 - \exp(-\tau/\lambda)}{\tau/\lambda} - \exp(-\tau/\lambda) \right]$$

In the above equation, the yield y for a particular maturity is the sum of several components. Now let us interpret these components. The parameter λ controls the exponential decay rate; small values of λ lead to fast decay and can better fit the curve at short maturities, while large values of λ slow the decay and can better fit the curve at long maturities. β_1 is independent of time to maturity, and it is often interpreted as the long-term yield level, i.e. $\lim_{\tau \to \infty} f(\tau) = \beta_1$. Constant β_2 is weighted by an

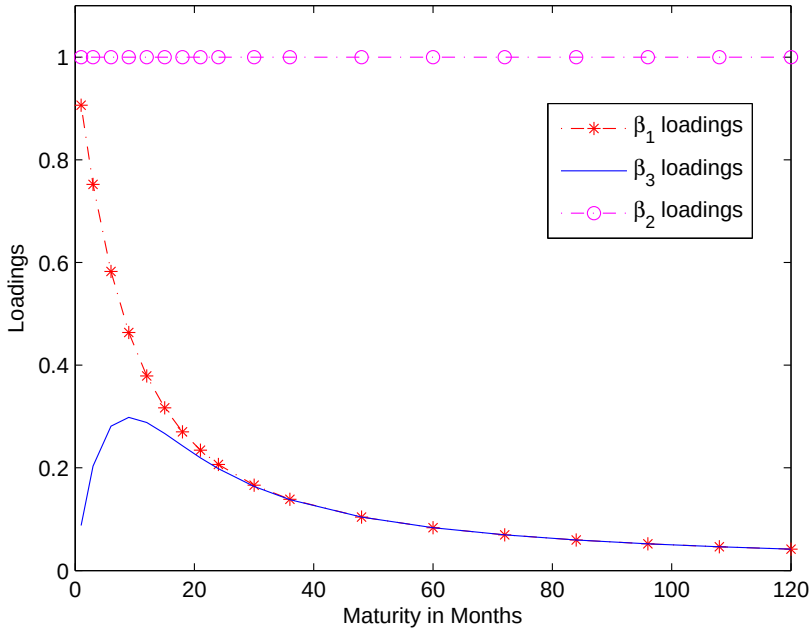

Figure 4.1: Factor loadings in Nelson-Siegel model with $\lambda = 5$.

exponential decay function of time to maturity with unity for $\tau = 0$ and decaying to zero as τ growing. Hence it may be viewed as a short-term factor. The decay function loading on β_3 is zero at $\tau = 0$, increases and then decreases back to zero as τ grows. It adds a hump to the curve and may be viewed as a medium-term factor. Figure 4.1 displays the three factor loadings with $\lambda = 5$. To guarantee the yield obtained from Nelson-Siegel model is nonnegative, we add the following constraints on the parameters

$$\beta_1 > 0, \quad \beta_2 + \beta_3 > 0, \quad \lambda > 0.$$

The Nelson-Siegel-Svensson model adds a second hump term to the Nelson-Siegel model. The forward rate at time t for maturity is given by

$$f_t(\tau) = \beta_1 + \beta_2 \exp(-\tau/\lambda_1) + \beta_3 \lambda_1 \exp(-\tau/\lambda_1) + \beta_4 \lambda_2 \exp(-\tau/\lambda_2) \tag{4.22}$$

Accordingly, the Nelson-Siegel-Svensson yield curve is formulated as

$$\begin{aligned} y(\tau) = \beta_1 &+ \beta_2 \left[\frac{1 - \exp(-\tau/\lambda_1)}{\tau/\lambda_1} \right] \\ &+ \beta_3 \left[\frac{1 - \exp(-\tau/\lambda_1)}{\tau/\lambda_1} - \exp(-\tau/\lambda_1) \right] \\ &+ \beta_4 \left[\frac{1 - \exp(-\tau/\lambda_2)}{\tau/\lambda_2} - \exp(-\tau/\lambda_2) \right]. \end{aligned} \tag{4.23}$$

The two parameters λ_2 and β_4 are analogous to λ_1 and β_3 determining the decay rate and the magnitude and direction of the second hump respectively. Hence, we need to estimate six parameters: β_1, β_2, β_3, β_4, λ_1, and λ_2 subject to constraints

$$\beta_1 > 0, \quad \beta_1 + \beta_2 > 0, \quad \lambda_1 > 0 \quad \lambda_2 > 0.$$

Generally, the parameters of the models can be estimated by minimizing the difference between the model rates y and observed rates y^O. An optimization problem can be stated as

$$\min_{\beta,\lambda} \sum \left(y - y^O\right)^2$$

subject to the constraints given above.

Next, we shall estimate the parameters in Nelson-Siegel and Nelson-Siegel-Svensson models by the treasury yield data described in Section 3.4.2. To solve the constraint optimization problem, we again apply multi-start optimization solver of MATLAB to find the optimal parameters in these two models.

We first estimate the parameters in Nelson-Siegel and Nelson-Siegel-Svensson models on the observed yield at four dates: 3/31/1989, 7/31/1989/, 5/30/1997, and 8/31/1998 and then plot fitted yield curves together with with actual yields in Figure 4.2. Apparently, both the Nelson-Siegel model and Nelson-Siegel-Svensson model

are capable of replicating a variety of yield curve shapes: upward sloping, downward sloping, humped and inverted humped. However, both models have difficulties at dispersed yield data. The root mean square error of in sample fit of NS and NSS models is presented in Table 4.1. Overall, the error indicates a good fit and Nelson-Siegel-Svesson model fits better than Nelson- Siegel model does. Thus we will use Nelson-Siegel-Svesson model to fit current yield and forward curves.

Table 4.1: RMSE of In Sample Fit NS vs NSS

Date	3/31/1989	7/31/1989	5/30/1997	8/31/1998
NS	0.0568	0.0517	0.0437	0.0688
NSS	0.0256	0.0214	0.0157	0.0217

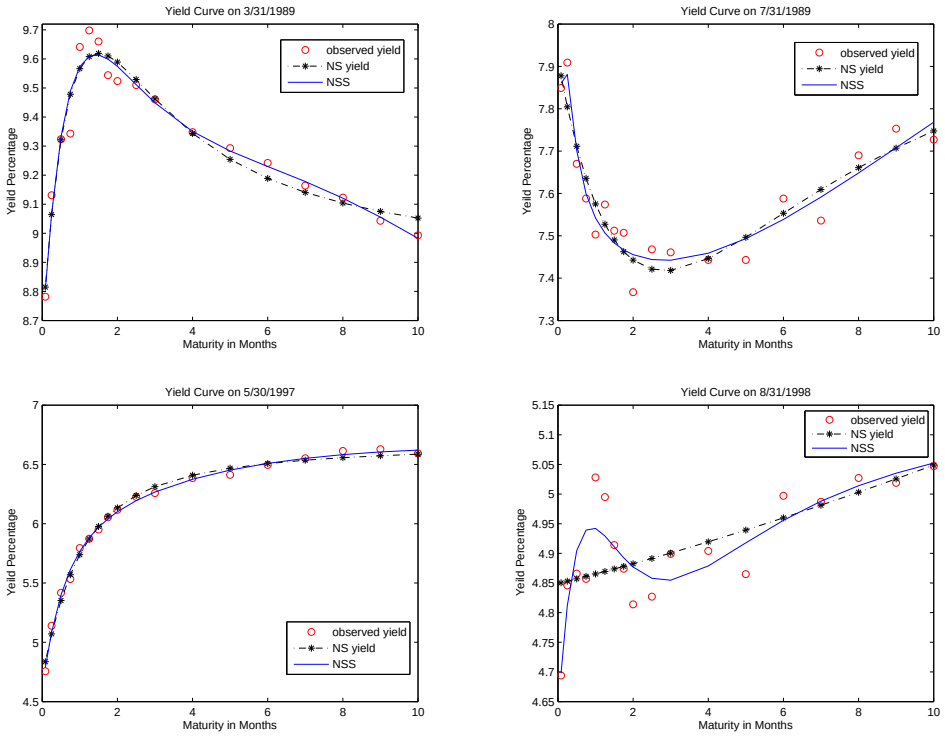

Figure 4.2: NS and NNS fitted yield curves

4.4 Calibration to the Current Term Structure

In this section, we shall calibrate the models in previous sections to the current term structure. We assume the current date is 31 January 1990 with the current yield data given in Table 4.2 and the past ten years' (January 1980 to December 1989) yield data are available. The parameter estimation results in Tables 3.4 and 3.5 are used to compare the in-sample fit of the Vasicek, CIR, and quadratic models with constant parameters. Then we also use these estimated parameters together with the current yield data to calibrate the Hull-white, extended CIR, and quadratic models with time-varying parameters. Furthermore, the out-of-sample forecasting performance of all these models will be shown and compared in Tables 4.5 and 4.6.

Table 4.2: Observed yields on 31 January 1990

M	1	3	6	9	12	15	18	21	24
r	7.648	37.922	7.964	7.996	8.081	8.145	8.202	8.169	8.103
M	30	36	48	60	72	84	96	108	120
r	8.211	8.172	8.220	8.250	8.303	8.218	8.308	8.338	8.279

M: Maturities in months, r: spot rates in percentage.

Table 4.3: NSS estimation results on 31 January 1990

β_1	β_2	β_3	β_4	λ_1	λ_2
0.0775	0.0039	0.0141	−0.0162	17.5766	0.0438

Figure 4.3 shows the models' in-sample fit performances of recovering the observed yield rates of four different dates. The left column figures compare the Vasicek model and CIR model and the right column figures compare the quadratic model estimated by extended Kalman filter, unscented Kalman filter, and quadratic Kalman filter. Table 4.4 shows the in-sample fit root mean square errors of Vasicek model, CIR model, and quadratic model from January 1980 to December 1989 with different maturities. Table 4.4 demonstrates that all models in general fit the data well. Figure 4.3 shows that the quadratic model estimated by quadratic Kalman filter captures the yield rate's movements better than the rest models in randomly selected dates. Table 4.5 and 4.6 present and compare the out-of-sample within one year forecasting root mean square errors of models with time-independent parameters and time-dependent parameters respectively. Figure 4.4 shows the models' in-sample fit performances of forecasting the yield rates of three randomly selected dates in the same manner as Figure 4.3. Figures 4.6 – 4.7 compare out-of-sample forecasting performance of the time-dependent models to the corresponding time-independent models. Both the time-dependent models and time-independent models generates very similar root mean square errors and the time-dependent models do not have an obvious advantage over the time-independent models in our testing period.

Table 4.4: In-sample fit RMSE: January 1980 – December 1989

Maturity	Vasicek	CIR	Quadratic
1-month	0.005	0.005	0.005
3-month	0.009	0.010	0.009
6-month	0.011	0.014	0.011
9-month	0.012	0.014	0.013
1-year	0.013	0.014	0.013
1.25-year	0.014	0.014	0.015
1.5-year	0.015	0.014	0.015
1.75-year	0.015	0.015	0.015
2-year	0.015	0.014	0.015
2.5-year	0.015	0.013	0.015
3-year	0.015	0.013	0.015
4-year	0.015	0.012	0.015
5-year	0.015	0.012	0.014
6-year	0.015	0.011	0.015
7-year	0.014	0.011	0.014
8-year	0.014	0.011	0.014
9-year	0.014	0.011	0.014
10-year	0.014	0.010	0.014

Table 4.5: Out-of-sample forecasting RMSE: January 1990 – December 1991–I

Maturity	Vasicek	CIR	Quadratic
1-month	0.003	0.007	0.006
3-month	0.003	0.010	0.010
6-month	0.003	0.011	0.012
9-month	0.002	0.011	0.013
1-year	0.002	0.011	0.014
1.25-year	0.002	0.012	0.015
1.5-year	0.002	0.011	0.016
1.75-year	0.002	0.011	0.016
2-year	0.001	0.010	0.015
2.5-year	0.001	0.009	0.016
3-year	0.001	0.008	0.016
4-year	0.002	0.007	0.015
5-year	0.002	0.005	0.015
6-year	0.003	0.004	0.015
7-year	0.004	0.003	0.015
8-year	0.004	0.002	0.014
9-year	0.005	0.002	0.014
10-year	0.006	0.003	0.014

Table 4.6: Out-of-sample forecasting RMSE: January 1990 – December 1991–II

Maturity	Hull-White	Extended CIR	Quadratic
1-month	0.004	0.010	0.003
3-month	0.005	0.009	0.003
6-month	0.005	0.006	0.004
9-month	0.005	0.007	0.004
1-year	0.007	0.006	0.003
1.25-year	0.006	0.008	0.005
1.5-year	0.010	0.007	0.005
1.75-year	0.009	0.006	0.005
2-year	0.008	0.006	0.004
2.5-year	0.005	0.007	0.005
3-year	0.005	0.006	0.006
4-year	0.005	0.005	0.006
5-year	0.006	0.007	0.007
6-year	0.006	0.008	0.007
7-year	0.007	0.009	0.004
8-year	0.007	0.010	0.005
9-year	0.008	0.011	0.006
10-year	0.010	0.009	0.006

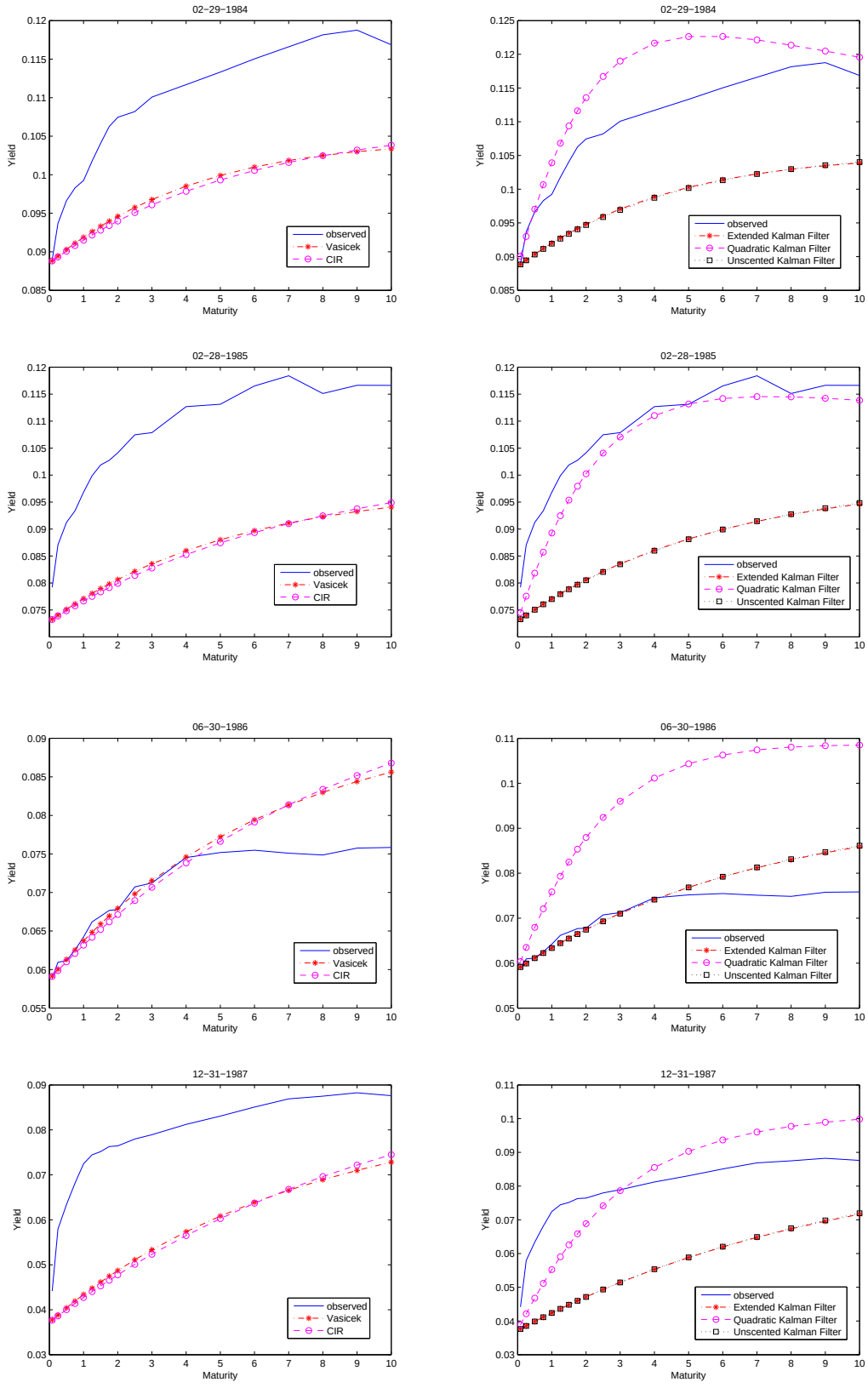

Figure 4.3: In-sample fit comparison: Vasicek, CIR, and Quadratic

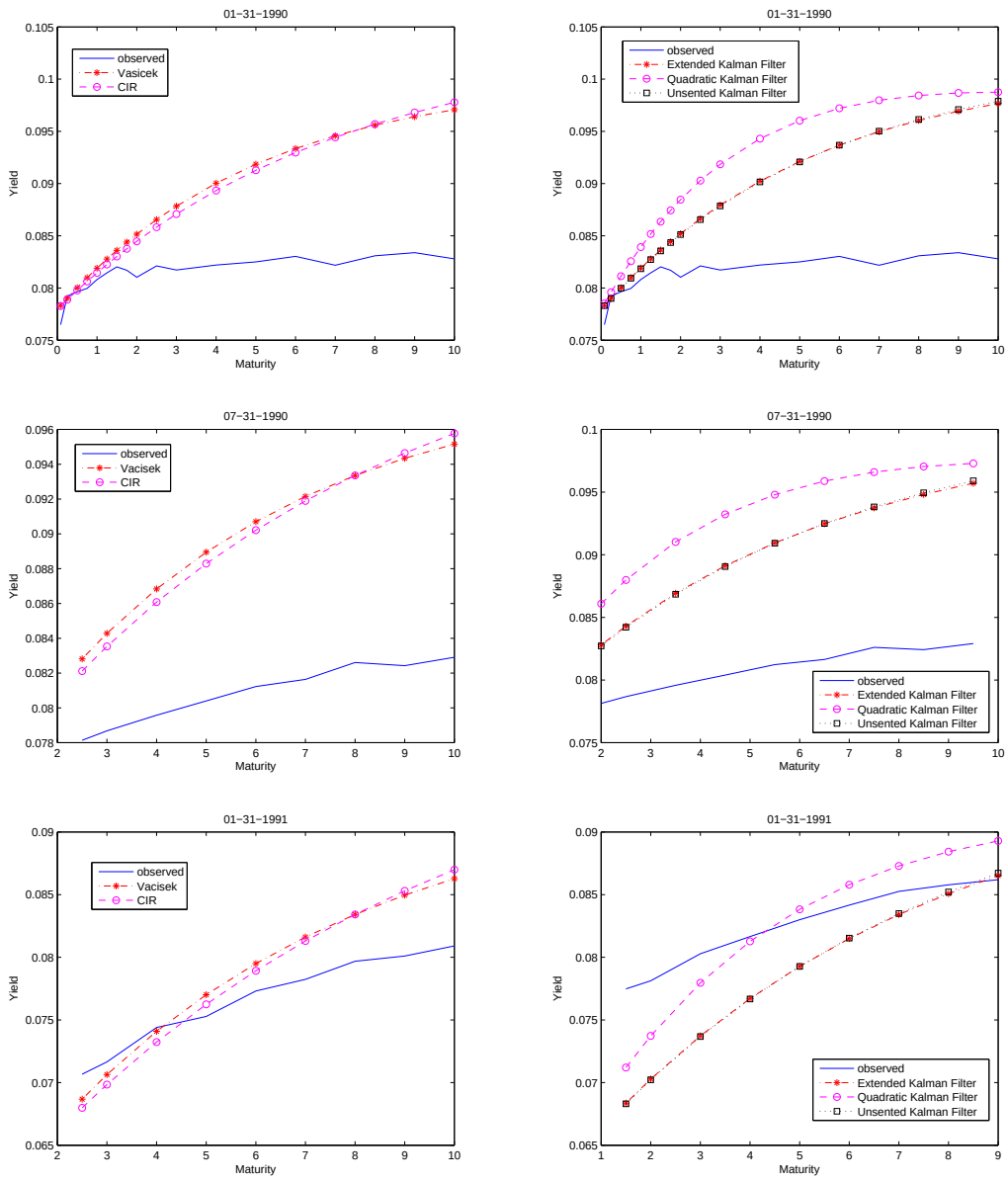

Figure 4.4: Out-of-sample fit comparison: Vasicek, CIR, and Quadratic

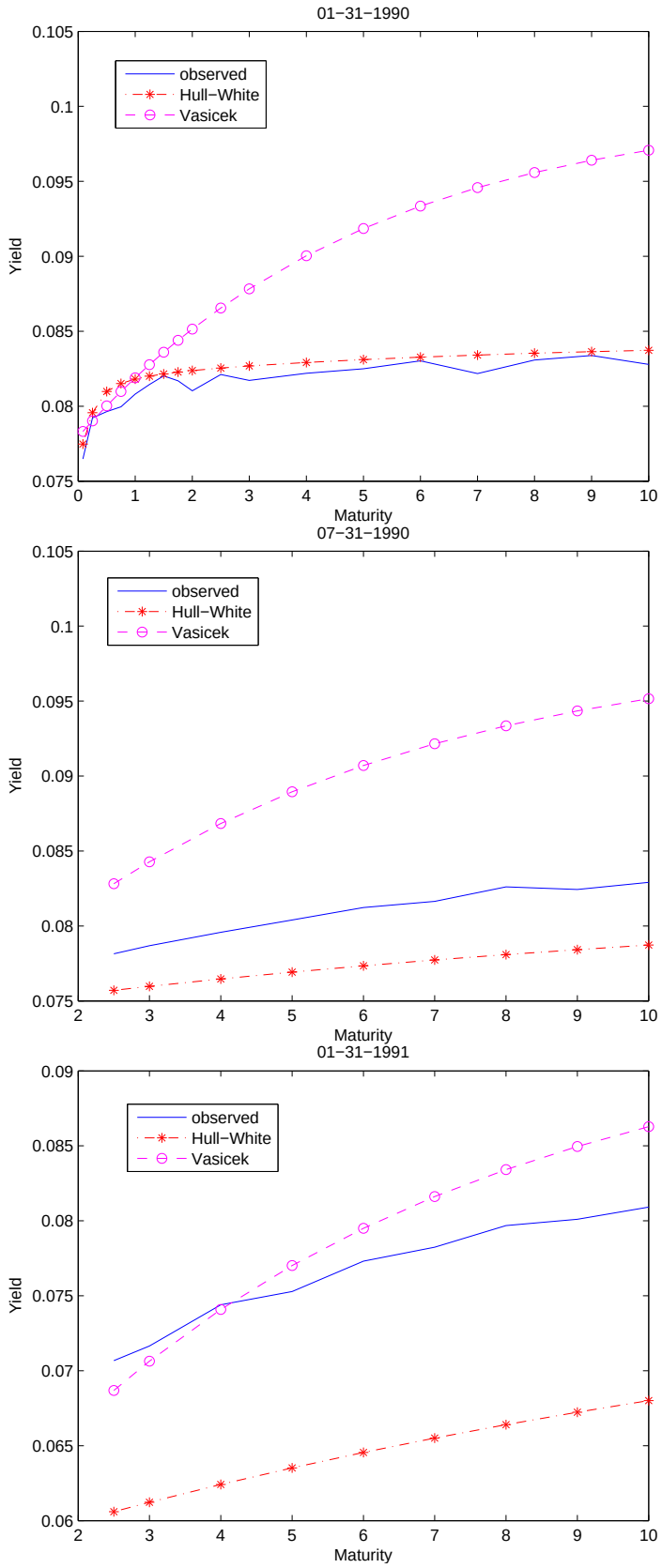

Figure 4.5: Out-of-sample fit comparison: Vasicek and Hull-White

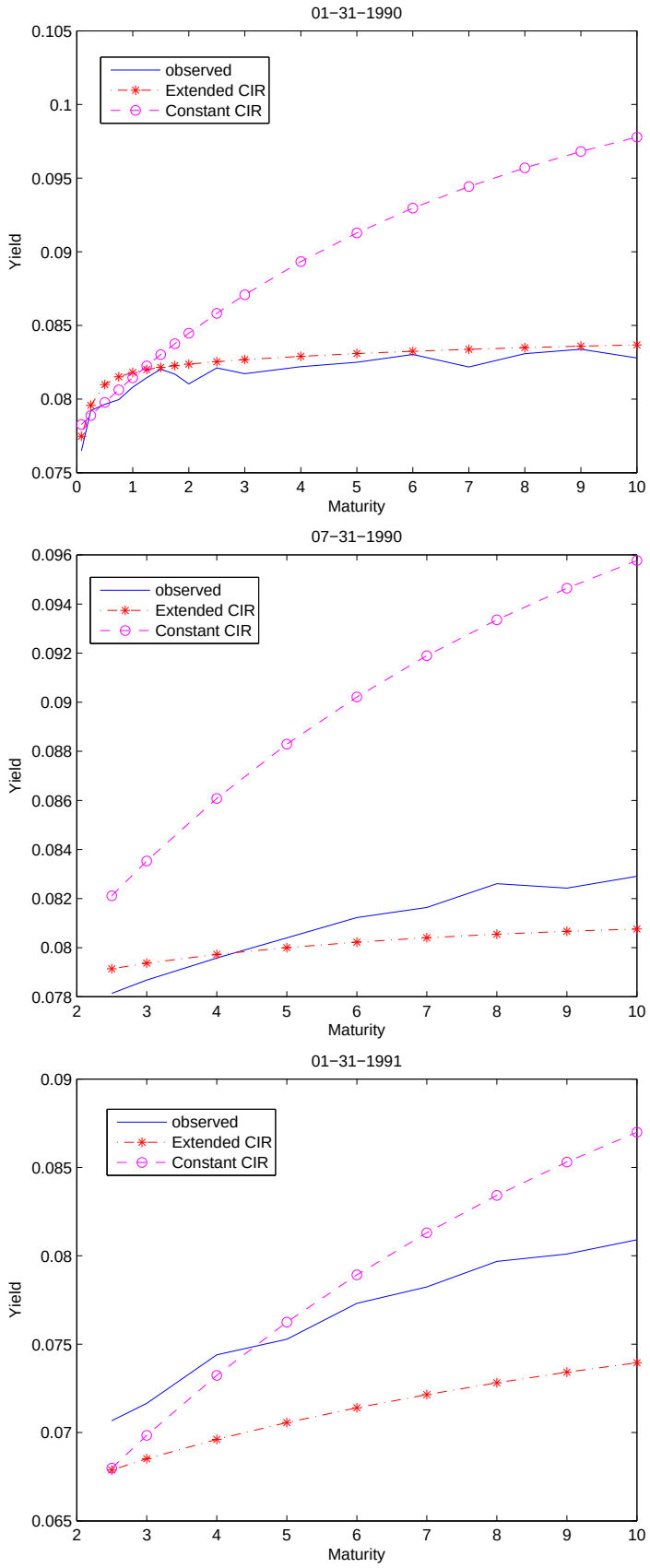

Figure 4.6: Out-of-sample fit comparison: CIR and Extended CIR

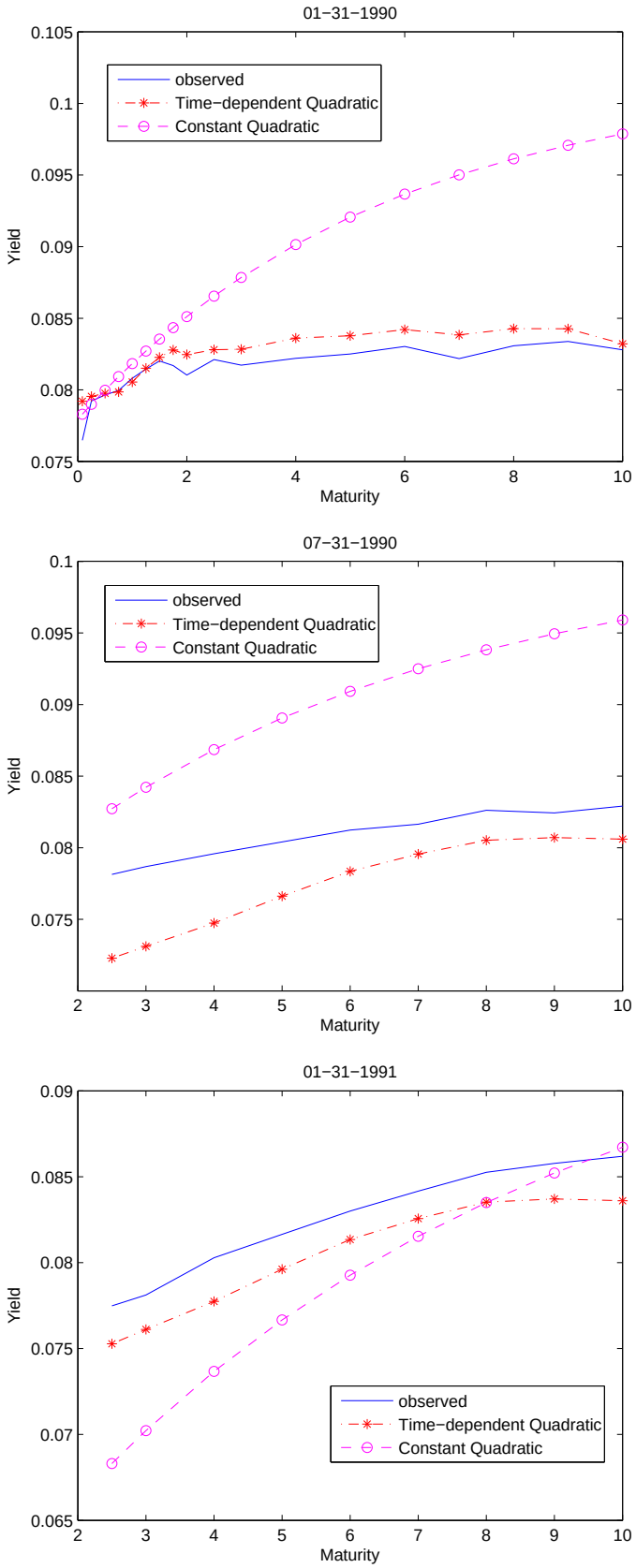

Figure 4.7: Out-of-sample fit comparison: Quadratic and Time-dependent Quadratic

CHAPTER 5

A CLASS OF ONE-FACTOR MODELS

5.1 Introduction

Let $X(t)$ follow the following Ornstein-Uhlenbeck process:

$$dX(t) = (\phi(t) - \psi(t)X(t))dt + \sigma(t)dW(t),$$

where $\phi(t)$, $\psi(t)$ and $\sigma(t)$ are some known functions of t and $W(t)$ is a standard Brownian motion under the risk-neutral measure. Then a class of one-factor models of the short interest rate process is given by

$$r(t) = \zeta(X(t))$$

where $\zeta(x)$ is an invertible function $\zeta(x)$ on $(-\infty, +\infty)$. Especially, for $\zeta(x) = x$ and $\zeta(x) = e^x$, we have the Hull-White model ([32]) and the Black-Karasinski model ([11]), respectively. It should be pointed out that $\zeta(x)$ can chosen to be a bounded function from $\mathbb{R} = (-\infty, +\infty)$ to $(0, 1)$, for example,

$$\zeta(x) = \frac{1}{2}\left(1 + \frac{2}{\pi}\arctan(x)\right) \quad \text{or} \quad \zeta(x) = \frac{e^x}{1+e^x}.$$

In this way, the interest rates will not take unrealistic values greater than 1.

Let us consider a zero-coupon bond with face value \$1 and maturity date T. By the Fundamental Theorem of Asset Pricing, the bond price $P(x, t; T)$ is the solution

of the following final value problem:

$$P_t + \frac{1}{2}\sigma(t)^2 P_{xx} + (\phi(t) - \psi(t)x)P_x - \zeta(x)P = 0, \quad x \in \mathbb{R}, \quad 0 \leq t \leq T, \tag{5.1}$$

$$P(x, T; T) = 1, \quad x \in \mathbb{R}, . \tag{5.2}$$

5.2 Homotopy Perturbation Method

Homotopy perturbation method (HPM) is a popular technique to find an approximated series solutions of nonlinear problems. The homotopy perturbation method was initially proposed by He [46, 47, 48]. The essential idea of this method is to introduce a homotopy parameter p. When $p = 0$, we have a simple equation which admits an analytic solution. As p grows to 1, the homotopy goes though a sequence of deformations, the solution for each of which is close to the one at the previous stage of deformation. When $p = 1$, the homotopy takes the original form of the equation and the final stage of deformation gives the desired solution. One of the the most attractive features of homotopy perturbation method is that a few perturbation terms are sufficient to obtain a reasonably accurate solution.

To illustrate the basic idea of He's homotopy perturbation method, we consider the following nonlinear differential equation

$$A(u) - f(r) = 0, \quad r \in \Sigma \tag{5.3}$$

with boundary conditions

$$B\left(u, \frac{du}{dn}\right) = 0, \quad r \in \Gamma, \tag{5.4}$$

where A is a general differential operator, B is a boundary operator, u is a known analytic function, and Γ is the boundary of the domain Σ. The operator A can be

divided into linear part L and nonlinear part N. Therefore, equation 5.3 can be rewritten as follows

$$L(u) + N(u) - f(r) = 0. \tag{5.5}$$

Consider the following homotopy $w(r,p) : \Sigma \times [0,1] \to R$:

$$H(w,p) = (1-p)[L(w) - L(u_0)] + p[A(w) - f(r)] = 0, \tag{5.6}$$

where $r \in \Gamma$ and $p \in [0,1]$ is an homotopy parameter, u_0 is an initial approximation of 5.3, which satisfies the boundary conditions. It is apparent that equation (5.6) becomes the original nonlinear equation (5.3) for $p = 1$, while it is the following linear equation for $p = 0$:

$$H(w,0) = L(w) - L(u_0) = 0. \tag{5.7}$$

The changing process of p from 0 to 1 deforms $H(w,p)$ from $L(w) - L(u_0)$ to $A(w) - f(r)$. In topology, $L(w) - L(u_0)$ and $A(w) - f(r)$ are called homotopic. If the embedding parameter p is considered as a "small parameter", applying the classical perturbation technique, we can assume the solution of equation (5.6) can be given by the power series of p:

$$w = w_0 + pw_1 + p^2 w_2 + \cdots . \tag{5.8}$$

Then the solution u of the original nonlinear equation can be obtained by letting $p \to 1$:

$$u = \lim_{p \to 1} w = w_0 + w_1 + w_2 + \cdots$$

The convergence of the above series has been proved by He [49].

5.3 A HPM for bond prices

In this section, we apply a variable transformation to PDE (5.1). Let $\bar{X}(t)$ be the expected value of $X(t)$, then $\bar{X}(t) := \mathbf{E}\{X(t)|X(0) = 0\}$ and $\bar{X}$ satisfies the deterministic process

$$\frac{d\bar{X}}{dt} = \phi(t) - \psi(t)\bar{X},$$
$$\bar{X}(0) = 0.$$

The above ordinary differential equation can be easily solve and its solution is

$$\bar{X}(t) = \lambda(t) \int_0^t \frac{\phi(\tau)}{\lambda(\tau)} d\tau$$

where $\lambda(t) = e^{-\int_0^t \psi(\tau)d\tau}$.

Let us introduce a new state variable Y such that $X(t) = \bar{X}(t) + \lambda(t)Y(t)$. The interest rate r now becomes a function of y, i.e. $r = f(\bar{X}(t) + \lambda(t)y)$. Following Itó's lemma, it is easy to justify that Y is a martingale that follows the process

$$dY = \frac{\sigma(t)}{\lambda(t)} dW,$$
$$Y(0) = 0.$$

The zero-coupon bond pricing problem becomes

$$P(y, t, T) = \mathbb{E}\left\{ e^{-\int_t^T r(\tau, Y(\tau))d\tau} \Big| Y(t) = y \right\}.$$

By Feynman-Kac Theorem, zero-coupon bond price $P(y, t, T)$ can be solve from the the following partial differential equation:

$$P_t + \frac{1}{2}\left(\frac{\sigma(t)}{\lambda(t)}\right)^2 P_{yy} - rP = 0,$$
$$P(y, T, T) = 1, \forall y \in \mathbb{R}.$$

Since the bond price P is always positive, we can express $P(y,t,T) = e^{-h(y,t,T)}$ for some function $h(y,t,T)$. It is not difficult to see that h solves the following nonlinear problem:

$$h_t + \frac{1}{2}\left(\frac{\sigma}{\lambda}\right)^2 (h_{yy} - h_y^2) + r = 0, \quad x \in \mathbb{R}, \quad 0 \le t < T, \tag{5.9}$$

$$h(y,T,T) = 0, \quad \forall y \in \mathbb{R}. \tag{5.10}$$

The homotopy of the above problem with the embedding parameter q is as follows:

$$\frac{\partial h}{\partial t} - \frac{\partial h_0}{\partial t} = q\left(\frac{1}{2}\left(\frac{\sigma}{\lambda}\right)^2 (h_y^2 - h_{yy}) - r - \frac{\partial h_0}{\partial t}\right). \tag{5.11}$$

Assume that the solution of equation (5.9) has a power series expansion of q:

$$h = h_0 + qh_1 + q^2 h_2 + \cdots \tag{5.12}$$

Substituting (5.12) into equation (5.11) and combining terms of the same power of q gives:

$$\begin{aligned}
&q^0 : \frac{\partial h_0}{\partial t} - \frac{\partial h_0}{\partial t} = 0,\\
&q^1 : \frac{\partial h_1}{\partial t} = \frac{1}{2}\left(\frac{\sigma}{\lambda}\right)^2 \left((\frac{\partial h_0}{\partial y})^2 - \frac{\partial^2 h_0}{\partial y^2}\right) - r - \frac{\partial h_0}{\partial t},\\
&q^2 : \frac{\partial h_2}{\partial t} = \frac{1}{2}\left(\frac{\sigma}{\lambda}\right)^2 \left((\frac{\partial h_1}{\partial y})^2 - \frac{\partial^2 h_1}{\partial y^2}\right),\\
&q^3 : \frac{\partial h_3}{\partial s} = \frac{1}{2}\left(\frac{\sigma}{\lambda}\right)^2 \left((\frac{\partial h_2}{\partial y})^2 - \frac{\partial^2 h_2}{\partial y^2}\right),\\
&\vdots
\end{aligned}$$

All the linear equations above can be solved, and we get all the solutions.

$$\begin{aligned}
&h_0 = 0,\\
&h_1 = R(y,t,T),\\
&h_2 = \frac{1}{2}\int_t^T \nu(\tau)\left(R_{yy}(y,\tau,T) - R_y(y,\tau,T)^2\right) d\tau,
\end{aligned}$$

$$h_3 = \frac{1}{4}\int_t^T \left[\int_\tau^T \nu(\tau)\xi(y,u,T)du - \left(\int_\tau^T \nu(\tau)\eta(y,u,T)du\right)^2\right] d\tau$$
$$\vdots$$

where

$$R(y,t,T) = \int_t^T r(y,\tau)d\tau,$$
$$\nu(\tau) = \left(\frac{\sigma(\tau)}{\lambda(\tau)}\right)^2,$$
$$\xi(y,u,T) = R_{yyyy}(y,u,T) - 2R_{yy}(y,u,T)^2 - 2R_y(y,u,T)R_{yyy}(y,u,T),$$
$$\eta(y,u,T) = R_{yyy}(y,u,T) - 2R_y(y,u,T)R_{yy}(y,u,T).$$

5.4 Examples

Case I. The Hull-White model. In this case, we have

$$\lambda(t) = e^{-\psi t},$$
$$\bar{X}(t) = e^{-\psi t}\int_0^t e^{\psi\tau}\phi(\tau)d\tau,$$
$$r(y,t) = e^{-\psi t}(y + \int_0^t e^{\psi\tau}\phi(\tau)d\tau),$$
$$R(y,t,T) = \int_t^T e^{-\psi\tau}\left(y + \int_0^\tau e^{\psi u}\phi(u)du\right)d\tau,$$
$$R_y(y,t,T) = -(e^{-\psi T} - e^{-\psi t}),$$
$$\frac{\partial^i R(y,t,T)}{\partial y^i} = 0, \quad \text{for} \quad i \geq 2.$$

Then we can obtain by simple calculation

$$h_0 = 0,$$
$$h_1 = \frac{1}{\psi}(e^{-\psi T} - e^{-\psi t})y + \int_t^T \left(e^{-\psi\tau}\int_0^\tau e^{\psi u}\phi(u)du\right)d\tau,$$

$$h_2 = \frac{\sigma^2}{2}\int_t^T \left(e^{-\psi(T-s)} - 1\right)^2 ds = \frac{1}{2\psi}(1 - e^{-2\psi\tau}) - \frac{2}{\psi}(1 - e^{-\psi\tau}) + \tau,$$

$$h_i = 0, \quad \text{for} \quad i \geq 2$$

where $\tau = T - t$. Therefore,

$$h = h_0 + h_1 + h_2 = \frac{1}{\psi}(e^{-\psi T} - e^{-\psi t})y$$
$$+ \int_t^T \left(e^{-\psi\tau}\int_0^\tau e^{\psi u}\phi(u)du\right)d\tau + \frac{1}{2\psi}(1 - e^{-2\psi\tau}) - \frac{2}{\psi}(1 - e^{-\psi\tau}) + \tau,$$

which will give us the exact solution of Hull-White model if y is replaced in terms of x.

Case II. The Black-Karasinski model (BKM). In this case, we have

$$\lambda(t) = e^{-\psi t},$$
$$\bar{X}(t) = e^{-\psi t}\int_0^t e^{\psi\tau}\phi(\tau)d\tau,$$
$$r(y,t) = e^{\bar{x}(t)+\lambda(t)y},$$
$$R(y,t,T) = \int_t^T e^{\bar{x}(\tau)+\lambda(\tau)y}d\tau,$$
$$\frac{\partial^i R(y,t,T)}{\partial y^i} = \int_t^T \lambda(\tau)^i e^{\bar{x}(\tau)+\lambda(\tau)y}d\tau, \quad \text{for} \quad i \geq 1.$$

Then we can obtain

$$h_0 = 0,$$
$$h_1 = \int_t^T r(y,\tau)d\tau,$$
$$h_2 = \frac{\sigma^2}{2}\int_t^T e^{2\psi\tau}\left[\int_\tau^T e^{-2\psi u}r(y,u)du - \left(\int_\tau^T e^{-\psi u}r(y,u)du\right)^2\right]d\tau,$$
$$h_3 = \frac{\sigma^4}{4}\int_t^T e^{2\psi\tau}\left(h_3^a(y,\tau,T) + h_3^b(y,\tau,T)\right)d\tau$$
$$\vdots$$

where

$$h_3^a(y,\tau,T) = \int_\tau^T e^{2\psi u}\left(w_4(u,T,y) - 2\left(w_2(u,T,y)\right)^2 - 2w_1(u,T,y)w_3(u,T,y)\right)du,$$
$$h_3^b(y,\tau,T) = \left(\int_\tau^T e^{2\psi u}\left(w_3(u,T,y) - 2w_1(u,T,y)w_2(u,T,y)\right)du\right)^2,$$
$$w_j(u,T,y) = \int_u^T e^{-j\psi s} r(y,s)ds, \quad j = 1,2,3,4.$$

When ψ is also a constant, we have the following results by some tedious calculations:

$$\lambda(t) = e^{-\psi t},$$
$$\bar{X}(t) = \frac{\phi}{\psi}(1 - e^{-\psi t}),$$
$$r(y,t) = e^{e^{-\psi t}y + \frac{\phi(1-e^{\psi t})}{\psi}},$$
$$h_1(y,t,T) = \int_t^T e^{e^{-\psi\tau}y + \frac{\phi(1-e^{\psi\tau})}{\psi}}d\tau,$$
$$h_2(y,t,T) = \frac{\sigma^2}{2}\int_t^T e^{2\psi\tau} I_2(y,\tau,T)d\tau,$$
$$h_3(y,t,T) = \frac{\sigma^4}{4}\int_t^T e^{2\psi\tau}\left(\int_\tau^T e^{2\psi u} I_3^a(y,u,T)du + \left(\int_\tau^T e^{2\psi u} I_3^b(y,u,T)du\right)^2\right)d\tau,$$

where

$$I_2(y,\tau,T) = -\frac{\left(e^{\frac{e^{-\psi T}(\phi(e^{\psi T}-1)+\psi y)}{\psi}} - e^{\frac{e^{-\psi\tau}(\phi(e^{\psi\tau}-1)+\psi y)}{\psi}}\right)^2}{(\phi - y\psi)^2} + \frac{e^{\phi/\psi}}{(\phi - y\psi)^2}$$
$$\left[e^{e^{-\psi T}(y-\psi/\psi)-\psi T}\left(\phi + \psi(e^{\psi T} - y)\right)e^{e^{-\psi\tau}(y-\psi/\psi)-\psi\tau}\left(\phi + \psi(e^{\psi\tau} - y)\right)\right],$$
$$I_3^a(y,u,T) = \frac{e^{\frac{e^{-\psi T}(\phi(e^{\psi T}-1)+\psi y)}{\psi}} - e^{\frac{e^{-\psi u}(\phi(e^{\psi u}-1)+\psi y)}{\psi}}}{\phi - y\psi}$$
$$-\frac{2e^{\phi/\psi}(1 + (\phi - \psi y))}{(\phi - y\psi)^4}\left(e^{\frac{e^{-\psi T}(\phi(e^{\psi T}-1)+\psi y)}{\psi}} - e^{\frac{e^{-\psi u}(\phi(e^{\psi u}-1)+\psi y)}{\psi}}\right)$$
$$\left[e^{e^{-\psi T}(y-\phi/\psi)-2\psi T}\left(\phi^2 + 2\phi\psi(e^{\psi T} - y)\right) + \psi^2\left(2e^{2\psi T} - 2e^{\psi T}y + y^2\right)\right.$$
$$\left.-e^{e^{-\psi u}(y-\phi/\psi)-2\psi u}\left(\phi^2 + 2\phi\psi(e^{\psi u} - y)\right) + \psi^2\left(2e^{2\psi u} - 2e^{\psi u}y + y^2\right)\right],$$

$$
\begin{aligned}
I_3^b(y,u,T) = &-\frac{2e^{\phi/\psi}\left(e^{\frac{e^{-\psi T}(\phi(e^{\psi T}-1)+\psi y)}{\psi}} - e^{\frac{e^{-\psi u}(\phi(e^{\psi u}-1)+\psi y)}{\psi}}\right)}{(\phi-\psi y)^3} \\
&\left[e^{e^{-\psi T}(y-\phi/\psi)-\psi T}\left(\phi+\psi(e^{\psi T}-y)\right) - e^{e^{-\psi u}(y-\phi/\psi)-\psi u}\left(\phi+\psi(e^{\psi u}-y)\right)\right] \\
&+\frac{e^{\phi/\psi}}{(\phi-\psi y)^3}\left[e^{e^{-\psi T}(y-\phi/\psi)-2\psi T}\left(\phi^2+2\phi\psi(e^{\psi T}-y)\right)\right. \\
&+\psi^2(2e^{2\psi T}-2e^{\psi T}y+y^2)e^{e^{-\psi u}(y-\phi/\psi)-2\psi u} \\
&\left.\left(\phi^2+2\phi\psi(e^{\psi u}-y)\right)+\psi^2(2e^{2\psi u}-2e^{\psi u}y+y^2)\right].
\end{aligned}
$$

Case III. $\zeta(x) = \frac{e^x}{1+e^x}$. In this case, we have

$$
\begin{aligned}
&\lambda(t) = e^{-\psi t}, \\
&\bar{X}(t) = e^{-\psi t}\int_0^t e^{\psi\tau}\phi(\tau)d\tau, \\
&r(y,t) = \frac{e^{\bar{X}(t)+\lambda(t)y}}{1+e^{\bar{X}(t)+\lambda(t)y}}, \\
&R(y,t,T) = \int_t^T \frac{e^{\bar{X}(\tau)+\lambda(\tau)y}}{1+e^{\bar{X}(\tau)+\lambda(\tau)y}}d\tau, \\
&\frac{\partial R(y,t,T)}{\partial y} = \int_t^T \frac{\lambda(\tau)e^{\bar{X}(\tau)+\lambda(\tau)y}}{1+e^{\bar{X}(\tau)+\lambda(\tau)y}}d\tau, \\
&\frac{\partial^2 R(y,t,T)}{\partial y^2} = \int_t^T -\frac{\lambda(\tau)^2 e^{\bar{X}(\tau)+\lambda(\tau)y}\left(e^{\bar{X}(\tau)+\lambda(\tau)y}-1\right)}{\left(1+e^{\bar{X}(\tau)+\lambda(\tau)y}\right)^3}d\tau, \\
&\frac{\partial^3 R(y,t,T)}{\partial y^3} = \int_t^T \frac{\lambda(\tau)^3 e^{\bar{X}(\tau)+\lambda(\tau)y}\left(e^{2(\bar{X}(\tau)+\lambda(\tau)y)}-4e^{\bar{X}(\tau)+\lambda(\tau)y}+1\right)}{\left(1+e^{\bar{X}(\tau)+\lambda(\tau)y}\right)^3}d\tau, \\
&\frac{\partial^4 R(y,t,T)}{\partial y^4} = \int_t^T -\frac{\lambda(\tau)^4 e^{\bar{X}(\tau)+\lambda(\tau)y}p\left(e^{\bar{X}(\tau)+\lambda(\tau)y}\right)}{\left(1+e^{\bar{X}(\tau)+\lambda(\tau)y}\right)^4}d\tau, \\
&p(z) = z^3-11z^2-11z-1.
\end{aligned}
$$

Then we can obtain

$$
\begin{aligned}
&h_0 = 0, \\
&h_1 = \int_t^T r(y,\tau)d\tau,
\end{aligned}
$$

$$h_2 = \frac{\sigma^2}{2}\int_t^T e^{2\psi\tau}\left[\int_\tau^T \frac{\lambda(u)^2 e^{\bar{X}(\tau)+\lambda(u)y}\left(1-e^{\bar{X}(u)+\lambda(u)y}\right)}{1+e^{\bar{X}(u)+\lambda(u)y}}du\left(\int_\tau^T \frac{\lambda(u)e^{\bar{X}(u)+\lambda(u)y}}{1+e^{\bar{X}(u)+\lambda(u)y}}du\right)^2\right]d\tau,$$

$\vdots$

Case IV. $\zeta(x) = \frac{1}{2}\left(1+\frac{2}{\pi}\arctan(x)\right)$. In this case,we have

$$\begin{aligned}
&\lambda(t) = e^{-\psi t},\\
&\bar{X}(t) = e^{-\psi t}\int_0^t e^{\psi\tau}\phi(\tau)d\tau,\\
&r(y,t) = \frac{1}{2}\left(1+\frac{2}{\pi}\arctan(\bar{X}(t)+\lambda(t)y)\right),\\
&R(y,t,T) = \frac{1}{2}\int_t^T\left(1+\frac{2}{\pi}\arctan(\bar{X}(\tau)+\lambda(\tau)y)\right)d\tau,\\
&\frac{\partial R(y,t,T)}{\partial y} = \frac{1}{\pi}\int_t^T \frac{\lambda(\tau)}{1+(\bar{X}(\tau)+\lambda(\tau)y)^2}d\tau,\\
&\frac{\partial^2 R(y,t,T)}{\partial y^2} = -\frac{2}{\pi}\int_t^T \frac{\lambda(\tau)^2(\bar{X}(\tau)+\lambda(\tau)y)}{(1+(\bar{X}(\tau)+\lambda(\tau)y)^2)^2}d\tau,\\
&\frac{\partial^3 R(y,t,T)}{\partial y^3} = \frac{1}{\pi}\int_t^T \frac{6\lambda(\tau)^3(\bar{X}(\tau)+\lambda(\tau)y)^2-2\lambda(\tau)^3}{(1+(\bar{X}(\tau)+\lambda(\tau)y)^2)^3}d\tau,\\
&\frac{\partial^4 R(y,t,T)}{\partial y^4} = \frac{1}{\pi}\int_t^T \frac{24\lambda(\tau)^4(\bar{X}(\tau)+\lambda(\tau)y)\left(1-(\bar{X}(\tau)+\lambda(\tau)y)^2\right)}{(1+(\bar{X}(\tau)+\lambda(\tau)y)^2)^3}d\tau.
\end{aligned}$$

Then we have

$$\begin{aligned}
&h_0 = 0,\\
&h_1 = \int_t^T r(y,\tau)d\tau = \frac{1}{2}\int_t^T\left(1+\frac{2}{\pi}\arctan(\bar{X}(\tau)+\lambda(\tau)y)\right)d\tau,\\
&h_2 = \frac{\sigma^2}{2}\int_t^T e^{2\psi\tau}\left[-\frac{2}{\pi}\int_\tau^T \frac{\lambda(u)^2(\bar{X}(u)+\lambda(u)y)}{(1+(\bar{X}(u)+\lambda(u)y)^2)^2}du\right]\\
&\quad-\left(\frac{1}{\pi}\int_\tau^T \frac{\lambda(u)}{1+(\bar{X}(u)+\lambda(u)y)^2}du\right)^2\Bigg]d\tau,\\
&\vdots
\end{aligned}$$

5.5 Numerical Tests

In this section, we shall test the numerical accuracy of the homotopy perturbation method (HPM) compared to the binomial tree method (BIN). The parameters for the three models (the Black-Karasinski model, Case III and Case IV) are $\psi = 0.04$, $\sigma = 0.06$, $r_0 = 0.06$, $F = \$100$, and $\phi = -0.08,\ -0.16,\ -0.08$, respectively. Table 5.1, 5.2, and 5.3 display the bond prices, yield rates and the errors between the two methods for various maturities. Here the errors are the maximum absolute errors for interest rates $r = 0.01, 0.02, \ldots, 0.2$. The step size for the binomial method is $1.0e-6$. The results in the tables show that HPM with a few terms can produce very accurate approximations of the bond prices. Hence, we may be able to estimate the model parameters by combining the HPM with the Kalman filter method in Chapter 3, which will be our future work.

Table 5.1: Numerical Comparison for BKM

Maturity	Bond ($)			Yield (%)		
	BIN	HPM	Error ($\times 10^{-3}$)	BIN	HPM	Error ($\times 10^{-4}$)
$T = 0.25$	95.1312	95.1264	4.7126	19.9654	19.9853	1.9816
$T = 0.5$	90.5150	90.5080	7.0095	19.9309	19.9463	1.5489
$T = 1$	81.9864	81.9865	0.1208	19.8617	19.8616	0.0147
$T = 2$	67.4036	67.4044	0.7189	19.7236	19.7230	0.0533
$T = 5$	38.0718	38.0707	1.1170	19.3139	19.3145	0.0587

Table 5.2: Numerical Comparison for Case III

Maturity	Bond ($)			Yield (%)		
	BIN	HPM	Error ($\times 10^{-3}$)	BIN	HPM	Error ($\times 10^{-4}$)
$T = 0.25$	95.1715	95.1670	4.4550	19.7959	19.8146	1.8724
$T = 0.5$	90.6672	90.6612	5.9979	19.5949	19.6081	1.3231
$T = 1$	82.5288	82.5317	2.8300	19.2022	19.1988	0.3429
$T = 2$	69.1382	69.1459	7.6990	18.4531	18.4476	0.5568
$T = 5$	43.8846	43.8935	8.9027	16.4721	16.4681	0.4057

Table 5.3: Numerical Comparison for Case IV

Maturity	Bond ($)			Yield (%)		
	BIN	HPM	Error ($\times 10^{-3}$)	BIN	HPM	Error ($\times 10^{-4}$)
$T = 0.25$	95.1437	95.1387	4.9908	19.9126	19.9336	2.0983
$T = 0.5$	90.5623	90.5542	8.0189	19.8265	19.8442	1.7710
$T = 1$	82.1533	82.1497	3.5576	19.6583	19.6627	0.4331
$T = 2$	67.9269	67.9164	10.4833	19.3369	19.3446	0.7717
$T = 5$	39.6953	39.6609	34.3502	18.4788	18.4961	1.7314

CHAPTER 6

CONCLUSION

In this dissertation, we have studied a front-fixing finite element method for American put option on zero-coupon bond under quadratic term structure model. We also employed the efficient algorithm to solve the two-boundary complimentary system resulting from discretizing the partial differential equations. The numerical results are presented and compared to the ordinary finite element method as well as trinomial tree approach. Our numerical results show that the front-fixing element method outperform both the finite difference method and trinomial tree method in terms of accurate approximations of early exercise interest rates.

Next, we have applied the Kalman filter to estimate the parameters in the affine models and the CIR model while extended Kalman filter and unscented Kalman filter were used to estimate the quadratic short rate model. Due to the second order polynomial form of yield in the quadratic model, we have proposed a variation of Kalman filter which reduces the non-linearity by increasing one more dimension. Monte Carlo simulation results show that all these Kalman filter related estimators did a fine job in estimating the parameters. We further calibrate both time invariant Vasiceck model, CIR model and the quadratic model and the time varying parameters

in extended Vasiceck model, extended CIR model and quadratic short rate model with time-varying parameters to the Treasury zero-coupon yield data. Numerical results show that models with time varying parameters did a better job of both in-sample fitting and out-of-sample fitting than models with invariant parameters. Among models with time-varying parameters, quadratic short rate model outperforms the affine models due to its ability to capture the non-linearity in the data.

We also have studied a general one-factor model which allows that the short rates only take realistic value in $(0, 1)$. Homotopy perturbation methods (HPM) are employed to solve the partial differential equations resulting from the zero-coupon bond pricing. The numerical accuracy of the HPM is compared to the binomial approach for pricing zero-coupon bonds and yield rates under the artificial parameters. The test results show that the accuracy of the HPM is very good. Our future work is to combine the HPM and the unscented Kalman filter to estimate the model parameters by fitting it with treasury yield data.

BIBLIOGRAPHY

[1] D. H. Ahn, R. F. Dittmar, and A. R. Gallant, *Quadratic term structure models: theory and evidence*, Review of Financial Studies, 15 (2002), pp. 243–288.

[2] Y. Aït-Sahalia and R. Kimmel, *Estimating affine multifactor term structure models using closed-form likelihood expansions*, Journal of Financial Economics 98 (2010), pp. 113–144.

[3] T. Andersen and J. Lund, *Estimating continuous time stochastic volatility models of the short term interest rate*, Journal of Econometrics 77 (1997), pp. 343–377

[4] S. Babbs and K. Nowman, *Kalman Filtering of Generalized Vasicek Term Structure Models*, Journal of Financial and Quantitative Analysis, 34.01 (1999), pp. 115–130.

[5] F. Black, E. Derman, and W. Toy, *A one-factor model of interest rates and its applications to Treasury bond options*, Financial Analysis Journal 2, 1990, pp. 33–39.

[6] D. Bolder, *Affine Term-Structure Models: Theory and Implementation*, Bank of Canada, 2001.

[7] M. Brandt and P. He, *Simulated likelihood estimation of affine term structure moddels from panel data*, Working paper, Univeristy of Pennsylvania.

[8] D. R. Beaglehole and M. Tenney, *General solution of some interest rate contingent claim pricing equations*, J. Fixed Income, 1 (1991), pp. 69–83.

[9] D. R. Beaglehole and M. Tenney, *A Nonlinear equilibrium model of term structures of interest rates: corrections and additions*, J. Fin. Econ, 32 (1992), pp. 345–354.

[10] T. Björk, *Interest Rate Theory*, Lecture Notes in Math. 1656, Springer, New York, 1996, pp. 53–122.

[11] F. Black and P. Karasinski, *Bond and option pricing when short rates are lognormal*, Financial Analysts Journal 47 (1991), pp. 52–59.

[12] T. Bollerslev and J. Wooldridge, *Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances*, Econometric reviews 11.2 (1992) pp. 143-172.

[13] K. C. Chan, G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, *An empirical comparison of the short-term interest rate*, Journal of Finance 1 (1992), pp. 1309–1227.

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

[15] P. Christoffersen, C. Dorion, K, Jacobs and L. Karoui, *Nonlinear Kalman Filtering in Affine Term Structure Models*, Working Paper: University of Toronto.

[16] J. C. Cox, J. E. Ingorsoll, and S. A. Ross, *A theory of the term structure of interest rates*, Economica, 53 (1985), pp. 385–407.

[17] C. W. Cryer, *The efficient solution of linear complementarity problems for tridiagonal Minkowski matrices*, ACM Transactions on Mathematical Software, 9 (1983), pp. 199–214.

[18] Q. Dai and K. Singleton, *Simulated Moments Estimation of Markov Models of Asset Prices*, Econometrica, 61 (1993), pp. 929–952.

[19] Q. Dai and K. Singleton, *Specification analysis of affine term structure models*, Journal of Finance 55 (2000) pp. 1943–1978.

[20] F. De Jong, *Time-Series and Cross-Section Information in Affine Term Structure Models*, Journal of Financial and Quantitative Analysis, 31 (2000) pp .131–157.

[21] M. De Pooter, *Examining the Nelson-Siegel class of term structure models*, Tinbergen Institute Discussion Papers. Tinbergen Institute (2007).

[22] F. Diebold and C. Li, *Forecasting the Term Structure of Government Bond Yields*, Journal of Econometrics (2006), pp. 337–364

[23] C. Doz, D. Giannone, and L. Reichlin, *A quasi maximum likelihood approach for large approximate dynamic factor models* (2006).

[24] J. Duan and J. Simonato, *Estimating and testing exponential-affine term structure models by Kalman filter*, Review of Quantitative Finance and Accounting 13 (1999), pp. 102-127

[25] G. Duffee and R.H. Standon, *Estimation of dynamic term structure models*, The Quarterly Journal of Finance 2 (2012).

[26] R. Gallant and G. Tauchen, *Which moments to match?*, Econometric Theory 12 (1996), pp. 657–681.

[27] A. Geyer and S. Pichler, *A state-space approach to estimate and test multifactor Cox-Ingersoll-Ross models of the term structure*, Journal of Financial Research 23 (1999), pp. 107–130.

[28] M. Gill, S. Große, and E. Schumann, *Calibrating the Nelson-Siegel-Svensson model*, COMISEF Working Paper Series 31 (2010).

[29] R. Gimenoa and J. Nave, *A Genetic Algorithm Estimation of the Term Structure of Interest Rates*, Computational Statistics & Data Analysis, 53 (2009), pp. 2236–2250.

[30] R. Gürkaynak, B. Sack, and J. Wright, *The US Treasury yield curve: 1961 to the present*, Journal of Monetary Economics 54.8 (2007), pp. 2291–2304.

[31] L. Hansen, *Large sample properties of generalized method of moments estimators*, Econometrica: Journal of the Econometric Society (1982) pp. 1029–1054.

[32] Hull, J. and White, A., *Pricing interest-rate-derivative securities*, The Review of Financial Studies **3** (1990), 573–592

[33] J. Hull and A. White, *One-factor interest-rate models and the valuation of interest-rate derivative securities*, J. Fin. Quan. Anal., 28 (1993), pp. 235–254.

[34] B. Ingram and B. Lee, *Simulation Estimation of Time Series Models*, Journal of Economietrics, 47 (1991), pp .197–205.

[35] F. Jamshidian, *Bond, futures and option valuation in the quadratic interest rate model*, Appl. Math. Fin., 3 (1996), pp .93–115.

[36] P. Jaillet, E. Lamberton and B. Lapeyre, *Variational inequalities and the pricing of American options*, Acta App. Math., 21 (1990), pp. 263–289.

[37] S. J. Julier and J. K. Uhlmann, *A New Extension of the Kalman Filter to Nonlinear Systems*, Proc. AeroSense: 11th Int. Symp. Aerospace/Defense Sensing, Simulation and Controls, 1997, pp. 182–193.

[38] R. Kalman, *A new approach to linear filtering and prediction problems*, Journal of basic Engineering 82 (1960), pp. 35–45.

[39] T. C. Langetieg, *A multivariate model of the term structure*, The Journal of Finance **35** (1980), 71–97.

[40] M. Leippold and L. Wu, *Asset pricing under the quadratic class*, J. Fin. Quant. Anal., 37 (2002), pp. 271–295.

[41] M. Leippold and L. Wu, *Design and estimation of quadratic term structure models*, European Finance Review, 7 (2003), pp. 47–73.

[42] P. Linz, Analytical and Numerical Methods for Volterra Equations, SIAM Philadelphia, 1985.

[43] F. A. Longstaff, *A nonlinear general equilibrium model of the term structure of interest rates*, J. Fin. Econ., 23 (1989), pp. 195–224.

[44] F. A. Longstaff and E. S. Schwartz, *Interest rate volatility and the term structure: a two-factor general equilibrium model*, J. Finance, 47 (1992), pp. 1259–1282.

[45] F. A. Longstaff and E. S. Schwartz, *Implementation of the Longstaff–Schwartz interest rate model*, J. Fixed Income, 3 (1993), pp. 7–14.

[46] J. He, *The homotopy pertubation method for nonlinear oscillators with discontinuities*, Applied Mathematics and Computation 151 (2004), pp. 287–292.

[47] J. He, *Homotopy perturbation technique*, Computational Methods in Applied Mechanics and Engineering 178 (1999), pp. 257–252

[48] J. He, *A coupling method of a homotopy technique and a perturbation technique for nonlinear problems*, International Journal of Non-linear Mechanics 35 (2000), pp. 37–43.

[49] J. He, *Homotopy pertubation method for bifurcation of nonlinear problems*, International Journal of Nonlinear Science and Numerical Simuloation 6 (2005), pp. 207-208.

[50] S. Julier and J. Uhlmann, *A New Extension of the Kalman Filter to Nonlinear Systems*, In Proc. of AeroSense: The 11th Int. Symp. on Aerospace/Defence Sensing, Simulation and Controls, 3068 (1997), pp. 182–193.

[51] S. Julier and J. Uhlmann, *The scaled unscented transformation*, The Proc. Amer. Control Conf.,(2002), pp. 4555–4559.

[52] S. Julier, *"The spherical simplex unscented transformation*, The Proc. Amer. Control Conf., 3 (2003), pp. 2430–2434.

[53] M. Leippold and L. Wu, *Design and Estimation of Multi-currency Quadratic Models*, Review of Finance 11 (2007), pp. 167–207.

[54] J. Lund, *Econometric Analysis of Continous-Time Arbitrage-Free Models of the Term Structure of Interest Rates*, Working paper (1997).

[55] Merton, R. C., Continuous-Time Finance, Basil Blackwell, Oxford and Cambridge, 1990.

[56] M. Musiela and M. Rutkowski, *Martingale Methods in Financial Modelling (Sochastic Modelling and Applied Probability)*, Springer, 2nd edition, 2011

[57] A.Monfort, J. P. Renne and G. Roussellet, *The quadratic kalman filter*, mimeo (2013).

[58] K. Nyholm and V. Rositsa, *Nelson-Siegel, Affine and Quadratic Yield Curve Specifications: Which One is Better at Forecasting?*, Journal of Forecasting, 31(2012), pp. 540–564.

[59] C. Nelson and A. Siegel, *Parsimonious modeling of yield curves*, Journal of Business, 60(4), pp. 473–489.

[60] A. Pedersen, *A new approach to maximum likelihood estimation of stochastic differential equations based on discrete observations*, Scandinavian Journal of Statistics 22 (1995), pp. 55–71.

[61] N. Pearson and T. Sun, *Exploiting the conditional density in estimating the term structure: An application to the Cox, Ingersoll, and Ross model*, Journal of Finance 49, pp. 1279–1304.

[62] E. Ruiz, *Quasi-maximum likelihood estimation of stochastic volatility models*, Journal of econometrics 63.1 (1994) pp. 289–306.

[63] P. Santa-Clara,*Simulated Likelihood Estimation of Diffusion with an Application to the Short Term Interest Rate*, Mimeo (1995), UCLA.

[64] L. Svensson, *Estimating and Interpreting Forward Interest Rates: Sweden 1992-1994*, Sveriges Riksbank Quarterly Review 3 (1995), pp. 13–26.

[65] G. Terejanu, *Discrete Kalman filter tutorial*, University at Buffalo Department of Computer Science and Engineering, Buffalo, NY 14260.

[66] O. Vasicek, *An equilibrium characterization of the term structure*, Journal of Financial Ecom. 5 (1977), pp. 177–188.

[67] R. Van der Merwe, *Sigma-Point Kalman Filters for Probabilistic Inference in Dynamic State-Space Models*, Technical report, 2003.

[68] E. Wan and R. Van der Merwe, *The unscented Kalman filter for nonlinear estimation*, Proc. Symp. Adaptive Syst. Signal Process., Commun. Contr., 2000.

[69] H. Yang, *Calibration of the extended CIR model*, SIAM J. Appl. Math. 66 (2006), pp. 721–735.

[70] H. Zhou, *Finite sample properties of EMM, GMM, QMLE, and MLE for a square-root interest rate diffusion model*, Journal of Computational Finance 5 (2001), pp. 89–122.

VITA

Graduate College
University of Nevada, Las Vegas

Xudong Sun

Degrees:

Bachelor of Science in Mathematics, 2006
Dalian University of Technology, China

Master of Science – Mathematical Sciences, 2008
Dalian University of Technology, China

Publications:

M. Neda, X. Sun, and L. Yu, Increasing Accuracy and Efficiency for Regularized Navier–Stokes Equations, Acta Appl. Math. 118 (2012), pp. 57–79

Dissertation Title:

Empirical Studies on Interest Rate Derivatives

Dissertation Examination Committee:

Chairperson, Hongtao Yang, Ph.D.
Committee Member, Jichun Li, Ph.D.
Committee Member, Michael Marcozzi, Ph.D.
Committee Member, Monika Neda, Ph.D.
Committee Member, Pengtao Sun, Ph.D.
Graduate Faculty Representative, Jianzhong Zhang, Ph.D.