UNIVERSITÉ DU QUÉBEC À MONTRÉAL

INFÉRENCE POUR DES PROCESSUS AFFINES BASÉE SUR DES OBSERVATIONS À TEMPS DISCRET

MÉMOIRE

PRÉSENTÉ

COMME EXIGENCE PARTIELLE

DE LA MAÎTRISE EN MATHÉMATIQUES

PAR

MARYAM LOLO

DÉCEMBRE 2009

ACKOWLEDGMENTS

I am heartily thankful to my supervisor, René Ferland, whose encouragement, guidance and support enabled me to develop an understanding of the subject.

My parents, Parvin and Ahmad, and my sister, Shaghayegh have been a constant source of support, and this thesis would certainly not existed without their supports and I thank them. Lastly, I offer my regards and blessings to all of those who supported me in any respect during the completion of the project.

CONTENTS

LIST OF TABLES v

LIST OF FIGURES vi

RÉSUMÉ x

ABSTRACT xi

INTRODUCTION 1

CHAPTER I
STOCHASTIC DIFFERENTIAL EQUATIONS 3

1.1 ODE as a modeling tool 3

1.2 Brownian motion and Wiener integrals 4

1.3 Linear stochastic differential equation 6

1.4 Nonlinear stochastic differential equations 9

1.5 Systems of SDEs 10

1.6 Affine models and zero-coupon formulas 11

1.6.1 Single-factor models 12

1.6.2 Multi-factor models 13

CHAPTER II
MAXIMUM LIKELIHOOD ESTIMATION 15

2.1 General principles 15

2.2 MLE for the Vasicek process 17

2.3 MLE for the square root process 21

2.4 Numerical examples 23

CHAPTER III
ZERO-COUPON DATA 34

3.1 Zero-coupon prices 34

3.2 Single-factor models 36

3.3 Multi-factor models 37

3.4 Maximum likelihood estimation 39

3.5 Numerical analysis 40

CHAPTER IV
KALMAN FILTERING 50

4.1 Introduction to Kalman filtering 50

4.2 The state-space formulation 54

4.2.1 Vasicek model 54

4.2.2 The CIR model 57

4.3 The Kalman filter implementation 59

4.4 Maximum likelihood estimation 61

4.5 Numerical analysis 62

CONCLUSION 71

APPENDIX A
SIMULATION RESULTS FOR MAXIMUM LIKELIHOOD METHOD 72

APPENDIX B
SIMULATION RESULTS FOR INDIRECT MAXIMUM LIKELIHOOD METHOD 84

APPENDIX C
SIMULATION RESULTS USING KALMAN FILTERING 96

APPENDIX D
MATLAB CODE FOR SIMULATION AND ESTIMATION 108

D.1 Code for Chapter II 108

D.1.1 Vasicek Model 108

D.1.2 CIR Model 111

D.2 Code for Chapter III 113

D.2.1 Vasicek Model 113

D.2.2 CIR Model 115

D.3 Code for Chapter IV 117

D.3.1 Vasicek Model 117

D.3.2 CIR Model 120

REFERENCES 125

LIST OF TABLES

2.1 A simulation analysis of the MLEs with $n = 100$ observations 31

2.2 A simulation analysis of the MLEs with $n = 500$ observations 32

2.3 A simulation analysis of the MLEs with $n = 1000$ observations 33

3.1 A simulation analysis of the MLEs for zero-coupon data with $n = 100$ observations 47

3.2 A simulation analysis of the MLEs for zero-coupon data with $n = 500$ observations 48

3.3 A simulation analysis of the MLEs for zero-coupon data with $n = 1000$ observations 49

4.1 A simulation analysis using Kalman filter for $n = 100$ observations 68

4.2 A simulation analysis using Kalman filter for $n = 500$ observations 69

4.3 A simulation analysis using Kalman filter for $n = 1000$ observations 70

LIST OF FIGURES

2.1 Generated sample paths for the Vasicek process 25

2.2 Generated sample paths for Cox-Ingersoll-Ross process 26

2.3 Empirical distribution of $\hat{\kappa}$ for the Vasicek model with parameters ϕ_1 27

2.4 Empirical distribution of $\hat{\kappa}$ for the CIR model with parameters ϕ_1 28

2.5 Empirical distribution of $\hat{\theta}$ for the CIR model with parameters ϕ_2 29

2.6 Empirical distribution of $\hat{\sigma}$ for the CIR model with parameters ϕ_2 30

3.1 Generated prices for the Vasicek model 41

3.2 Generated prices for the Cox-Ingersoll-Ross model 42

3.3 Empirical distribution of $\hat{\kappa}$ with zero-coupon data for the Vasicek model with parameters ϕ_2 43

3.4 Empirical distribution of $\hat{\sigma}$ with zero-coupon data for the Vasicek model with parameters ϕ_2 44

3.5 Empirical distribution of $\hat{\kappa}$ with zero-coupon data for the CIR model with parameters ϕ_2 45

3.6 Empirical distribution of $\hat{\theta}$ with zero-coupon data for the CIR model with parameters ϕ_1 46

4.1 Empirical distribution of $\hat{\sigma}$ for zero-coupon data and Kalman filtering (Vasicek model with parameter ϕ_1). 64

4.2 Empirical distribution of $\hat{\sigma}$ for zero-coupon data and Kalman filtering (Vasicek model with parameter ϕ_2). 65

4.3 Empirical distribution of $\hat{\theta}$ for zero-coupon data and Kalman filtering (CIR model with parameter ϕ_1). 66

4.4 Empirical distribution of $\hat{\kappa}$ for zero-coupon data and Kalman filtering (CIR model with parameter ϕ_2). 67

A.1 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model. 72

A.2 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model. 73

A.3 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model. 74

A.4 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model. 75

A.5 Empirical distribution of MLE $\widehat{\sigma}$ for Vasicek's model. 76

A.6 Empirical distribution of MLE $\widehat{\sigma}$ for Vasicek's model. 77

A.7 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model. 78

A.8 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model. 79

A.9 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model. 80

A.10 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model. 81

A.11 Empirical distribution of MLE $\widehat{\sigma}$ for the CIR model. 82

A.12 Empirical distribution of MLE $\widehat{\sigma}$ for the CIR model. 83

B.1 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model (zero-coupon data). 84

B.2 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model (zero-coupon data). 85

B.3 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model (zero-coupon data). 86

B.4 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model (zero-coupon data). 87

B.5 Empirical distribution of MLE $\widehat{\sigma}$ for Vasicek's model (zero-coupon data). 88

B.6 Empirical distribution of MLE $\widehat{\sigma}$ for Vasicek's model (zero-coupon data). 89

B.7 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model (zero-coupon data). 90

B.8 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model (zero-coupon data). 91

B.9 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model (zero-coupon data). 92

B.10 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model (zero-coupon data). 93

B.11 Empirical distribution of MLE $\widehat{\sigma}$ for the CIR model (zero-coupon data). 94

B.12 Empirical distribution of MLE $\widehat{\sigma}$ for the CIR model (zero-coupon data). 95

C.1 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for Vasicek's model. 96

C.2 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for Vasicek's model. 97

C.3 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for Vasicek's model. 98

C.4 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for Vasicek's model. 99

C.5 Empirical distribution of Quasi-MLE $\widehat{\sigma}$ for Vasicek's model. 100

C.6 Empirical distribution of Quasi-MLE $\widehat{\sigma}$ for Vasicek's model. 101

C.7 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for the CIR model. 102

C.8 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for the CIR model. 103

C.9 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for the CIR model. 104

C.10 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for the CIR model. 105

C.11 Empirical distribution of Quasi-MLE $\widehat{\sigma}$ for the CIR model. 106

C.12 Empirical distribution of Quasi-MLE $\hat{\sigma}$ for the CIR model. 107

RÉSUMÉ

Dans ce mémoire, on étudie la distribution empirique des estimateurs de vraisemblance maximale pour des processus affines, basés sur des observations à temps discret. On examine d'abord le cas où le processus est directement observable. Ensuite, on regarde ce qu'il advient lorsque seule une transformation affine du processus est observable, une situation typique dans les applications financières. Deux approches sont alors considérées: maximisation de la vraisemblance exacte ou maximisation d'une quasi-vraisemblance obtenue du filtre de Kalman.

Mots-clés: estimation de vraisemblance maximale, processus affines, obligation à l'escompte, quasi-vraisemblance, filtre de Kalman.

ABSTRACT

In this dissertation, we study the empirical distribution of the maximum likelihood estimator for affine processes, based on discrete sample data. We first study the case where the process is directly observable. Then we look at what happens if only an affine transformation of the data is observable, a typical situation in financial applications. Two approaches are considered: maximisation of the exact likelihood or maximisation of a quasi-likelihood computed via Kalman filtering.

Keywords: maximum likelihood estimation, affine processes, zero-coupon bond, quasi-likelihood, Kalman's filter.

INTRODUCTION

This dissertation is concerned with the parameter estimation of affine processes used to describe the term structure of interest rates. Many methods have been proposed for this purpose, most of them being essentially variations of maximum likelihood estimation. Our presentation is divided into two parts. The first part provides an overview of concepts required to understand the term structure literature used in finance, which uses stochastic differential equations (SDEs) to show the dynamics of a zero-coupon bond. This development begins with a single-factor model and then generalizes to higher dimensions. Further, we study the relation between state variables and bond prices. Since the term structure of interest rates is never observed in continuous time, we consider an estimation method based on discrete-time data. The second part discusses and compares different approaches to find an accurate method that can be used to estimate the unknown parameter set at each of these term-structure models. We conduct an empirical study for each of these methods.

The first chapter includes an introduction to the theory and implementation of the class of affine term structure models. A variety of models exist, including those suggested by Duffie and Kan (1996), Vasicek (1977), Cox, Ingersoll and Ross (1985a), Longstaff and Schwartz (1992a) and Chen (1995). This dissertation focuses on the theoretical formulation of the Vasicek and CIR single and multi-factor affine models.

In the second chapter we present maximum likelihood estimation for stochastic differential equations based on a direct observation of the process itself. For the SDEs considered here, this is rather straightforward. For nonlinear SDEs the problem is more difficult and will not be addressed here. However, for a more detailed discussion of these issues, one can look at Pederson (1995), Cox, Ingersoll and Ross (1985b), Lo (1988) and Gourieroux and Jasiak (2001).

In the third chapter we consider maximum likelihood estimation for the term structure models introduced in Chapter I, and we study the empirical performance of the suggested estimator on two examples. The study begins with the single factor models and then generalizes these concepts into a multi-dimensional setting. Empirical papers that explore this issue in detail include Cox, Ingersoll and Ross (1985b), Geyer and Pichler (1999) and Bolder (2001).

The final chapter summarizes the material on the state-space models and introduces another technique to estimate unknown parameters, a technique called Kalman filtering. We then apply Kalman filtering to the simulated data of the previous chapter and compare the results.

CHAPTER I

STOCHASTIC DIFFERENTIAL EQUATIONS

1.1 ODE as a modeling tool

Differential equations consist of an unknown function of one or several variables which relates to the values of the function itself and of its derivatives of various order. Many laws in Physics, Chemistry, Engineering and Economics can be expressed in the simplest way by differential equations. For example, let $X(t)$ represent one coordinate of the position of a particle in space; then $X'(t)$ and $X''(t)$ represent the velocity and the acceleration of the particle at time t. Let m denote the mass of the particle and let $F(t)$ denote the force acting on the particle at time t. By Newton's law,

$$F(t) = mX''(t). \tag{1.1.1}$$

Usually, $F(t)$ consist of three types of forces:

a) a frictional force $-fX'(t)$,

b) a restoring force $-kX(t)$,

c) an external force $\xi(t)$, which is independent of the motion.

Therefore, we may write $F(t)$ with respect to these types of forces as

$$F(t) = -fX'(t) - kX(t) + \xi(t). \tag{1.1.2}$$

By combining (1.1.1) and (1.1.2) we obtain the differential equation

$$mX''(t) + fX'(t) + kX(t) = \xi(t).$$

If the differential equation contains functions of only one independent variable, and one or more of its derivatives with respect to that variable, we call it an ordinary differential equation (ODE); thus, the above example is an ODE.

ODEs frequently appear in the natural and physical models but in reality some parameters may not be deterministic and we should consider them as random variables. For example, if we suppose that the external force is due to some random effect, then in the above case we can think of $\xi(t)$ as a collection of random variables which change by time t.

1.2 Brownian motion and Wiener integrals

So far, we found that for modeling a natural event we can use an ordinary differential equation which has a random process component. Now the question is, how can we model this random process?

To answer this question let us introduce Brownian motion. Particles suspended in a fluid exhibit a random motion, called Brownian motion. Many mathematical models for this physical process have been proposed. The model usually used for such motion is called the *Wiener process.* A Wiener process, with parameter σ^2, is a collection $\{W(t), t \geq 0\}$ of random variables which satisfies the following properties:

i) $W(0) = 0$,

ii) $W(t) - W(s)$ has a normal distribution with mean 0 and variance $\sigma^2(t-s)$ for $s < t$,

iii) for $t_1 < t_2 < \cdots < t_n$, the variables

$$W(t_2) - W(t_1), \quad W(t_3) - W(t_2), \quad \ldots, W(t_n) - W(t_{n-1})$$

are independent,

iv) the function $t \mapsto W(t)$ is continuous.

From the properties of the Wiener process, we can immediately conclude that the mean of the random variable $W(t)$ is 0 and its variance is $\sigma^2 t$. Also a Wiener process has the property that any finite linear combination of random variables $W(t)$, $t \in [0,\infty)$, is normally distributed. A stochastic process having this property is called a Gaussian process.

In the equation (1.1), if the external force is due to molecular bombardment, then physical reasoning leads to the conclusion that this external force should be the *derivative* of the Brownian motion, so that $\xi(t) = W'(t)$; then we can rewrite the equation as,

$$mX''(t) + fX'(t) + kX(t) = W'(t). \tag{1.2.3}$$

In (1.2.3) we are interested to find the solution $X(t)$. But in order to do so, or even merely prove existence results, we have to calculate $\int W'(t)\,dt$, or at least give a meaning to it. Although the function $t \mapsto W(t)$ is continuous, it is almost surely not differentiable (in a probabilistic sense), so in general the integral

$$\int_a^b f(t)W'(t)\,dt$$

does not exist in the usual sense, where a and b are finite numbers and f is a continuously differentiable function on the closed interval $[a, b]$. But we are able to give meaning to this integral. One way of doing so is to define it as

$$\lim_{\varepsilon \to 0} \int_a^b f(t) \left(\frac{W(t+\varepsilon) - W(t)}{\varepsilon} \right) dt,$$

provided the indicated limit exists. To see that this limit actually exists and to evaluate it explicitly, we observe that

$$\int_a^b f(t) \left(\frac{W(t+\varepsilon) - W(t)}{\varepsilon} \right) dt = \int_a^b f(t) \frac{d}{dt} \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} W(s)\,ds \right) dt.$$

Integrating by parts the right hand side of this equation, we conclude that

$$\begin{aligned} \int_a^b f(t) \left(\frac{W(t+\varepsilon) - W(t)}{\varepsilon} \right) dt &= \left[f(t) \frac{1}{\varepsilon} \int_t^{t+\varepsilon} W(s)\,ds \right]_a^b \\ &\quad - \int_a^b f'(t) \left(\frac{1}{\varepsilon} \int_t^{t+\varepsilon} W(s)\,ds \right) dt. \end{aligned} \tag{1.2.4}$$

Since a Wiener process has continuous sample paths, it follows that the right hand side of (1.2.4) converges to

$$f(t)W(t)\Big|_a^b - \int_a^b f'(t)W(t)\,dt.$$

Thus, we are led to define

$$\int_a^b f(t)\,dW(t)$$

as the limit of right side of (1.2.4) as $\varepsilon \to 0$, that is, by the formula

$$\int_a^b f(t)\,dW(t) = f(b)W(b) - f(a)W(a) - \int_a^b f'(t)W(t)\,dt. \tag{1.2.5}$$

The "formal" derivative of a Wiener process is called *white noise*. Since a Wiener process is a Gaussian process, it follows from (1.2.5) that

$$\int_a^b f(t)dW(t)$$

is normally distributed with mean zero and variance

$$\mathrm{Var}\left[\int_a^b f(t)dW(t)\right] = \sigma^2 \int_a^b f^2(t)\,dt. \tag{1.2.6}$$

We will see later that white noise is widely used in sciences and finance.

1.3 Linear stochastic differential equation

A stochastic process is a collection of random variables indexed by some parameters, usually time. A *stochastic differential equation*, or SDE for short, is a differential equation in which one or more of the terms is a stochastic process. For instance, an n^{th} order linear stochastic differential equation taking the form

$$a_0 X^{(n)}(t) + a_1 X^{(n-1)}(t) + \cdots + a_n X(t) = W'(t), \tag{1.3.7}$$

where $a_0, \ldots, a_n$ are real constants with $a_0 \neq 0$, and $W'(t)$ is white noise with parameter σ^2. However, we will only consider first order equations and focus on them. We first define precisely what is meant by a solution to such a differential equation. Consider the differential equation (1.3.7) where $n = 1$,

$$a_0 X'(t) + a_1 X(t) = W'(t). \tag{1.3.8}$$

In order to solve (1.3.8) we proceed through a series of reversible steps. We first divide both side of (1.3.8) by a_0 and integrate from t_0 to t:

$$X(t)+\frac{a_1}{a_0}\int_{t_0}^{t}X(s)\,ds=X(t_0)-\frac{W(t_0)}{a_0}+\frac{W(t)}{a_0}.$$

Multiplying both sides by $e^{-\alpha t}$, where $\alpha=-a_1/a_0$, we have

$$e^{-\alpha t}X(t)-\alpha e^{-\alpha t}\int_{t_0}^{t}X(s)\,ds=\left(X(t_0)-\frac{W(t_0)}{a_0}\right)e^{-\alpha t}+\frac{e^{-\alpha t}}{a_0}W(t),$$

which we rewrite as

$$\frac{d}{dt}\left(e^{-\alpha t}\int_{t_0}^{t}X(s)\,ds\right)=\left(X(t_0)-\frac{W(t_0)}{a_0}\right)e^{-\alpha t}+\frac{e^{-\alpha t}}{a_0}W(t).$$

Integrating both sides of this equation from t_0 to t, we conclude that

$$\int_{t_0}^{t}X(s)\,ds=\left(X(t_0)-\frac{W(t_0)}{a_0}\right)\frac{(e^{\alpha(t-t_0)}-1)}{\alpha}+e^{\alpha t}\int_{t_0}^{t}\frac{e^{-\alpha s}}{a_0}W(s)ds.$$

By differentiating we see that,

$$\begin{aligned}X(t) &= (X(t_0)-\frac{W(t_0)}{a_0})e^{\alpha(t-t_0)}+\frac{W(t)}{a_0}+\frac{\alpha}{a_0}\int_{t_0}^{t}e^{\alpha(t-s)}W(s)\,ds.\\ &= X(t_0)e^{\alpha(t-t_0)}+\frac{1}{a_0}\int_{t_0}^{t}e^{\alpha(t-s)}\,dW(s),\end{aligned}$$

which we can rewrite as

$$X(t)=\frac{X(t_0)}{\alpha}e^{\alpha(t-t_0)}+\frac{1}{a_0}\int_{t_0}^{t}e^{\alpha(t-s)}\,dW(s)\,dt.$$

This process is a Gaussian process and we can easily compute its mean and variance:

$$\begin{aligned}\mu_X(t) &= x_0e^{\alpha t},\\ Var(X(t)) &= \frac{\sigma^2}{2a_0a_1}(1-e^{2\alpha t}).\end{aligned}$$

We illustrated above the techniques for handling equation (1.3.8). Further, we will consider a stochastic differential equation which is used for interest rate modeling in finance. It is written as

$$X'(t)+\kappa X(t)=\kappa\overline{r}+\sigma W'(t) \tag{1.3.9}$$

or equivalently (in differential notation)

$$dX(t) = \kappa(\overline{r} - X(t)) + \sigma dW(t).$$

It represents the behavior of a short term interest rate $X(t)$ moving randomly around a long term target $\overline{r}$. Equation (1.3.9) is also easy to solve; First we multiply by the integrating factor $e^{\kappa t}$ which gives

$$e^{\kappa t} dX(t) + e^{\kappa t} \kappa X(t) = e^{\kappa t} \kappa \overline{r} + e^{\kappa t} \sigma dW(t)$$

or, equivalently,

$$d(e^{\kappa t} X(t)) = \overline{r} d(e^{\kappa t}) + \sigma e^{\kappa t} dW(t).$$

Integrating from 0 to t, we get

$$\begin{aligned} e^{\kappa t} X(t) - e^{\kappa 0} X(0) &= \int_0^t d\left(e^{\kappa u} X(u)\right) \\ &= \int_0^t \overline{r} d\left(e^{\kappa t}\right) + \sigma \int_0^t e^{\kappa u} dW(u) \\ &= \overline{r}\left[e^{\kappa u}\right]_0^t + \sigma \int_0^t e^{\kappa u} dW(u) \\ &= \overline{r}\left(e^{\kappa t} - 1\right) + \sigma \int_0^t e^{\kappa u} dW(u) \end{aligned}$$

or

$$X(t) = \overline{r} + e^{-\kappa t}(x_0 - \overline{r}) + \sigma \int_0^t e^{-\kappa(t-s)} dW(s). \tag{1.3.10}$$

In the next chapter, we will study the problem of estimating the parameters of equation (1.3.9) using the trajectory of past values of the process. For this purpose, we shall need to know the conditional distribution of $X(t)$, given $X(s) = x$, for $s < t$. We have

$$X(s) = \overline{r} + e^{-\kappa s}(x_0 - \overline{r}) + \sigma e^{-\kappa s} \int_0^s e^{\kappa u} dW(u). \tag{1.3.11}$$

From the integral properties we have

$$\int_0^t e^{\kappa u} dW(u) = \int_0^s e^{\kappa u} dW(u) + \int_s^t e^{\kappa u} dW(u). \tag{1.3.12}$$

We substitute it in (1.3.10) to obtain

$$X(t) = \overline{r} + e^{-\kappa t}(x_0 - \overline{r}) + e^{-\kappa(t-s)}\left(\sigma e^{-\kappa s} \int_0^s e^{\kappa u} dW(u) + \sigma e^{-\kappa s} \int_s^t e^{\kappa u} dW(u)\right). \tag{1.3.13}$$

We can rewrite equation (1.3.11) as follow:

$$\sigma e^{-\kappa s}\int_0^s e^{\kappa u}\,dW(u) = X(s) - \overline{r} - e^{-\kappa s}(x_0 - \overline{r}), \tag{1.3.14}$$

and by substituting equation (1.3.14) in (1.3.13) we get

$$\begin{aligned} X(t) &= \overline{r} + e^{-\kappa t}(x_0 - \overline{r}) + e^{-\kappa(t-s)}\left(X(s) - \overline{r} - e^{-\kappa s}(x_0 - \overline{r}) + \sigma e^{-\kappa s}\int_s^t e^{\kappa u}\,dW(u)\right) \\ &= \overline{r} + e^{-\kappa(t-s)}(X(s) - \overline{r}) + \sigma e^{-\kappa t}\int_s^t e^{\kappa u}\,dW(u). \end{aligned}$$

So given $X(s) = x$, we have

$$X(t) = \overline{r} + e^{-\kappa(t-s)}(x - \overline{r}) + \sigma\int_s^t e^{-\kappa(t-u)}\,dW(u).$$

Therefore, conditionally on $X(s) = x$, the variable $X(t)$ is normally distributed with mean

$$\mu_{X(t)} = \overline{r} + e^{-\kappa(t-s)}(x - \overline{r})$$

and since white noise is normally distributed,

$$\mathrm{Var}\big(X(t)\big) = \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa(t-s)}\right).$$

1.4 Nonlinear stochastic differential equations

In the previous section we studied how to solve a stochastic differential equation, when the random process component is linear. But in many cases this component may not be linear. Consider the following nonlinear stochastic differential equation

$$dX(t) = a(t, X(t))\,dt + b(t, X(t))\,dW(t), \tag{1.4.15}$$

with $X_0 = x_0$ a specified initial value, and $a(t,x)$, $b(t,x)$ possibly nonlinear in x.

In that case we can not, in general, find an explicit formula for the solution, like the one in the previous section. So we typically need to use numerical methods to determine solutions approximately. Even then, we should first know that the equation actually does have a solution, a unique one preferably, for a given initial value. We can show the existence of this solution by an *existence and uniqueness theorem.*

Theorem 1 *Suppose that:*

1. *The functions $a(t,x)$ and $b(t,x)$ are measurable with respect to $t \in [0,T]$ and $x \in \Re$.*
2. *There exists a constant $K > 0$ such that for all $t \in [0,T]$ and all $x, y \in \Re$,*
   a) $|a(t,x) - a(t,y)| + |b(t,x) - b(t,y)| \leq K|x-y|$,
   b) $|a(t,x)|^2 + |b(t,x)|^2 \leq K^2(1+|x|^2)$.
3. *X_0 is independent of $W(t)$, $t > 0$, and $E\left[X_0^2\right] < \infty$.*

Gard (1988), Kloeden and Platen (1995) have described and developed this theorem with more details. Then there is a solution $X(t)$ of (1.4.15) defined on $[0,T]$ which is continuous with probability 1, and such that

$$\sup_{t\in[0,T]} E\left[X^2(t)\right] < \infty$$

Furthermore, a solution with these properties is path-wise unique, that is, if X and Y are two such solutions, then

$$\Pr\left\{ \sup_{t\in[0,T]} |X(t) - Y(t)| = 0 \right\} = 1.$$

Thus, if the drift $a(t,x)$ and the diffusion coefficient $b(t,x)$ of an equation satisfy the *Lipschitz* condition 2-a) and the *growth* condition 2-b), then we can conclude that the SDE has a unique solution.

1.5 Systems of SDEs

In many applications, we need to simultaneously solve several SDEs, with possibly linked drift and diffusion coefficients. It is what we call a system of SDEs. To give a precise definition, consider an m-dimensional Wiener process $W = \{W_t,\ t \geq 0\}$ with components $W_t^1, W_t^2, \ldots, W_t^m$, which are independent scalar Wiener processes with respect to a common filtration. Then we take a d-dimensional vector function $a : [0,T] \times \Re^d$ and a $d \times m$-matrix function $b : [0,T] \times \Re^d \to \Re^{d\times m}$ to form a d-dimensional *vector stochastic differential equation:*

$$dX(t) = a(t, X(t))dt + b(t, X(t))dW(t).$$

By this definition, the general form of a d-dimensional *linear stochastic differential equation* is

$$dX(t) = (A(t)X(t) + a(t))\,dt + \sum_{i=1}^{m} (B^i(t)X(t) + b^i(t))\,dW^i(t),$$

where $A(t), B^1(t), B^2(t), \ldots, B^m(t)$ are $d \times d$- matrices functions and $a(t), b^1(t), b^2(t), \ldots, b^m(t)$ are d-dimensional vector functions.

The existence and uniqueness of the $X(t)$, the solution of the vector stochastic differential equation can be obtained as in the previous section by using the *existence and uniqueness* theorems. see Ikeda and Watanabe (1981) for more details.

1.6 Affine models and zero-coupon formulas

A zero-coupon bond is a bond that pays one unit of account to the holder at maturity date, T and, before this date, no payment is made to the holder. The relation between the zero-coupon interest rate and their time to maturity is called term structure. Interest rate term structure modeling is one of the most important problems in financial literature. Duffie and Kan (1996) introduced the class of affine term structure models, extending Vasicek (1977) and Cox, Ingersoll and Ross (1985a) models. These models are formulated by assuming that the future dynamics of the term structure of interest rates depend on some observed and unobserved factors, called state variables. Affine term-structure models are constructed by assuming that the bond yields are linear functions of the underlying state-variables. Although interest rates change randomly over time, most popular models are based on the concept that it is possible to divide the changes into two parts using a stochastic differential equation. The first part in this modeling is a non-random deterministic component, termed drift; the second part is the random or noise part, termed diffusion.

Affine models are a class of SDEs for which the drift coefficient and the square of the diffusion coefficient are affine functions of x. They are very popular in financial engineering because they lead to closed-form formulas for default-risk free bonds. As a result, affine modeling has become the dominant framework to study the term structure of interest rates since 1980s.

In the chapters to come, we will study how to estimate the coefficients of an affine model for the short rate, using discrete-time observations either of the rate itself, or of financial asset prices derived from the model. Specifically, we will consider default-risk free bonds that make a single payment at a pre-specified future date, and which are called zero-coupon bonds. Zero coupon bonds are particularly important because they represent the basic discount rates in all financial claims that make payments through time. Two special case of affine models, Vasicek and CIR will be considered in detail.

1.6.1 Single-factor models

The Vasicek (1977) model is a one factor partial equilibrium model which assumes that the short rate evolves as an ornstein-uhlenbeck process:

$$dr_t = \kappa(\theta - r_t)dt + \sigma dW_t$$

where $\kappa, \theta > 0$, while $\sigma > 0$ is the unconditional instantaneous volatility of the process, the noise in the diffusion part is a Wiener process. The conditional and unconditional distributions of interest rate changes are Gaussian in this model.

In the single-factor CIR term structure model, the short rate evolves as

$$dr_t = \kappa(\theta - r_t)dt + \sigma\sqrt{r_t}dW_t$$

where $\kappa, \theta > 0$ and $\sigma > 0$ have the same interpretation in the Vasicek case, but the short rate is no longer Gaussian. The parameter restriction $2\kappa\theta \geq \sigma^2$ is imposed in order to ensure that the short rate process does not get trapped at zero. The rate r_t has a conditional non-central chi-square distribution.

Independent of any specific model for the short rate, it is always possible to express the price of a zero coupon bond with time to maturity T, at time t as follow,

$$P_t(\tau) = \widetilde{E}_t\left[e^{-\int_0^\tau r_s ds}\right]$$

where $\tau = T - t$ and $\widetilde{E}_t$ denotes the expected value at time t under the so-called "risk-neutral measure". The latter is obtained from the underlying model measure by adding a risk premium to the drift coefficient of the short rate. The main feature of an affine model for r_t

is the fact that an explicit expression for the price $P_t(\tau)$ is available. For the Vasicek (1977) model, one has

$$P_t(\tau) = e^{A(\tau)-B(\tau)r_t}$$

where

$$B(\tau) = \frac{1}{\kappa}(1 - e^{-\kappa\tau}),$$
$$A(\tau) = \frac{\gamma(B(\tau)-\tau)}{\kappa^2} - \frac{\sigma^2 B^2(\tau)}{4\kappa},$$

and

$$\gamma = \kappa^2(\theta - \frac{\sigma\lambda}{\kappa}) - \frac{\sigma^2}{2}.$$

A similar formula holds for the CIR model with

$$B(\tau) = \frac{2(e^{\gamma\tau}-1)}{(\gamma+\kappa+\lambda)(e^{\gamma\tau}-1)+2\gamma},$$
$$A(\tau) = \ln\left(\frac{2\gamma e^{\frac{(\gamma+\kappa+\lambda)\tau}{2}}}{(\gamma+\kappa+\lambda)(e^{\gamma\tau}-1)+2\gamma}\right)^{\frac{2\kappa\theta}{\sigma^2}},$$

and

$$\gamma = \sqrt{(\kappa+\lambda)^2+2\sigma^2}.$$

The use of a single-state variable or factor, might not be enough to explain the random future movement of the term structure of interest rates. This inadequacy comes from the fact that the dynamics of the term structure of interest rates are too complicated to be summarized by a single source of uncertainty. Because of that, in the next section we present multi-factor term structure models.

1.6.2 Multi-factor models

Now we generalize the single-factor models to higher dimensions. The basic format is similar to the one-factor case, though we need to consider the covariance structure between the diffusion terms. In these models, we typically assume that short rate is a linear combination of n correlated state variables, or factors, which we will denote $y_1, y_2, \ldots y_n$, since there is a relation between the short rate and the factors:

$$r = \sum_{i=1}^{n} y_i$$

Then the stochastic differential equations by using Vasicek model is

$$dy_1(t) = \kappa_1(\theta_1 - y_1(t))dt + \sigma_1 dW_1(t)$$

$$dy_2(t) = \kappa_2(\theta_2 - y_2(t))dt + \sigma_2 dW_2(t)$$

$$\vdots$$

$$dy_n(t) = \kappa_n(\theta_n - y_n(t))dt + \sigma_n dW_n(t)$$

where $W_i(t)$ is a standard scalar Wiener process. For the CIR model there are some restrictions, because the analytic solution exists only when the underlying Brownian motions driving each state variable are independent. Hence, although the model ensures that the interest rates can not be negative, the desire for tractability implies that we give up the correlation between state variables. The multi-factor CIR model is

$$dy_1(t) = \kappa_1(\theta_1 - y_1(t))dt + \sigma_1\sqrt{y_1(t)}dW_1(t)$$

$$\vdots$$

$$dy_n(t) = \kappa_n(\theta_n - y_n(t))dt + \sigma_n\sqrt{y_n(t)}dW_n(t)$$

where $W_1, \ldots, W_n$ are independent standard scalar Wiener processes. For both models, the price $P_t(\tau)$ is given by

$$P_t(\tau) = \exp\left\{ A(\tau) + \sum_{i=1}^{n} B_i(\tau) y_i(t) \right\}$$

where $A, B_1, \ldots, B_n$ are the solutions of (numerically or analytically solvable) ODEs.

CHAPTER II

MAXIMUM LIKELIHOOD ESTIMATION

2.1 General principles

In this section we consider one of the methods, commonly used, to estimate unknown parameters in stochastic differential equations. *Maximum Likelihood Estimation* (MLE) is a classical popular method to find the value of one or more parameters for a probability distribution from a given data set.

Suppose we have sample data

$$X_1, X_2, \ldots, X_n$$

and some of probabilistic model for the data, and we want to estimate the parameters of a model. Consider a family of probability distributions, D_θ, parameterized by an unknown parameter θ (which could be a vector), associated with either a known probability density function (continuous distribution) or a known probability mass function (discrete distribution), denoted f_θ. We draw a sample $x_1, x_2, \ldots, x_n$ of n values from this distribution, and then, by using f_θ we compute the probability density $f_\theta(x_1, x_2, \ldots, x_n)$ associated with our observed data.

As a function of θ with $x_1, x_2, \ldots, x_n$ fixed, the *likelihood* function is

$$L(\theta) = f_\theta(x_1, x_2, \ldots, x_n)$$

The method of *maximum likelihood* estimates θ by finding the value of θ that maximizes

$L(\theta)$. The *maximum likelihood* estimator of θ should be

$$\hat{\theta} = \operatorname{argmax}_\theta L(\theta)$$

In many situations one may assume that the given data are independent identically distributed (i.i.d.), which simplifies the problem because the *likelihood* can then be written as a product of n univariate probability densities:

$$L(\theta) = \prod_{i=1}^{n} f_\theta(x_i), \tag{2.1.1}$$

and since maximization is unaffected by monotone transformations, one can take the logarithm of this expression to turn it into a sum:

$$\ell(\theta) = \log L(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i). \tag{2.1.2}$$

In this work we are interested in estimating the parameters of a SDE based on the observation of the solution at a finite set of times. Hence the sample $X_0, \ldots, X_n$ is in fact of the form $X(t_0), \ldots, X(t_n)$ with $t_0 < t_1 < \cdots < t_n$, where X is this solution of the SDE. In that case $L(\theta)$ is not given by a product like (2.1.1). However, using conditioning, one can write

$$\begin{aligned} L(\theta) &= f_\theta(x_0, \ldots, x_n) \\ &= f_\theta(x_0, \ldots, x_{n-1}) f_\theta(x_n \mid x_0, \ldots, x_{n-1}) \\ &\vdots \\ &= f_\theta(x_0) \prod_{i=1}^{n} f_\theta(x_i \mid x_0, \ldots, x_{i-1}). \end{aligned}$$

Moreover, for almost all SDEs used in practice (and the ones considered in this work), the solution X is *Markovian* which means that

$$f_\theta(x_i \mid x_0, \ldots, x_{i-1}) = f_\theta(x_i \mid x_{i-1}).$$

As we consider only SDEs with fixed initial conditions, we may assume that $t_0 = 0$ and x_0 is given. Hence, only the product of the conditional densities is used for estimation and we end up with a log-likelihood $\ell(\theta)$ quite similar to (2.1.2):

$$\ell(\theta) = \sum_{i=1}^{n} \log f_\theta(x_i \mid x_{i-1}), \tag{2.1.3}$$

where the marginal densities are replaced with *transition* (or conditional) densities.

Unfortunately, for typical nonlinear SDEs, the exact formula of the transition density is unknown. As a result, several approaches towards approximating the transition density have been proposed by using various numerical procedures to estimate the *likelihood* function. For example, Pederson (1995) suggests a simulation-based approach when one splits the time interval into short pieces, and integrates unobserved variables out of a joint Euler density. At the end of the procedure the maximum *likelihood* estimate can be found numerically by using various optimization algorithms. In this work we shall consider only two special cases of the so-called affine processes, and for these, the analytical expression of the transition density is available.

2.2 MLE for the Vasicek process

Consider the following differential equation

$$\begin{cases} dX(t) &= \kappa X(t))dt + \sigma dW(t), \\ x(0) &= x_0, \end{cases} \tag{2.2.4}$$

which is actually the Vasicek equation, described in the previous Chapter, with $\bar{r} = 0$. We use the maximum likelihood principle to estimate the unknown parameters (κ and σ^2). Based on the method given in Chapter I, we can see that, if $X(s) = x$ is given then $X(t)$ is normally distributed with

$$E[X(t)] = e^{-\kappa(t-s)}x$$

and

$$\mathrm{Var}[X(t)] = \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa(t-s)}).$$

Hence the transition density in the Vasicek model is Gaussian:

$$f_{\kappa,\sigma^2}(y \mid x) = \frac{1}{\sqrt{2\pi \mathrm{Var}[X(t)]}} \exp\left\{-\frac{(y - E[X(t)])^2}{2\mathrm{Var}[X(t)]}\right\}.$$

If $x_0, x_1, x_2, \ldots, x_n$ is a sample of $X(t)$ at equally spaced times $0 = t_0 < t_1 < t_2 < \cdots < t_n$ then its *likelihood* is given by

$$L_{\kappa,\sigma^2,(x_0,x_1,\ldots,x_n)} = \prod_{i=1}^{n} \sqrt{\frac{\kappa}{\pi\sigma^2(1 - e^{-2\kappa\Delta})}} \exp\left\{\frac{-\kappa(x_i - e^{-\kappa\Delta}x_{i-1})^2}{\sigma^2(1 - e^{-2\kappa\Delta})}\right\}$$

where $\Delta = t_{j+1} - t_j$, $j = 0, \ldots, n-1$. Taking the logarithm from both sides of the above equation, we have

$$\ell\left(\kappa, \sigma^2; x_0, x_1, \ldots, x_n\right) = \frac{n}{2}\log\left(\frac{\kappa}{\pi\sigma^2(1-e^{-2\kappa\Delta})}\right) - \sum_{i=1}^{n}\frac{\kappa(x_i - e^{-\kappa\Delta}x_{i-1})^2}{\sigma^2(1-e^{-2\kappa\Delta})}.$$

In order to maximize this *log likelihood* with respect to κ and σ^2 we differentiate it with respect to κ and σ^2:

$$\begin{cases} \dfrac{\partial\ell\left(\kappa, \sigma^2; x_0, x_1, \ldots, x_n\right)}{\partial\kappa} &= 0, \\ \dfrac{\partial\ell\left(\kappa, \sigma^2; x_0, x_1, \ldots, x_n\right)}{\partial\sigma^2} &= 0, \end{cases} \tag{2.2.5}$$

and by solving the above system of equations we get the MLEs:

$$\widehat{\kappa} = -\frac{1}{\Delta}\log\left[\frac{\sum_{i=1}^{n} x_i x_{i-1}}{\sum_{i=1}^{n} x_{i-1}^2}\right]$$

and

$$\widehat{\sigma}^2 = \frac{2\kappa\sum_{i=1}^{n}(x_i - e^{-\kappa\Delta}x_{i-1})^2}{n(1-e^{-2\kappa\Delta})}.$$

Further, let us consider the full Vasicek model. The dynamics of the process is described by the stochastic differential equation

$$dr(t) = \underbrace{\kappa(\theta - r(t))dt}_{\text{drift}} + \underbrace{\sigma dW(t)}_{\text{diffusion}}, \tag{2.2.6}$$

where $W(t)$ is the standard Brownian motion and $\bar{r} = \theta$; $t \in [0, T]$. To solve this equation we can start from a simple equation without noise, that is, we take out only the drift term and denote it by

$$y(t) = \kappa(\theta - r(t)) \tag{2.2.7}$$

The above equation is equation (2.2.6) with $\sigma = 0$. This equation is an ordinary differential equation and is linear in $y(t)$; therefore, its general solution is

$$y(t) = Ce^{-\kappa t} + \theta$$

where C is an arbitrary constant. In order to solve the main equation (2.2.6), let us first introduce Itô's formula, which we are going to use in our next step. The formula states that if $x(t)$ is an Itô diffusion process satisfying

$$dx(t) = a(x(t), t)dt + b(x(t), t)dz(t)$$

where $z(t)$ is a Wiener process, then, for any twice continuously differentiable function G, the process $y(t) = G(x(t), t)$ is again an Itô process and it solves the SDE:

$$dG(x(t),t) = \left(\frac{\partial G}{\partial t} + \frac{\partial G}{\partial x}a(x(t),t) + \frac{1}{2}\frac{\partial^2 G}{\partial x^2}b(x(t),t)^2\right)dt + \frac{\partial G}{\partial x}b(x(t),t)dz(t),$$

In the Vasicek case, we can solve the SDE explicitly by using a change of variable, that is, we change $y(t)$ to $f(r(t), t)$ where

$$f(r(t),t) = e^{\kappa t}y(t) = e^{\kappa t}(\kappa(\theta - r(t))).$$

This gives a function that depends on the stochastic process $r(t)$. To apply Itô's formula, we need to compute the partial derivatives, with respect to t, r and r^2,

$$\frac{\partial f(r,t)}{\partial t} = \kappa e^{\kappa t}(\kappa(\theta - r)), \tag{2.2.8}$$

$$\frac{\partial f(r,t)}{\partial r} = -\kappa e^{\kappa t}, \tag{2.2.9}$$

$$\frac{\partial^2 f(r,t)}{\partial r^2} = 0 \tag{2.2.10}$$

It is obvious that because of the factor $e^{\kappa t}$ we still have $y(t)$ in derivatives. Now by Itô's formula we have

$$df(r(t),t) = \left(\frac{\partial f(r(t),t)}{\partial t} + \frac{\partial f(r(t),t)}{\partial r}y(t) + \frac{1}{2}\frac{\partial^2 f(r(t),t)}{\partial r^2}\sigma^2\right)dt + \sigma\frac{\partial f(r(t),t)}{\partial r}dW(t)$$

or, equivalently,

$$\begin{aligned} f(r(t),t) - f(r(0),0) &= \int_0^t \frac{\partial f(r(s),s)}{\partial s}ds + \int_0^t \frac{\partial f(r(s),s)}{\partial r}y(s)\,ds + \int_0^t \frac{\partial f(r(s),s)}{\partial r}\sigma\,dW(s) \\ &= \int_0^t \kappa e^{\kappa s}y(s)\,ds - \int_0^t \kappa e^{\kappa s}y(s)\,ds - \int_0^t \kappa e^{\kappa s}\sigma\,dW(s) \\ &= -\int_0^t \kappa e^{\kappa s}\sigma\,dW(s). \end{aligned}$$

Replacing $f(r(t), t)$ by its value gives

$$e^{\kappa t}y(t) - e^{\kappa 0}y(0) = -\int_0^t \kappa e^{\kappa s}\sigma\,dW(s) \tag{2.2.11}$$

or

$$e^{\kappa t}(\theta - r(t)) - \kappa(\theta - r(0)) = -\int_0^t \kappa e^{\kappa s}\sigma\,dW(s).$$

Hence,

$$-\kappa e^{\kappa t} r(t) = -\kappa\theta e^{\kappa t} + \kappa\theta - \kappa r(0) - \int_0^t \kappa e^{\kappa s}\sigma\, dW(s)$$

and the solution for $r(t)$ is

$$r(t) = e^{-\kappa t} r(0) + \theta\left(1 - e^{-\kappa t}\right) + \sigma \int_0^t e^{-\kappa(t-s)}\, dW(s).$$

We have a recursive expression for $r(t)$ in terms of its previous value. Now if we subdivide the interval $[0, T]$ into n subintervals and let $t_i = i\frac{T}{n}$ for $i = 0, \ldots, n$, we can denote each time-step as $\Delta t = t_i - t_{i-1}$; $i = 1, \ldots, n$. Therefore, for $r(t_1), \ldots, r(t_n)$, where $0 = t_0 < t_1 < t_2 < \cdots < t_n = T$ we can write

$$r(t_i) = e^{-\kappa} r(t_{i-1}) + \theta\left(1 - e^{-\kappa\Delta t}\right) + \sigma \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i - s)}\, dW(s).$$

We can define

$$\varepsilon(t_i) = \sigma \int_{t_{i-1}}^{t_i} e^{-\kappa(t_i - u)}\, dW(u)$$

so the recursive expression for $r(t_i)$ can be written as follows,

$$r(t_i) = e^{-\kappa\Delta t} r(t_{i-1}) + \theta[1 - e^{-\kappa\Delta t}] + \varepsilon(t_i) \tag{2.2.12}$$

where $\varepsilon(t_i)$ is in fact a Gaussian random variable. According to the properties of the Wiener process given in Chapter I, we can conclude that

$$E[\varepsilon(t_i) \mid \varepsilon(t_{i-1})] = 0$$

and the variance is calculated as follows,

$$\begin{aligned} \mathrm{Var}\left(\varepsilon(t_i) \mid \varepsilon(t_{i-1})\right) &= E\left[\varepsilon^2(t_i) \mid \varepsilon(t_{i-1})\right] \\ &= E\left[\sigma^2 \int_{t_{i-1}}^{t_i} e^{-2\kappa(t_i - u)}\, du\right] \\ &= \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta t}\right). \end{aligned}$$

In general, we can write

$$\varepsilon(t_i) \mid \varepsilon(t_{i-1}) \sim N\left(0, \frac{\sigma^2}{2\kappa}(1 - e^{-2\kappa\Delta t})\right)$$

In other words, we have the first two moments of the Gaussian transition density of $r(t_i)$ given $r(t_{i-1})$, specifically:

$$\begin{aligned} E[r(t_i) \mid r(t_{i-1})] &= \theta\left(1-e^{-\kappa\Delta t}\right)+e^{-\kappa\Delta t}r_{t_{i-1}} \\ \mathrm{Var}(r(t_i) \mid r(t_{i-1})) &= \frac{\sigma^2}{2\kappa}\left(1-e^{-2\kappa\Delta t}\right) \end{aligned}$$

thus,

$$r(t_i) \mid r(t_{i-1}) \sim N\left(\theta[1-e^{-\kappa\Delta t}]+e^{-\kappa\Delta t}r_{t_{i-1}}, \frac{\sigma^2}{2\kappa}\left(1-e^{-2\kappa\Delta t}\right)\right).$$

If $p\left(r(t_i) \mid r(t_{i-1});\phi\right)$ denotes the density function of the previous normal distribution, and $\phi=(\kappa,\theta,\sigma^2)$ is the set of unknown parameters, then the (conditional) likelihood function is

$$L\left(\phi; r(t_0),\ldots,r(t_n)\right)=\prod_{i=1}^{n} p\left(r(t_i) \mid r(t_{i-1});\phi\right).$$

By taking logarithms of both sides of the above equation,

$$\ell\left(\phi; r(t_0),\ldots,r(t_n)\right)=\sum_{i=1}^{n}\log p\left(r(t_i) \mid r(t_{i-1});\phi\right) \tag{2.2.13}$$

and the maximum likelihood of estimator ϕ is

$$\hat{\phi}=\arg\max_{\phi}\ell\left(\phi; r(t_0),\ldots,r(t_n)\right).$$

Instead of trying to write down the explicit formula of $\hat{\phi}$, maximization can be performed numerically. We use the **fminsearch** function in MATLAB. We implement this method in section 2.4.

2.3 MLE for the square root process

The SDE described earlier in section 1.6.1 of Chapter I was introduced by Cox, Ingersoll and Ross (1985a) to represent the dynamics of the short-rate interest rate :

$$dr(t)=\kappa(\theta-r(t))dt+\sigma\sqrt{r(t)}dW(t) \tag{2.3.14}$$

with $r(0)=r_0$. The drift is an affine function of $r(t)$ as for the Vasicek process. However, the diffusion coefficient is the square root of $r(t)$ and for this reason the CIR model is often called the *square root process*. Unlike (2.2.6), the equation (2.3.14) can not be solved explicitly

using Itô's formula. We have to find another way to get the transition density. The *Laplace transform* $\widetilde{f}(\lambda; r(t_{i-1}), \phi)$ of the transition density is useful for that purpose:

$$\widetilde{f}(\lambda; r(t_{i-1}), \phi) = E\left[e^{-\lambda r(t_i)} \mid r(t_{i-1}), \phi\right].$$

This is because, for the CIR process, the logarithm of $\widetilde{f}$ is an affine function of $r(t_{i-1})$ (hence the name of the class of processes to which it belongs). Let

$$L = \frac{\sigma^2}{4\kappa}(1 - e^{-\kappa \Delta t})$$

where $\Delta t = t_i - t_{i-1}$; then the conditional Laplace transform of $x(t_i) = r(t_i)/L$ is given by

$$E\left[e^{-\lambda x(t)} \mid r(t_{i-1})\right] = \frac{1}{(2\lambda + 1)^{\frac{2\theta}{\kappa\sigma^2}}} \exp\left\{-\frac{\lambda}{2\lambda + 1}\frac{4r(t_{i-1})\kappa}{\sigma^2(e^{\kappa \Delta t} - 1)}\right\}$$

(see Lamberton and Lapeyre (1991), page 121, for a proof). It follows that the conditional distribution of $x(t_i) = r(t_i)/L$, given $r(t_{i-1})$, is a non-central χ^2 distribution with parameters

$$\nu = \frac{4\kappa\theta}{\sigma^2},$$

$$\delta = \frac{4\kappa r(t_{i-1})}{\sigma^2(e^{\kappa \Delta t} - 1)},$$

where ν denotes the degrees of freedom and δ the non-centrality parameter. The non-central $\chi^2(\nu, \delta)$ distribution has the following density function:

$$f(x) = \frac{1}{2}e^{-\frac{(x+\delta)}{2}}\left(\frac{x}{\delta}\right)^{\left(\frac{\nu}{4} - \frac{1}{2}\right)} I_{\frac{\nu}{2}-1}(\sqrt{\delta x}),$$

where I_a is a modified Bessel function of the first kind given by

$$I_a(x) = \left(\frac{\lambda}{2}\right)^a \sum_{j=1}^{\infty} \frac{\left(\frac{x^2}{4}\right)^j}{j!\Gamma(a + j + 1)}.$$

This result is valid when ν and δ are positive or, equivalently, if $\kappa, \theta > 0$. The SDE (2.3.14) has a unique positive solution when $2\kappa\theta \geq \sigma^2$ but the rather complicated form of the transition density makes it impossible to find explicit formulas for the MLEs of the parameters κ, θ, and σ^2. One can still compute the log-likelihood function numerically. The (conditional) log-likelihood for the interest rate with $n + 1$ observations is

$$\ell(\phi; r(t_0), r(t_1), \ldots, r(t_n)) = \sum_{i=1}^{n} \log p(r(t_i), r(t_{i-1}); \phi)$$

$$\sum_{i=1}^{n} \log\left(\frac{\chi^2(r(t_i)/L, \nu, \delta)}{L}\right)$$

and the MLE $\widehat{\phi}$ is

$$\hat{\phi} = \arg\max_{\phi} \ell(\phi; r(t_0), r(t_1), \ldots, r(t_n)).$$

The first two moments of the non-central χ^2 distribution are

$$\mu_1 = \theta\left(1 - e^{-\kappa\Delta t}\right) + e^{\kappa\Delta t} r(t_{i-1})$$

$$\mu_2 = \frac{\theta\sigma^2}{2\kappa}\left(1 - e^{-\kappa\Delta t}\right)^2 + \frac{\sigma^2}{\kappa}\left(e^{-\kappa\Delta t} - e^{-2\kappa\Delta t}\right) r(t_{i-1}).$$

Ball and Torous (1996) showed that, over small time intervals, the transition density can be reasonably approximated by a normal density. Therefore, we can use the first two moments of the non-central χ^2 distribution and assume, alternatively, that

$$r(t_i) = e^{-\kappa\Delta t} r(t_{i-1}) + \theta\left(1 - e^{-\kappa\Delta t}\right) + \varepsilon(t_i), \tag{2.3.15}$$

with

$$\varepsilon(t_i) \sim N\left(0, \frac{\theta\sigma^2}{2\kappa}\left(1 - e^{-\kappa\Delta t}\right)^2 + \frac{\sigma^2}{\kappa}\left(e^{-\kappa\Delta t} - e^{-2\kappa\Delta t}\right) r(t_{i-1})\right).$$

We shall use this approximation in Chapter IV.

2.4 Numerical examples

In this section we apply our estimation techniques for the Vasicek and CIR processes. To this end we generate different sets of data using (i) Vasicek and (ii) CIR model with different parameters and examine these data to see how effective is the maximum likelihood technique in estimating the parameters.

First, we consider the Vasicek model, starting from an arbitrary set of parameters for each sample path where κ, $\theta > 0$ and σ is the unconditional instantaneous volatility of the process. That is $\phi_i = (\kappa_i, \theta_i, \sigma_i)$, then we will have, $r_i(t_1), \ldots, r_i(t_n)$ (in our case $i = 1, 2$), where n denotes the number of observations. In our case we use weekly observations ($\Delta = 1/52$) and the simulation is repeated $N = 1000$ times for each set of parameters. Since we use Gaussian density to produce the white noises of the model, sometimes $r_i(t_j)$ may become negative. In such cases the program simulates new data.[1] However, this occurs rarely so we can

[1] This was done so that, in the chapters to come, we could use the same simulated data.

keep the i.i.d. assumption of our data. The percentage occurrence in Vasicek models were, 13.09 (for ϕ_1) and 0.34 (for ϕ_2). For the generated data used in CIR models, these percentages were 0 in both cases. Figure 2.1 shows, for each set of parameters, the first simulated sample path of the Vasicek model, with $n = 1000$ observations.

A similar procedure is performed to generate two sets of trajectories for the CIR processes. Moreover, in the CIR case we should also apply the condition $2\kappa\theta \geq \sigma^2$. Figure 2.2 shows, for each set of parameters, the first simulated sample path of the CIR model, with $n = 1000$ observations.

The tables and figures in the next pages summarize the results of the simulation exercise for the Vasicek and CIR model using the maximum likelihood method. The plots for all the results are outlined in Appendix A. The convergence of the estimated parameters to the true values as n increases shows that the results are asymptotically unbiased.

Tables 2.1, 2.2, and 2.3 summarize the results of the simulation implementation with different numbers of observations. These tables include the true values, mean estimates over $N = 1000$ simulations and the associated standard deviations of the estimates. The mean provides the information about any bias in the estimation technique, while the standard deviation is useful in assessing the accuracy of the technique. Also we should mention that we observed outliers in results, especially when the number of observations was small. We used a *box-and-whisker* method to identify these outliers and ignore them afterwards in the computations.

Overall, based on the results, we might conclude that maximum likelihood method is not a very successful method for determining the parameters. In the two next chapters we will study two different techniques.

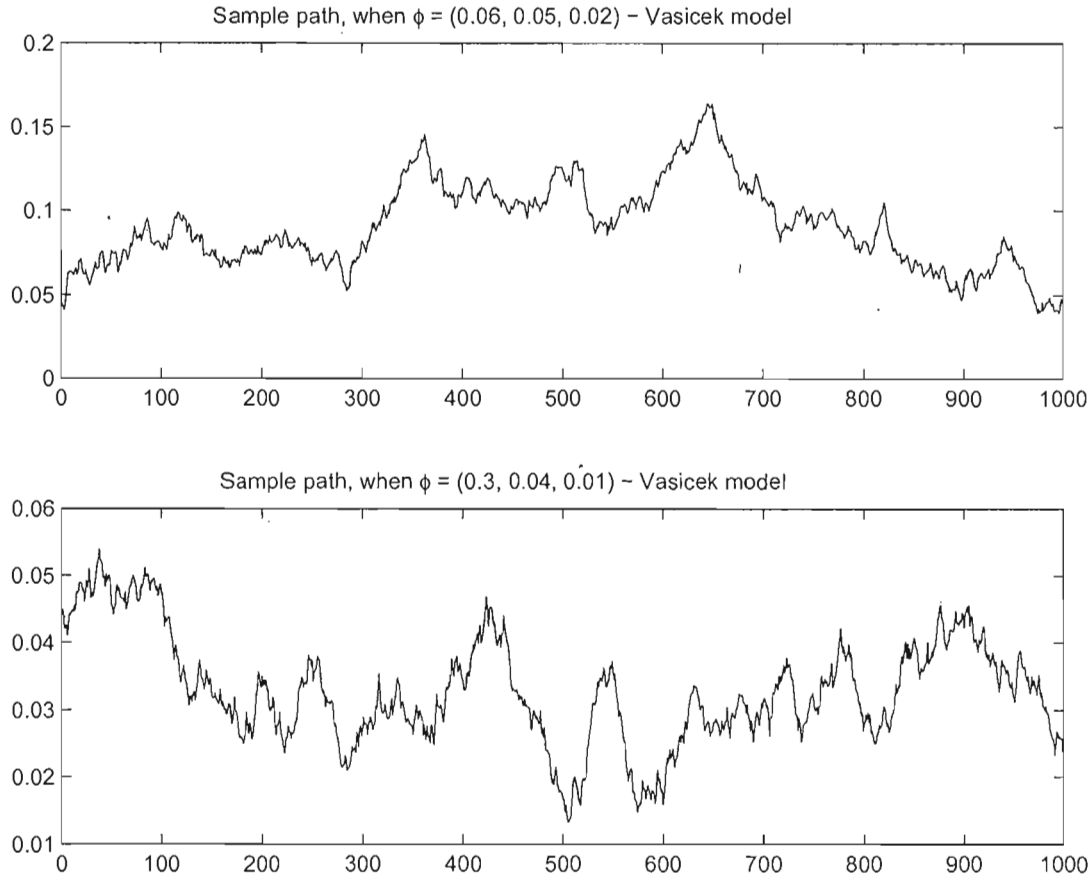

Figure 2.1 Generated sample paths for the Vasicek process

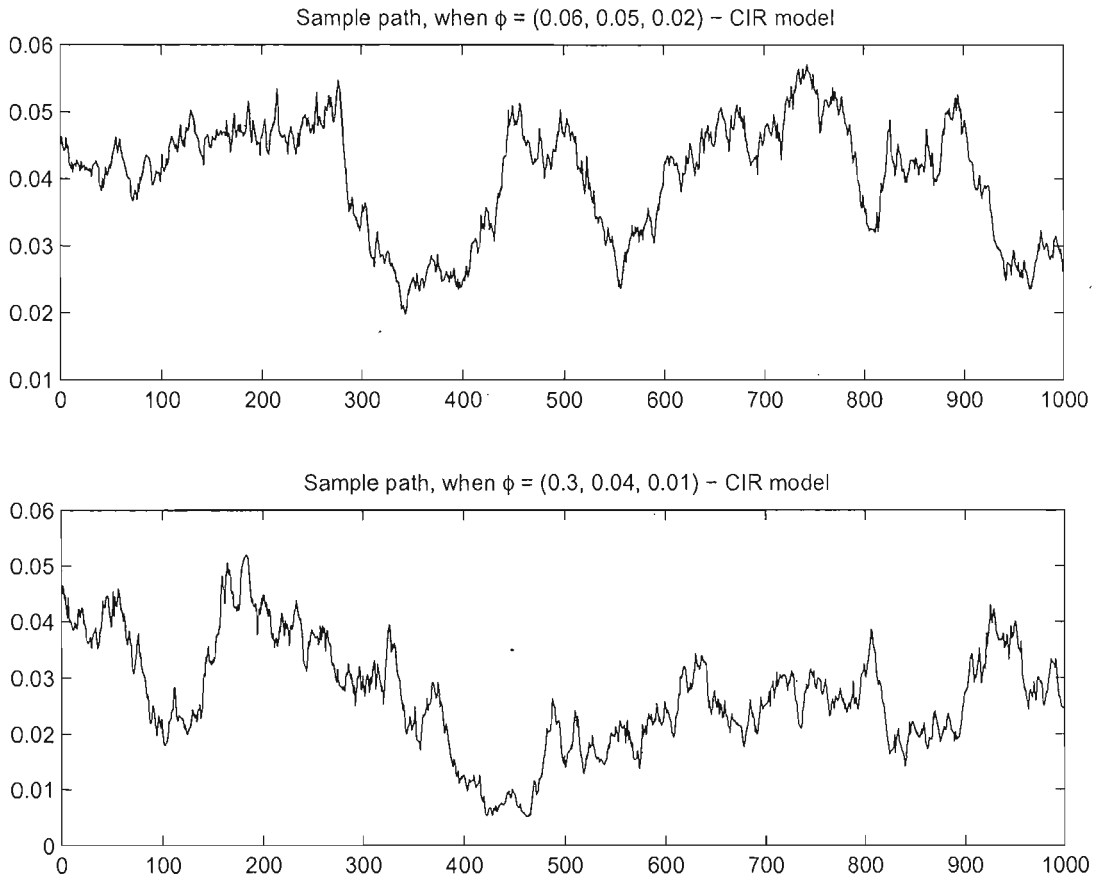

Figure 2.2 Generated sample paths for Cox-Ingersoll-Ross process

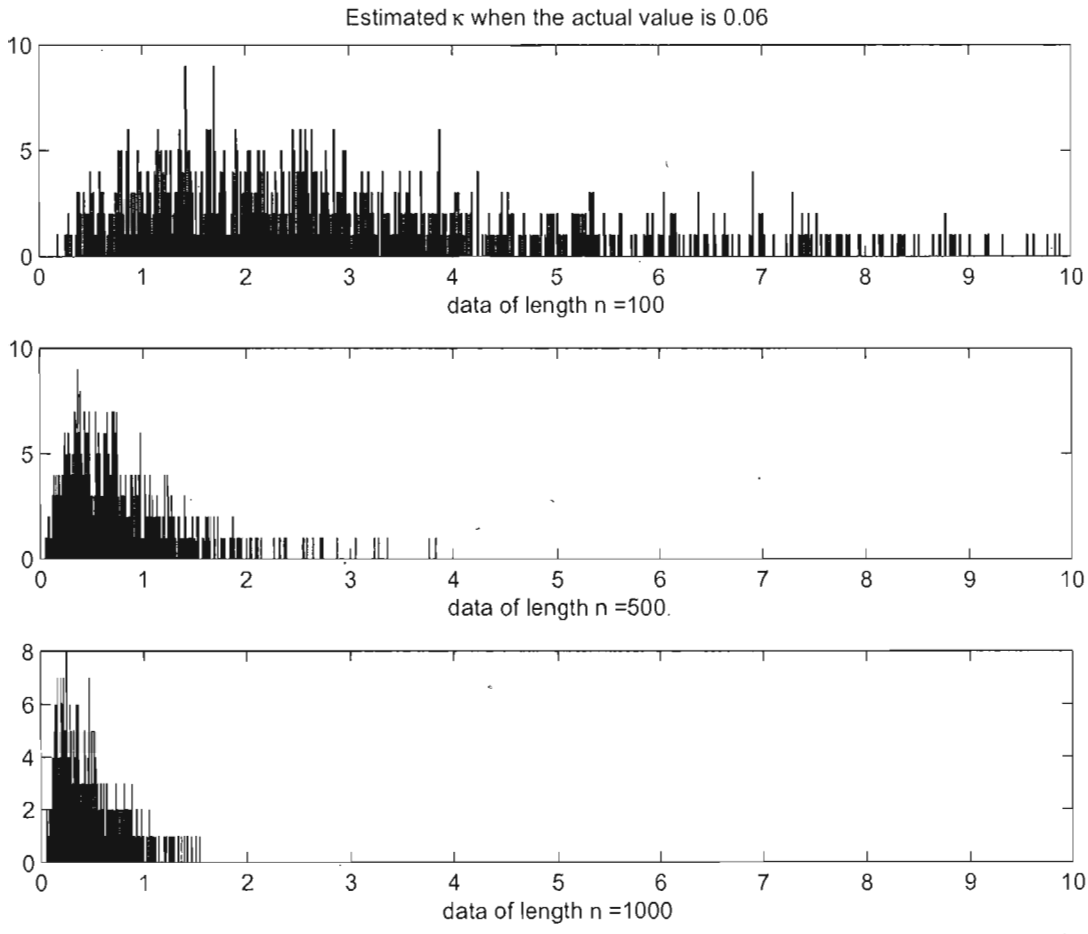

Figure 2.3 Empirical distribution of $\hat{\kappa}$ for the Vasicek model with parameters ϕ_1

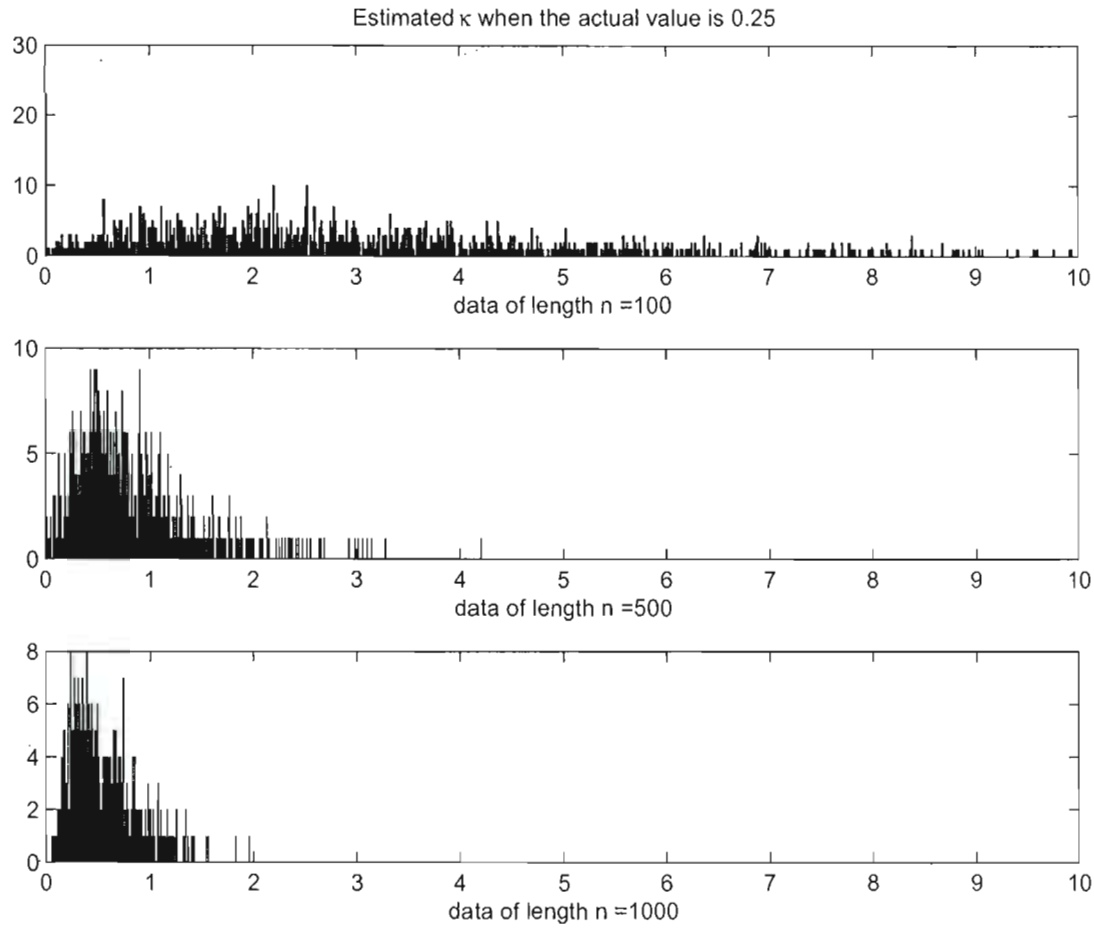

Figure 2.4 Empirical distribution of $\hat{\kappa}$ for the CIR model with parameters ϕ_1

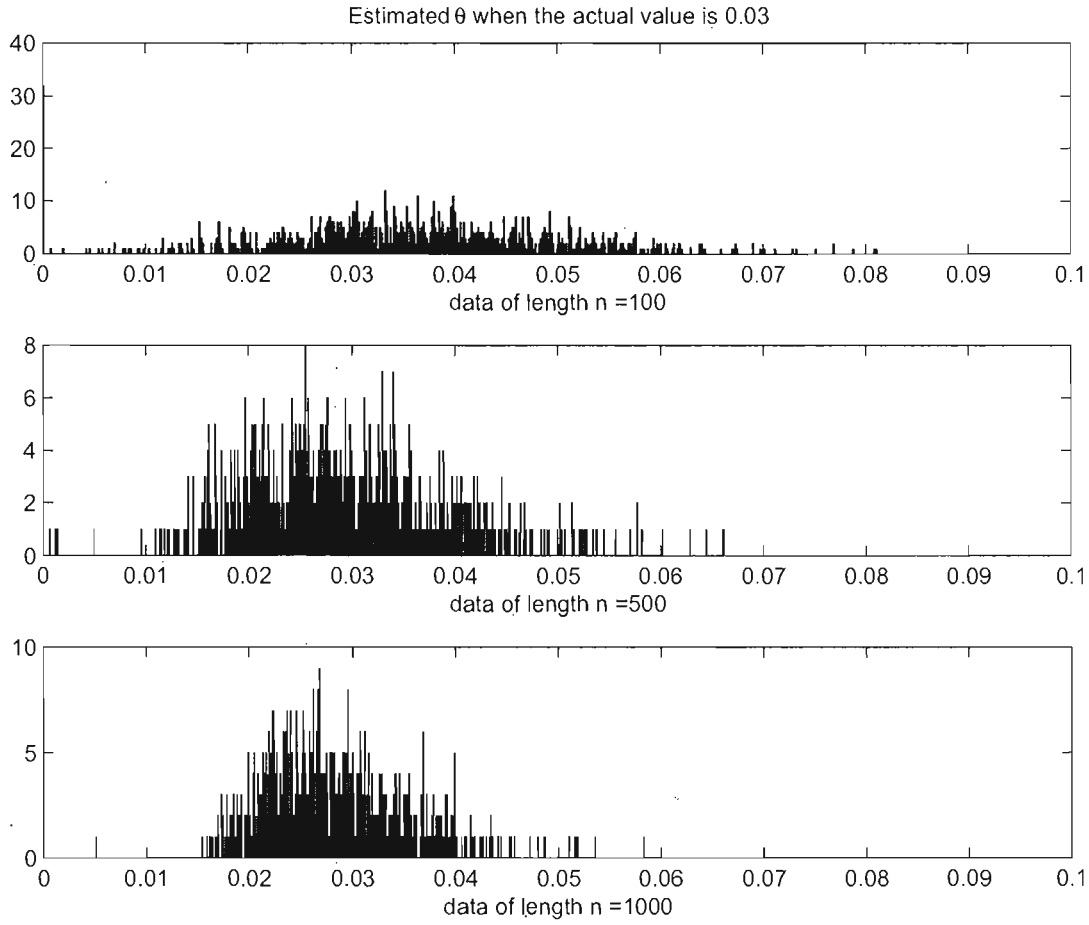

Figure 2.5 Empirical distribution of $\hat{\theta}$ for the CIR model with parameters ϕ_2

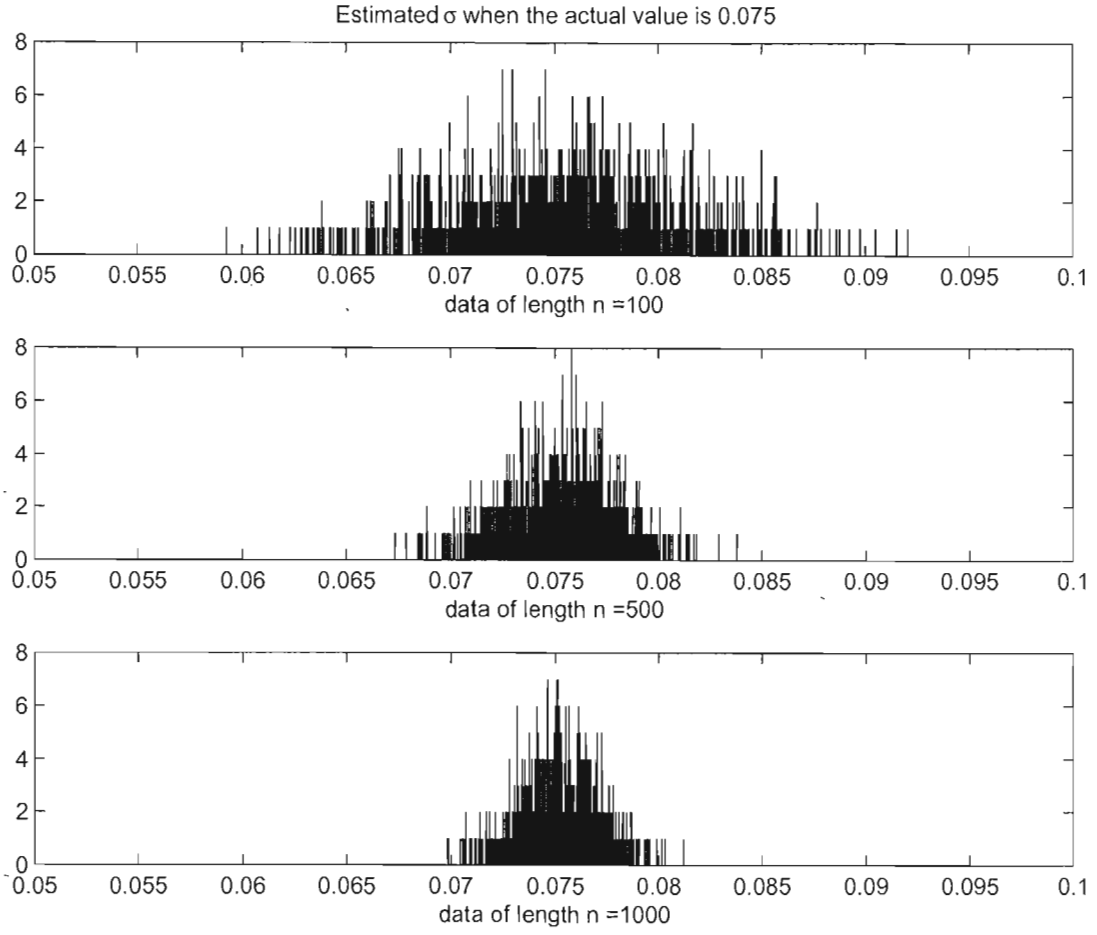

Figure 2.6 Empirical distribution of $\hat{\sigma}$ for the CIR model with parameters ϕ_2

Table 2.1 A simulation analysis of the MLEs with $n = 100$ observations

	Vasicek			CIR		
Parameters	Actual values	Mean estimates	Standard deviations	Actual values	Mean estimates	Standard deviations
κ_1	0.060	3.182	2.337	0.250	3.126	2.360
κ_2	0.300	3.409	2.492	0.450	3.299	2.414
θ_1	0.050	0.052	0.020	0.050	0.047	0.019
θ_2	0.040	0.044	0.009	0.030	0.036	0.016
σ_1	0.020	0.020	0.000	0.050	0.050	0.004
σ_2	0.010	0.010	0.000	0.075	0.075	0.005

Table 2.2 A simulation analysis of the MLEs with $n = 500$ observations

	Vasicek			CIR		
Parameters	Actual values	Mean estimates	Standard deviations	Actual values	Mean estimates	Standard deviations
κ_1	0.060	0.746	0.537	0.250	0.803	0.544
κ_2	0.300	0.860	0.540	0.450	0.933	0.508
θ_1	0.050	0.077	0.029	0.050	0.049	0.023
θ_2	0.040	0.040	0.008	0.030	0.029	0.009
σ_1	0.020	0.019	0.000	0.050	0.050	0.001
σ_2	0.010	0.010	0.000	0.075	0.075	0.002

Table 2.3 A simulation analysis of the MLEs with $n = 1000$ observations

	Vasicek			CIR		
Parameters	Actual values	Mean estimates	Standard deviations	Actual values	Mean estimates	Standard deviations
κ_1	0.060	4.271	0.271	0.250	0.516	0.287
κ_2	0.300	0.055	0.282	0.450	0.676	0.296
θ_1	0.050	0.082	0.027	0.050	0.048	0.016
θ_2	0.040	0.041	0.006	0.030	0.028	0.006
σ_1	0.020	0.020	0.000	0.050	0.050	0.001
σ_2	0.010	0.010	0.000	0.075	0.075	0.002

CHAPTER III

ZERO-COUPON DATA

3.1 Zero-coupon prices

Very often in finance we need to estimate the unknown parameter sets of state variables, when the random variates specified in the model are not directly observable. For example in the financial markets we can only see the prices of interest rate instruments, while we would need to know the interest rates (short rates in discrete time) for performing the estimation.

A different approach for the estimation of the SDE-type models has been introduced in the financial literature, by finding a function which defines the relation between the observable data and the non-observable values, and then fitting this function to observable values (bond prices for example) at a specific period of time.

In the literature, as discussed in Chapter II, the term structure of the interest rates, is assumed to follow a diffusion process; Vasicek and CIR processes are examples of this kind. However, these rates are not observable, but we can use the affine term structure to relate the observable prices to unobserved state variables. In other words, we are interested in a model that is numerically and empirically tractable, but sometimes the random variables of the model are not directly observable. For example, in our case the generated $r(t)$ of the previous chapter doesn't exist in reality and can not be considered an instantaneous interest rate. However, the available data (such as bond prices) are often the result of some transformation of these unobservable rates.

In this Chapter, we develop a maximum likelihood method for dealing with parameter estimation of zero-coupon interest rates which is an example of a problem of this type. The first step is to establish some relation between interest rates and the observable prices of the bonds. We denote the value (or price) of a risk-free-zero-coupon bond as the function $P(t, T)$ where t refers to the current time, while the T represents the coupon's maturity date; therefore it is obvious that $t_i < T$ for $i = 1, \ldots, n$ where n is the number of periods left to the maturity date. The zero-coupon bond pays one unit of account to the holder at maturity date T ($t_i = T, i = n$), in other word $P(T, T) = 1$.

Cox, Ingersoll and Ross (1985b) suggest the following relation between the price of the zero coupon and the continuous associated data, spot rate of interest, which is denoted by $z(t, T)$:

$$P(t, T) = \exp(-(T - t)z(T, t))$$

that is,

$$z(t, T) = -\frac{\log P(t, T)}{T - t}.$$

Thus, $z(t, T)$ can be regarded as a risk-free rate of interest in a fixed period of time $T - t$. The *short rate* is the spot interest rate with instantaneous maturity, i.e.

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

and from the above equation we have

$$\begin{aligned} r(t) = \lim_{T \to t} z(t, T) &= \lim_{T \to t} \left\{ -\frac{\log P(t, T)}{T - t} \right\} \\ &= -\left[\frac{\partial \log P(t, T)}{\partial T} \right]_{T=t} \\ &= -\left[\frac{1}{P(t, T)} \frac{\partial P(t, T)}{\partial T} \right]_{T=t} \\ &= -\frac{\partial P(t, t)}{\partial T} \end{aligned}$$

The affine term structure model is a key procedure for calculating the zero-coupon rate from a given time to maturity, $p(t, T)$, by having only the value of the instantaneous rate

on interest, $r(t)$. We consider a class of models, called exponential affine, where the prices $\{P(t,T), t \in [0,T]\}$ are of the following general form:

$$P(t,T) = e^{A(t,T)-B(t,T)r(t)} \tag{3.1.1}$$

with deterministic functions $A(t,T)$ and $B(t,T)$. Since $P(T,T) = 1$, we should have the following boundary conditions:

$$A(T,T) = 0, \quad B(T,T) = 0.$$

In the next sections we will describe the single-factor development of the above affine term structure model, and then generalize it to higher dimensions (multi-factor models).

3.2 Single-factor models

Vasicek (1977) assumes that $r(t)$ follows the SDE (2.2.6), and uses this process and the assumption of a constant market price risk, λ, to derive a bond pricing model like (3.1.1) for $\tau = T - t$, where,

$$B(t,T) = B(\tau) = \frac{1}{\kappa}(1 - e^{-\kappa\tau}),$$

$$A(t,T) = A(\tau) = \frac{\gamma(B(\tau)-\tau)}{\kappa^2} - \frac{\sigma^2 B^2(\tau)}{4\kappa}, \tag{3.2.2}$$

with

$$\gamma = \kappa^2(\theta - \frac{\sigma\lambda}{\kappa}) - \frac{\sigma^2}{2}.$$

Let us denote $\phi = (\kappa, \theta, \sigma)$ the vector of unknown parameters, $\tau = T - t$, the time to maturity ($\tau = T - t$ is usually considered as a weekly, monthly or yearly point of observation).

Many authors use the Vasicek model in their pricing models, although there are many others who prefer to work with the model suggested by Cox, Ingersoll and Ross (1985b). The exponential affine formula (3.1.1) is still valid with A and B given by

$$B(\tau) = \frac{2(e^{\gamma\tau} - 1)}{(\gamma + \kappa + \lambda)(e^{\gamma\tau} - 1) + 2\gamma},$$

$$A(\tau) = \ln\left(\frac{2\gamma e^{\frac{(\gamma+\kappa+\lambda)\tau}{2}}}{(\gamma + \kappa + \lambda)(e^{\gamma\tau} - 1) + 2\gamma}\right)^{\frac{2\kappa\theta}{\sigma^2}}, \tag{3.2.3}$$

and

$$\gamma = \sqrt{(\kappa + \lambda)^2 + 2\sigma^2}.$$

The complete calculation for finding A and B in the single factor model for both Vasicek and CIR processes is done in Bolder (2001).

3.3 Multi-factor models

In these models one assumes that the short rate is in fact a linear combination of N correlated state variable (factors) which we denote by $y_1, y_2, \ldots, y_N$. Therefore, we have the following equation,

$$r(t) = \sum_{i=1}^{N} y_i(t),$$

and the associated price function is,

$$\begin{aligned} P(t,T) &= P(t,T;y_1,\ldots,y_N) \\ &= \exp\left\{ A(t,T) - \sum_{i=1}^{N} B_i(t,T)y_i(t) \right\}. \end{aligned}$$

According to Bolder (2001), we have the following solution for the Vasicek model with:

$$B_i(\tau) = \frac{1}{\kappa_i}(1 - e^{-\kappa_i \tau}),$$

$$A(\tau) = \sum_{i=1}^{N} \frac{\gamma(B_i(\tau) - \tau)}{\kappa_i^2} - \frac{\sigma_i^2 B_i^2(\tau)}{4\kappa_i} + \sum_{i,j\,;i\neq j} \frac{\sigma_{ij}}{2\kappa_i\kappa_j}(\tau - B_i(\tau) - B_j(\tau) + \frac{1}{\kappa_i + \kappa_j}(1 - e^{-(\kappa_i+\kappa_j)(\tau)})) \tag{3.3.4}$$

and

$$\gamma_i = \kappa_i^2(\theta_i - \frac{\sigma_i\lambda_i}{\kappa_i}) - \frac{\sigma_i^2}{2}.$$

By comparing the above equation with the single-factor case (3.2.2), we realize that, in addition to replacing the right-hand side in (3.2.2) with a sum of N terms, we also have an extra covariance term.

However, in the CIR multi-factor model, things get more difficult and the analytical solution exists only when the Ricatti equation arising from the partial differential equation (the

so-called *term structure equation*) can be reduced to independent one-dimensional equations, which implies that the underlying Brownian motions, $W_1, \ldots, W_N$ are independent. In a similar way, the solution to the N-factor CIR partial differential equation has the form

$$\begin{aligned} P(t,T) &= P(t,T;y_1,\ldots,y_n) \\ &= \exp\left\{\sum_{i=1}^{N} A_i(t,T) - B_i(t,T)y_i(t)\right\} \end{aligned} \tag{3.3.5}$$

where

$$dy_1(t) = \kappa_1(\theta_1 - y_1(t))dt + \sigma_1\sqrt{y_1(t)}dW_1(t)$$

$$\vdots$$

$$dy_N(t) = \kappa_N(\theta_N - y_N(t))dt + \sigma_N\sqrt{y_N(t)}dW_N(t),$$

and the Brownian motions are independent. Again, the function $A_i(t,T)$ and $B_i(t,T)$ are of the same form as in the one-dimensional case, since we assume that the state variables are not correlated (in Chapter I, we also assumed that $W_1, \ldots, W_n$ were independent); this gives

$$\begin{aligned} B_i(\tau) &= \frac{2(e^{\gamma_i\tau}-1)}{(\gamma_i+\kappa_i+\lambda_i)(e^{\gamma_i\tau}-1)+2\gamma_i}, \\ A_i(\tau) &= \ln\left(\frac{2\gamma_i e^{\frac{(\gamma_i+\kappa_i+\lambda_i)\tau}{2}}}{(\gamma_i+\kappa_i+\lambda_i)(e^{\gamma_i\tau}-1)+2\gamma_i}\right)^{\frac{2\kappa_i\theta_i}{\sigma_i^2}}, \end{aligned} \tag{3.3.6}$$

where

$$\gamma_i = \sqrt{(\kappa_i+\lambda_i)^2+2\sigma_i^2}.$$

In both cases the boundary conditions for $i = 1, \ldots, N$, are defined as

$$A_1(0) = \cdots = A_N(0) = B_1(0) = \cdots = B_N(0) = 0.$$

According to Bolder (2001), the previous derivation of continuous-time affine term structure models exists also for the discrete-time classes.

The theoretical development required to represent the bond prices as an affine function of the underlying state variables of the two specific Vasicek and CIR models is completed and we can introduce techniques to estimate the unknown parameters of the affine term-structure models. In this Chapter we will only describe the maximum likelihood approach, and leave the *Kalman filter* methodology for the next Chapter.

3.4 Maximum likelihood estimation

We view equation (3.1.1), as a transform function which relates short rates to zero-coupon prices:

$$r(t) \mapsto P(t,T) = e^{A(t,T)-B(t,T)r(t)}.$$

For $0 = t_0 < t_1 < \ldots < t_n$ and a fixed set of parameters, we can then derive an inverse relation:

$$\hat{r}(t_i) = \frac{A(t_i,T) - \log P(t_i,T)}{B(t_i,T)}. \tag{3.4.7}$$

To obtain the log-likelihood function for the transformed data, we employ the following classic theorem.

Theorem 2 *If the transformation from X to Y is on an element-by-element basis, i.e.,* $y_i = T_i(x_i;\phi)$ *for all* i*, then*

$$\ell(y_0,\ldots,y_n;\phi) = \ell(\hat{x}_i(\phi), i = 0,\ldots,n;\phi) - \sum_{i=1} n \log\left|\frac{\partial T_i(\hat{x}_i(\phi);\phi)}{\partial x_i}\right|$$

where

$$\hat{x}_i(\phi) = T_i^{-1}(y_i;\phi).$$

The proof of this theorem can be found in Duan (1994). In our specific case, $T(X,\phi)$ represents the pricing function (3.1.1) and T^{-1} will be the equation (3.4.7). Therefore, we can calculate the log-likelihood function for the prices bond,

$$\begin{aligned}\ell(P(t_1,T),\ldots,P(t_n,T);\phi) &= \ell(\hat{r}(t_1),\ldots,\hat{r}(t_n);\phi) - \sum_{i=1}^{n} \log\left|\frac{\partial P(\hat{x}_i(\phi);\phi)}{\partial x_i}\right| \\ &= \ell(\hat{r}(t_1),\ldots,\hat{r}(t_n);\phi) - \sum_{i=1}^{n} \log|B(t_i,T)| \end{aligned} \tag{3.4.8}$$

where $\ell(\hat{r}(t_1),\ldots,\hat{r}(t_n);\phi)$ is in fact the calculated log-likelihood function for the Vasicek (or CIR model) in the previous chapter. The minimization of minus the log-likelihood function (3.4.8) can be achieved by the **fminsearch** function in MATLAB.

3.5 Numerical analysis

In this section, to illustrate our method, we apply the preceding theoretical discussion to the previous Chapter generated data, and apply equation 3.1.1 to collect the observable prices related to these instantaneous unobservable short rates (in reality).

In particular, we compute these values by using weekly observations ($\Delta = 1/52$) over a 20-year time horizon. The generated prices illustration for a 10-year zero-coupon bond ($\tau = 0.5$) using Vasicek and CIR models with the assumption of a constant risk premium $\lambda = 1$ are shown in figures 3.1 and 3.2 respectively.

Then we assume that the only data we have is the set of observed prices and use the inverse equation (3.4.7) to achieve estimation of short rates. By computing $\hat{r}_0, \ldots, \hat{r}_n$ from $P_0, \ldots, P_n$ starting with an initial value $\hat{\phi}_0$ and applying the maximum likelihood method to the log-likelihood function of prices (3.4.8), we can estimate the parameters of our model, the vector $\phi = (\kappa, \theta, \sigma)$. The tables and figures in the following pages summarize the results of the simulation exercise for the Vasicek and CIR models using the maximum likelihood method with zero-coupon prices. The plots of all the results are outlined in Appendix B.

The convergence of the estimators to the true values n increases shows that the results are asymptotically unbiased. Tables 3.1, 3.2, and 3.3, summarize the results of the simulation implementation with different numbers of observations. The estimators in the indirect method tend to the real values but the convergence is slow. Since in this method we are not directly getting the maximum likelihood estimates of the short rates, we were expecting to get these results.

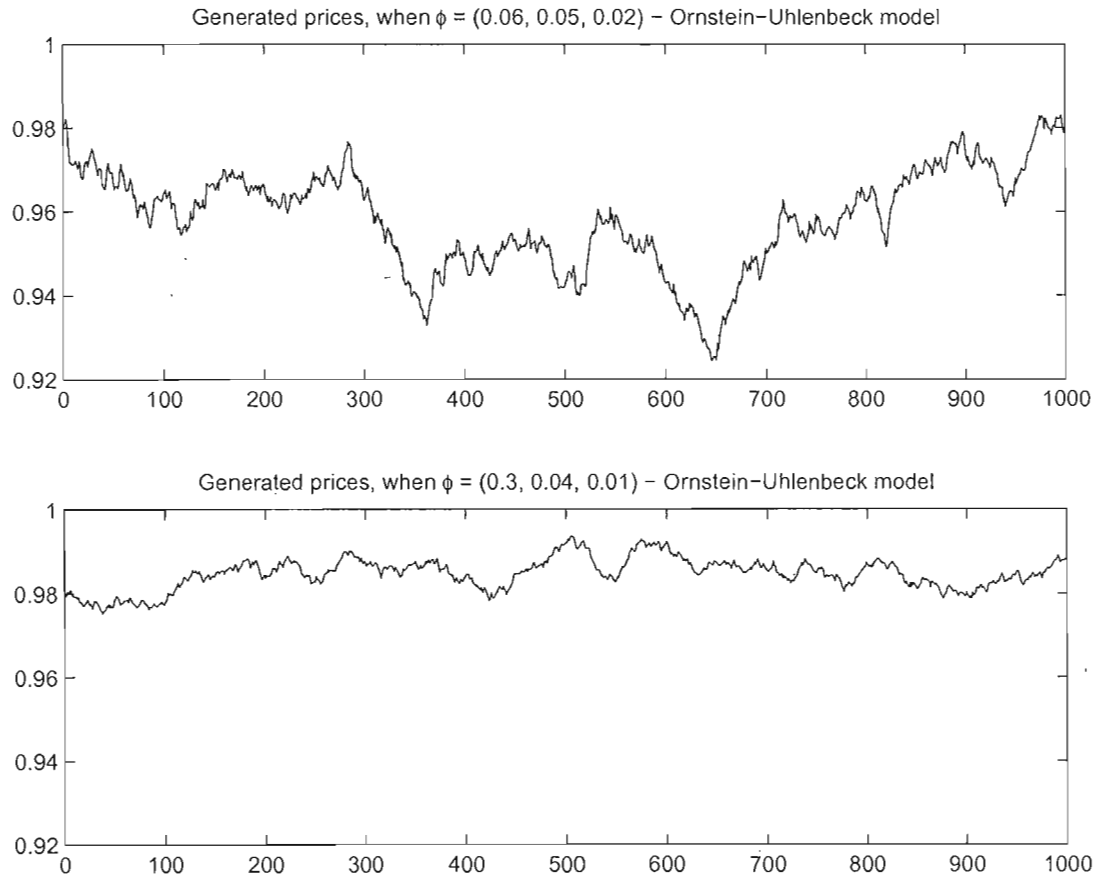

Figure 3.1 Generated prices for the Vasicek model

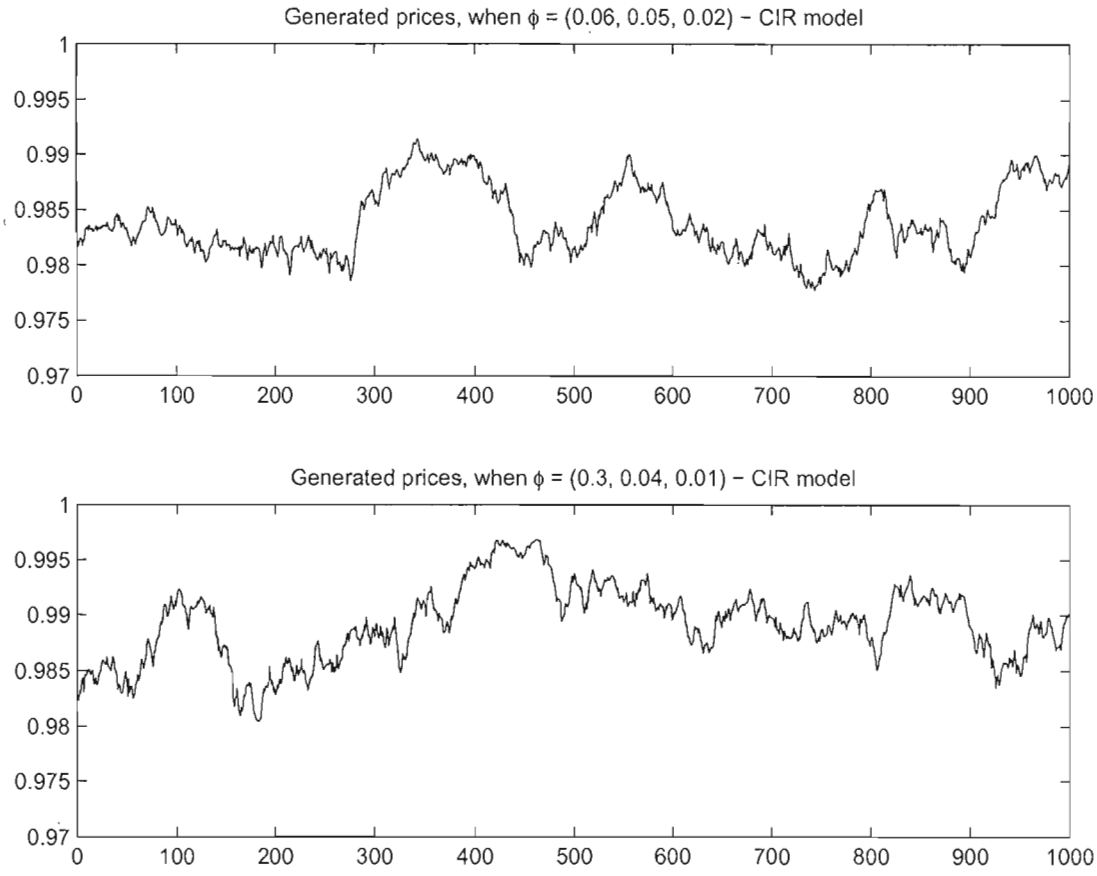

Figure 3.2 Generated prices for the Cox-Ingersoll-Ross model

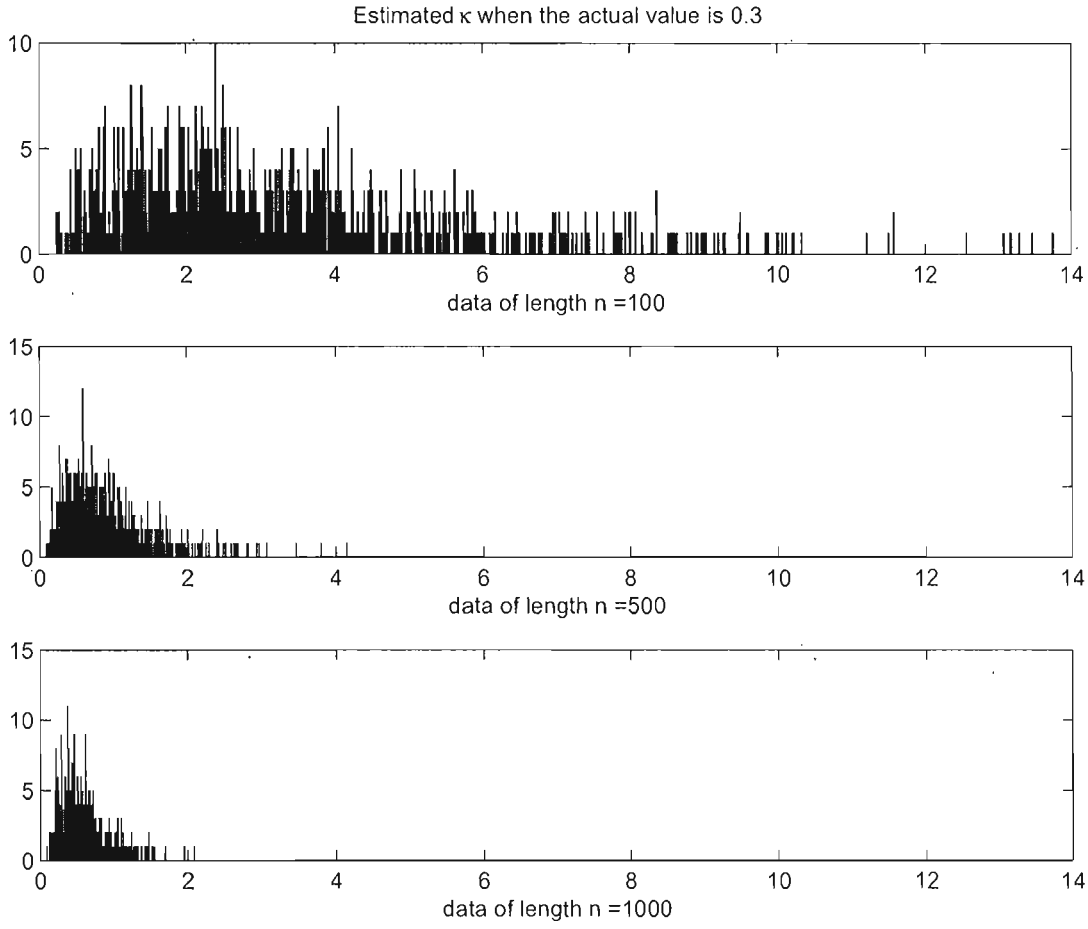

Figure 3.3 Empirical distribution of $\hat{\kappa}$ with zero-coupon data for the Vasicek model with parameters ϕ_2

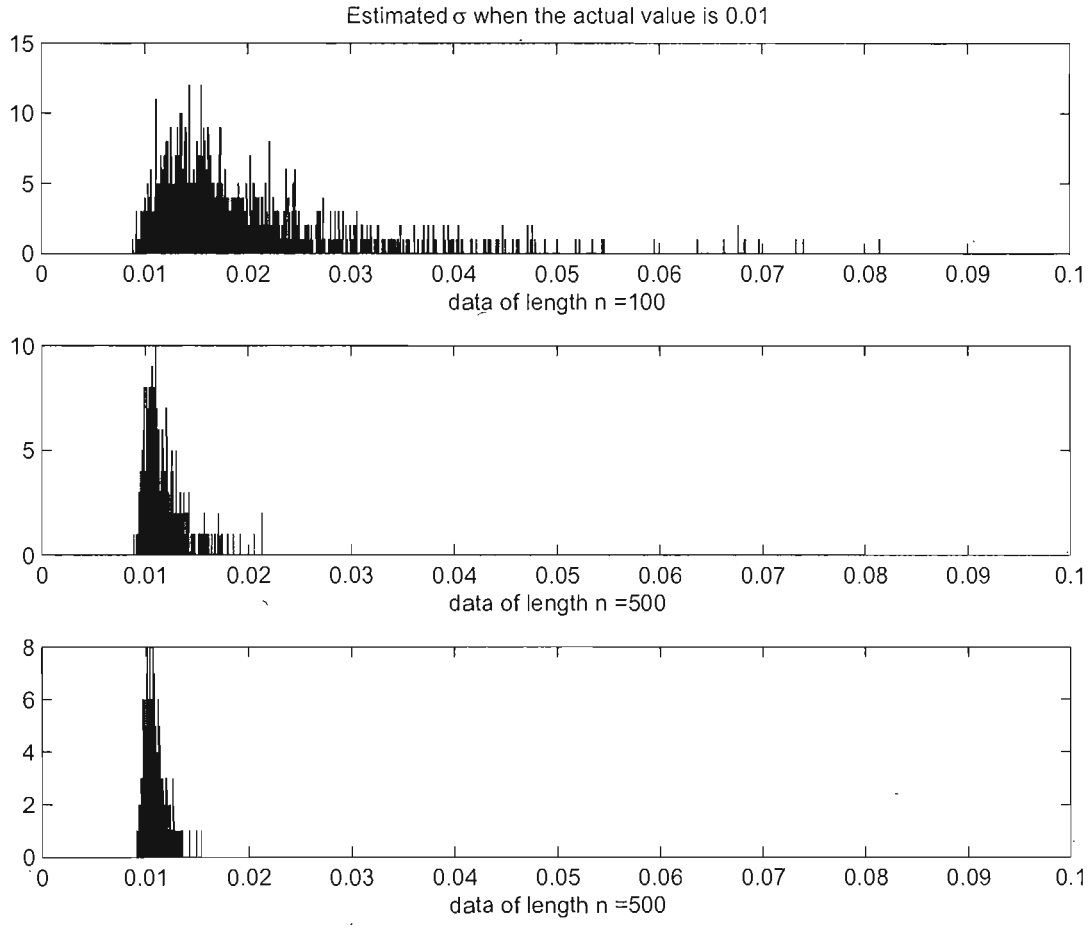

Figure 3.4 Empirical distribution of $\hat{\sigma}$ with zero-coupon data for the Vasicek model with parameters ϕ_2

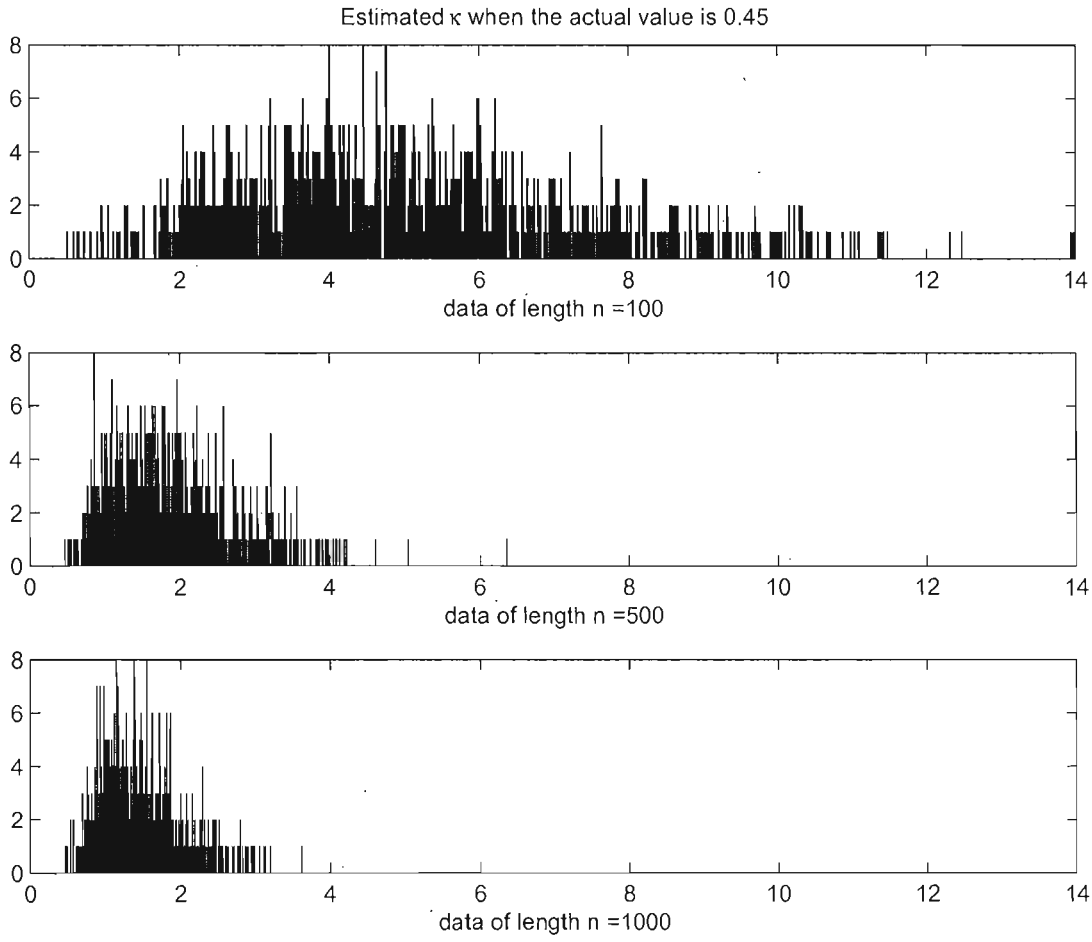

Figure 3.5 Empirical distribution of $\hat{\kappa}$ with zero-coupon data for the CIR model with parameters ϕ_2

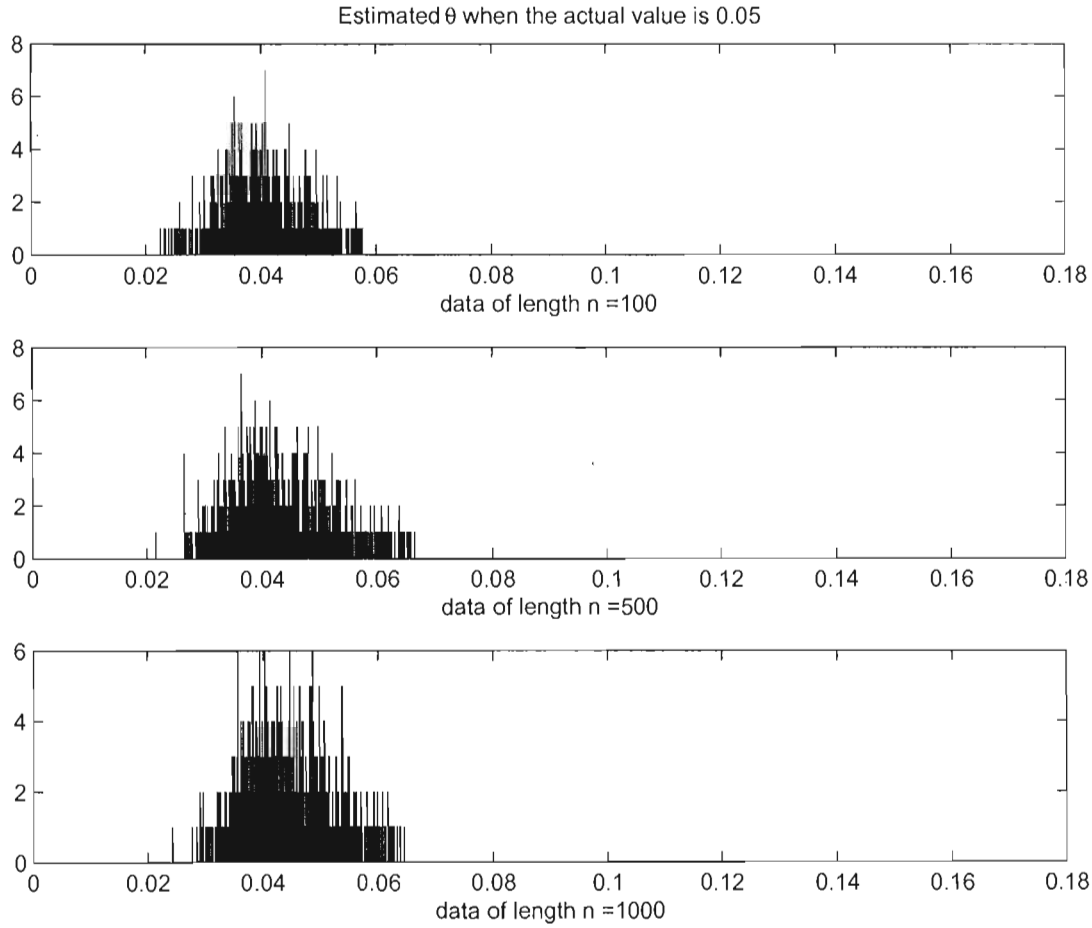

Figure 3.6 Empirical distribution of $\hat{\theta}$ with zero-coupon data for the CIR model with parameters ϕ_1

Table 3.1 A simulation analysis of the MLEs for zero-coupon data with $n = 100$ observations

	Vasicek		CIR	
Parameters	Actual values	Mean estimates	Actual values	Mean estimates
κ_1	0.060	3.177	0.250	5.946
κ_2	0.300	3.405	0.450	5.060
θ_1	0.050	0.053	0.050	0.040
θ_2	0.040	0.044	0.030	0.032
σ_1	0.020	0.041	0.050	0.152
σ_2	0.010	0.020	0.075	0.187

Table 3.2 A simulation analysis of the MLEs for zero-coupon data with $n = 500$ observations

	Vasicek		CIR	
Parameters	Actual values	Mean estimates	Actual values	Mean estimates
κ_1	0.060	0.746	0.250	2.136
κ_2	0.300	0.856	0.450	1.881
θ_1	0.050	0.076	0.050	0.043
θ_2	0.040	0.041	0.030	0.027
σ_1	0.020	0.026	0.050	0.079
σ_2	0.010	0.011	0.075	0.104

Table 3.3 A simulation analysis of the MLEs for zero-coupon data with $n = 1000$ observations

	Vasicek		CIR	
Parameters	Actual values	Mean estimates	Actual values	Mean estimates
κ_1	0.060	0.427	0.250	1.447
κ_2	0.300	0.554	0.450	1.434
θ_1	0.050	0.082	0.050	0.044
θ_2	0.040	0.041	0.030	0.025
σ_1	0.020	0.022	0.050	0.067
σ_2	0.010	0.010	0.075	0.094

CHAPTER IV

KALMAN FILTERING

4.1 Introduction to Kalman filtering

In 1960 R. R. Kalman published his famous paper describing a recursive solution to the discrete data linear filtering problem. In our context, the Kalman filter technique has been used by Duan and Simonato (1999), Geyer and Pichler (1999) and Babbs and Nowman (1999). The nature of the application of the Kalman filter depends on whether the term structure model is Gaussian such as in the Vasicek model or non-Gaussian, such as in the CIR model.

A Gaussian distribution is fully characterized by its first two moments and the exact likelihood function is obtained as a by-product of the Kalman filter algorithm. An example of the Gaussian case is provided in Babbs and Nowman (1999), who estimated a two-factor generalized Vasicek model. Babbs and Nowman (1999) observed eight interest rates with maturities between one and ten years.

When using non-Gaussian models the exact likelihood function is not available in closed form, however a quasi-maximum likelihood estimator can be constructed from the first and second conditional moments of the state variables. Examples of the non-Gaussian CIR model, may be found in Duan and Simonato (1999) and Geyer and Pichler (1999). De-Jong (2000) provides an empirical analysis of the affine class of term structure models proposed by Duffie and Kan (1996) using a quasi-maximum likelihood estimator.

The Kalman filter is a set of mathematical equations that provides an efficient compu-

tational (recursive) means to estimate the state of a process, such as to minimize the mean squared error. In this section we provide a short practical introduction to the discrete Kalman filter.

The Kalman filter is in fact a recursive algorithm for calculating estimates of unobserved state variables based on observations that depend on these state variables. The principle of the Kalman filter is to use a time series of observable data to infer the values of state variables. This technique is useful when there is a linear dependency of the observable data upon the state variables. In our case the affine model term structure satisfies this condition.

In the present case, the Kalman filter uses a state space formulation to recursively make inferences about the unobserved values of the state variables (transition system) by conditioning on the observed market zero-coupon prices (measurement system). In fact, it forms an optimal predictor of the unobserved state variable given its previously estimated value. This prediction is obtained by using the transition distribution of the state variables and updating it with the information provided by the observed variables. The Kalman filter recursion updates the transition system once a new observation is available.

We begin with a simple state space model and then we develop it to cover our case. For this reason, first consider a system that is described through the following linear model,

$$z_t = H_t x_t + \upsilon_t$$

which represents the measurement equation while the transition system is given by

$$x_{t+1} = F x_t + \varepsilon_t$$

where υ_t and ε_t are the measurement noise and of the the process, respectively. They are assumed to be independent with normal distributions,

$$\upsilon_t \sim N(0, R),$$

$$\varepsilon_t \sim N(0, Q),$$

The Kalman filter recursion is a set of equations that allows an estimator to be updated once a new observation becomes available. This estimator should be linear, recursive and unbi-

ased. The Kalman filter's equations falls into two groups, time update equations (*a priori*) and measurement update equations (*a posteriori*). According to Welch and Bishot (2005) we define $\hat{x}_t^- = (x_t|Z_{t-1})$ to be the *a priori* estimate at step t, given the knowledge of the process prior to step t, ($Z_{t-1} = (z_{t-1}, z_{t-2}, \ldots, z_1)$) and $\hat{x}_t = (x_t|Z_t)$ to be the *a posteriori* state estimate at step t given information about z_t at time t. Now we can define the *a priori* and *a posteriori* estimate errors respectively as

$$\begin{aligned} e_t^- &= \hat{x}_t^- - x_t, \\ e_t &= \hat{x}_t - x_t. \end{aligned}$$

The *a priori* estimate error covariance is then,

$$P_t^- = E\left[e_t^-(e_t^-)'\right]$$

and the *a posteriori* estimate error covariance is,

$$P_t = E\left[e_t e_t'\right].$$

The goal of Kalman filtering is to derive an equation that computes the *a posteriori* state estimate, as a combination of the *a priori* estimate and a weighted difference between an actual measurement z_t and a measurement prediction $z_t = H\hat{x}_t^-$. Let's say we can write,

$$\hat{x}_t = M_t\hat{x}_{t-1} + K_t z_t. \tag{4.1.1}$$

The unbiased criteria means that $\hat{x}_t$ has the same expected value as x_t,

$$E[\hat{x}_t - x_t] = 0,$$

by taking expectations in (4.1.1),

$$E[M_t\hat{x}_{t-1} + K_t z_t - x_t] = 0.$$

Adding and subtracting two terms, and replacing z_t by the measurement equation formula, and x_t by the transition system formula leads to,

$$\begin{aligned} E[\hat{x}_t - x_t] &= E[M_t(\hat{x}_{t-1} - x_{t-1}) + K_t(H_t(Fx_{t-1} + \varepsilon_{t-1}) + \upsilon_t) - (Fx_{t-1} + \varepsilon_{t-1}) + M_t x_{t-1}] \\ &= M_t E[\hat{x}_{t-1} - x_{t-1}] + (K_t H_t F - A + M_t)E[x_{t-1}]. \end{aligned}$$

The estimation will be unbiased when

$$\begin{aligned} K_t H_t F - A + M_t &= 0 \\ M_t = F - K_t H_t F &= (I - K_t H_t) F. \end{aligned}$$

By substituting M_t in (4.1.1),

$$\hat{x}_t = (I - K_t H_t) F \hat{x}_{t-1} + K_t z_t$$

defining the *a priori* estimate as $\hat{x}_t^- = F\hat{x}_{t-1}$,

$$\hat{x}_t = (I - K_t H_t)\hat{x}_t^- + K_t z_t = \hat{x}_t^- + K_t(z_t - H_t \hat{x}_t^-).$$

The difference $(z_t - H\hat{x}_t^-)$ is called the measurement innovation, the prediction error, or the residual. A residual of zero means that the predicted value at time $t-1$ is equal to the observed measurement at next time step, t. The matrix K, called the gain matrix, minimizes the posterior covariance of the estimation error. To find K consider the posterior error as,

$$\begin{aligned} e_t &= \hat{x}_t - x_t = \hat{x}_t^- + K_t(z_t - H_t\hat{x}_t^-) - x_t \\ &= \hat{x}_t^- + K_t(H_t x_t + \upsilon_t - H_t\hat{x}_t^-) - x_t \\ &= (I - K_t H_t)(\hat{x}_t^- - x_t) + K_t\upsilon_t = (I - K_t H_t)e_t^- + K_t\upsilon_t. \end{aligned}$$

Then, the posterior covariance would be,

$$\begin{aligned} P_t &= E[e_t e_t'] = E[((I - K_t H_t)e_t^- + K_t\upsilon_t)((I - K_t H_t)e_t^- + K_t\upsilon_t)'] \\ &= E[(I - K_t H_t)e_t^- e_t^{-'}(I - K_t H_t)' + (I - K_t H_t)e_t^- \upsilon_t' K_t' + K_t\upsilon_t e_t^{-'}(I - K_t H_t)' + K_t\upsilon_t\upsilon_t' K_t'] \\ &= (I - K_t H_t)P_t^-(I - K_t H_t)' + K_t R K_t' \\ &= (P_t^- - K_t H_t P_t^- P_t^- H_t' K_t' + K_t H_t P_t^- H_t' K_t') + K_t R K_t'. \end{aligned}$$

Our purpose is to minimize this covariance. We will use the Jacobian to minimize error; then the trace of P_t is,

$$\mathrm{tr}(P_t) = \mathrm{tr}(P_t^-) - 2\mathrm{tr}(K_t H_t P_t^-) + \mathrm{tr}(K_t H_t P_t^- H_t' K_t') + \mathrm{tr}(K_t R K_t'),$$

and to find the minimum error referring to the element of the gain matrix, we calculate the derivative of $\mathrm{tr}(P_t)$ with respect to K_t and set it equal to zero:

$$\begin{aligned}
\frac{\partial \mathrm{tr}(P_t)}{\partial K_t} &= -2P_t^- + H_t' + 2K_t H_t P_t^- H_t' + 2K_t R = 0, \\
P_t^- H_t' &= K_t(H_t P_t^- H_t' + R), \\
K_t &= P_t^- H_t'(H_t P_t^- H_t' + R)^{-1}.
\end{aligned}$$

Hence, the measurement update equations are

$$K_t = P_t^- H_t'(H_t P_t^- H_t' + R)^{-1}, \tag{4.1.2}$$

$$\hat{x}_t = \hat{x}_t^- + K_t(z_t - H_t \hat{x}_t^-), \tag{4.1.3}$$

$$P_t = (I - K_t H_t) P_t^-. \tag{4.1.4}$$

Since we assumed $\hat{x}_t^- = F\hat{x}_{t-1}$ the *a priori* error estimate covariance will be,

$$\begin{aligned}
P_t^- &= E\left[e_t^-(e_t^-)'\right] \\
&= E\left[(\hat{x}_t^- - x_t)(\hat{x}_t^- - x_t)'\right] \\
&= E\left[(F\hat{x}_{t-1} - Fx_{t-1} - w_{t-1})(F\hat{x}_{t-1} - Fx_{t-1} - w_{t-1})'\right] \\
&= E\left[Fe_{t-1}e_{t-1}'F' + Fe_{t-1}w_{t-1}' + w_{t-1}Fe_{t-1} + w_{t-1}w_{t-1}'\right] \\
&= FP_{t-1}'F' + Q.
\end{aligned}$$

In the Kalman filter algorithm, at each step, we update the measurement using the above equations, and repeat the process with the previous posterior used to project the state ahead and obtain a new prior estimate. Using the filter, we can compute a likelihood function and find the optimal parameter set by numerical optimization. We begin a more detailed presentation of the Kalman filter in the subsequent section with the specifics of the state-space formulation.

4.2 The state-space formulation

4.2.1 Vasicek model

The idea of Kalman filtering is to express a dynamic system in a particular form called the state-space representation. Constructing the state-space form involves the specification of

the transition system and the measurement system. The unobserved system of equations is called the transition system. This system describes the dynamics of the state variables as they were formulated in the model. The second observed system of equations, termed the measurement system, represents the affine relation between the market zero-coupon rates $z(t,T)$, and the state variables. We begin our development with the m-factor Vasicek model as the transition system, when the state variables are assumed to be generated by Vasicek processes,

$$dy_1(t) = \kappa_1(\theta_1 - y_1(t))dt + \sigma_1 dW_1(t),$$

$$dy_2(t) = \kappa_2(\theta_2 - y_2(t))dt + \sigma_2 dW_2(t),$$

$$\vdots$$

$$dy_m(t) = \kappa_m(\theta_m - y_m(t))dt + \sigma_m dW_m(t)$$

and the instantaneous short-term interest rates are in fact a linear combination of the above correlated state variables,

$$r(t) = \sum_{i=1}^{m} y_i(t).$$

Using the recursive expression that was used in Chapter II for the Vasicek model, namely

$$y(t_i) = e^{-\kappa\Delta t}y(t_{i-1}) + \theta\left(1 - e^{-\kappa\Delta t}\right) + \varepsilon(t_i)$$

with

$$\varepsilon(t_i) \sim N\left(0, \frac{\sigma^2}{2\kappa}\left(1 - e^{-2\kappa\Delta t}\right)\right)$$

we can specify the transition system as follows:

$$\underbrace{\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \\ \vdots \\ y_m(t_i) \end{bmatrix}}_{y(t_i)} = \underbrace{\begin{bmatrix} e^{-\kappa_1\Delta t} & 0 & \dots & 0 \\ 0 & e^{-\kappa_2\Delta t} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & e^{-\kappa_m\Delta t} \end{bmatrix}}_{F} \underbrace{\begin{bmatrix} y_1(t_{i-1}) \\ y_2(t_{i-1}) \\ \vdots \\ y_m(t_{i-1}) \end{bmatrix}}_{y(t_{i-1})} + \underbrace{\begin{bmatrix} \theta_1(1 - e^{-\kappa_1\Delta t}) \\ \theta_2(1 - e^{-\kappa_2\Delta t}) \\ \vdots \\ \theta_m(1 - e^{-\kappa_m\Delta t}) \end{bmatrix}}_{C} + \underbrace{\begin{bmatrix} \varepsilon_1(t_i) \\ \varepsilon_2(t_i) \\ \vdots \\ \varepsilon_m(t_i) \end{bmatrix}}_{\varepsilon(t_i)}$$

where

$$\varepsilon(t_i) \sim N(0, Q).$$

Since in Chapter I we assumed that our state variables are i.i.d. for both the Vasicek and the CIR multi-factor models, the covariances between the state variables are zero and the process noise covariance matrix, Q is diagonal,

$$Q = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1}\left(1-e^{-2\kappa_1\Delta t}\right) & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2}\left(1-e^{-2\kappa_2\Delta t}\right) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_m^2}{2\kappa_m}\left(1-e^{-2\kappa_m\Delta t}\right) \end{bmatrix} \tag{4.2.5}$$

In general we can write

$$y(t_i) = C + Fy(t_{i-1}) + \varepsilon(t_i). \tag{4.2.6}$$

Further we assume that we have a sequence of N zero-coupon rates $z_1,\ldots,z_N$ with terms to maturity $T_1,\ldots,T_N$ respectively. In general we require one market zero-coupon rate for each factor used in the estimation. For example, if we were considering a two factor model, we would require only two observed zero-coupon bond yields. By adding market rates, however, we provide cross-sectional information about the term structure of interest rates at each observed point in time. This information is particularly helpful in specifying the market price of risk parameters (λ_i, $i = 1,\ldots,m$). In fact, each state variable has a risk parameter, λ_i, which is treated as a fixed parameter. To construct our measurement system, we need these N zero-coupon rates and the following relation between the zero-coupon yield and the price of a zero-coupon bond, which we described in the previous chapter,

$$z(t,T) = -\frac{\log P(t,T)}{T-t} = \frac{-A(t,T) + \sum_{i=1}^{m} B_i(t,T)\,y_i(t)}{T-t}$$

where $A(t,T)$ and $B_i(t,T)$ for $i = 1,\ldots,m$, are,

$$\begin{aligned} B_i(t,T) &= \frac{1}{\kappa_i}(1-e^{-\kappa_i(T-t)}), \\ A(t,T) &= \sum_{i=1}^{m} \frac{\gamma(B_i(t,T)-(T-t))}{\kappa_i^2} - \frac{\sigma_i^2 B_i^2(t,T)}{4\kappa_i} \\ &\quad + \sum_{i,j\,;i\neq j} \frac{\sigma_{ij}}{2\kappa_i\kappa_j}\left(T-t-B_i(t,T)-B_j(t,T)+\frac{1}{\kappa_i+\kappa_j}\left(1-e^{-(\kappa_i+\kappa_j)(T-t)}\right)\right) \end{aligned}$$

and

$$\gamma_i = \kappa_i^2\left(\theta_i - \frac{\sigma_i\lambda_i}{\kappa_i}\right) - \frac{\sigma_i^2}{2}.$$

We assume that the measurement errors in the interest rates are additive and normally distributed. The measurement system is then,

$$z(t_i) = \hat{A} + Hy(t_i) + v(t_i) \tag{4.2.7}$$

or, in more detail,

$$\underbrace{\begin{bmatrix} z(t_i,T_1) \\ z(t_i,T_2) \\ \vdots \\ z(t_i,T_N) \end{bmatrix}}_{z(t_i)} = \underbrace{\begin{bmatrix} \frac{A(t_i,T_1)}{T_1-t_i} \\ \frac{A(t_i,T_2)}{T_2-t_i} \\ \vdots \\ \frac{A(t_i,T_N)}{T_N-t_i} \end{bmatrix}}_{\hat{A}} + \underbrace{\begin{bmatrix} \frac{B_1(t_i,T_1)}{T_1-t_i} & \frac{B_2(t_i,T_1)}{T_1-t_i} & \cdots & \frac{B_m(t_i,T_1)}{T_1-t_i} \\ \frac{B_1(t_i,T_2)}{T_2-t_i} & \frac{B_2(t_i,T_2)}{T_2-t_i} & \cdots & \frac{B_m(t_i,T_2)}{T_2-t_i} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{B_1(t_i,T_N)}{T_N-t_i} & \frac{B_2(t_i,T_N)}{T_N-t_i} & \cdots & \frac{B_m(t_i,T_N)}{T_N-t_i} \end{bmatrix}}_{H} \underbrace{\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \\ \vdots \\ y_m(t_i) \end{bmatrix}}_{y(t_i)} + \underbrace{\begin{bmatrix} v_1(t_i) \\ v_2(t_i) \\ \vdots \\ v_N(t_i) \end{bmatrix}}_{v(t_i)}.$$

This model is affine in the state vector $y(t_i)$, but the A and B's are non-linear functions of the underlying parameters. So we have to assume that our measurement system has errors, $v(t_i)$, in the measurement system. We assume that the measurement errors in the bond yields are additive and normally distributed, giving,

$$v(t_i) \sim N(0,R).$$

Moreover the number of observed bonds and their associated maturities do not change over time. Therefore, R has namely constant dimension $N \times N$ and is assumed to be a diagonal matrix, such as,

$$R = \begin{bmatrix} r_1^2 & 0 & \cdots & 0 \\ 0 & r_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & r_N^2 \end{bmatrix}.$$

4.2.2 The CIR model

This section presents the reformulation of the CIR model. The filter for the CIR model is not linear and may be biased. The exact transition density for the CIR model is a non-central $\chi^2(v,\delta)$, with v degrees of freedom and δ the non-centrality parameter.

$$v = \frac{4\kappa\theta}{\sigma^2}\delta = \frac{4\kappa y(t_{i-1})}{\sigma^2(e^{\kappa\Delta t}-1)}.$$

However, to satisfy the assumption of normality in the Kalman filter model's maximum likelihood estimation, substituting the exact transition density by a Gaussian or normal density is often done. Therefore, in the CIR model we will assume that variables of y are approximately of the following Gaussian distribution,

$$y(t_i) \sim N\left(e^{-\kappa\Delta t}y(t_{i-1}) + \theta\left(1-e^{-\kappa\Delta t}\right), \frac{\theta\sigma^2}{2\kappa}\left(1-e^{-\kappa\Delta t}\right)^2 + \frac{\sigma^2}{\kappa}\left(e^{-\kappa\Delta t}-e^{-2\kappa\Delta t}\right)y(t_{i-1})\right).$$

With this assumption, the transition system for the CIR model has the following characteristics,

$$y(t_i) = C + Fy(t_{i-1}) + \varepsilon(t_i) \tag{4.2.8}$$

where, the form of C and F for the CIR model is almost the same as that described by the Vasicek model, but the $\varepsilon(t_i)$ are be different, since the variance of the system is dependent on the state of the process,

$$\varepsilon(t_i) \sim N(0, Q_{t_i})$$

$$Q_{t_i} = \begin{bmatrix} \xi_1 & 0 & \dots & 0 \\ 0 & \xi_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \xi_m \end{bmatrix}$$

where, for each $i = 1, \dots, n$ and $j = 1, \dots, N$,

$$\xi_j = \frac{\theta_j\sigma_j^2}{2\kappa_j}\left(1-e^{-\kappa_j\Delta t}\right)^2 + \frac{\sigma_j^2}{\kappa_j}\left(e^{-\kappa_j\Delta t}-e^{-2\kappa_j\Delta t}\right)y(t_{i-1}).$$

For the measurement system we have,

$$z(t,T) = -\frac{\ln P(t,T)}{T-t} = \sum_{i=1}^{m}\frac{-A_i(t,T) + B_i(t,T)y_i(t)}{T-t} \tag{4.2.9}$$

where according to Chapter III we have,

$$B_i(\tau) = \frac{2(e^{\gamma_i\tau}-1)}{(\gamma_i+\kappa_i+\lambda_i)(e^{\gamma_i\tau}-1)+2\gamma_i},$$

$$A_i(\tau) = \log\left(\frac{2\gamma_i e^{\frac{(\gamma_i+\kappa_i+\lambda_i)\tau}{2}}}{(\gamma_i+\kappa_i+\lambda_i)(e^{\gamma_i\tau}-1)+2\gamma_i}\right)^{\frac{2\kappa_i\theta_i}{\sigma_i^2}},$$

and

$$\gamma_i = \sqrt{(\kappa_i + \lambda_i)^2 + 2\sigma_i^2}.$$

Then the measurement system for the CIR model is as follows,

$$z(t_i) = \hat{A} + Hy(t_i) + v(t_i). \tag{4.2.10}$$

The form of the measurement system is almost the same as the one for the Vasicek model, the only difference is in the $\hat{A}$ matrix, which in this case is:

$$\hat{A} = \begin{bmatrix} \sum_{i=1}^{n} \frac{A_i(t_i,T_1)}{T_1 - t_i} \\ \sum_{i=1}^{n} \frac{A_i(t_i,T_2)}{T_2 - t_i} \\ \vdots \\ \sum_{i=1}^{n} \frac{A_i(t_i,T_N)}{T_N - t_i} \end{bmatrix}.$$

4.3 The Kalman filter implementation

The Kalman Filter recursion is a set of equations which allows an estimator to be updated when a new observation becomes available. These equations provide the minimum variance estimator over all unbiased estimators. The resulting estimator is thus linear, recursive and unbiased. In the previous subsection the model have been reconstructed in state space form, now the Kalman filter can be used to obtain information about space variables from observed zero-coupon yields.

In Kalman filtering we begin with the initial values for the state variables and then we proceed to use these initial values to predict the value of the measurement equation when we actually are able to observe in the next period of time. Using these observed value we can update the inferences about the current value of the state variables in the next period of time. These updated values are then used to forecast the next state variables for the next period of time. This procedure is repeated for all periods of time.

We define Z_i for $i = 0, \ldots, n$, to determine the period of time we are working on. In fact $Z_i = z(t_i), z(t_{i-1}), \ldots, z(t_0)$ denotes the information we have observed before the time t_i.

Consider the unconditional mean and variance of the state variables at $t_0 = 0$, for the Vasicek model with a Gaussian transition density, given by

$$E\left[\, y(t_0) \mid Z_0 \,\right] = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_m \end{bmatrix},$$

$$\mathrm{Var}\left(y(t_0) \mid Z_0\right) = \begin{bmatrix} \frac{\sigma_1^2}{2\kappa_1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2}{2\kappa_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_n^2}{2\kappa_m} \end{bmatrix}.$$

The second step in predicting the next state variable $y(t_i)$ given $y(t_{i-1})$, is to forecast the measurement equation of the next level with respect to the estimated state variables, which according to (4.2.7) will be,

$$\begin{aligned} E\left[z(t_i) \mid Z_{i-1}\right] &= E\left[\hat{A} + Hy(t_i) + v(t_i) \mid Z_{i-1}\right] \\ &= E\left[\hat{A}\right] + E[H]E\left[\, y(t_i) \mid Z_{i-1}\right] + E[v(t_i)] \\ &= \hat{A} + HE\left[y(t_i)\right] \\ \mathrm{Var}\left(z(t_i) \mid Z_{i+1}\right) &= \mathrm{Var}\left(\hat{A} + Hy(t_i) + v(t_i) \,\middle|\, Z_{i-1}\right) \\ &= H\mathrm{Var}\left(\, y(t_i) \mid Z_{i-1}\right) H' + R. \end{aligned}$$

Now that we have a conditional prediction of the measurement system, we can observe the next period of time's values for the measurement system and compare these two values; that is, we can compute the difference

$$\zeta(t_i) = z(t_i) - E[z(t_i) \mid Z_{i-1}],$$

where the $\zeta(t_i)$ is the residual of our prediction, also called the prediction error. Now that we have the observed values we can use the recursive equation to change the state variables related to the predicted measurement. We can update the transition system using (4.1.3), as

$$E\left[\, y(t_i) \mid Z_i \,\right] = E\left[\, y(t_i) \mid Z_{i-1} \,\right] + K(t_i)\zeta(t_i)$$

where $K(t_i)$ is called a gain matrix, or blending factor, where (4.1.2) suggests that,

$$K(t_i) = \mathrm{Var}\left(y(t_i) \mid Z_{i-1}\right) H' \mathrm{Var}\left(z(t_i) \mid Z_{i-1}\right)^{-1}.$$

The updated conditional variance (4.1.4) is,

$$\mathrm{Var}\left(y(t_i) \mid Z_i\right) = (I - K(t_i)H)\,\mathrm{Var}\left(y(t_i) \mid Z_{i-1}\right).$$

By having the updated mean and variance of state variables and using (4.2.6), we can forecast the unobserved state variables for the next time period,

$$\begin{aligned} E\left[y(t_{i+1}) \mid Z_i\right] &= E\left[C + Fy(t_i) + \varepsilon(t_i + 1) \mid z_i\right] \\ &= C + FE\left[y(t_i) \mid z_i\right] \\ \mathrm{Var}\left(y(t_{i+1}) \mid Z_i\right) &= \mathrm{Var}\left(y(t_i) \mid Z_{i-1}\right) - F\left(\mathrm{Var}\left(y(t_i) \mid Z_i\right)\right) F' + Q. \end{aligned}$$

This algorithm must be repeated for each discrete time step in the data sample. The same procedure is applied to the CIR state-space formulation. The difference is only in the state-space formulation and the initial value for the variance of state variables which indeed, in the CIR case the variance is,

$$\mathrm{Var}\left(y(t_1) \mid Z_0\right) = \begin{bmatrix} \frac{\sigma_1^2\theta_1}{2\kappa_1} & 0 & \cdots & 0 \\ 0 & \frac{\sigma_2^2\theta_2}{2\kappa_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\sigma_m^2\theta_m}{2\kappa_m} \end{bmatrix}.$$

The algorithm would be the same for the CIR model.

4.4 Maximum likelihood estimation

The (conditional) likelihood of a set of observations $z(t_0), z(t_1), z(t_2), \ldots, z(t_n)$ is

$$L(z(t_1), z(t_2), \ldots, z(t_n); z(t_0), \phi) = \prod_{i=1}^{n} p\left(z(t_i) \mid Z(t_{i-1})\right)$$

where $p\left(z(t_i) \mid Z(t_{i-1})\right)$ is the distribution of $z(t_i)$ conditional on the information set at time t_{i-1}, that is $Z(t_{i-1}) = \{z(t_{i-1}), z(t_{i-2}), \ldots, z(t_1)\}$. If the disturbances and initial state vector of

the model have a proper normal distributions, then the distribution $p(z(t_i)_Z(t_{i-1})))$, is itself multivariate normal, and the mean and covariance matrix of this conditional distribution are given directly by the Kalman filter, namelu

$$p(z(t_i) \mid Z(t_{i-1})) = \frac{1}{(2\pi)^{\frac{N}{2}} |\mathrm{Var}(z(t_i) \mid Z(t_{i-1}))|^{\frac{1}{2}}} \exp\left\{ -\frac{1}{2}\zeta(t_i)' \mathrm{Var}(z(t_i) \mid Z(t_{i-1}))^{-1} \zeta(t_i) \right\}.$$

The *likelihood* function is

$$\begin{aligned} L(z(t_1), z(t_2), \ldots, z(t_n); z(t_0), \phi) &= \prod_{i=1}^{n} p(z(t_i) \mid Z(t_{i-1})) \\ &= \prod_{i=1}^{n} (2\pi)^{-\frac{N}{2}} |\mathrm{Var}(z(t_i) \mid Z(t_{i-1}))|^{-\frac{1}{2}} e^{-\frac{1}{2}\zeta(t_i)' \mathrm{Var}(z(t_i)|Z(t_{i-1}))^{-1}\zeta(t_i)}, \end{aligned}$$

while the log-likelihood function can be written as

$$\begin{aligned} \ell(\phi) &= \sum_{i=1}^{n} \log\left((2\pi)^{-\frac{N}{2}} |\mathrm{Var}(z(t_i) \mid Z(t_{i-1})|^{-\frac{1}{2}} e^{-\frac{1}{2}\zeta(t_i)' \mathrm{Var}(z(t_i)|Z(t_{i-1}))^{-1}\zeta(t_i)} \right) \\ &= -\frac{nN\log(2\pi)}{2} - \frac{1}{2}\sum_{i=1}^{n} \left(\log\left|\mathrm{Var}(z(t_i) \mid Z(t_{i-1}))\right| + \zeta(t_i)' \mathrm{Var}(z(t_i) \mid Z(t_{i-1}))^{-1} \zeta(t_i) \right). \end{aligned}$$

The last step is to find the optimal parameter set ϕ. We use **fminsearch** as a nonlinear numerical optimization technique to find the maximum likelihood in MATLAB.

4.5 Numerical analysis

In this section, we apply Kalman filtering to the same problem as the one considered in the previous Chapter. That is, we put the Vasicek and CIR single-factor models, in the state-space form. In this simulation we use four sequences of zero coupon rates with terms to maturity $[T_1, T_2, T_3, T_4]$. In the estimation we use weekly observation $\Delta = \frac{1}{52}$ for the different zero-coupons $z_1, \ldots, z_4$ with terms to maturity $\tau = [T_1, \ldots, T_4] = [0.070.250.51]$ which indicate we have 2-year and half, 5-year, 10-year and 20-year bonds over a 20-year time horizon. However the state variables are simulated weekly over a 20-year time horizon.

We use computed price values from the previous chapter as observed variables, and by using equation (4.2.9), we obtain the real values of the zero-coupon yields for each term to maturity. According to Kalman filtering, for the first step we start from the first estimated values for interest rates using the conditional mean and variance of the transition system.

Later, we compute zero-coupon yields by using the measurement systems concerning the estimated interest rate.

In our case, we assumed that λ is constant and equal to 1, but in reality we should also consider λ as an unknown parameter. R is the error of measurement system that we add to the system, moreover we assume that there is no serial correlation in these measurement errors for zero coupon rates. Therefore, R has a small constant value and is assumed to be a diagonal matrix.

Now by comparing the the observed zero-coupon rates and the estimated ones, we try to improve the estimated interest rates. Finally, by using these improved values in the transition system we calculate the interest rates for the next period of time. This process continues for all time steps, giving a Gaussian distribution for all the price values. At the end we compute the log-likelihood function and try to optimize it.

This estimation procedure is repeated 1000 times. Although the procedure is lengthy, the results are quite accurate. The tables and figures illustrated in this section summarize the results of this simulation exercise for the Vasicek and CIR models. The rest of plots are outlined in Appendix C.

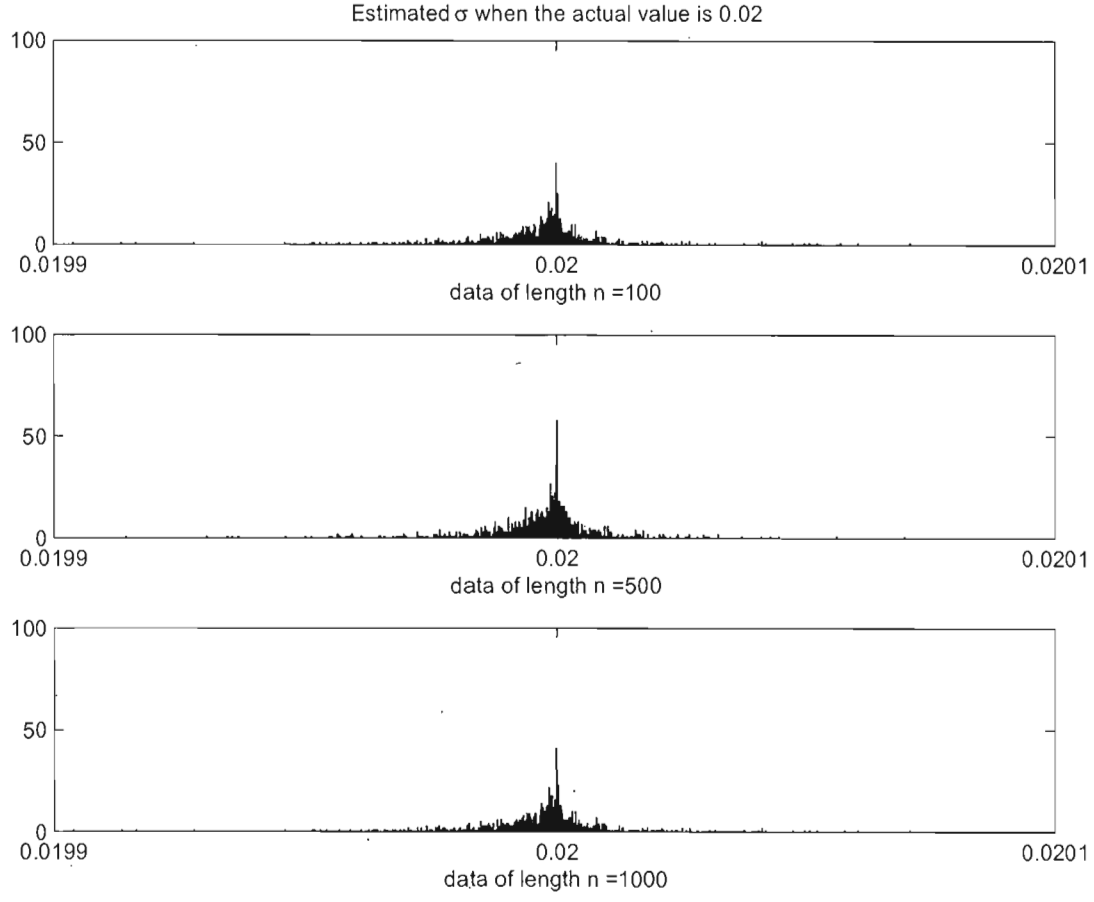

Figure 4.1 Empirical distribution of $\hat{\sigma}$ for zero-coupon data and Kalman filtering (Vasicek model with parameter ϕ_1).

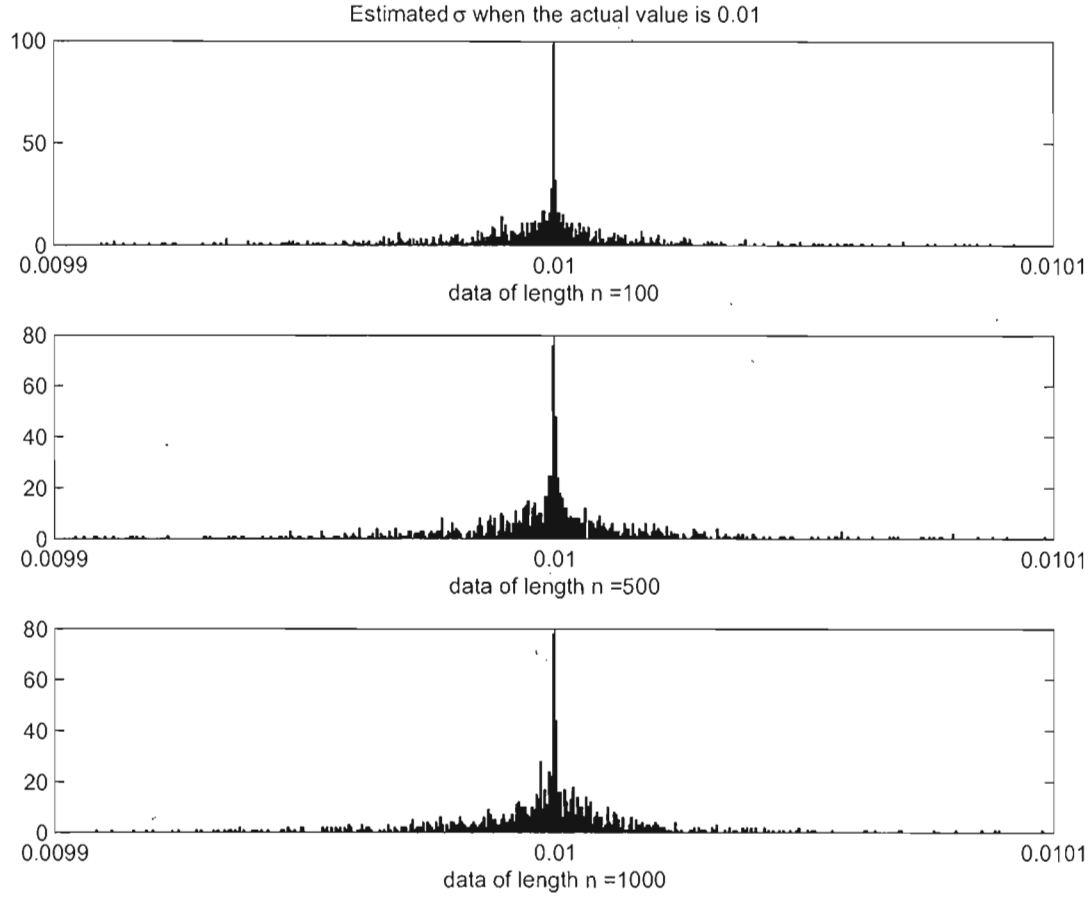

Figure 4.2 Empirical distribution of $\hat{\sigma}$ for zero-coupon data and Kalman filtering (Vasicek model with parameter ϕ_2).

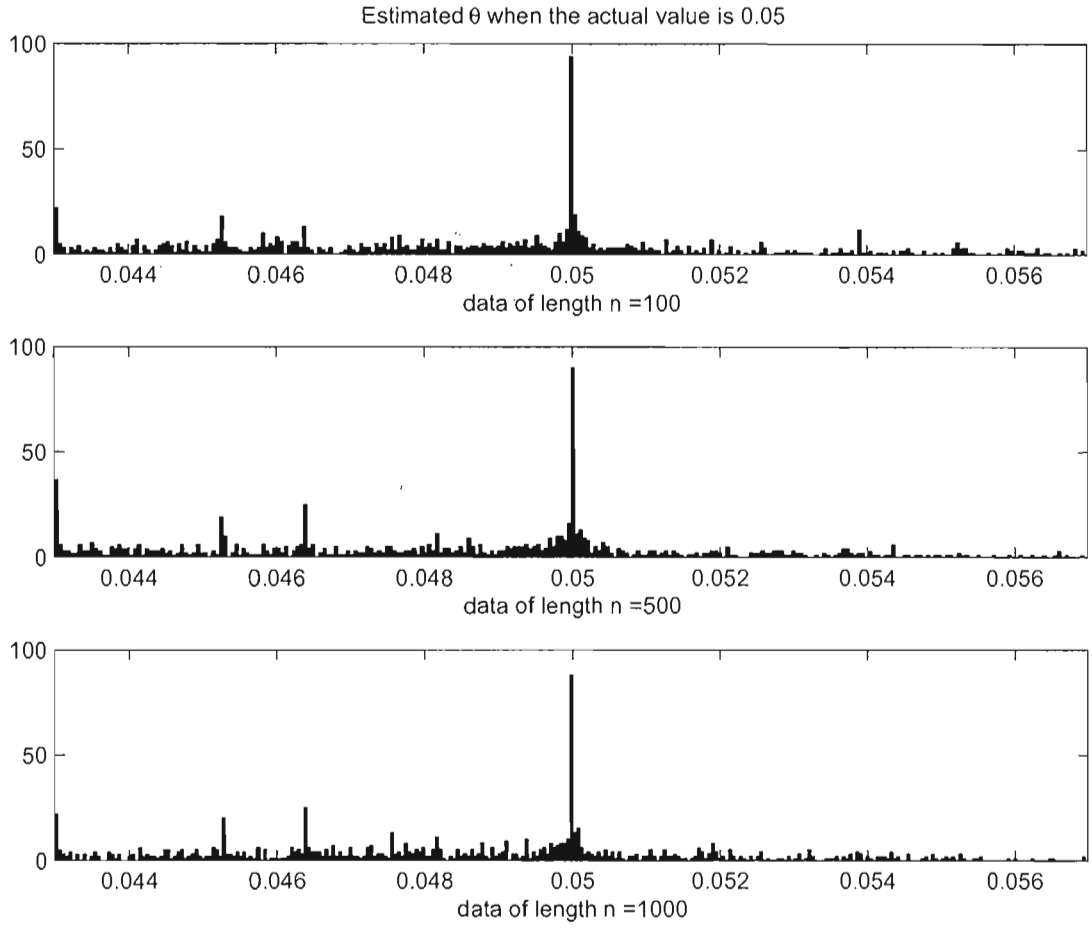

Figure 4.3 Empirical distribution of $\hat{\theta}$ for zero-coupon data and Kalman filtering (CIR model with parameter ϕ_1).

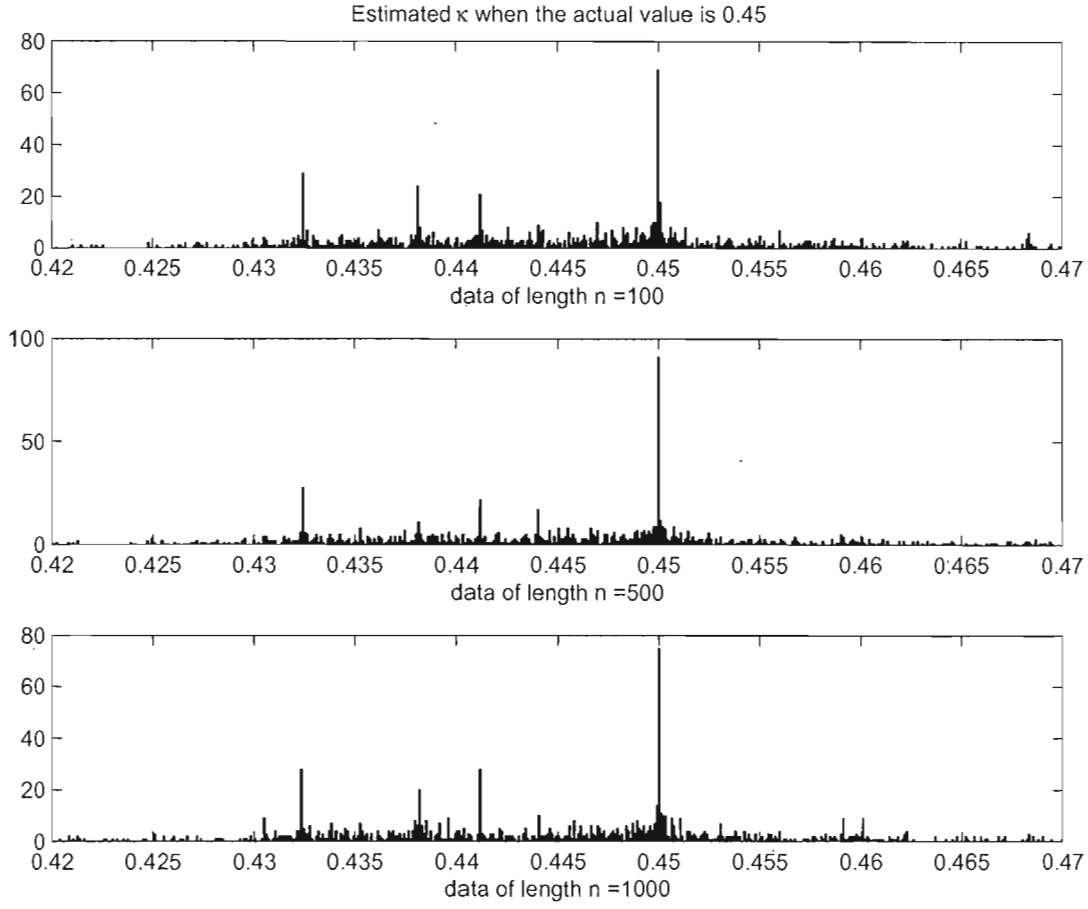

Figure 4.4 Empirical distribution of $\hat{\kappa}$ for zero-coupon data and Kalman filtering (CIR model with parameter ϕ_2).

Table 4.1 A simulation analysis using Kalman filter for $n = 100$ observations

	Vasicek			CIR		
Parameters	Actual values	Mean estimates	Standard deviations	Actual values	Mean estimates	Standard deviations
κ_1	0.060	0.059	0.0002	0.250	0.245	0.0109
κ_2	0.300	0.299	0.0000	0.450	0.446	0.0108
θ_1	0.050	0.050	0.0000	0.050	0.049	0.0047
θ_2	0.040	0.039	0.0000	0.030	0.029	0.0033
σ_1	0.020	0.020	0.0000	0.050	0.048	0.005
σ_2	0.010	0.010	0.0000	0.075	0.073	0.006

Table 4.2 A simulation analysis using Kalman filter for $n = 500$ observations

	Vasicek			CIR		
Parameters	Actual values	Mean estimates	Standard deviations	Actual values	Mean estimates	Standard deviations
κ_1	0.060	0.059	0.0002	0.250	0.246	0.0106
κ_2	0.300	0.299	0.0000	0.450	0.446	0.0115
θ_1	0.050	0.050	0.0000	0.050	0.049	0.0049
θ_2	0.040	0.039	0.0000	0.030	0.029	0.0032
σ_1	0.020	0.020	0.0000	0.050	0.049	0.004
σ_2	0.010	0.010	0.0000	0.075	0.073	0.005

Table 4.3 A simulation analysis using Kalman filter for $n = 1000$ observations

	Vasicek			CIR		
Parameters	Actual values	Mean estimates	Standard deviations	Actual values	Mean estimates	Standard deviations
κ_1	0.060	0.059	0.0002	0.250	0.245	0.0103
κ_2	0.300	0.299	0.0000	0.450	0.445	0.0112
θ_1	0.050	0.050	0.0000	0.050	0.049	0.0045
θ_2	0.040	0.039	0.0000	0.030	0.029	0.0033
σ_1	0.020	0.020	0.0000	0.050	0.043	0.004
σ_2	0.010	0.010	0.0000	0.075	0.073	0.006

CONCLUSION

An affine term structure model is a time series model used to describe the stochastic behavior of interest rates. We examined a variety of techniques to estimate the parameters of the SDEs of the underlying factors of the model. By comparing the estimated values of the parameter set of the problem using direct state variables (Chapter I) to the estimated values using state variables constructed by applying the *affine term structure model* to the prices, we conclude that these models can not accurately explain the dynamics of the term structure. However, in higher dimensions the results are better.

In contrast, by using Kalman filtering in these affine models we get reasonable results. Therefore it is important to choose an appropriate technique to estimate the parameters. Of course, this dissertation represents only a first step in estimating a parameter set of zero-coupons using different techniques.

A natural next step would be to consider the estimation of the market price of risk, a rather hard problem according to the finance literature. Another problem which arose during our empirical studies that deserves attention is the following: the numerical computation of the MLEs depends heavily on the proposed starting values for the parameters. That is especially true for zero coupon data. Is it possible to compute "good" starting values (for example with the EM algorithm)?

APPENDIX A

SIMULATION RESULTS FOR MAXIMUM LIKELIHOOD METHOD

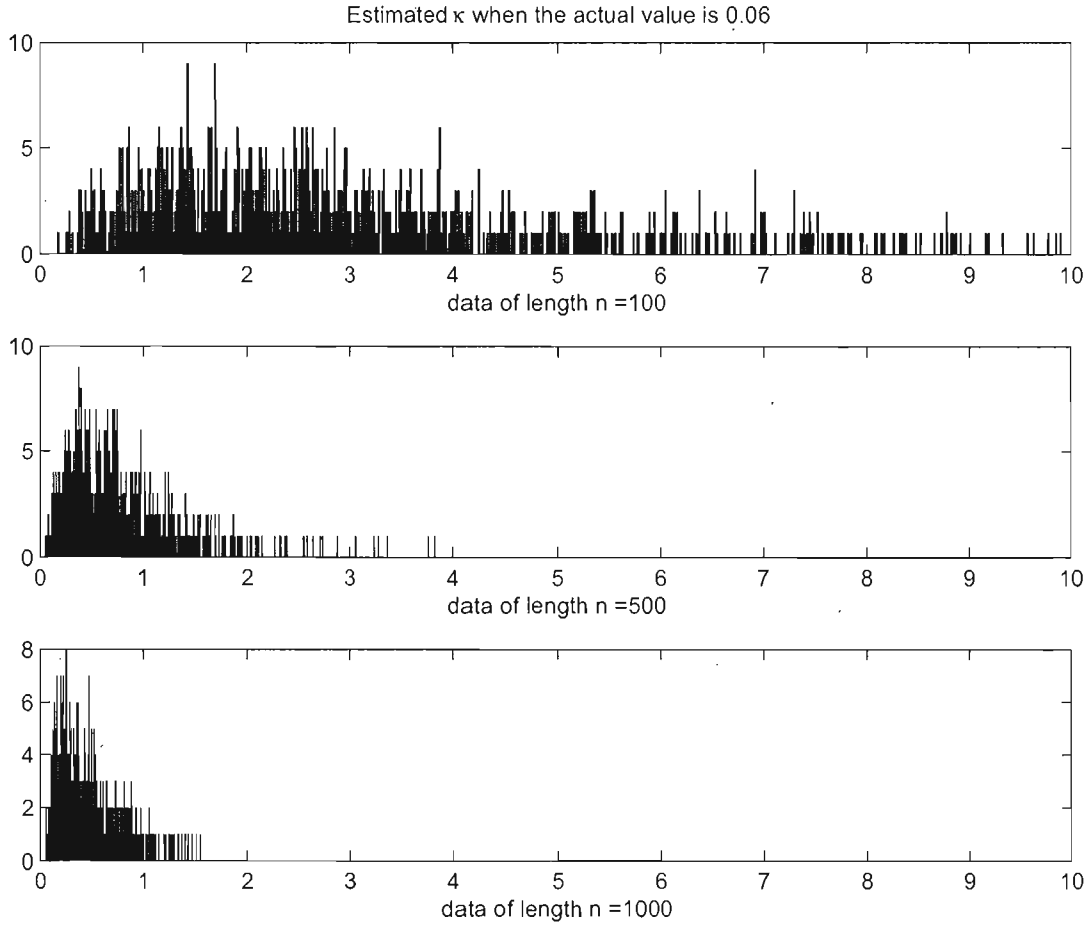

Figure A.1 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model.

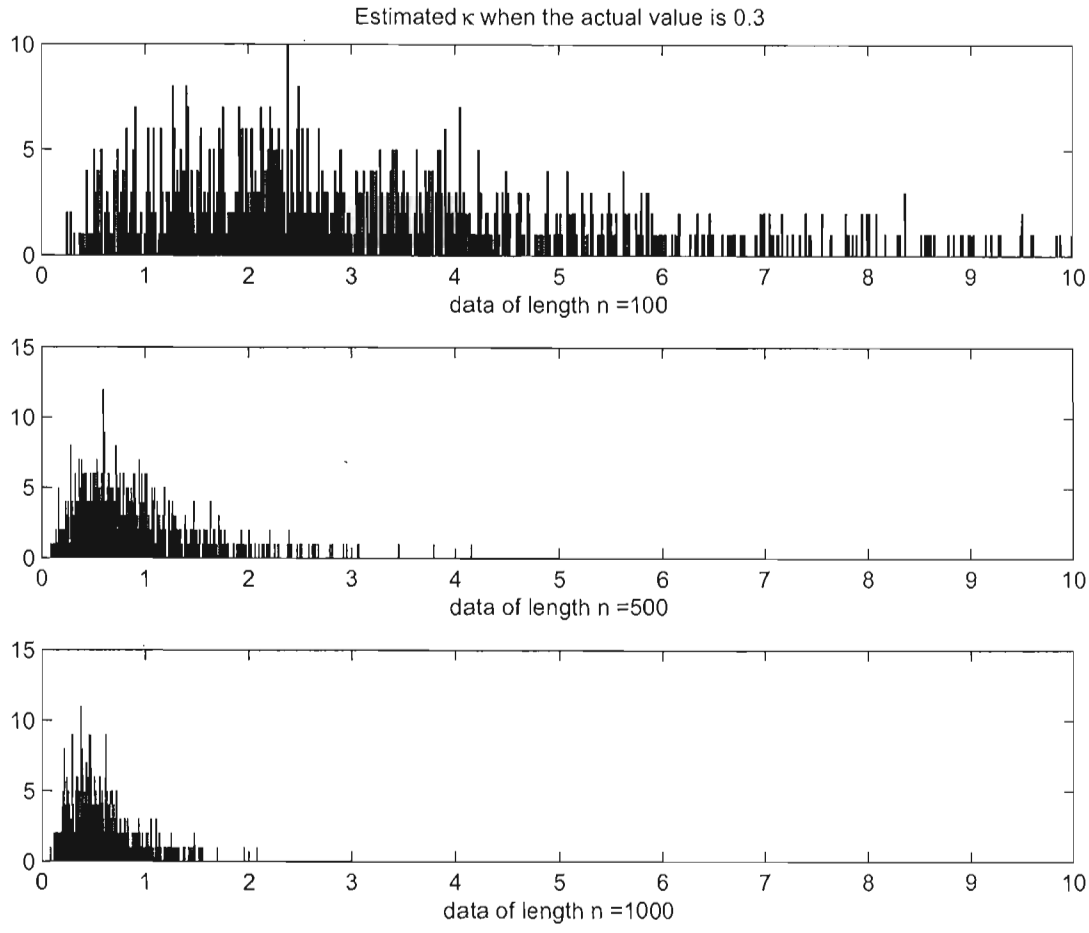

Figure A.2 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model.

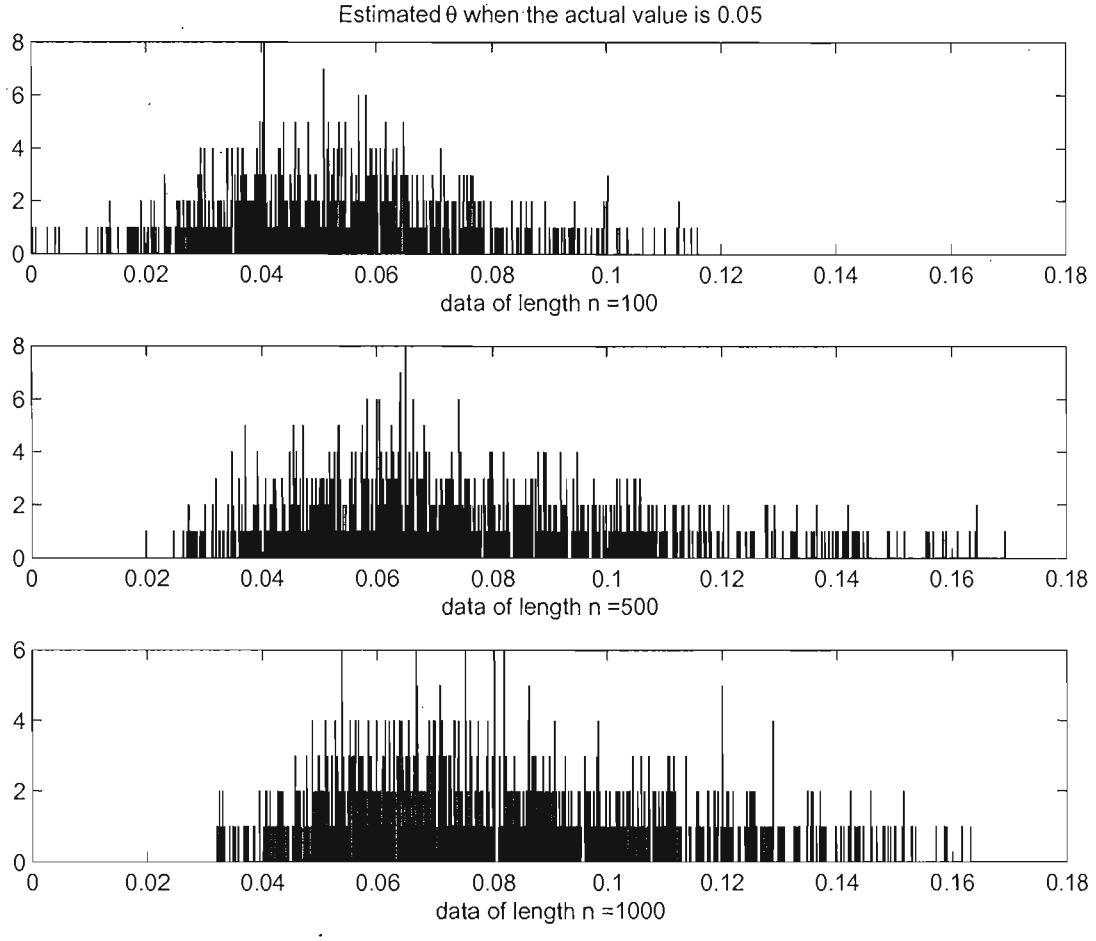

Figure A.3 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model.

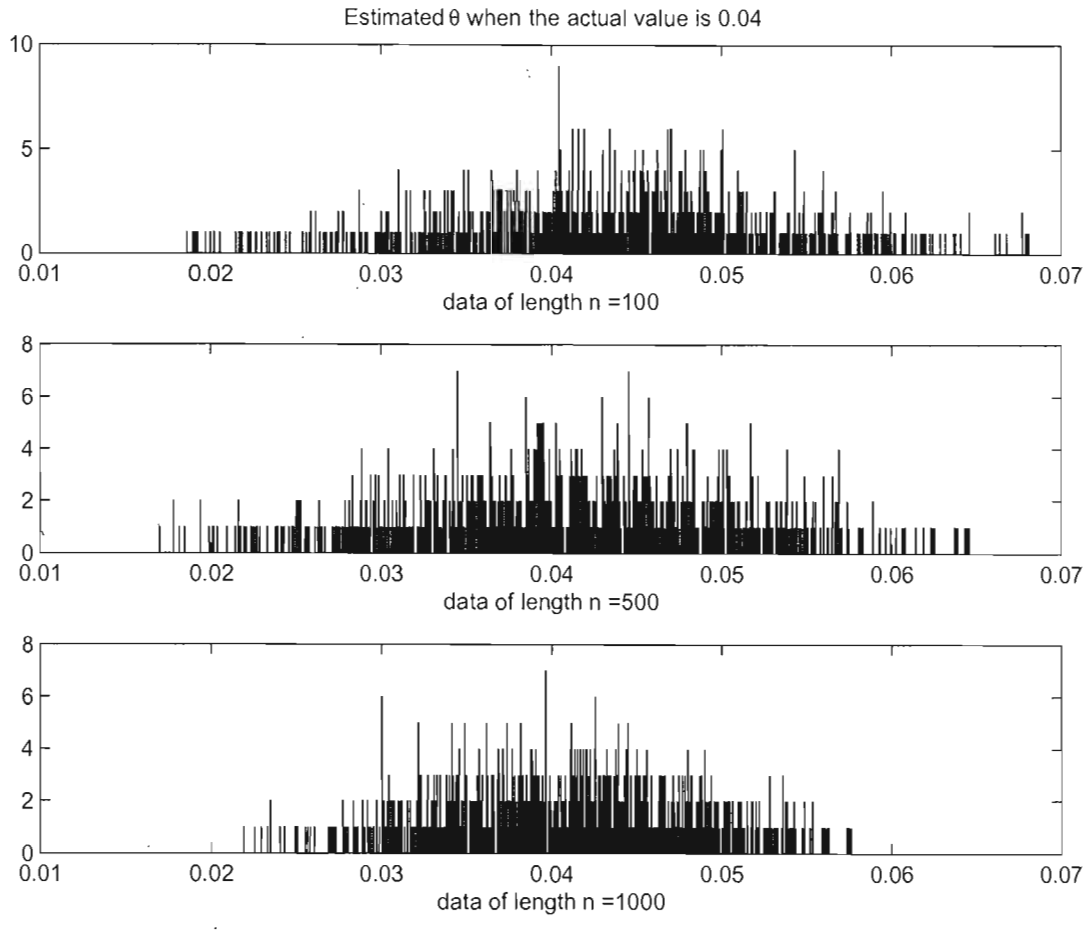

Figure A.4 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model.

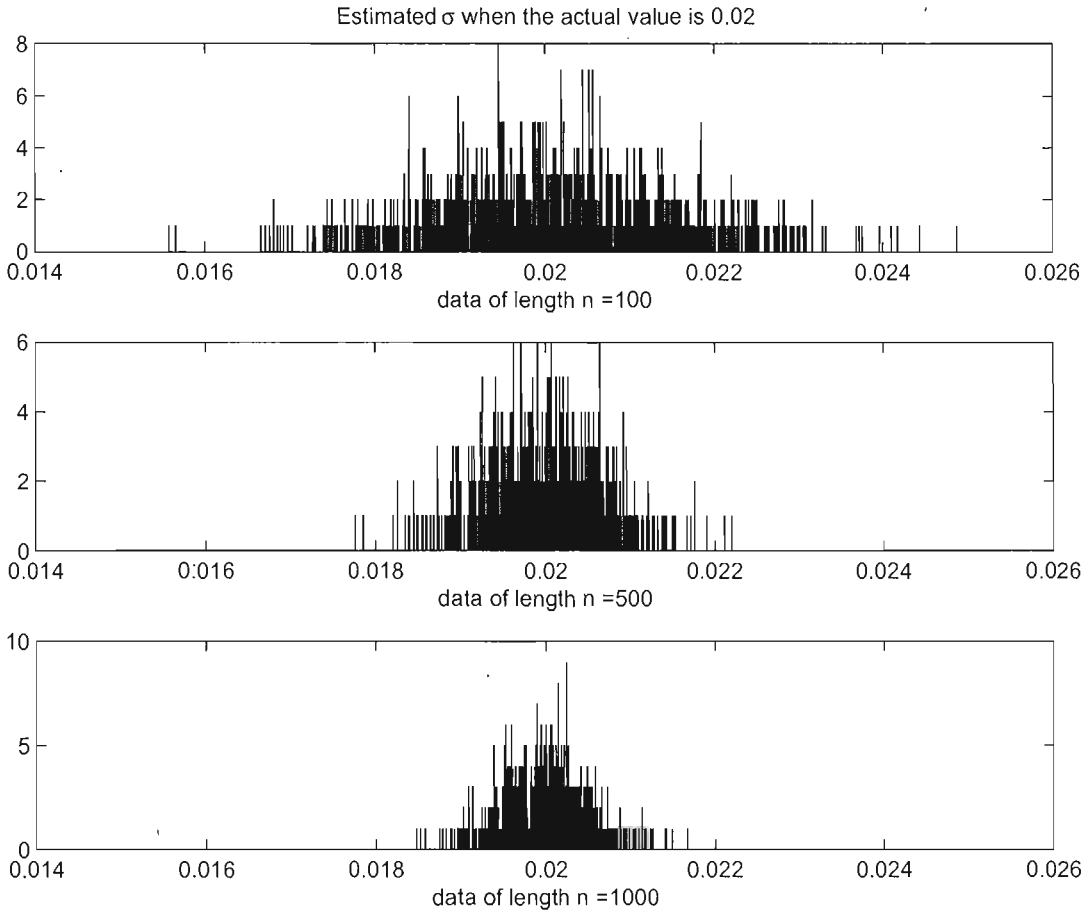

Figure A.5 Empirical distribution of MLE $\hat{\sigma}$ for Vasicek's model.

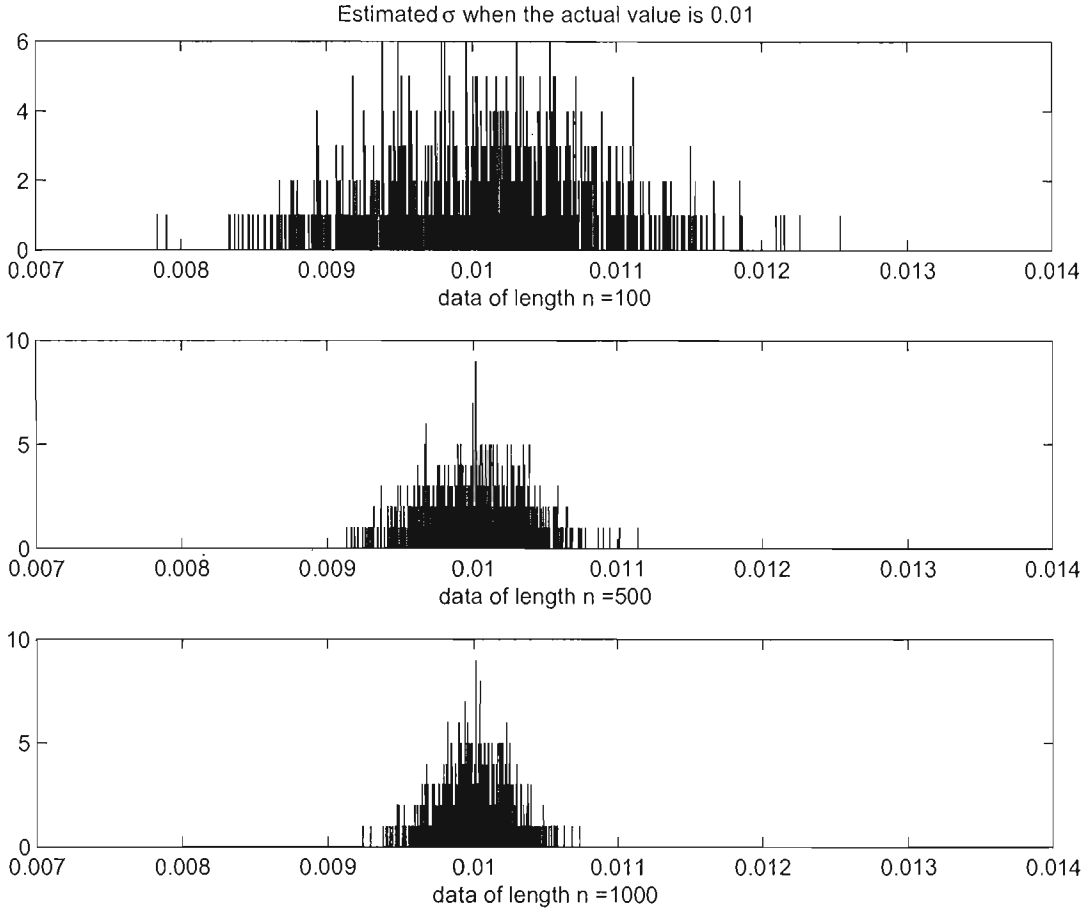

Figure A.6 Empirical distribution of MLE $\widehat{\sigma}$ for Vasicek's model.

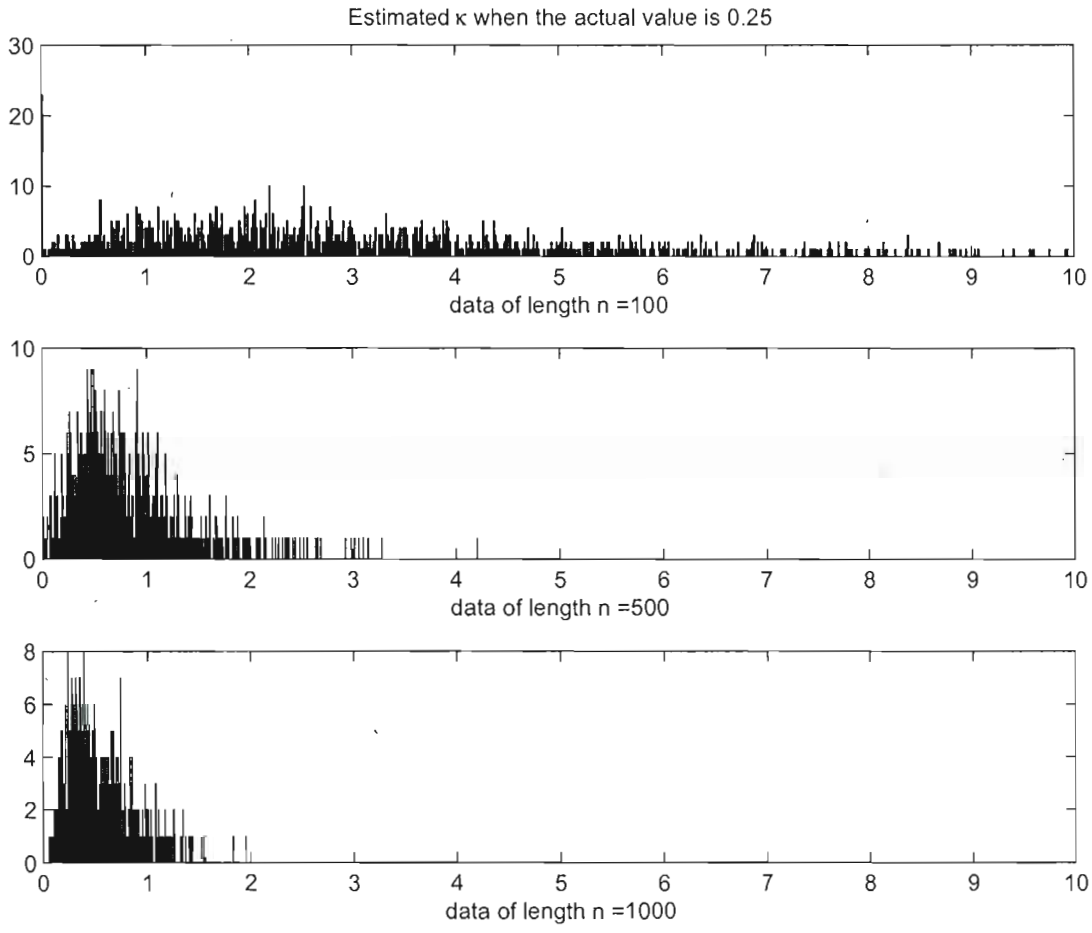

Figure A.7 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model.

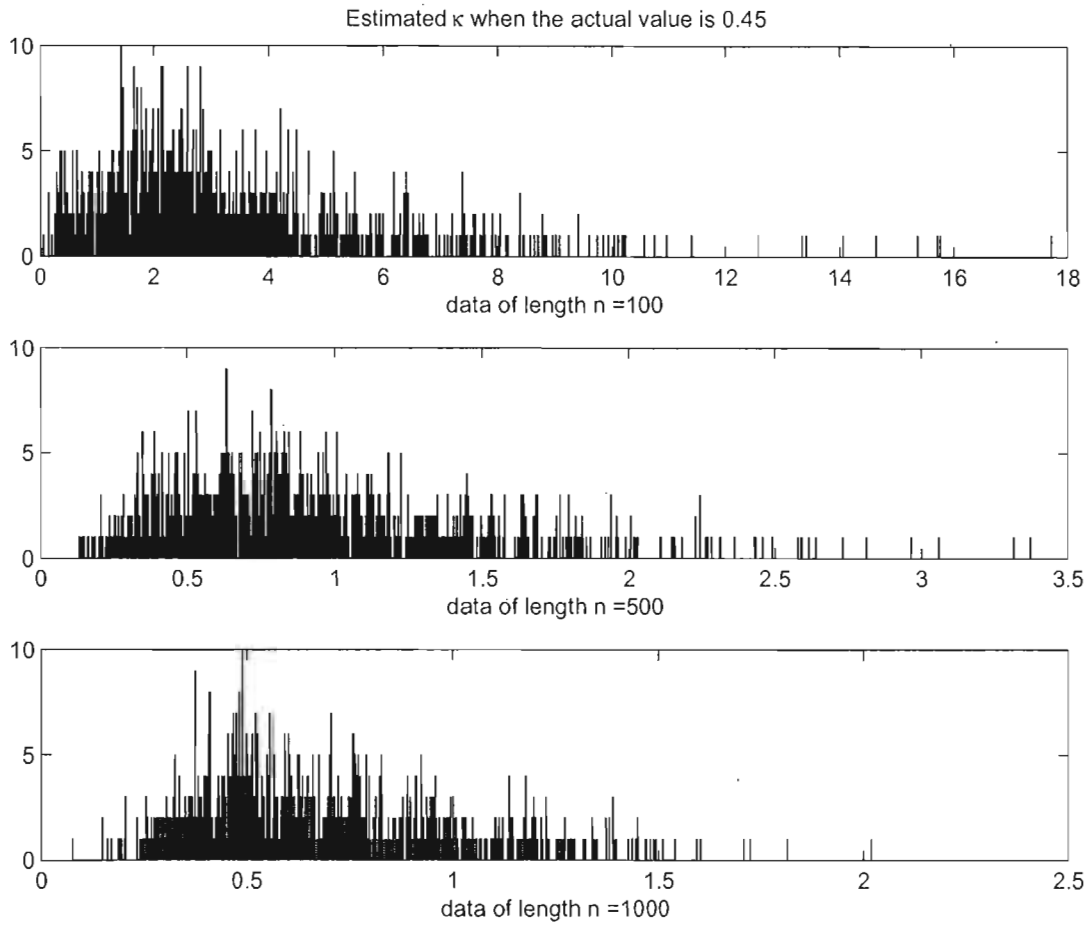

Figure A.8 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model.

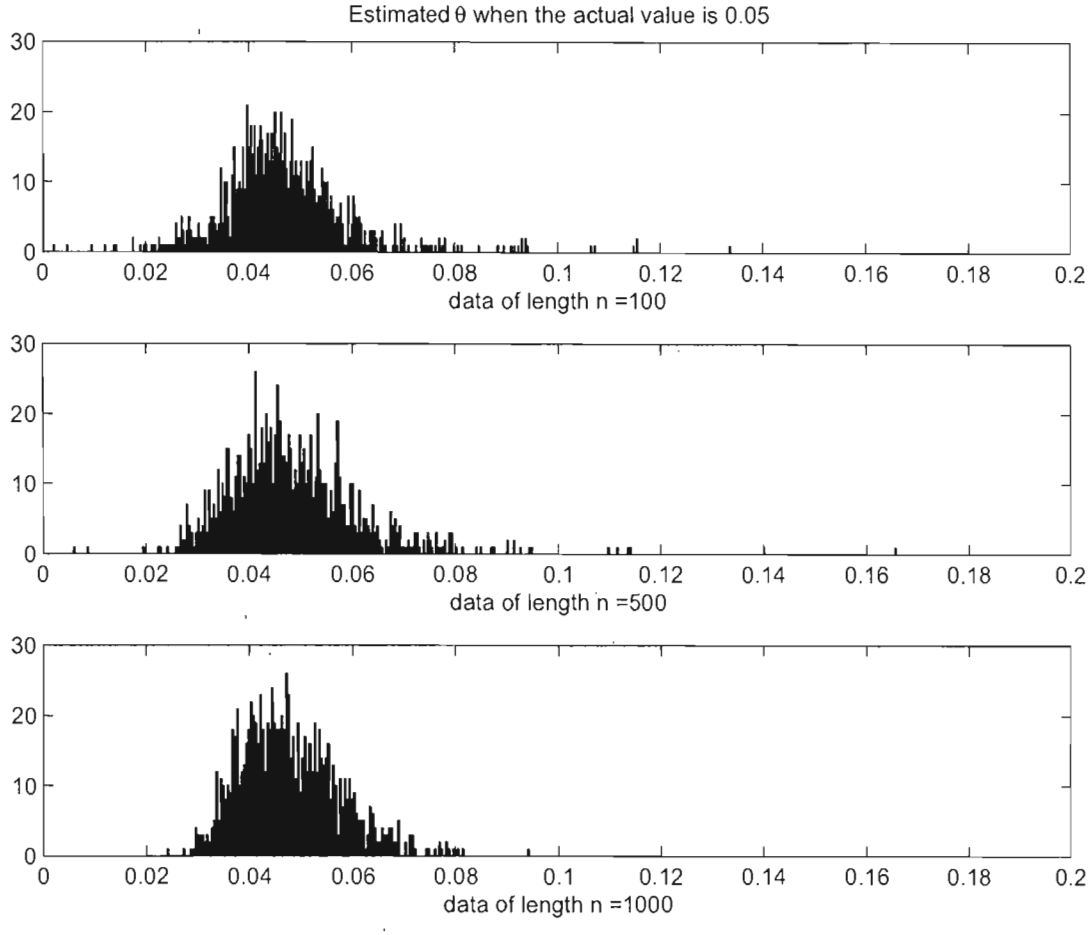

Figure A.9 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model.

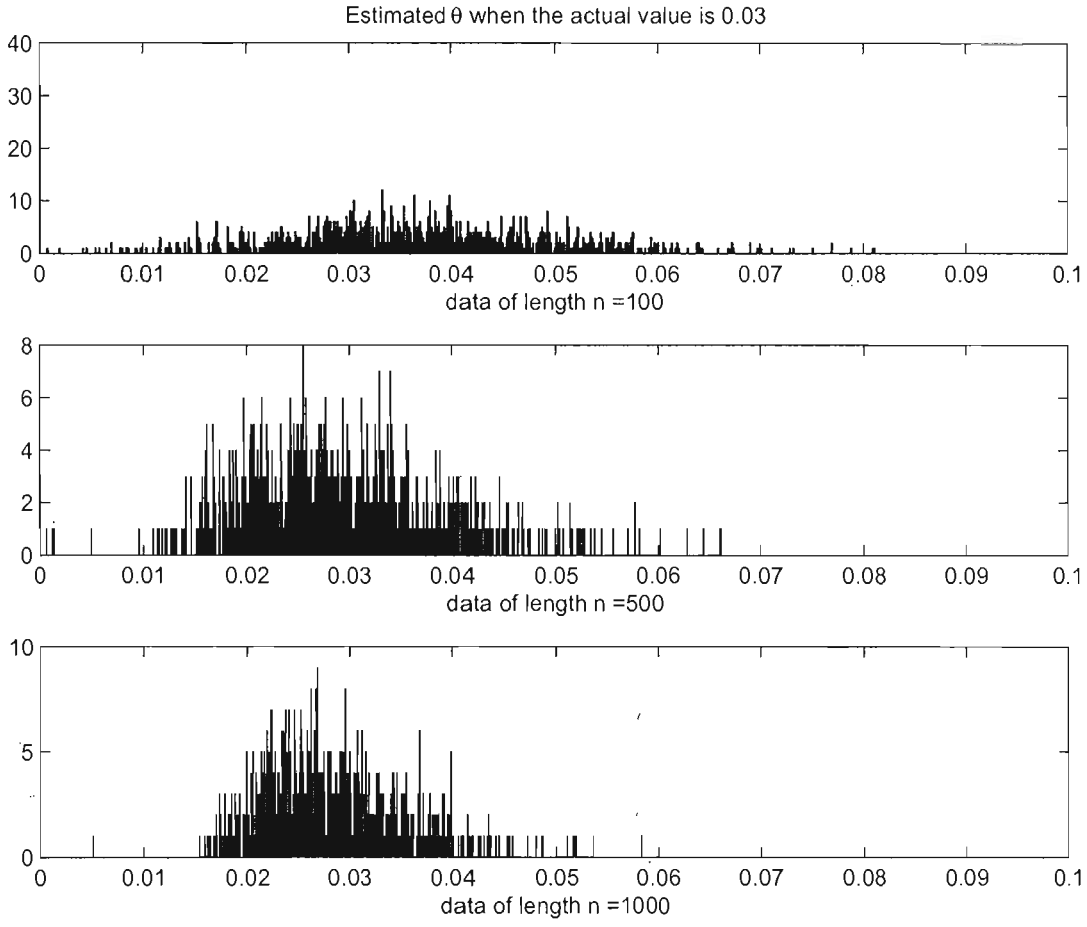

Figure A.10 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model.

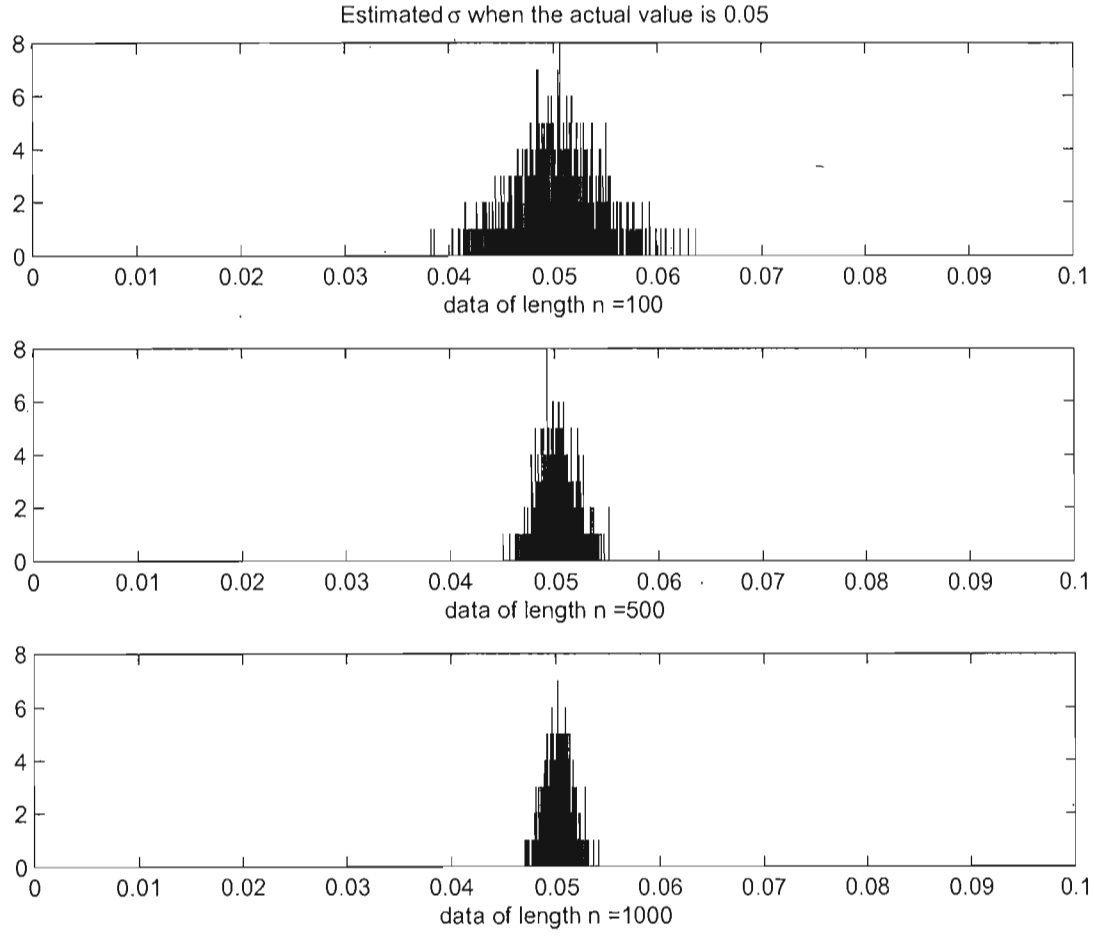

Figure A.11 Empirical distribution of MLE $\hat{\sigma}$ for the CIR model.

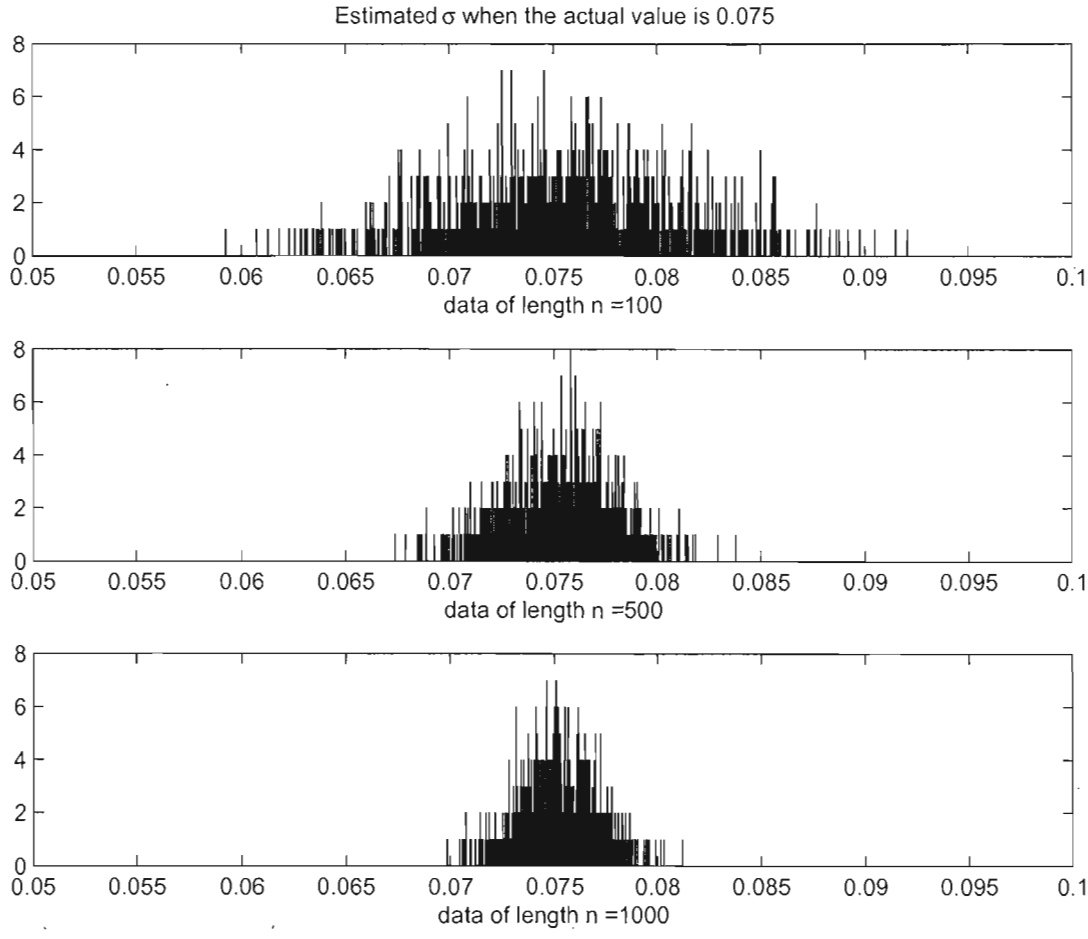

Figure A.12 Empirical distribution of MLE $\hat{\sigma}$ for the CIR model.

APPENDIX B

SIMULATION RESULTS FOR INDIRECT MAXIMUM LIKELIHOOD METHOD

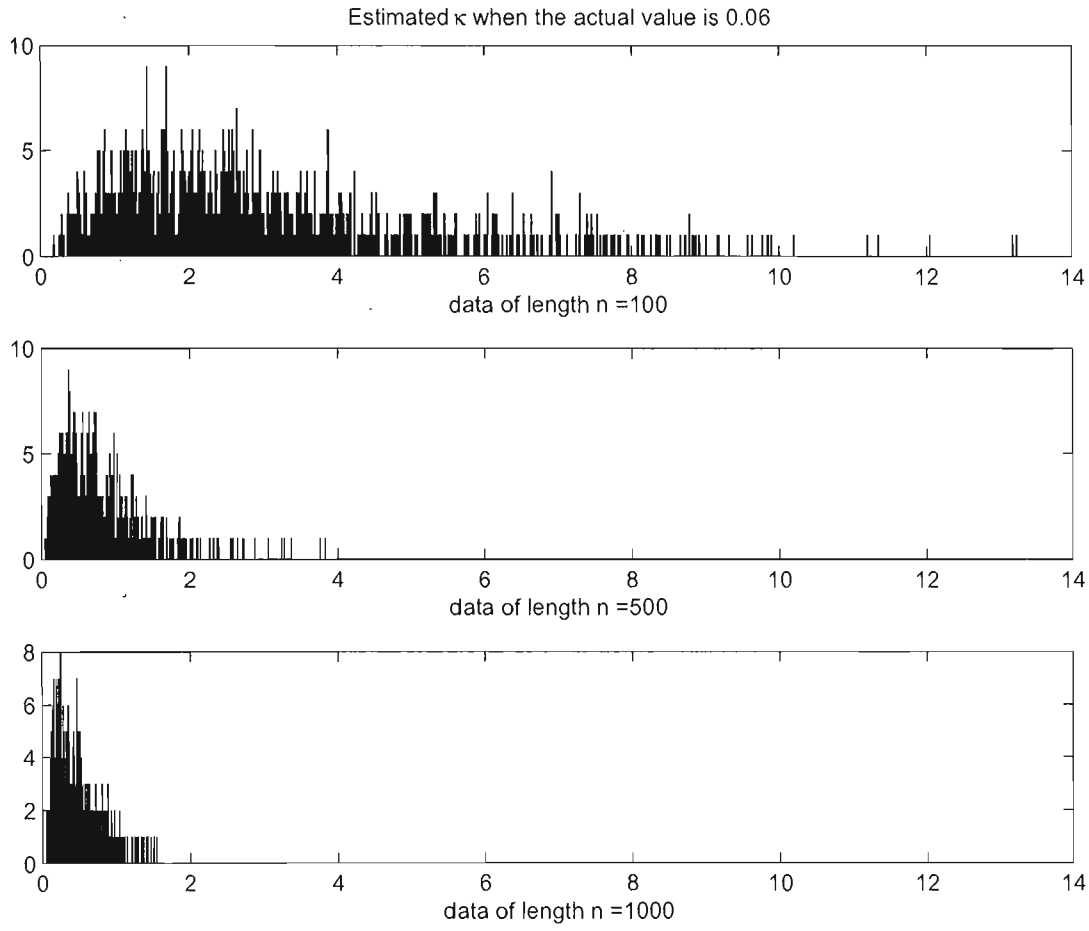

Figure B.1 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model (zero-coupon data).

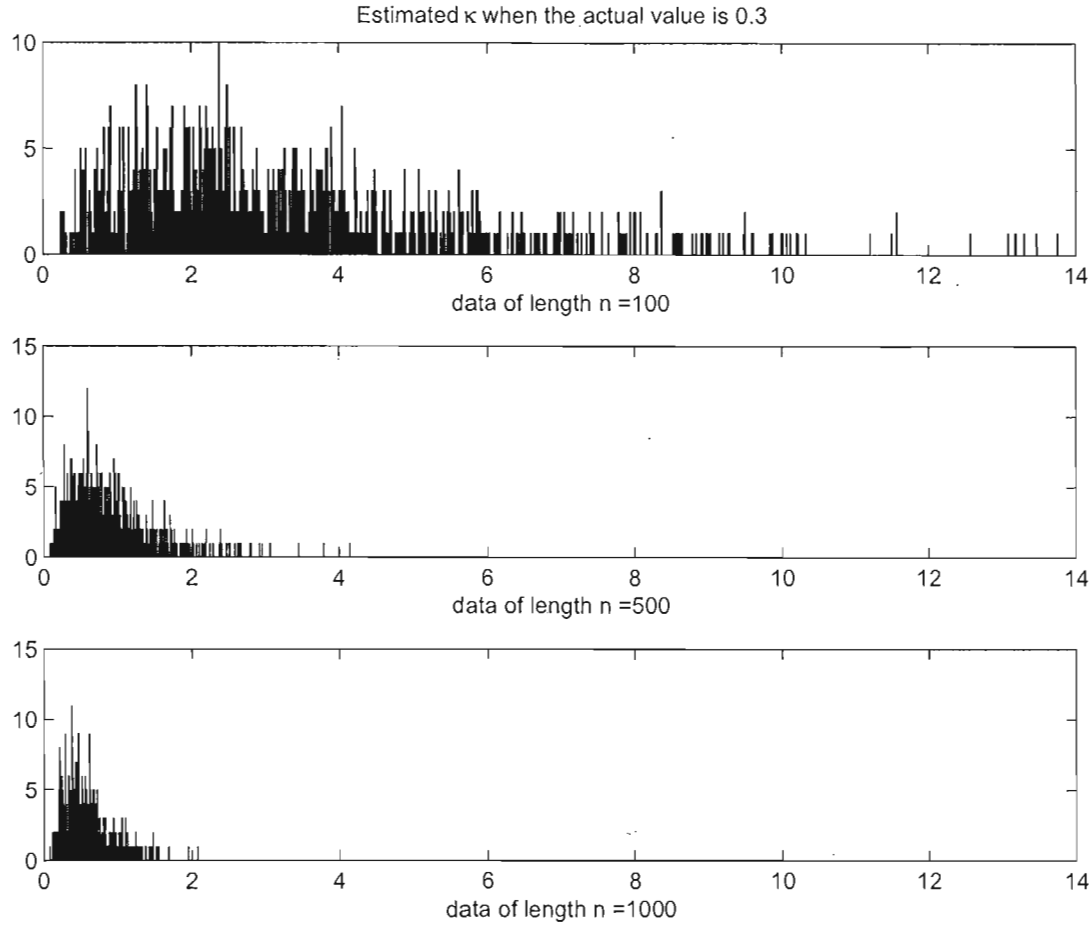

Figure B.2 Empirical distribution of MLE $\widehat{\kappa}$ for Vasicek's model (zero-coupon data).

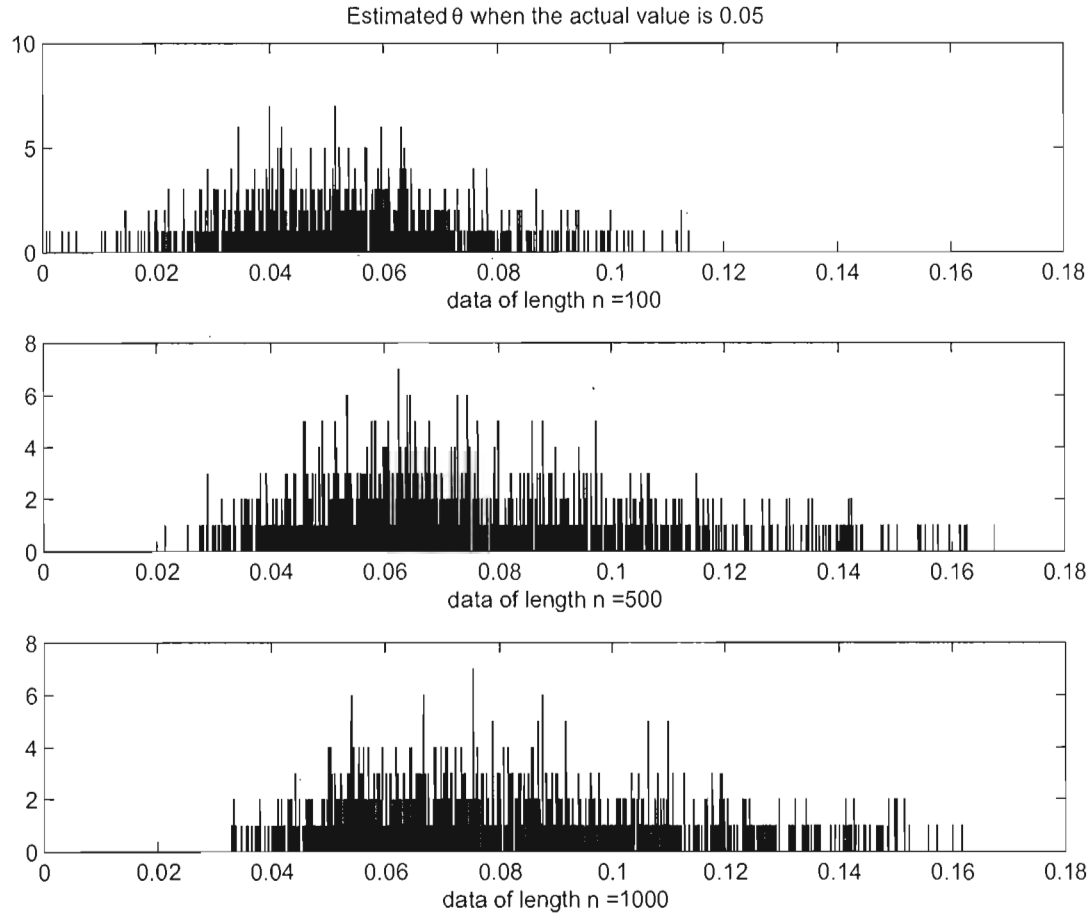

Figure B.3 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model (zero-coupon data).

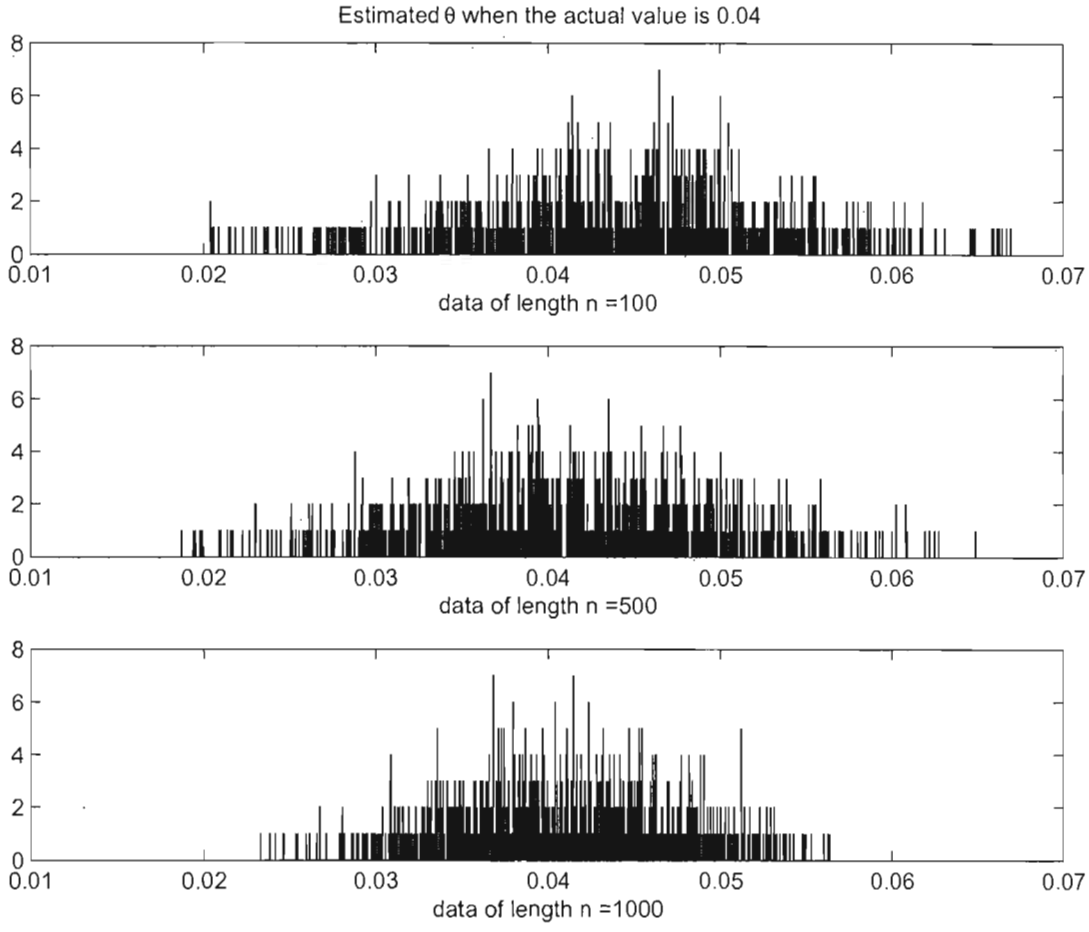

Figure B.4 Empirical distribution of MLE $\widehat{\theta}$ for Vasicek's model (zero-coupon data).

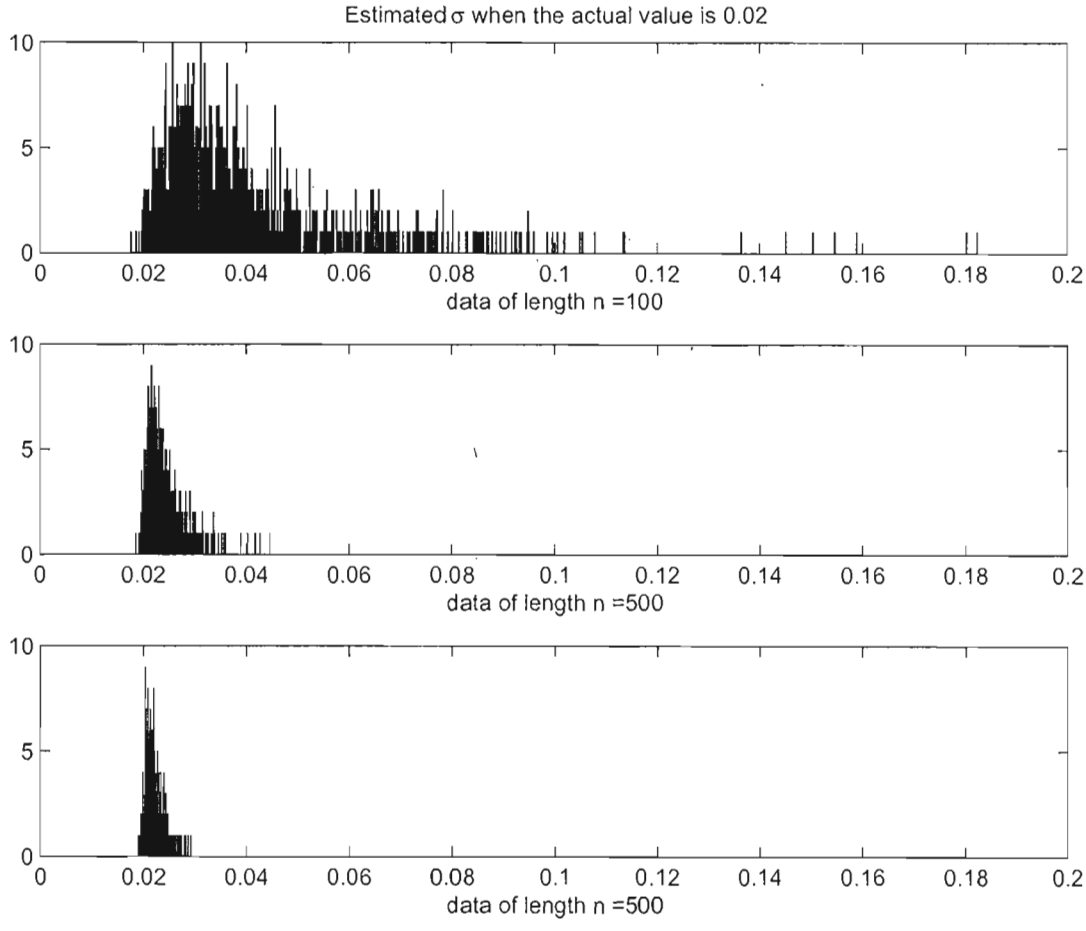

Figure B.5 Empirical distribution of MLE $\hat{\sigma}$ for Vasicek's model (zero-coupon data).

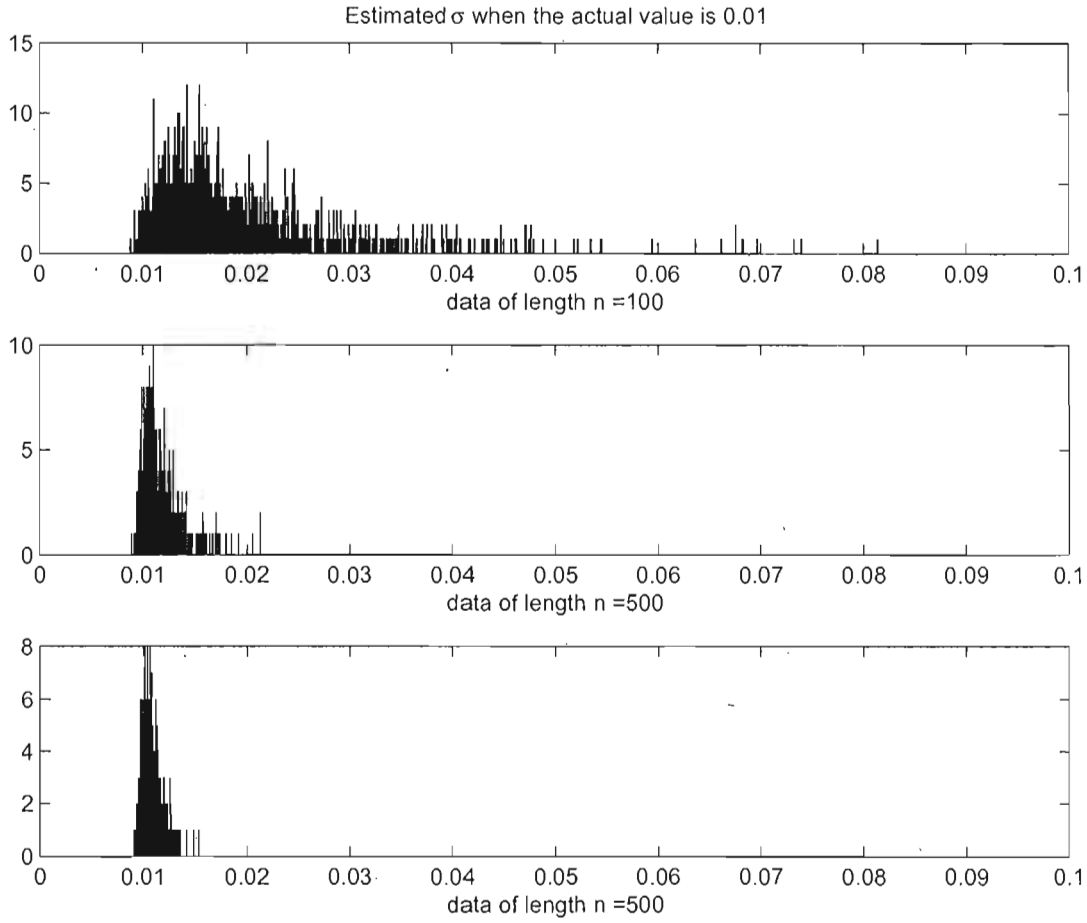

Figure B.6 Empirical distribution of MLE $\widehat{\sigma}$ for Vasicek's model (zero-coupon data).

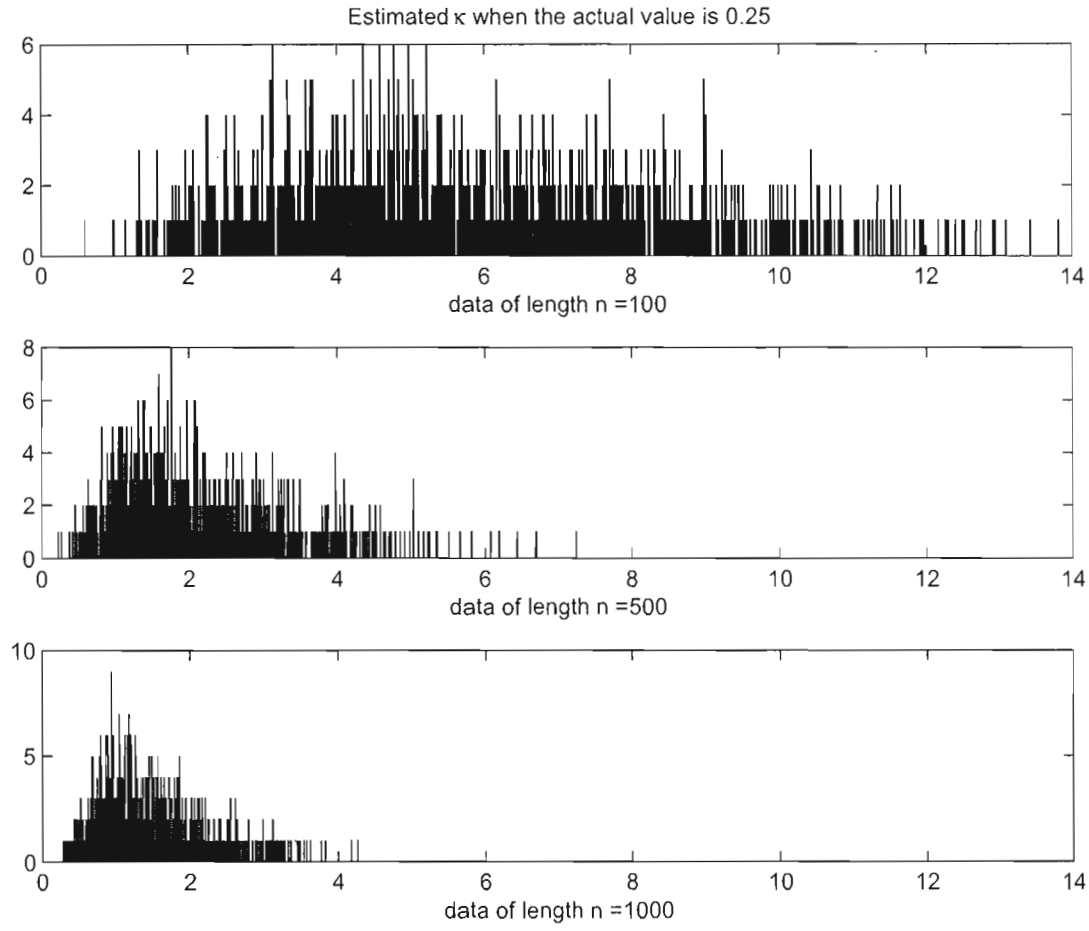

Figure B.7 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model (zero-coupon data).

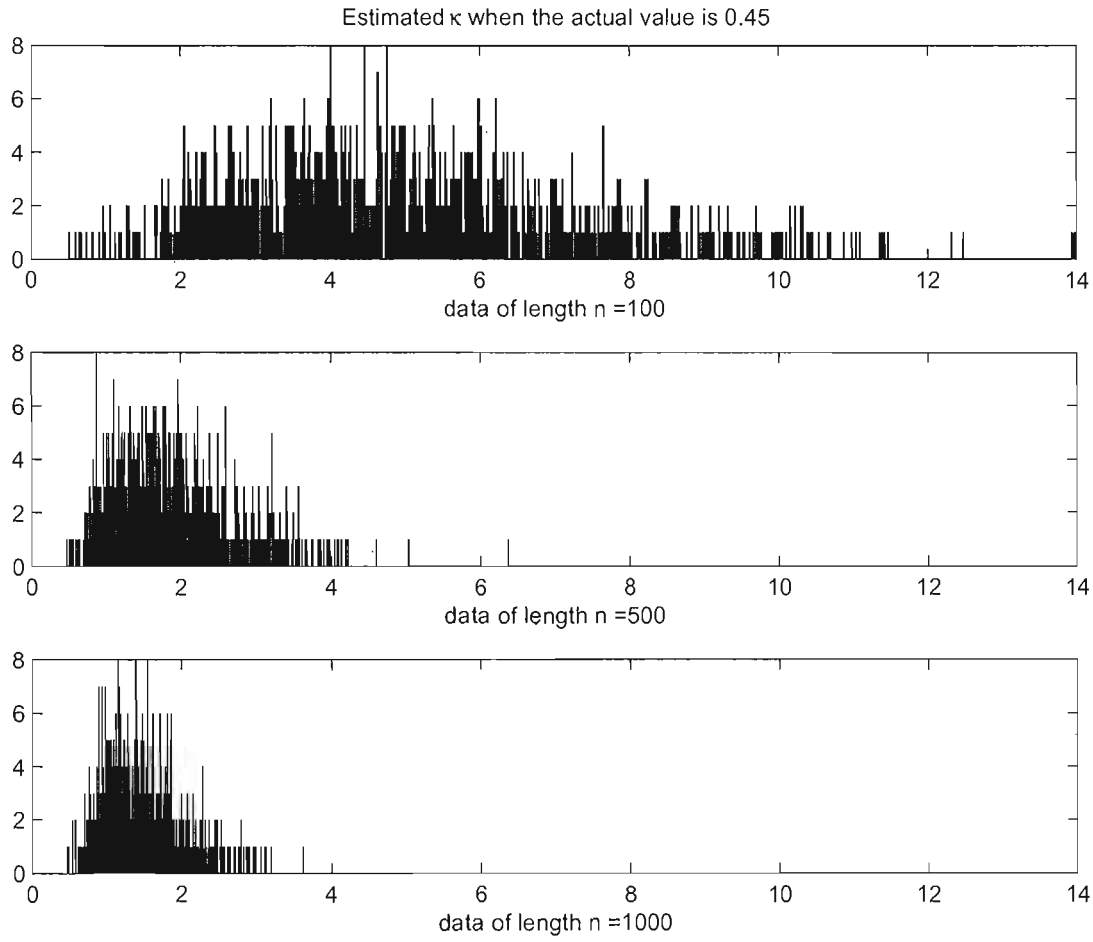

Figure B.8 Empirical distribution of MLE $\widehat{\kappa}$ for the CIR model (zero-coupon data).

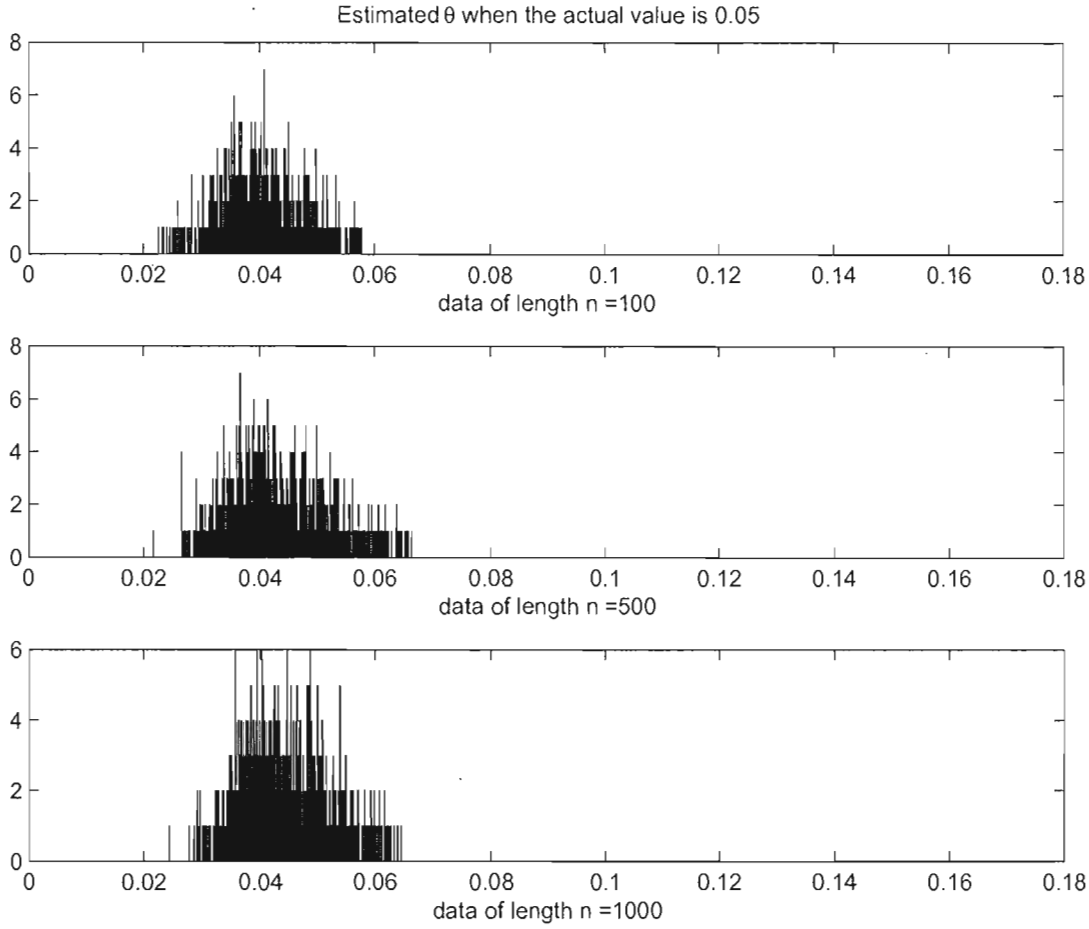

Figure B.9 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model (zero-coupon data).

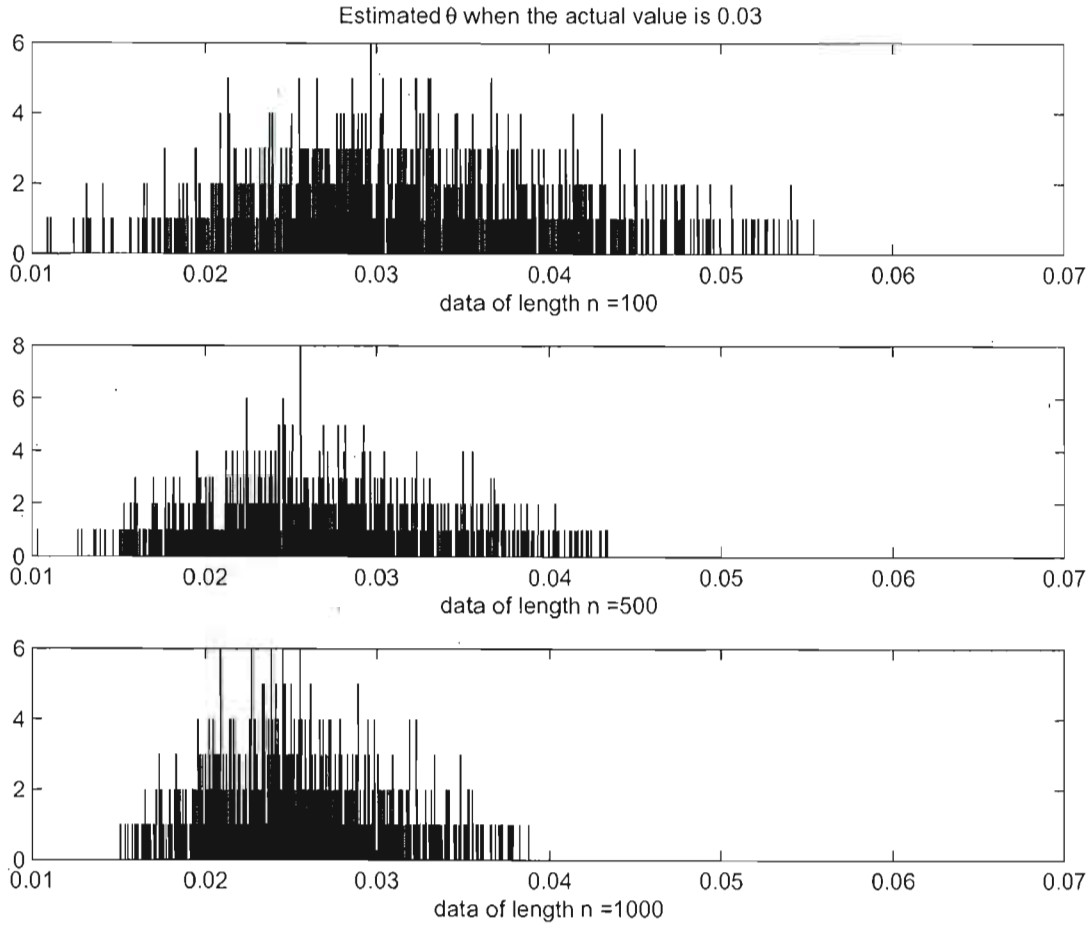

Figure B.10 Empirical distribution of MLE $\widehat{\theta}$ for the CIR model (zero-coupon data).

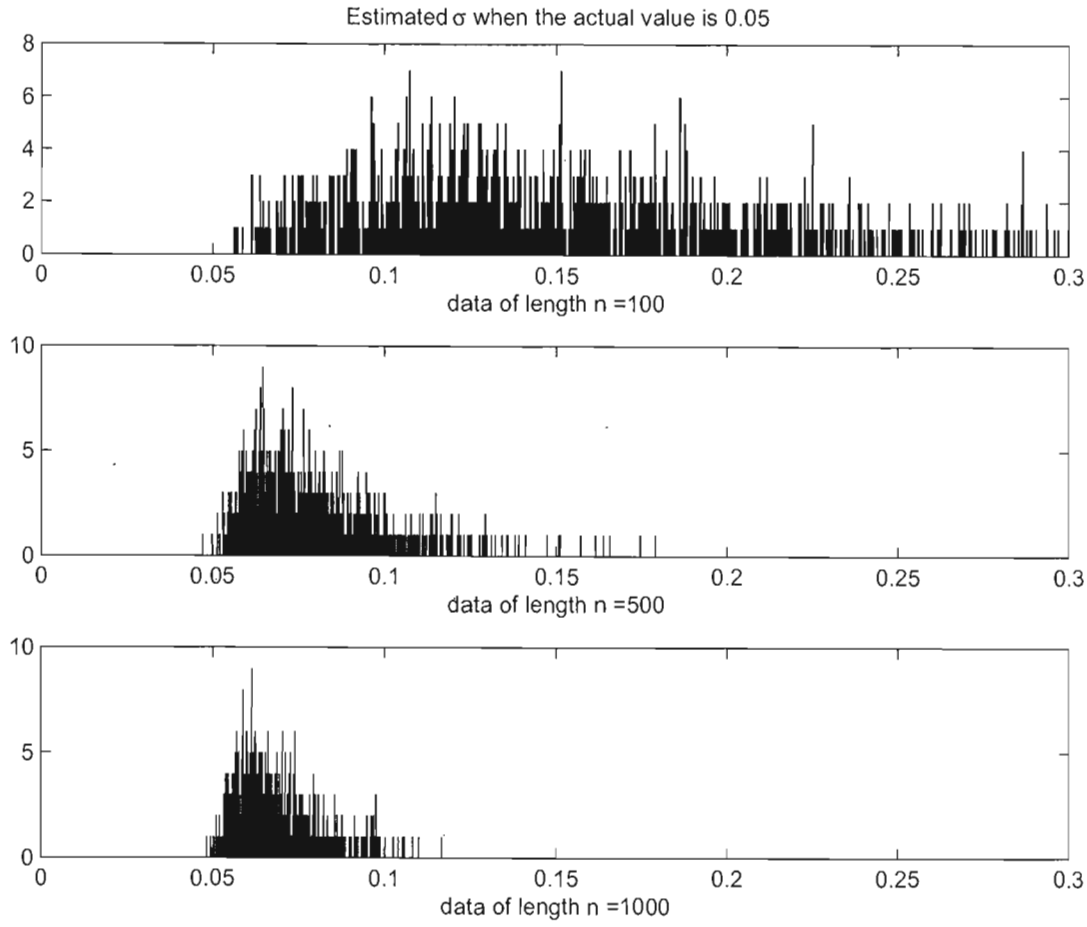

Figure B.11 Empirical distribution of MLE $\widehat{\sigma}$ for the CIR model (zero-coupon data).

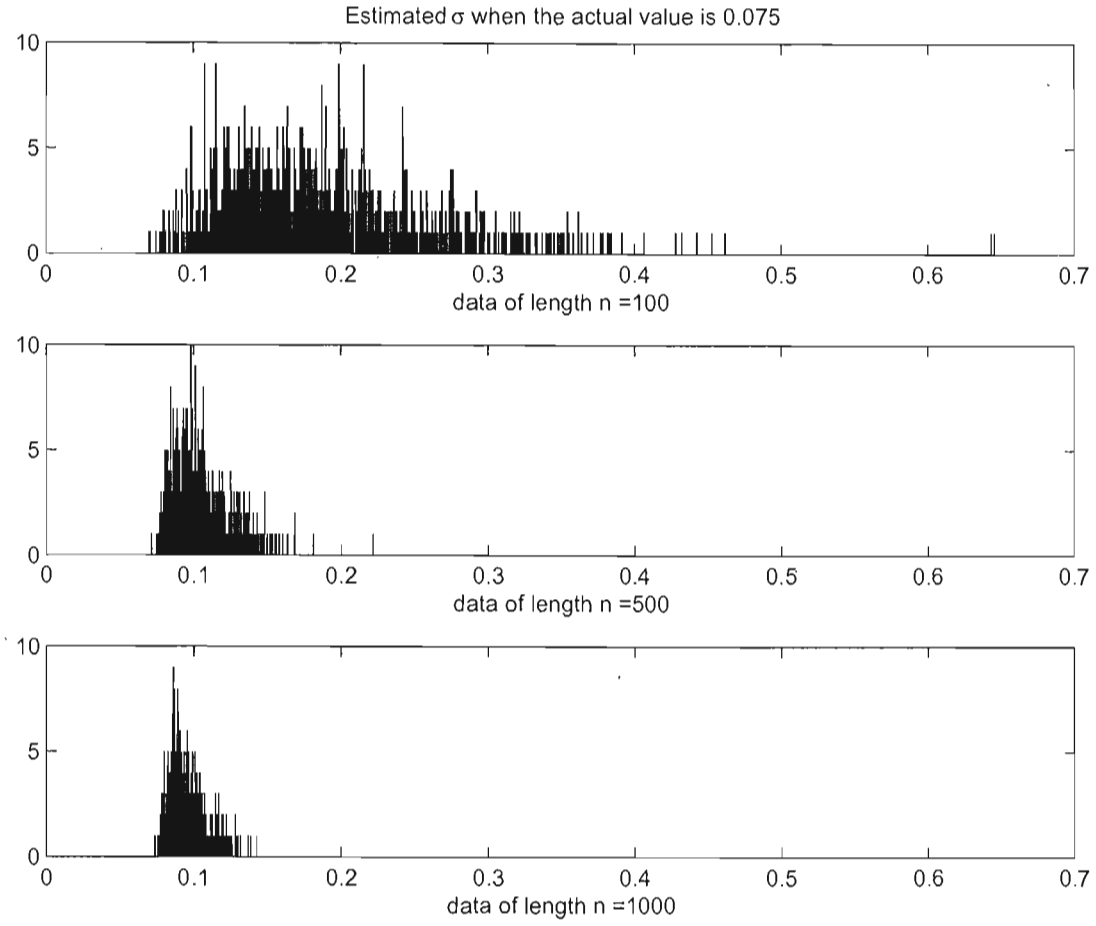

Figure B.12 Empirical distribution of MLE $\widehat{\sigma}$ for the CIR model (zero-coupon data).

APPENDIX C

SIMULATION RESULTS USING KALMAN FILTERING

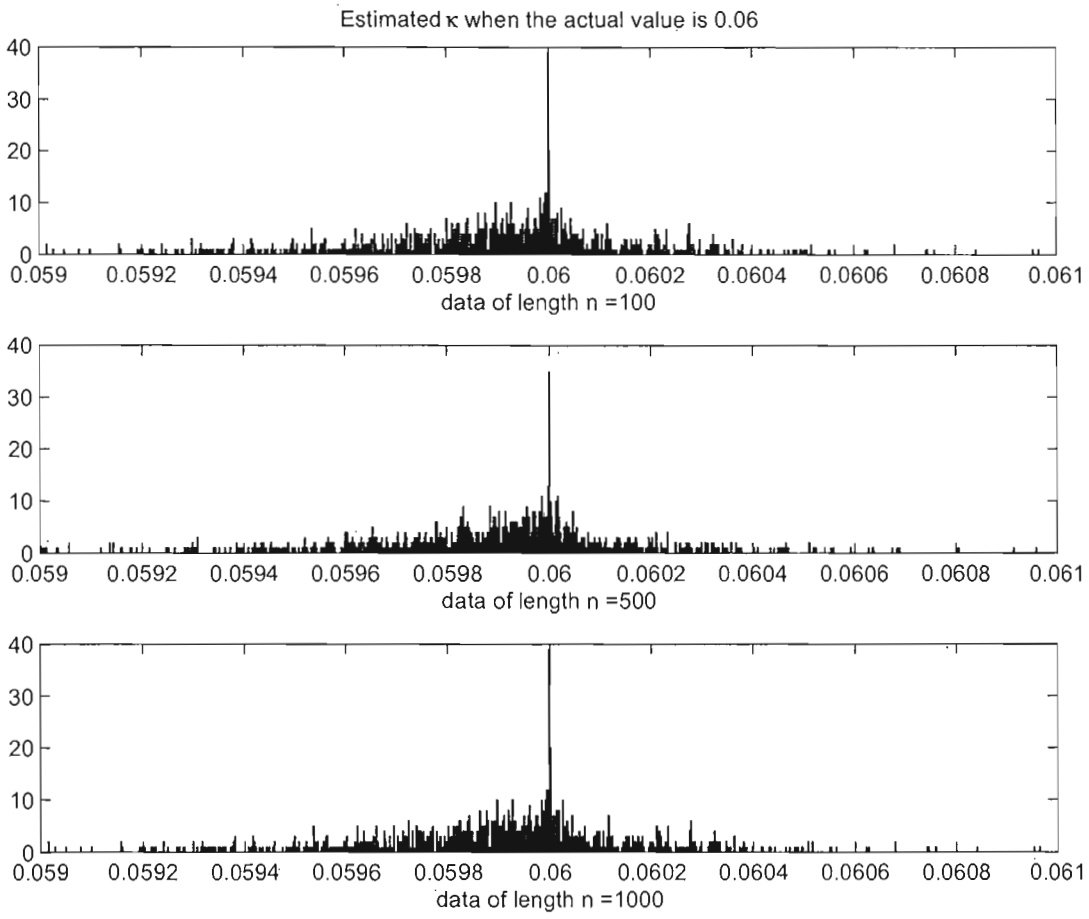

Figure C.1 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for Vasicek's model.

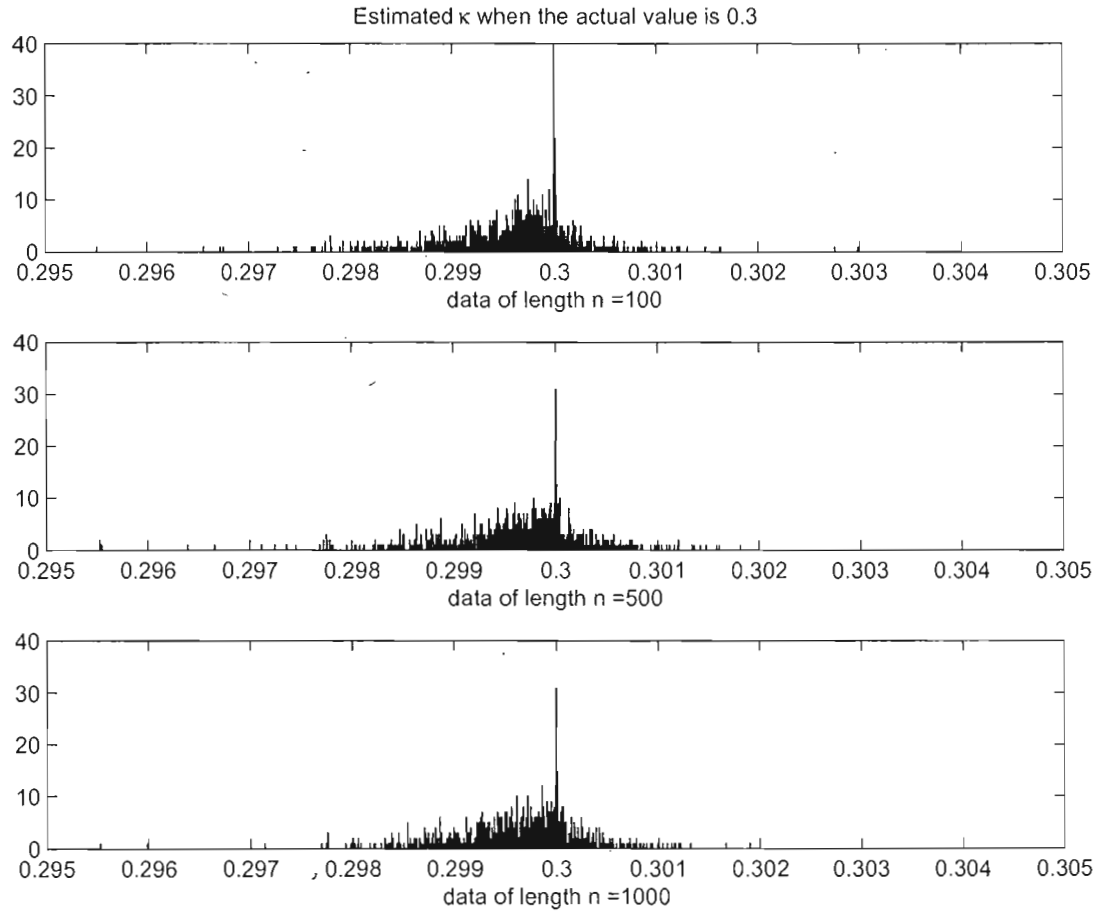

Figure C.2 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for Vasicek's model.

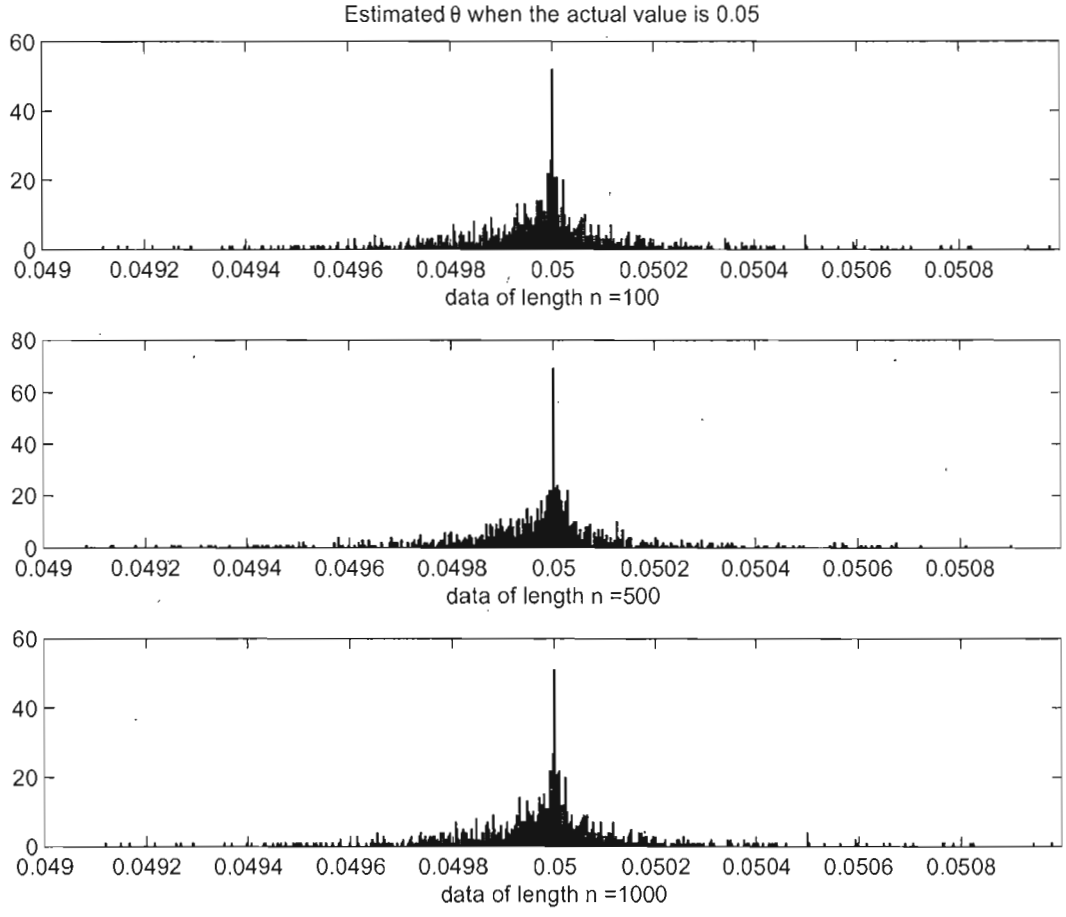

Figure C.3 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for Vasicek's model.

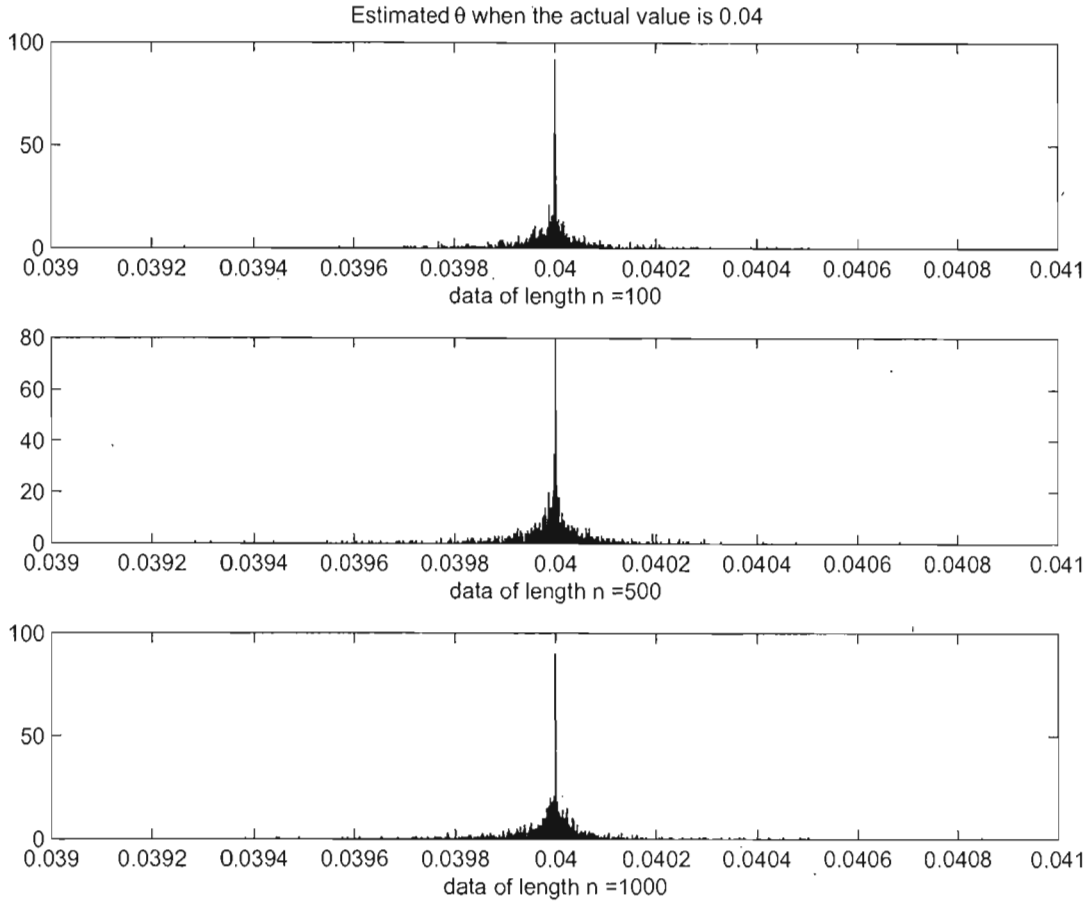

Figure C.4 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for Vasicek's model.

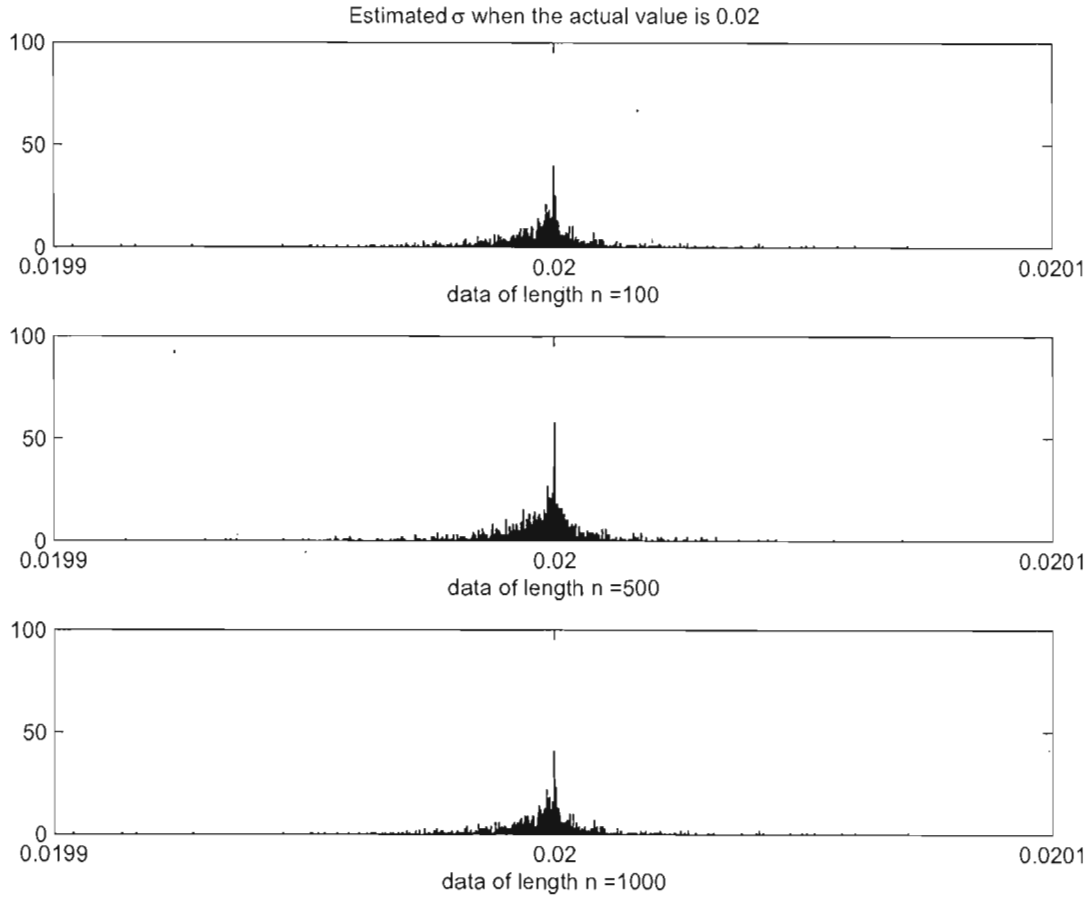

Figure C.5 Empirical distribution of Quasi-MLE $\hat{\sigma}$ for Vasicek's model.

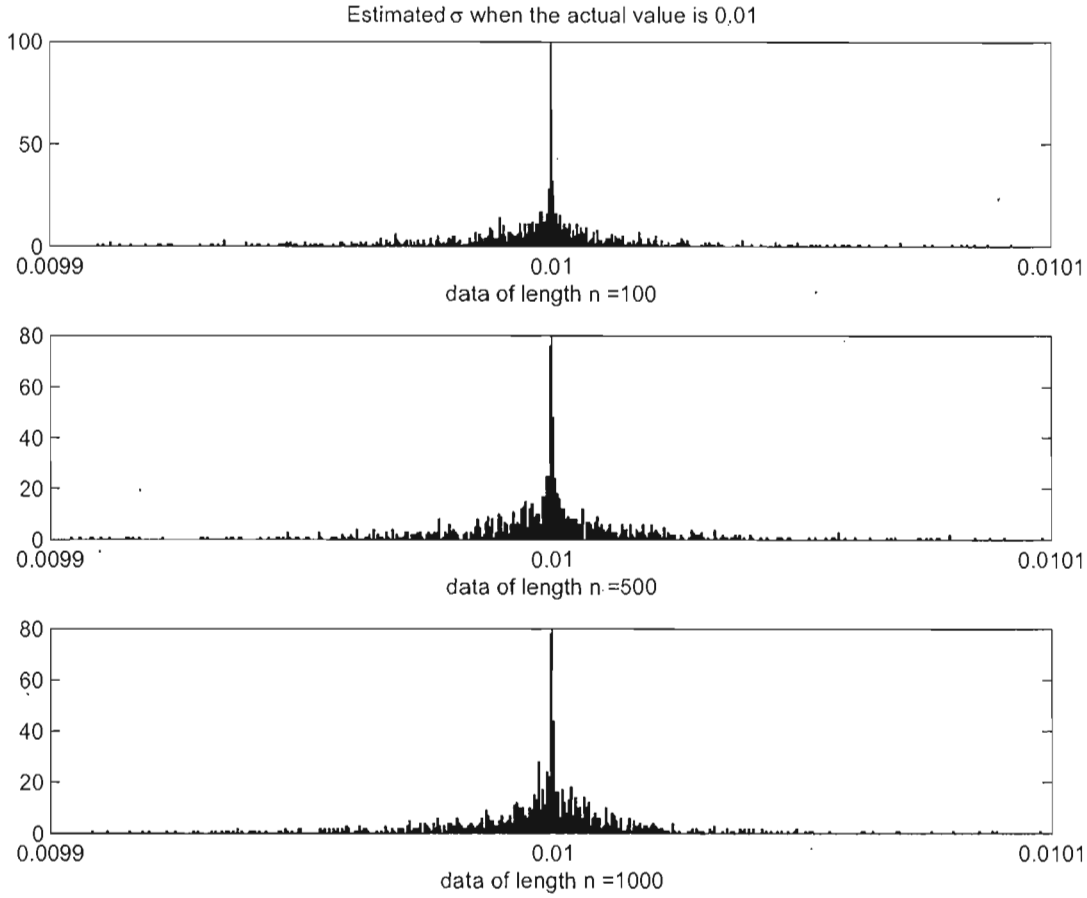

Figure C.6 Empirical distribution of Quasi-MLE $\hat{\sigma}$ for Vasicek's model.

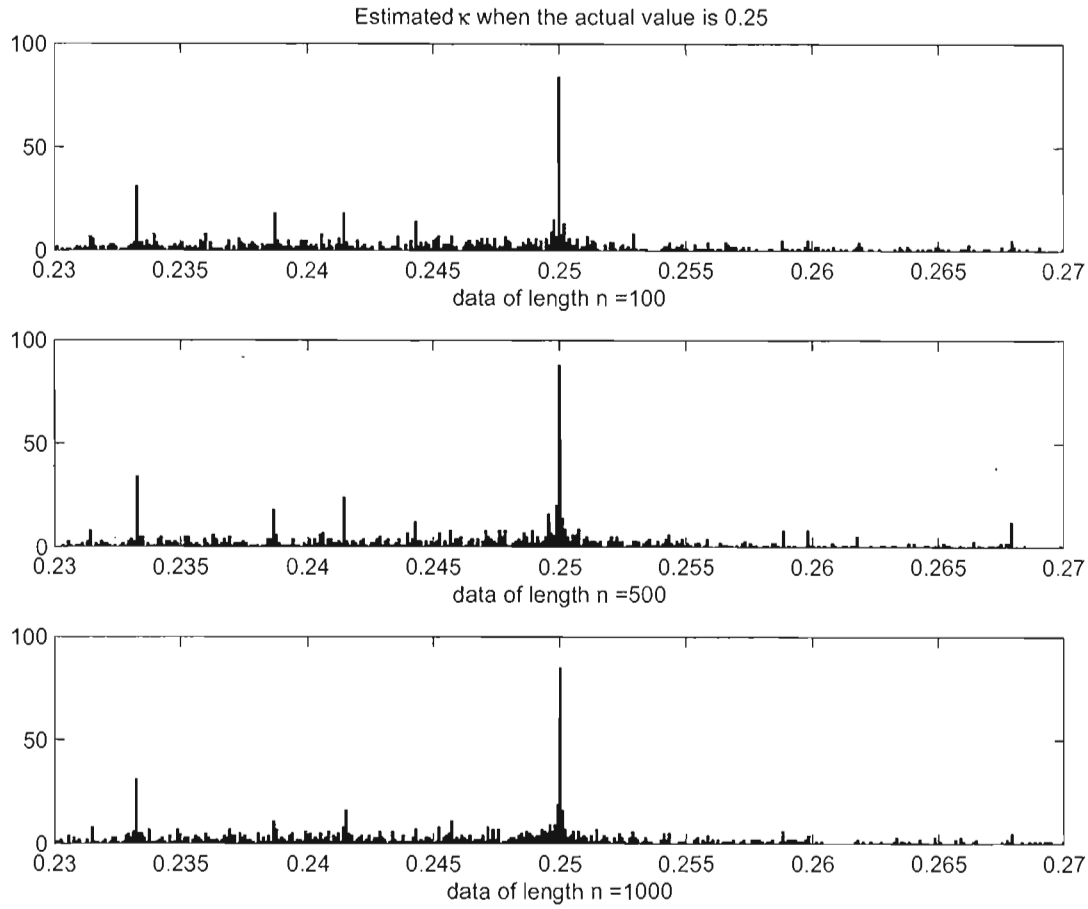

Figure C.7 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for the CIR model.

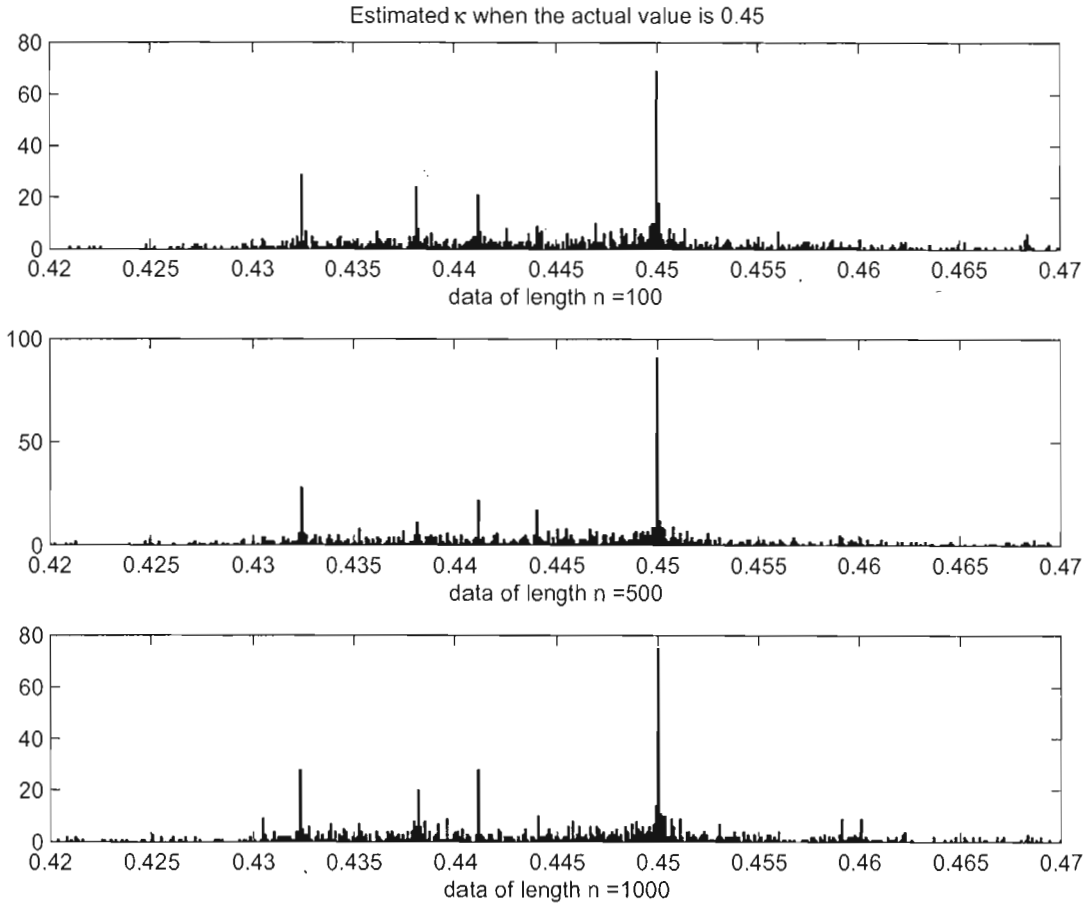

Figure C.8 Empirical distribution of Quasi-MLE $\widehat{\kappa}$ for the CIR model.

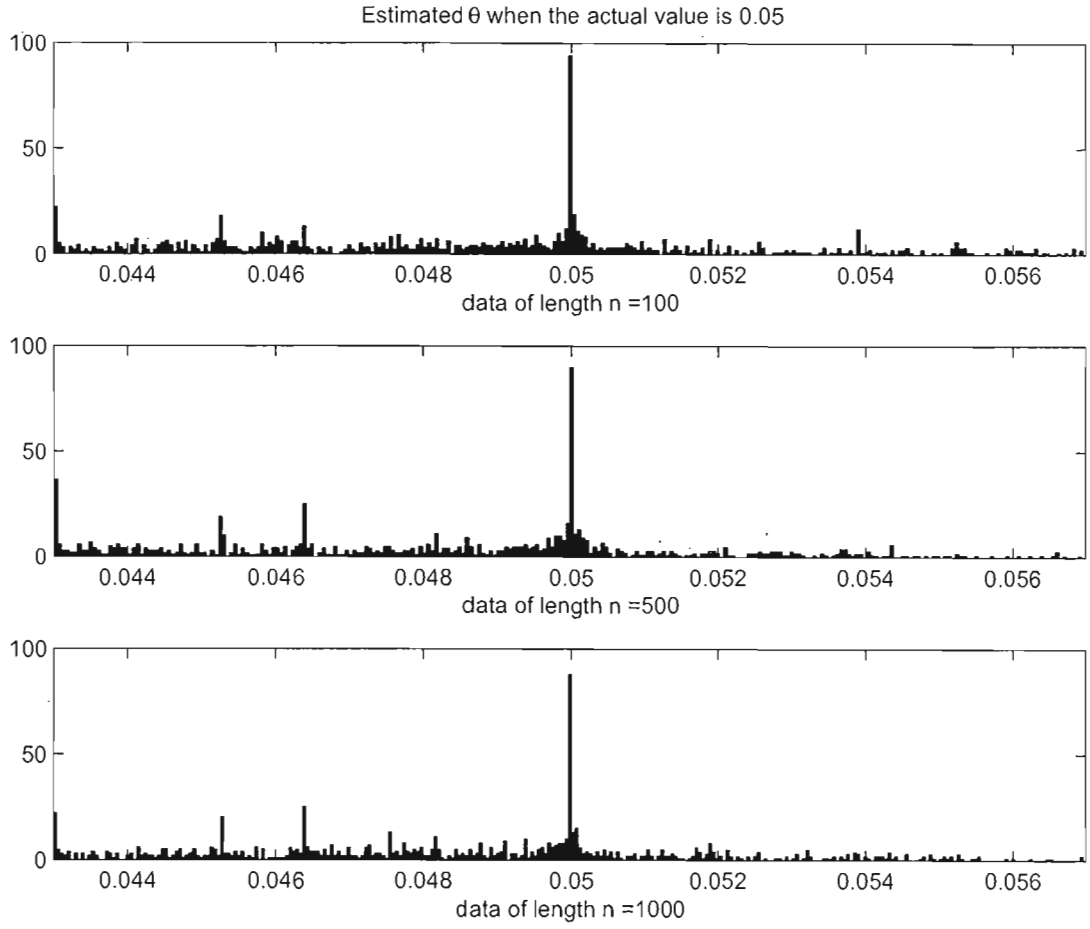

Figure C.9 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for the CIR model.

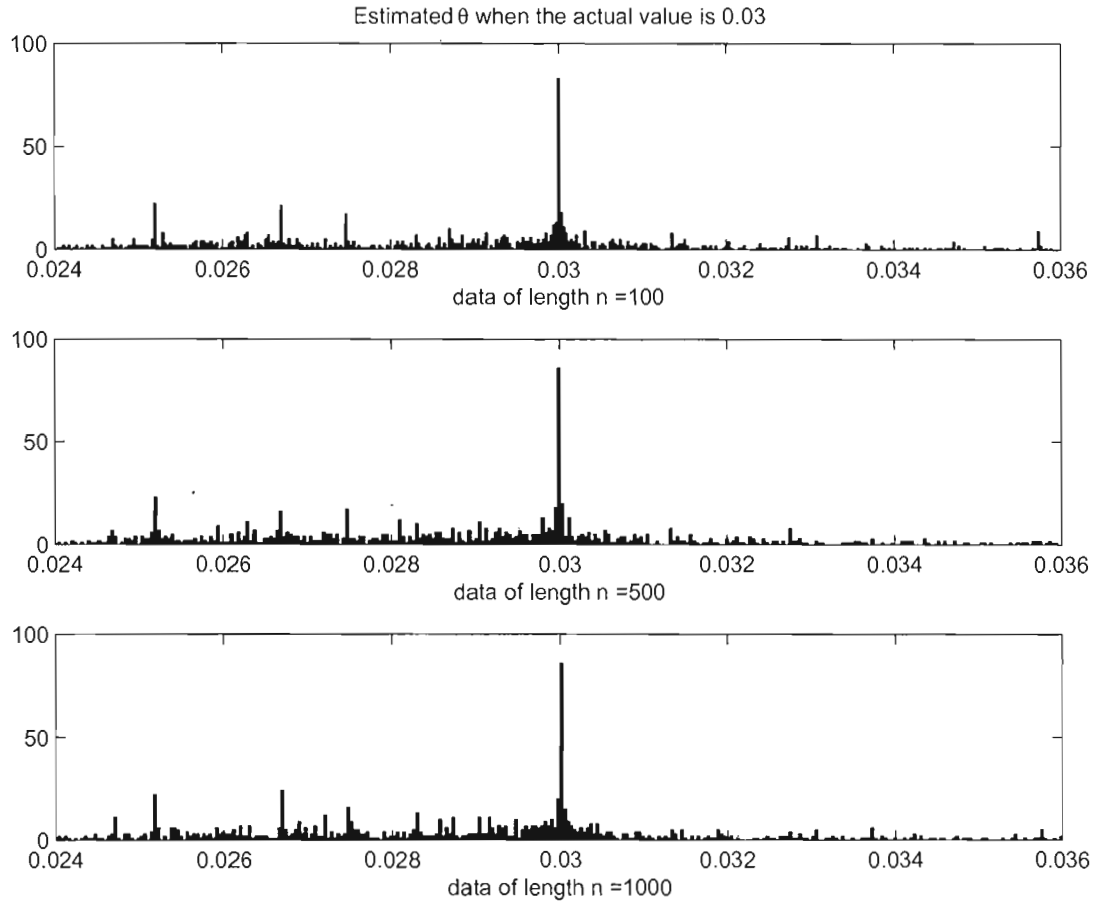

Figure C.10 Empirical distribution of Quasi-MLE $\widehat{\theta}$ for the CIR model.

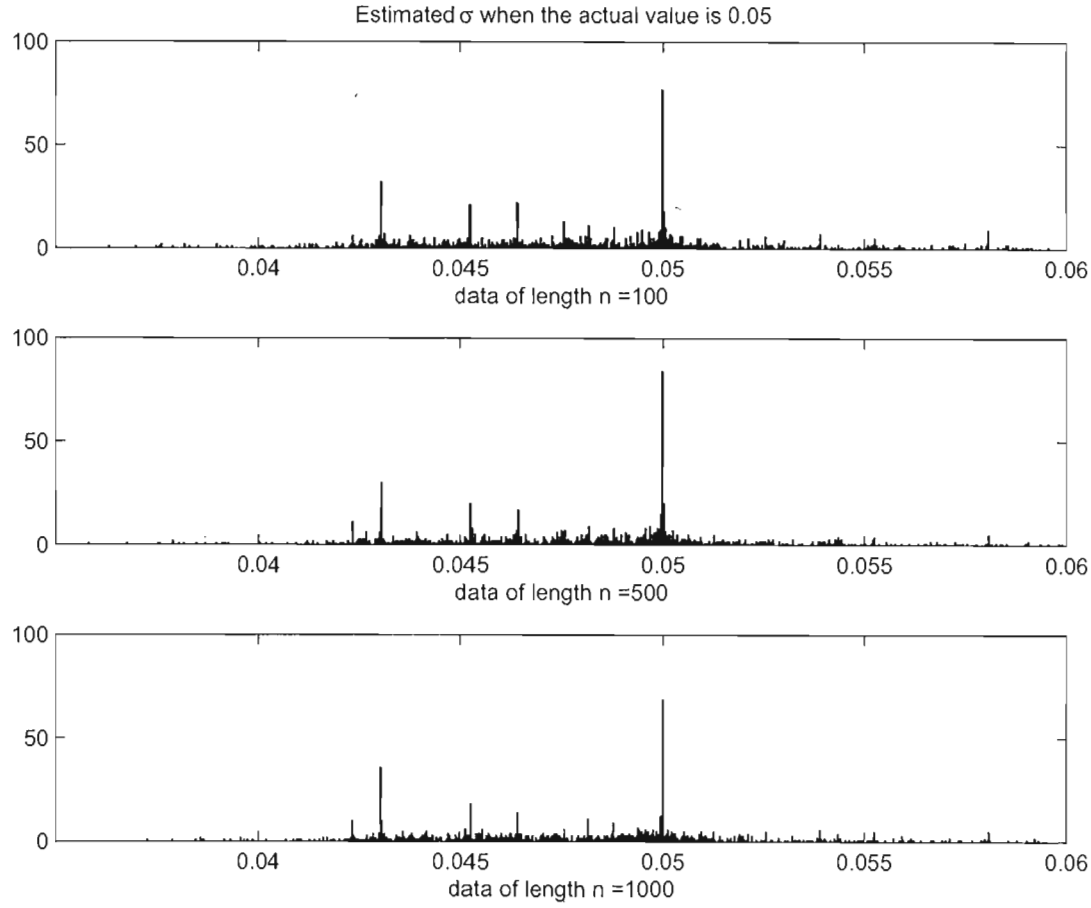

Figure C.11 Empirical distribution of Quasi-MLE $\hat{\sigma}$ for the CIR model.

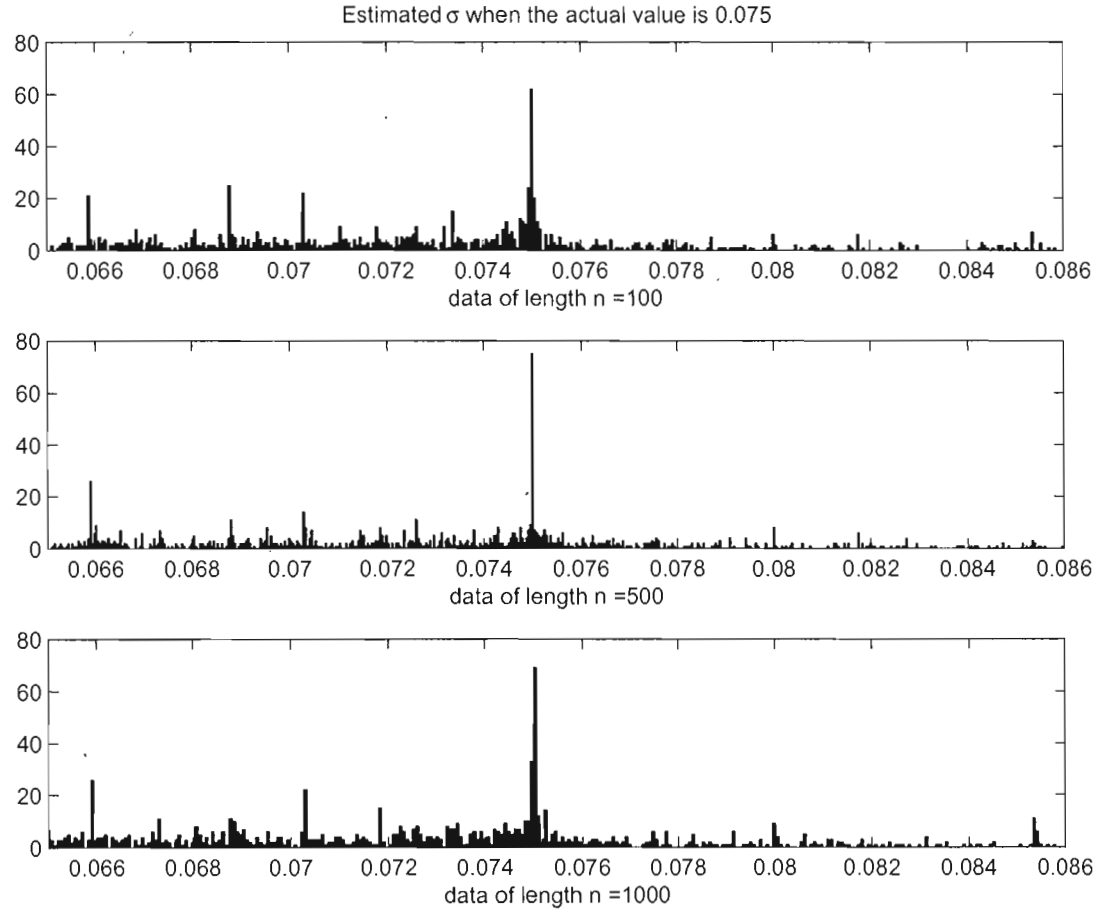

Figure C.12 Empirical distribution of Quasi-MLE $\hat{\sigma}$ for the CIR model.

APPENDIX D

MATLAB CODE FOR SIMULATION AND ESTIMATION

D.1 Code for Chapter II

D.1.1 Vasicek Model

```
% This function generates short rates by using Vasicek model
% Equation 2.2.12 - Page 20

function r = GenOU(n,r1,kappa,theta,sigma,Delta)

    r = zeros(n,1);
    r(1) = r1; % Initial value
    eps = randn(n,1);
    % The variance of the white noise
    v = sqrt((sigma^2/(2*kappa))*(1-exp(-2*kappa*Delta)));
    for i=2:n
        mu = theta+exp(-kappa*(Delta))*(r(i-1)-theta);
        r(i)= mu+v*eps(i-1);
    end

end


% Generate N=1000 Paths each with n=1000 observations
% Single Factor Vasicek Model

% Model parameters
kappa = 0.06; theta = 0.05; sigma = 0.02; % set 1
% kappa = 0.30; theta = 0.04; sigma = 0.01; % set 2

% Path parameters
Delta = 1/52;      % for weekly samples
                   % (for yearly, monthly and daily samples
                   %  Delta is 1, 1/12,1/250 respectively)
```

```
r1 = 0.045;        % Initial value for state space variable
N= 1000;           % Number of iterations
n = 1000;          % Number of points we observe the data in time

% Generate the path of interest rate
% To reset the randn function to its default setting in Matlat 2007a
% In Matlab 2008 we can use "reset(RandStream.getDefaultStream);"
randn('state',0);

n = 1000; % Number of points we observe the data in time
sg = 1;
while c <= N
    % Call the function GenOU to generate set of data
    % with Vasicek model
    r = GenOU(n,r1,kappa,theta,sigma,Delta);
    % To avoid generating negative paths
    if (min(r) > 0)
        Generated(c,:) = r;
        c= c+1;
    end
    sg = sg+1;
    if sg > N*5
        write 'Number of negative paths exceeds, ...
               the interest rates are no longer independent'
        break
    end
end

% the percentage of undesired paths (negative)
NegativePath = (sg-c)/100;

save 'set1OU.mat'


% Log-likelihood function for state variable using Vasicek model
% Equation 2.2.13 - Page 20

function y = DllhOU(x,r,Delta)

    % Starting points of optimization
    % Using exp(ln(initial values)) to avoid
    % negative estimated values
    kappa = exp(x(1));
    theta = exp(x(2));
    sigma = exp(x(3));
    n = max(size(r));

    % Constructing Log-Likelihood Function
    v = sqrt((sigma^2/(2*kappa))*(1-exp(-2*kappa*Delta)));
```

```
    mu = theta+exp(-kappa*(Delta))*(r(1:n-1)-theta);
    y = sum(log(normpdf(r(2:n),mu,v)));

    % To find Maximum log-likelihood we minimize the
    % negative of log-likelihood since Optimization
    % package use minimizing procedure.
    y = -y;

end


% MLE estimation for state variables generated by Vasicek model

load 'set1OU.mat'

for c=1:N % N is number of itteration
    r = Generated(c,1:n); % n is the number of observations
    % Using natural logarith of starting points
    % to avoid negative estimated values
    x0 = log([ kappa theta sigma ]); % starting points

    % Call the optimization routine
    options = optimset('TolFun',1e-15,'TolX',1e-15, ...
                       'MaxIter',1000,'MaxFunEvals',1000);
    % The minimization of negative log-likelihood function
    % can be achieved by using fminsearch
    [ x, fval, exitflag, output ] ...
        = fminsearch(@(x) DllhOU(x,r,Delta),x0,options);
    Result(c,:) = exp(x);
    Exitflag(c,:) = exitflag;
    FvaL(c,:) = fval;
    clear out fval ans x r
end

save 'res1OU-1000.mat'


% Removing outliers in results by using box-and-whisker diagram

function [x r]= remoutliers(x,r)

    h=boxplot(x); % Box Plot
    ou=get(h(7,2),'ydata');
    for i=1:max(size(ou))
        [row column] = find(x==ou(i));
        x(row,:)=[];
        r(row,:)=[];
    end
```

```
end
```

D.1.2 CIR Model

```
% This function generates short rates by using CIR model
% Equation 2.3.12 - Page 22

function r = GenCIR(n,r1,kappa,theta,sigma,Delta)

    r = zeros(n,1);
    r(1) = r1; % Initial value
    L= sigma*sigma/(4*kappa)*(1-exp(-kappa*Delta));
    % Degrees of freedom
    v = 4*kappa*theta/(sigma*sigma);
    for i=2:n
        % Non-centrality parameter
        nc = 4*kappa*r(i-1)/((sigma*sigma)*(exp(kappa*Delta)-1));
        r(i) = ncx2rnd(v,nc,1,1)*L;
    end

end


% Generate N=1000 Paths each with n=1000 observations
% Single Factor Vasicek Model

% Model parameters
kappa = 0.25; theta = 0.05; sigma = 0.05; % set 1
% kappa = 0.45; theta = 0.03; sigma = 0.075; % set 2

% Path parameters
Delta = 1/52; % for weekly samples
              % (for yearly, monthly and daily samples Delta
              % is 1, 1/12,1/250 respectively)
r1 = 0.045; % Initial value for state space variable
N= 1000; % Number of iterations
n = 1000; % Number of points we observe the data in time

% Generate the path of interest rate
% To reset the randn function to its default setting in Matlab 2007a
% In Matlab 2008 we can use "reset(RandStream.getDefaultStream);"

c=1;
sg = 1;
while c <= N
    % Call the function GenOU to generate set of data with CIR model
    r = GenCIR(n,r1,kappa,theta,sigma,Delta);
    % To avoid generating negative paths
```

```
    if (min(r) > 0)
        Generated(c,:) = r;
        c= c+1;
    end
    sg = sg+1;
    if sg > N*5
        write 'Number of negative paths exceeds, ...
               the interest rates are no longer independent'
        break
    end
end

% the percentage of undesired paths (negative)
NegativePath = (sg-c)/100;

save 'set2CIR.mat'


% Log-likelihood function for state variable using CIR model
% Equation 2.3.15 - Page 22

function y = DllhCIR(x,r,Delta)

    % Starting points of optimization
    % Using exp(ln(initial values)) to avoid
    % negative estimated values
    kappa = exp(x(1));
    theta = exp(x(2));
    sigma = exp(x(3));
    n = max(size(r));

    % Constructing Log-Likelihood Function
    % Using non-central chi-square distribution
    L = 2*kappa/(sigma*sigma*(1-exp(-kappa*Delta)));
    % Degrees of freedom
    v = 4*kappa*theta /(sigma*sigma);
    % Non-centrality parameter
    nc = 4*kappa*r(1:n-1)/((sigma*sigma)*(exp(kappa*Delta)-1));
    y = sum(log(ncx2pdf(r(2:n)/L, v, nc)/L));

    % To find Maximum log-likelihood we minimize
    % the negative of log-likelihood
    y=-y;
end

% MLE estimation for state variables generated by CIR model

load 'set1CIR.mat'
```

```
for c=1:N % N is number of iteration
    r = Generated(c,1:n); % n is the number of observations
    % Using natural logarithm of starting points
    % to avoid negative estimated values
    x0 = log([ kappa theta sigma ]);

    % Call the optimization routine
    options = optimset('TolFun',1e-15,'TolX',1e-15, ...
        'MaxIter',1000,'MaxFunEvals',1000);
    % The minimization of negative log-likelihood function
    % can be achieved by using fminsearch
    [ x, fval, exitflag, output ] ...
        = fminsearch(@(x) DllhCIR(x,r,Delta),x0,options);
    Result(c,:) = exp(x);
    Exitflag(c,:) = exitflag;
    FvaL(c,:) = fval;
    clear out fval ans x r
end

save 'res1CIR-100.mat'
```

D.2 Code for Chapter III

D.2.1 Vasicek Model

```
%
% Compute the prices from simulated state variables
% By using affine model introduced for Vasicek model
%

function P = PriceOU(r,kappa,theta,sigma,lambda,tau)

    % Equations 3.2.2 - page 33
    gam = kappa*kappa*(theta-(sigma*lambda/kappa))-(sigma*sigma)/2;
    B = (1-exp(-kappa*tau))/kappa;
    A = (gam*(B-tau)/(kappa*kappa))-(sigma*sigma*B*B/(4*kappa));
    % Equation 3.1.1 - page 32
    P = exp(A-B*r);

end


% Assuming that one can't observe state variables in reality,
% We use affine model to predict unobserved state variables
% When start points  are kappa, theta, sigma.

function rhat = FindROU(P,kappa,theta,sigma,lambda,tau)
```

```
    % Recalculate affine model elements with start point
    gam = kappa*kappa*(theta-(sigma*lambda/kappa))-(sigma*sigma)/2;
    B = (1-exp(-kappa*tau))/kappa;
    A = (gam*(B-tau)/(kappa*kappa))-(sigma*sigma*B*B/(4*kappa));
    % Equation 3.4.8 to find state variables from prices
    rhat = (A-log(P))/B;

end


% Log-likelihood function for prices using Vasicek model
% Equation 3.4.9 - Page 36

function y = ZCllhOU(x,P,lambda,tau,Delta)

    % Starting points of optimization
    % Using exp(ln(initial values)) to
    % avoid negative estimated values
    kappa = exp(x(1));
    theta = exp(x(2));
    sigma = exp(x(3));
    n = max(size(P));

    % Estimate unobserved state variable by using observed prices
    rh = FindROU(P,kappa,theta,sigma,lambda,tau);

    % Constructing Log-Likelihood Function using normal distribution
    v = sqrt((sigma^2/(2*kappa))*(1-exp(-2*kappa*Delta))); % variance
    mu = theta+exp(-kappa*(Delta))*(rh(1:n-1)-theta); % mean
    % log-likelihood function of estimated state variables
    y = sum(log(normpdf(rh(2:n),mu,v)));

    % log-likelihood function of prices
    B = (1-exp(-kappa*tau))/kappa;
    y = y-((n-1)*log(abs(B)));

    % To find Maximum log-likelihood we minimize
    % the negative of log-likelihood
    y = - y;

end


% MLE estimation via zero-coupon prices

load price2OU

for j=1:N
```

```
    P = Price(j,1:n);
    % Using natural logarithm of starting points
    % to avoid negative estimated values
    x0 = log([ kappa theta sigma ]); % starting points

    % Call the optimization routine
    options = optimset('TolFun',1e-15,'TolX',1e-15, ...
                       'MaxIter',1000,'MaxFunEvals',1000);
    % The minimization of negative log-likelihood function
    % can be achieved by using fminsearch
    [ x, fval, exitflag, output ] ...
        = fminsearch(@(x) ZCllhOU(x,P,lambda,tau,Delta),x0,options);
    Result(j,:) = exp(x);
    Exitflag(j,:) = exitflag;
    FvaL(j,:) = fval;

end

save res2OU1000p
```

D.2.2 CIR Model

```
% Compute the prices from simulated state variables
% By using affine model introduced for CIR model

function P = PriceCIR(r,kappa,theta,sigma,lambda,tau)

    % Equations 3.3.6 - page 35
    gam = sqrt((kappa+lambda)*(kappa+lambda)+2*sigma*sigma);
    B = (2*(exp(gam*tau)-1)) ...
        /((gam+kappa+lambda)*(exp(gam*tau)-1)+2*gam);
    A = log(((2*gam*exp((gam+kappa+lambda)*tau/2))/...
        ((gam+kappa+lambda)*(exp(gam*tau)-1) ...
         +2*gam))^(2*kappa*theta/(sigma*sigma)));
    % Equation 3.3.5 - page 34
    P = exp(A-B*r);

end


% Assuming that one can't observe state variables in reality,
% We use affine model to predict unobserved state variables
% When start points  are kappa, theta, sigma.

function rhat = FindRCIR(P,kappa,theta,sigma,lambda,tau)

    % Recalculate affine model elements with start point
    gam = sqrt((kappa+lambda)*(kappa+lambda)+2*sigma*sigma);
```

```
    B = (2*(exp(gam*tau)-1)) ...
        /((gam+kappa+lambda)*(exp(gam*tau)-1)+2*gam);
    A = log(((2*gam*exp((gam+kappa+lambda)*tau/2))/...
        ((gam+kappa+lambda)*...
        (exp(gam*tau)-1)+2*gam))^(2*kappa*theta/(sigma*sigma)));
    % Equation 3.4.8 to find state variables from prices
    rhat = (A-log(P))/B;

end


% Log-likelihood function for prices using CIR model
% Equation 3.4.9 - Page 36

function y = ZCllhCIR(x,P,lambda,tau,Delta)

    % Starting points of optimization
    % Using exp(ln(initial values))
    % to avoid negative estimated values
    kappa = exp(x(1));
    theta = exp(x(2));
    sigma = exp(x(3));
    n = max(size(P));

    % Estimating non-observable state
    %variables by using observable prices
    rh = FindRCIR(P,kappa,theta,sigma,lambda,tau);
    % Constructing Log-Likelihood Function
    % Using non-central chi-square distribution
    L= sigma*sigma/(4 * kappa)*(1-exp(-kappa*Delta));
    % Degrees of freedom
    v = 4*kappa*theta /(sigma*sigma);
    % Non-centrality parameter
    nc = 4*kappa*rh(1:n-1)/((sigma*sigma)*(exp(kappa*Delta)-1));
    % log-likelihood function of estimated state variables
    y = sum(log(ncx2pdf(rh(2:n)/L,v,nc)/L));
    % log-likelihood function of prices
    B = (1-exp(-kappa*tau))/kappa;
    y = y-((n-1)*log(abs(B)));
    % To find Maximum log-likelihood we minimize
    % the negative of log-likelihood
    y = -y;

end


% MLE estimation via zero-coupon prices

load price2CIR
```

```
for j=1:N
    P = Price(j,1:n);
    % Using natural logarithm of starting points
    % to avoid negative estimated values
    x0 = log([ kappa theta sigma ]); % starting points

    % Call the optimization routine
    options = optimset('TolFun',1e-15,'TolX',1e-15, ...
                       'MaxIter',1000,'MaxFunEvals',1000);
    % The minimization of negative log-likelihood function
    % can be achieved by using fminsearch
    [ x, fval, exitflag, output ] ...
      = fminsearch(@(x)ZCllhCIR(x,P,lambda,tau,Delta),x0,options);
    Result(j,:) = exp(x);
    Exitflag(j,:) = exitflag;
    FvaL(j,:) = fval;
    clear out fval ans x P exitflag

end

save res2CIR-1000p
```

D.3 Code for Chapter IV

D.3.1 Vasicek Model

```
% Calculating zero-coupon yield from observed price
% of zero-coupon bond, in time t.

function z = GenZOU(Price,t,m)

    % m is number of time to maturities
    % Calculating the vector of zero-coupon yields
    % in time t for m term to maturities
    z = - log(repmat(Price, m , 1))./t;

end


% Kalman Filter Implementation
% log-likelihood function of Vasicek Model

function llh = KFllhOU(x, z, t, lambda, Delta)

    % State space representation to be
    % forecasted by Kalman filter
```

```
% z(t) = A + H * y(t) + v(t)    --> Observed Variables
% v = normrnd(0,R)
% y(t) = C + F * y(t-1) + epsilon  --> Unobserved Variables
% epsilon ~ normrnd(0,Q)

% Starting points of optimization
% Using exp(ln(initial values))
% to avoid negative estimated values
kappa = exp(x(1));
theta = exp(x(2));
sigma = exp(x(3));

% Input Parameters - state space
n = max(size(z)); % number of time points
N = min(size(t)); % number of zero coupons
% measurement system error
R = (normrnd(0,0.001^2,N,N)).^2.*eye(N);
% Constructing Measurement System
affinegam = kappa*kappa*(theta-(sigma*lambda/kappa))...
            -(sigma*sigma)/2;
affineB = (1.-exp(-kappa.*t))./kappa;
affineA = (affinegam.*(affineB-t)/(kappa*kappa))...
          -(sigma*sigma*affineB.*affineB/(4*kappa));
A = -affineA./t ; % coefficient in Equation 4.2.7
H=affineB./t; % coefficient in Equation 4.2.7
% Constructing Transition System
C=theta*(1-exp(-kappa*Delta)); %coefficient in Equation 4.2.6
F=exp(-kappa*Delta); %coefficient in Equation 4.2.6
%variance of state space process
Q=sigma*sigma*(1-exp(-2*kappa*Delta))/(2*kappa);

%step 1: Intialize state space variable
% by using unconditional mean and variance of transition system
Ey(:,1) = theta;
Vary(:,:,1) = 1/2*sigma*sigma/kappa;
llh = 0;

for i=2:n
    % step 2: Forecasting the measurement equation
    Ez = A(:,i)+H(:,i)*Ey(:,i-1);
    Varz = H(:,i)*Vary(:,:,i-1)*H(:,i)'+R;
    % step 3: Updating the inference about the state vector
    % Observing the zero-coupon yield at time t_i
    eta = z(:,i)-Ez; % measurement system prediction error
    K = Vary(i-1)*H(:,i)'*inv(Varz); % Kalman gain
    % Updating state space variable at time t_{i-1}
    % considering the observed z
    Ey(:,i) = Ey(:,i-1)+K*eta;
    Vary(:,:,i) = (1-K*H(:,i))*Vary(:,:,i-1);
```

```
        % step 4: Forecasting the state vector
        Ey(:,i) = C+F*Ey(:,i);
        Vary(:,:,i) = Vary(:,:,i-1)-F*Vary(:,:,i)*F'+Q;
        % Calculating the concept used in
        % constructing log-likelihood function
        llh = llh+log(det(Varz))+eta'*inv(Varz)*eta;
    end

    %step 5: Constructing the likelihood function
    llh = -n*N*log(2*pi)/2 - llh;

    % We minimize the negative of log-likelihood
    llh = -llh;

end


% MLE estimation using Kalman Filtering

clear all
load price2OU.mat

n = 100; % number of time points
tau = [0.07 0.25 0.5 1]; % term to maturities - 4 zero coupon rates
m = max(size(tau));
lambda = 1; % Risk Parameter

% Calculating T-t
t(:,1) = tau';
for i=2:n
        t(:,i) = t(:,i-1) + Delta;
end

% MLE for N=1000 iteration
for i=1:N

   % Observing the zero-coupon yields
   z = GenZOU(Price(i,1:n), t, m);

   % Using natural logarithm of starting points
   % to avoid negative estimated values
   x0 = log([ kappa theta sigma ]);

   % Call the optimization routine
   options = optimset('TolFun',1e-15,'TolX',1e-15, ...
       'MaxIter',1000,'MaxFunEvals',1000);

   % The minimization of negative log-likelihood function
   % can be achieved by using fminsearch
```

```
    [ x, fval, exitflag, output ] ...
        = fminsearch(@(x)KFllhOU(x,z,t,lambda,Delta)',x0,options);

    Result(i,:) = exp(x);
    Exitflag(i) = exitflag;

end

save Result2OU-1000
```

D.3.2 CIR Model

```
% Calculating zero-coupon yield from observed price
% of zero-coupon bond, in time t.

function z = GenZCIR(Price,t,m)

    % m is number of time to maturities
    % Calculating the vector of zero-coupon yields
    % in time t for m term to maturities
    z = - log(repmat(Price, m , 1))./t;

end


% Kalman Filter Implementation
% log-likelihood function of CIR model

function llh = KFllh/CIR(x, z, t, lambda, Delta)

    % State space representation to be forecasted by kalman filter
    % z(t) = A + H * y(t) + v(t)    --> Observed Variables
    % v = normrnd(0,R)
    % y(t) = C + F * y(t-1) + epsilon  --> Unobserved Variables
    % epsilon ~ normrnd(0,Q)

    % Starting points of optimization
    % Using exp(ln(initial values))
    % to avoid negative estimated values
    kappa = exp(x(1));
    theta = exp(x(2));
    sigma = exp(x(3));

    % Input Parameters - state space
    n = max(size(z)); % number of time points
    N = min(size(t)); % number of zero coupons
    % measurement system error
    R = (normrnd(0,0.001^2,N,N)).^2.*eye(N);
```

```
% Constructing Measurement System
affinegam = sqrt((kappa+lambda)*(kappa+lambda)+2*sigma*sigma);
affineB = (2*(exp(affinegam.*t)-1))...
          ./((affinegam+kappa+lambda) ...
          .*(exp(affinegam.*t)-1)+2*affinegam);
affineA = log(((2*affinegam*...
          exp((affinegam+kappa+lambda).*t./2))./...
          ((affinegam+kappa+lambda).*(exp(affinegam.*t)-1)+ ...
          2*affinegam)).^(2*kappa*theta/(sigma*sigma)));
A = -affineA./t ; % coefficient in Equation 4.2.7
H=affineB./t; % coefficient Equation 4.2.7
% Constructing Transition System
C=theta*(1-exp(-kappa*Delta)); %coefficient in Equation 4.2.6
F=exp(-kappa*Delta); %coefficient in Equation 4.2.6

%step 1: Initialize state space variable
% by using unconditional mean and variance of transition system
Ey(:,1) = theta;
Vary(:,:,1) = 1/2*sigma*sigma*theta/kappa;
llh = 0;
y = zeros(1,n);
epsilon = zeros(1,n);
Q = zeros(1,n);
Q(1) = Vary(:,:,1);
epsilon(1) = normrnd(0,Q(1),1,1);
y(1) = theta + epsilon(1);
for i=2:n
    %variance of state space process
    Q(i)=theta*sigma*sigma*(1-exp(-kappa*Delta))...
         *(1-exp(-kappa*Delta))/(2*kappa) ...
         + sigma*sigma/kappa*(exp(-kappa*Delta)...
         -exp(-2*kappa*Delta))*y(i-1);
    epsilon(i) = normrnd(0,Q(i),1,1);
    y(i) = C + F*y(i-1) + epsilon(i);
end

for i=2:n
    % step 2: Forecasting the measurement equation
    Ez = A(:,i)+H(:,i)*Ey(:,i-1);
    Varz = H(:,i)*Vary(:,:,i-1)*H(:,i)'+R;
    % step 3: Updating the inference about the state vector
    % Observing the zero-coupon yield at time t_i
    eta = z(:,i)-Ez;
    K = Vary(i-1)*H(:,i)'*inv(Varz); % Kalman gain
    % Updating state space variable at time t_{i-1}
    % considering the observed z
    Ey(:,i) = Ey(:,i-1)+K*eta;
    Vary(:,:,i) = (1-K*H(:,i))*Vary(:,:,i-1);
    % step 4: Forecasting the state vector
```

```
        Ey(:,i) = C+F*Ey(:,i);
        Vary(:,:,i) = Vary(:,:,i-1)-F*Vary(:,:,i)*F'+Q(i);
        % Calculating the concept used in
        % constructing log-likelihood function
        llh = llh+log(det(Varz))+eta'*inv(Varz)*eta;
    end

    %step 5: Constructing the likelihood function
    llh = -n*N*log(2*pi)/2 - llh;
    % To find Maximum log-likelihood we minimize
    % the negative of log-likelihood
    llh = -llh;

end


% MLE estimation using Kalman Filtering

clear all
load price2CIR.mat

n = 100; % number of time points
tau = [0.07 0.25 0.5 1]; % term to maturities - 4 zero coupon rates
m = max(size(tau));
lambda = 1; % Risk Parameter

% Calculating T-t
t(:,1) = tau';
for i=2:n
        t(:,i) = t(:,i-1) + Delta;
end

% MLE for N=1000 iterations
for i=1:4

   % Observing the zero-coupon yields
   z = GenZCIR(Price(i,1:n), t,m);

   % Using natural logarith of starting points
   % to avoid negative estimated values
   x0 = log([ kappa theta sigma ]);

   % Call the optimization routine
   options = optimset('TolFun',1e-15,'TolX',1e-15, ...
       'MaxIter',1000,'MaxFunEvals',1000);

   % The minimization of negative log-likelihood function
   % can be achieved by using fminsearch
    [ x, fval, exitflag, output ] ...
```

```
      = fminsearch(@(x)KFllhCIR(x,z,t,lambda,Delta),x0,options);

    Result(i,:) = exp(x);
    Exitflag(i) = exitflag;

end

save Result2CIR-1000
```

REFERENCES

Ball, C. A., and W. N. Torous, 1996, "Unit Roots and Estimation of Interest Rate Dynamics," *Journal of Empirical Finance*, Vol. 3, No. 2, pp. 215–238.

Babbs, S. H., and K. B. Nowman, 1999, "Kalman Filtering of Generalized Vasicek Term Structure Models," *Journal of Financial and Quantitative Analysis*, Vol. 34, No. 1, pp. 115–130.

Bergstrom, A. R., 1983, "Gaussian Estimation of Structural Parameters in Higher Order Continuous Time Dynamic Models," *Econometrica*, Vol. 51, No. 1, pp. 117–152.

Bergstrom, Abraham R., 1976, *Statistical Inference in Continuous Time Economic Models*, North-Holland, Amsterdam.

Bolder, David J., 2001, *Affine Term-Structure Models: Theory and Implementation*, Working Paper 2001-15, Bank of Canada, Ottawa.

Chen, R. R., 1995, "A Two-Factor, Preference-Free Model for Interest Rate Sensitive Claims ," *Journal of Futures Markets*, Vol. 15, No. 3, pp. 345–372.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985a, "An Intertemporal General Equilibrium Model of Asset Prices," *Econometrica*, Vol. 53, No. 1, pp. 363–384.

Cox, J. C., J. E. Ingersoll, and S. A. Ross, 1985b, "A Theory of Term Structure of Interest Rates," *Econometrica*, Vol. 53, No. 1, pp. 385–407.

De Jong, F., 200, "Time Series and Cross Section Information in Affine Term Structure Models," *Journal of Business and Economic Statistics*, Vol. 18, No. 3, pp. 300–314.

Duan, J. C., and J. G. Simonato, 1999, "Estimating and Testing Affine Term Structure Models by Kalman Filter," *Review of Quantitative Finance and Accounting*, Vol. 13, No. 2, pp. 111–135.

Duan, J. C., 1994, "Maximum Likelihood Estimation Using Price Data of The derivative Contract," *mathematical Finance*, Vol. 4, No. 2, pp. 155–167.

Duffie, D., and R. Kan, 1996, "A Yield-Factor Model of Interest Rates," *Mathematical Finance*, Vol. 6, No. 4, pp. 379–406.

Ferland, R., 2000, "Mouvement brownien et calcul stochastique," In *Mathématiques, enseignement des mathématiques: l'apport du XX^{e} siècle* (Actes du 42^{e} congrès annuel de l'Association mathématique du Québec), Le Griffon d'argile, Sainte-Foy, pp. 104–116.

Gard, Thomas C., 1988, *Introduction to Stochastic Differential Equations*, M. Dekker, New York.

Geyer, A., and S. Pichler, 1999, "A State-Space Approach to Estimate and Test Multifactor Cox-Ingersoll-Ross Models of the Term Structure," *Journal of Financial Research*, Vol. 22, No. 1, pp. 107–130.

Gourieroux, Christian, and Joann Jasiak, 2001, *Financial Econometrics. Problems, Models, and Methods*, Princeton University Press, Princeton.

Harvey, Andrew C., 1989, *Forecasting, Structural Time Series Models and the Kalman Filter*, Cambridge Univessity Press, Cambridge, UK.

Heath, D., R. Jarrow, and A. Morton, 1992, "Bond Pricing and the Term Structure of Interest Rates: a New Methodology for Contingent Claims Valuation," *Econometrica*, Vol. 60, No. 1, pp. 77–105.

Hoel, Paul G., Sidney C. Port, and Charles J. Stone, 1972, *Introduction to Stochastic Processes*, Houghton Mifflin, Boston.

Ikeda, Noboyuki, and Shinzo Watanabe, 1981, *Stochastic Differential Equations amd Diffusion Processes*, North-Holland, Amsterdam.

Kloeden, Peter E., and Eckhard Platen, 1995, *Numerical Solution of Stochastic Differential Equations*, Springer-Verlag, New York.

Lamberton, Damien, and Bernard Lapeyre, 1991, *Introduction au calcul stohastique appliqué à la finance*, Ellipses, Paris.

Longstaff, F. A., and E. S. Schwartz, 1992, "Interest Rate Volatility and the Term Structure: A Two-Factor General Equilibrium Model," *Journal of Finance*, Vol. 47, No. 4, pp. 1259–1282

Lo, A., 1988, "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sample Data," *Econometric Theory*, Vol. 4, No. 2, pp. 231–247.

Pederson, A. R., 1995, "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, Vol. 22, No. 1, pp. 55–71.

Phillips, P. C. B., 1972, "The Structural Estimation of a Stochastic Differential Equation System," *Econometrica*, Vol. 40, No. 6, pp. 1021–1041.

Phillips, P. C. B., 1974, "The Estimation of Some Continuous Time Models," *Econometrica*, Vol. 42, No. 5, pp. 803–823.

Vasicek, O., 1977, "An Equilibrium Characterization of the Term Structure," *Journal of Financial Economics*, Vol. 5, No. 2, pp. 177–188

Welch, G., and G. Bishop, 2005, *An Introduction to the Kalman Filter*, Technical Report TR-95-041, Department of Computer Science, University of North Carolina at Chapel Hill, Chapel Hill, NC.