GOODNESS-OF-FIT TESTS FOR CONTINUOUS-TIME

FINANCIAL MARKET MODELS

YANG LONGHUI

NATIONAL UNIVERSITY OF SINGAPORE

2004

GOODNESS-OF-FIT TESTS FOR CONTINUOUS-TIME

FINANCIAL MARKET MODELS

YANG LONGHUI

(B.Sc. EAST CHINA NORMAL UNIVERSITY)

A THESIS SUBMITTED

FOR THE DEGREE OF MASTER OF SCIENCE

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY

NATIONAL UNIVERSITY OF SINGAPORE

2004

Acknowledgements

I would like to extend my eternal gratitude to my supervisor, Assoc. Prof. Chen SongXi, for all his invaluable suggestions and guidance, endless patience and encouragement during the mentor period. Without his patience, knowledge and support throughout my studies, this thesis would not have been possible.

This thesis, I would like to contribute to my dearest family who have always been supporting me with their encouragement and understanding in all my years. To He Huiming, my husband, thank you for always standing by me when the nights were very late and the stress level was high. I am forever grateful for your sacrificing your original easy life for companying with me in Singapore.

Special thanks to all my friends who helped me in one way or another for their friendship and encouragement throughout the two years. And finally, thanks are due to everyone at the department for making everyday life enjoyable.

Contents

Summary

Diffusion processes have wide applications in many disciplines, especially in modern finance. Due to their wide applications, the correctness of various diffusion models needs to be verified. This thesis concerns the specification test of diffusion models proposed by Aït-Sahalia (1996a). A serious doubt on Aït-Sahalia's test in general and the employment of the kernel method in particular has been cast by Pritsker (1998) by carrying out some simulation studies on the empirical performance of Aït-Sahalia's test. He found that Aït-Sahalia's test had very poor empirical size relative to nominal size of the test. However, we found that the dramatic size distortion is due to the use of the asymptotic normality of the test statistic. In this thesis, we reformulate the test statistic of Aït-Sahalia by a version of the empirical likelihood. To speed up the convergence, the bootstrap is employed to find the critical values of the test statistic. The simulation results show that the proposed test has reasonable size and power, which then indicate there is nothing wrong with using the kernel method in the test of specification of diffusion models. The key is how to use it.

List of Tables

List of Figures

Chapter 1

Introduction

1.1   A Brief Introduction To Diffusion Processes

The study of diffusion processes originally arises from the field of statistical physics,
but diffusion processes have widely applied in engineering, medicine, biology and
other disciplines. In these fields, they have been well applied to model phenomena
evolving randomly and continuously in time under certain conditions, for example
security price fluctuations in a perfect market, variations of population growth on
ideal condition and communication systems with noise, etc.

Karlin and Taylor (1981) summed up three main advantages for diffusion processes.
Firstly, diffusion processes model many physical, biological, economic and social
phenomena reasonably. Secondly, many functions can be calculated explicitly for
one-dimensional diffusion process. Lastly in many cases Markov processes can be
approximated by diffusion processes by transforming the time scale and renormal-

izing the state variable. In short, diffusion processes specify phenomena well and

possess practicability.

From the influential paper of Merton (1969), continuous-time methods on dif-

fusion models have become an important part of financial economics. Moreover,

it is said that modern finance would not have been possible without them. These

models are important to describe stock prices, exchange rates, interest rates and

portfolio selection which are certain core areas in finance. Although its develop-

ment is only about thirty years, continuous-time diffusion methods have proved

to be one of the most attractive ways to guide financial research and offer correct

economic applications.

What is diffusion processes? Here, we give the definition of the diffusion

processes derived from Karlin and Taylor (1981) and more details can be found

in their book. "A continuous time parameter stochastic process which possesses

the Markov property and for which the sample paths X_t are continuous functions

of t is called a diffusion process."

Generally continuous-time diffusion process $X_t, t \geq 0$ has the form

$$dX_t = \mu(X_t)dt + \sigma(X_t)d\mathrm{B_t} \tag{1.1}$$

where $\mu(\cdot)$ and $\sigma(\cdot) > 0$ are respectively the drift and diffusion functions of the

process, and $\mathrm{B_t}$ is a standard Brownian motion. Generally, the functions are para-

meterized:

$$\mu(x) = \mu(x, \theta) \ and \ \sigma^2(x) = \sigma^2(x, \theta), \ where \ \theta \in \Theta \subset R^K. \tag{1.2}$$

where Θ is a compact parameter space (see the appendix of Aït-Sahalia (1996a) for more details).

1.2 Notation

Before we review part of the works on diffusion processes in financial economics, we first present some notations on the marginal density and the transition density of a diffusion process in this thesis. For easy reference, from now the marginal density function and the transition density function for a diffusion process described in (1.1) are denoted as $f(\cdot, \theta)$ and $p_\theta(\cdot, \cdot | \cdot, \cdot)$ respectively. Here the transition density $p_\theta(y, s|x, t)$ is the probability density that $X_s = y$ at time s given that $X_t = x$ at time t for $t < s$. If the diffusion process is stationary, we have $p_\theta(y, s|x, t) = p_\theta(y, s-t|x, 0)$ which is denoted as $p_\theta(y|x, s-t)$. The marginal density $f(x, \theta)$ denotes the unconditional probability density. In fact, the relationship between the transition density and the marginal density is

$$f(x, \theta) = lim_{s \to \infty} p_\theta(y, s|x, t). \tag{1.3}$$

This was implied by Pritsker (1998).

From the two different densities, different information about the process can be obtained. The transition density shows that $X_s = y$ at time s depends on $X_t = x$ at time t when the time between the observations is finite. It is clear that the transition density describes the short-run time-series behavior of the diffusion process. Therefore, the transition density captures the full dynamics of the diffusion

process. From the relationship indicated in (1.3), we know that the marginal density describes the long-run behavior of the diffusion process.

1.3   Commonly Used Diffusion Models

The seminal contributions by Black and Scholes (1973) and Merton (1969) are always mentioned in the development of continuous-time methods in finance. Their works on options pricing signify a new and promising stage of research in financial economics. The Black-Scholes (B-S) model proposed by Fisher Black and Myron Scholes (1973) is often cited as the foundation of modern derivatives markets. It is the first model that provided accurate price options. Merton (1973) investigated B-S model and derived B-S model under weaker assumptions and this model is indeed more practical than the original B-S model.

The term structure of interest rates is one of core areas in finance where continuous-time methods made a great impact. Most research works focus on finding the suitable expressions for drift and diffusion functions of the diffusion process (1.1). Table 1.1 is driven from Aït-Sahalia (1996a) who collected commonly used diffusion models in the literature for the drift and the instantaneous variance of the short-term interest rate. Merton (1973) derived a model of discount bond prices and the diffusion process he considered is simply a Brownian motion with drift. The Vasicek model has a linear drift function and a constant diffusion function. This model is widely applied to value bond options, futures options, etc. Jamshid-

ian (1989) derived a closed-form solution for European options on pure discount bonds using the Vasicek (1977) model. Gibson and Schwartz (1990) applied the model to derive oil-linked assets.

Table 1.1: Alternative specifications of the spot interest rate process

$$dX_t = \mu(X_t)dt + \sigma(X_t)dB_t$$

$\mu(X)$	$\sigma(X)$	Stationary	Reference
β	σ	Yes	Merton(1973)
$\beta(\alpha - X)$	σ	Yes	Vasicek(1977)
$\beta(\alpha - X)$	$\sigma X^{1/2}$	Yes	Cox-Ingersoll-Ross(1985b),
			Brown-Dybvig(1986),
			Gibbons-Ramaswamy(1993)
$\beta(\alpha - X)$	σX	Yes	Courtadon(1982)
$\beta(\alpha - X)$	σX^λ	Yes	Chan et al.(1992)
$\beta(\alpha - X)$	$\sqrt{\sigma + \gamma X}$	Yes	Duffie-Kan(1993)
$\beta r(\alpha - ln(X))$	σX	Yes	Brennan-Schwartz(1979)[one-factor]
$\alpha X^{(-1-\delta)} + \beta X$	$\sigma X^{\delta/2}$	Yes	Marsh-Rosenfeld(1983)
$\alpha + \beta X + \gamma X^2$	$\sigma + \gamma X$	Yes	Constantinides(1992)

Cox-Ingersoll-Ross (1985) (CIR) specified that the instantaneous variance is a linear function of the level of the spot rate X, namely $\sigma^2(x, \theta) = \sigma^2 x$. Applying the CIR model, Cox-Ingersoll-Ross (1985) derived the discount bond option and

Ramaswamy and Sundaresan (1986) evaluated the floating-rate notes. Longstaff (1990) extended the CIR model and derived closed-form expressions for the values of European calls. Courtadon (1982) studied the pricing of options on default-free bonds using the CIR model.

These diffusion models have simple drift and diffusion functions and have closed forms for the transition density and marginal density in theory. However it is generally thought that their performances are poor in empirical tests to capture the dynamics of the short-term interest rate. Chan, Karolyi, Longstaff and Sanders (1992) presented a parametric model that the diffusion function $\sigma^2(x, \theta) = \sigma^2 x^{2\lambda}$, where $\lambda > 1/2$ (If $\lambda = 1/2$, it is the CIR model). Using annualized monthly Treasury Bill Yield from June, 1964 to December, 1989 (306 observations), Chan et al. applied Generalized Method of Moments (GMM) to estimate their diffusion model as well as other eight different diffusion models such as the Merton (1973) model, the Vasicek (1977) model, the CIR (1982) model and so on. They also formulated a test statistic which is asymptotically distributed χ^2 with k degrees of freedom and compared these variety diffusion models. They found that the value of λ in their model was the most important feature differentiating these diffuion models. At last, they concluded that these models, which allow $\lambda \geq 1$, capture the dynamics of the short-term interest rate, better than those where the parameter $\lambda < 1$. Brennan and Schwartz (1979) expressed the term structure of interest rates as a function of the longest and shortest maturity default free instruments which follow a Gauss-Wiener process and the model was applied to derive the bond price.

Marsh-Rosenfeld (1983) considered a mean-reverting constant elasticity of variance diffusion model which was nested within the typical diffusion-poisson jump model and examined these models for nominal interest rate changes. Constantinides (1992) developed a model of the nominal term structure of interest rate and derived the closed form expression for the prices of discount bonds and European options on bonds.

1.4   Parameter Estimation

These different parametric models of short rate process attempt to capture particular features of observed interest rate movements in real market. However, there are unknown parameters or unknown functions in these models. Generally, they are estimated from observations of the diffusion processes. Kasonga (1988) showed that the least squares estimator of the drift function derived from the diffusion model is strongly consistent under some mild conditions. Dacunha-Castelle and Florens-Zmirou (1986) estimated the parameters of the diffusion function from a discretized stationary diffusion process. Dohnal (1987) considered the estimation of a parameter from a diffusion process observed at equidistant sampling points only and proved the local asymptotic mixed normality property of the volatility function. Genon-Catelot and Jacod (1993) constructed the estimation of the diffusion coefficient for multi-dimensional diffusion processes and studied their asymptotic. Furthermore, they also considered a general sampling scheme. Here, we review two

main parametric estimation strategies for diffusion models, Maximum likelihood

methods (MLE) and Generalized Method of Moments (GMM).

Recall the diffusion model expression in (1.1). If the functions μ and σ are given,

the transition density $p_\theta(y, s|x, t)$ satisfies the Kolmogorov forward equation,

$$\frac{\partial p_\theta(y, s|x, t)}{\partial s} = -\frac{\partial}{\partial y} \left[\mu(y, \theta) p_\theta(y, s|x, t) \right] + \frac{1}{2} \frac{\partial^2}{\partial y^2} \left[\sigma^2(y, \theta) p_\theta(y, s|x, t) \right] \qquad (1.4)$$

and the backward equation (see Øksendal,1985)

$$-\frac{\partial p_\theta(y, s|x, t)}{\partial t} = \mu(x, \theta) \frac{\partial}{\partial x} \left[p_\theta(y, s|x, t) \right] + \frac{1}{2} \sigma^2(x, \theta) \frac{\partial^2}{\partial x^2} \left[p_\theta(y, s|x, t) \right]. \qquad (1.5)$$

In some applications, the marginal and transition densities can be expressed in

closed forms. For example, the marginal and transition densities for the Vasicek

(1977) model are all Gaussian and the transition density of the CIR (1985) model

follows non-central chi-square. In such situations, MLE is often selected to estimate

the parameters of the diffusion process.

Lo (1988) discussed the parametric estimation problem for continuous-time sto-

chastic processes using the method of maximum likelihood with discretized data.

Pearson and Sun (1994) applied the MLE method to estimate the two-factor CIR

(1985) model using data on both discount and coupon bonds. Chen and Scott

(1993) extended the CIR model to a multifactor equilibrium model of the term

structure of interest rate and presented a maximum likelihood estimation for one-,

two-, and three-factor models of the nominal interest rate. As a result, they as-

sumed that a model with more than one factor is necessary to explain the changes

over time in the slope and shape of the yield curve.

However, most of transition densities of the diffusion models have no closed form expression. Therefore, researchers estimate the likelihood function by Monte Carlo simulation methods (see Lo (1988) and Sundaresan (2000)). Recently, Aït-Sahalia (1999) investigated the maximum-likelihood estimation with unknown transition functions. He applied a Hermite expansion of the transition density around a normal density up to order K and generated closed-form approximations to the transition function of an arbitrary diffusion model, and then used them to get approximate likelihood functions.

Another important estimation method is the Generalized Method of Moments (GMM) proposed by Hansen (1982). The method is often applied when the likelihood function is too complicated especially for the nonlinear diffusion model or where we only have interest on certain aspects of the diffusion processe. Hansen and Scheinkman (1995) discussed ways of constructing moment conditions which are implied by stationary Markov processes by using infinitesimal generators of the processes. The Generalized Method of Moments estimators and tests can be constructed and applied to discretized data obtained by sampling Markov processes. Chen et. al (1992) used Generalized Method of Moments to estimate a variety of diffusion models.

1.5 Nonparametric Estimation

Parametric estimation methods for diffusion models are well developed to specify features of observed interest rate movements. However, the inference statistics of a diffusion process rely on the parametric specifications of the diffusion model. If the parametric specification of the diffusion model is misspecified, the inference statistics of the diffusion process are misleading. Hence, some researchers have used nonparametric techniques to reduce the number of arbitrary parametric restrictions imposed on the underlying process. Florens-Zmirou (1993) proposed an estimator of volatility function nonparametrically based on discretized observations of the diffusion processes and described the asymptotic behavior of the estimator. Aït-Sahalia (1996b) estimated the diffusion function nonparametrically and gave a linear specification for the drift function. Stanton (1997) constructed kernel estimators of the drift and diffusion functions based on discretized data.

The results of these studies for nonparametric estimation showed that the drift function has substantial nonlinearity. Stanton (1997) also pointed out that there was the evidence of substantial nonlinearity in the drift. As maintained out by Ahn and Gao (1999), the linearity of the drift imposed in the literature appeared to be the main source of misspecification.

Aït-Sahalia (1996a) considered testing the specification of a diffusion process. His work may be the first and the most significant one on specifying the suitability of a parametric diffusion model. Let the true marginal density be $f(x)$. In order to

test whether both the drift and the diffusion functions satisfy certain parametric forms, he checked if the true density of the diffusion process is the same as the parametric one which is determined by the drift and diffusion functions. As a matter of fact, once we know the drift and the diffusion functions, the marginal density is determined according to

$$f(x, \theta) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} exp\{\int_{x_0}^{x} \frac{2\mu(u, \theta)}{\sigma^2(\mu, \theta)} du\} \tag{1.6}$$

where x_0 the lower bound of integration in the interior of $D = (\underline{x}, \overline{x})$ for given $\underline{x}, \overline{x}$ such that $\underline{x} < \overline{x}$. The constant $\xi(\theta)$ is applied so that the marginal density integrates to one. However the true marginal density is unknown and Aït-Sahalia (1996a) applied the nonparametric kernel estimator to replace the true marginal density. Therefore, the test statistic proposed by Aït-Sahalia (1996a) is based on a differece between the parametric marginal density $f(x, \theta)$ and the kernel estimator of the same density $\hat{f}(x)$. For a daily short-rate data of 22 years, he strongly rejected all the well-known one factor diffusion models of the short interest rate except the model which has non-linear drift function. Aït-Sahalia (1996a) maintained that the linearity of the drift was the main source of the misspecification.

However, Pritsker (1998) carried out the simulation on Aït-Sahalia's (1996a) test and discovered that Aït-Sahalia's test had very poor empirical size relative to the nominal size of the test. Aiming to find the reason of the poor performance of Aït-Sahalia's (1996a) test, Pritsker(1998) considered the finite sample of Aït-Sahalia's test of diffusion models properties. He pointed out the main reasons for

the poor performance were that the nonparametric kernel estimator based test was unable to differentiate between independent and dependent series as the limiting distributions were the same. Furthermore, the interest rate is highly persistent and the nonparametric estimators converged very slowly. Particularly, in order to attain the accuracy of the kernel density estimator implied by asymptotic distribution with 22 years of data generated from the Vasicek (1977) model, 2755 years of data are required.

There is no doubt that the observation of Pritsker (1998) is valid. However, the poor performance of Aït-Sahalia's (1996a) test is not because of the nonparametric kernel density estimator. As a matter of fact, the test statistic proposed by Aït-Sahalia (1996a) is a U-statistic, which is known for slow convergence even for independent observations.

In this thesis, we propose a test statistic based on the bootstrap in conjunction with an empirical likelihood formulate. We find that the empirical likelihood goodness-of-fit test proposed by us has reasonable properties of size and power even for time span of 10 years and our results are much better than those reported by Pritsker (1998).

Chapman and Pearson (2000) carried out a Monte Carlo study of the finite sample properties of the nonparametric estimators of Aït-Sahalia (1996a) and Stanton (1997). They pointed out that there were quantitatively significant biases in kernel regression estimators of the drift advocated by Stanton (1997). Their empirical results suggested that nonlinearity of the short rate drift is not a robust stylized

fact. The studies of Chapman and Pearson (2000) and Pritsker (1998) cast serious doubts on the nonparametric methods applied in finance because the interest rate and many other high frequency financial data are usually dependent with high persistence.

Recently, Hong and Li (2001) proposed two nonparametric transition density-based specification tests for testing transition densities in continuous–time diffusion models and showed that nonparametric methods were a reliable and powerful tool in finance area. Their tests are robust to persistent dependence in data by using an appropriate data transformation and correcting the boundary bias caused by kernel estimators.

1.6 Methodology And Main Results

In this thesis, we consider the nonparametric specification test to reformulate Aït-Sahalia's (1996a) test statistic via a version of the empirical likelihood (Owen, 1988). This empirical likelihood formulation is designed to put the discrepancy measure which is used in Aït-Sahalia's original proposal by taking into account of the variation of the kernel estimator. But the discrepancy measure is the difference between the nonparametric kernel density and the smoothed parametric density in order to avoid the bias associated with the kernel estimator. Then we use a bootstrap procedure to profile the finite sample distribution of the test statistic. Since it is well-known that both the bootstrap and the full empirical likelihood are

time-consuming, the least squares empirical likelihood introduced by Brown and Chen (1998) is applied in this thesis instead of the full empirical likelihood.

We carry out a simulation study of the same five Vasicek diffusion models as in Pritsker (1998) study and find that the proposed bootstrap based empirical likelihood test had reasonable size for time spans of 10 years to 80 years.

1.7   Chapter Development

This thesis is organized as follows:

In Chapter 2, we present the misspecification of parametric methods and the misspecification may be caused in applications of diffusion models. Then, the details about Aït-Sahalia (1996a) test and asymptotic distribution of the test statistic are introduced. We then describe Pritsker's (1998) simulation studies on Aït-Sahalia's (1996a) test and his findings based on his simulation results.

Our main task in Chapter 3 is to propose the empirical likelihood goodness-of-fit test for the marginal density. At the beginning, the empirical likelihood is presented. It includes the empirical likelihood for mean parameter and the full empirical likelihood. Then we describe a version of the empirical likelihood for the marginal density which employed in this thesis. The empirical likelihood goodness-of-fit test is discussed in the last section.

Chapter 4 focus on simulation results for the empirical likelihood goodness-of-fit test. We discuss some practical issues in formulating the test, for example

parameters estimator, bandwidth selection, the diffusion process generation, etc. In the part of result, we first report the result of the goodness-of-fit test for IID case to make sure that the new method works. Then we show the simulation result on the empirical size and power for the least square empirical likelihood goodness-of-fit test of the marginal density. Lastly, we implement Aït-Sahalia (1996a) test again which is similar to Pritsker's (1998) simulation studies.

In Chapter 5, we employ the proposed empirical likelihood specification test to evaluate five popular diffusion models for the spot interest rate. We measure the goodness-of-fit of these five models for the interest rate first. After that, we present the test statistic and p-values of these diffusion models.

Chapter 2

Existing Tests For Diffusion

Models

2.1   Introduction

As mentioned in Chapter 1, most researchers studied continuous-time diffusion models in order to capture the term structure of important economic variables, such as exchange rates, stock prices and interest rates. Among them, most of the works focused on selecting suitable parametric drift and diffusion families which determine the diffusion models. There are so many parametric models that we might have no idea which model to choose. In fact, the statistical inference of diffusion processes rest entirely on the parametric specifications of the diffusion models. If the parametric specification is misspecified, not only the performance of the model is poor but also the results of inference may be misleading. Therefore,

determining the suitability of a parametric diffusion model is important and this is the focus of this thesis.

Among the research works to determine the suitability of a parametric diffusion model, the test proposed by Aït-Sahalia (1996a) is one of the most influential tests. Although some papers have pointed out that the performance of the test statistic proposed by Aït-Sahalia was poor, Aït-Sahalia's test was the first one to make such idea into reality and many later research works were based on Aït-Sahalia's idea. In this chapter, we outline the details of Aït-Sahalia's test first. At the same time, the nonparametric kernel estimator applied by Aït-Sahalia (1996a) is described. Lastly, we show the asymptotic distribution of the test statistic.

Pritsker (1998) studied the performance of the finite sample distribution of Aït-Sahalia (1996a) test. Pritsker found Aït-Sahalia's test had very poor empirical size relative to the nominal size of the test. In particular, he found that 2755 years of data were required for obtaining a reasonable agreement between the empirical size and the nominal size. Actually, the cause of poor performance he believed is that the nonparametric kernel estimator based test was unable to differentiate between independent and dependent series as their limiting distributions are the same.

In this thesis, we propose a test based on the least square empirical likelihood via the bootstrap. We carry the same simulation study as Pritsker (1998) and compare the performance between these two tests. Therefore, it is necessary for us to know the details of the Pritsker (1998) study as well. To this end, a detail of Pritsker (1998) study is outlined in Section 2.3.

2.2 Aït-Sahalia's Test

2.2.1 Test Statistic

Suppose that the stationary diffusion process with dynamics represented by a diffusion equation (1.1) is $\{X_t, t \geq 0\}$. The joint parametric family of the drift and diffusion is

$$P \equiv \{(\mu(\cdot, \theta), \sigma^2(\cdot, \theta)) | \theta \in \Theta\}, \tag{2.1}$$

where θ is a parameter within the parametric space Θ. The null and alternative hypotheses described by Aït-Sahalia's (1996a) are

$$H_0 \quad : \quad \mu(\cdot, \theta_0) = \mu_0(\cdot) \quad and \quad \sigma^2(\cdot, \theta_0) = \sigma_0^2(\cdot) \quad for \quad some \quad \theta_0 \in \Theta,$$

$$H_1 \quad : \quad (\mu_0(\cdot), \sigma_0^2(\cdot)) \notin P, \tag{2.2}$$

where $(\mu_0(\cdot), \sigma_0^2(\cdot))$ are the "true" drift and diffusion functions for diffusion equation (1.1).

Aït-Sahalia (1996a) proposed a test for the specification of a diffusion model based on the marginal density which is the focus of this thesis. As mentioned before, once we know the drift and diffusion functions as specified in H_0 of (2.2), the marginal density is determined according to

$$f(x, \theta) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} exp\{ \int_{x_0}^{x} \frac{2\mu(u, \theta)}{\sigma^2(\mu, \theta)} du \}, \tag{2.3}$$

where x_0 the lower bound of integration in the interior of $D = (\underline{x}, \overline{x})$ for given $\underline{x}, \overline{x}$ such that $\underline{x} < \overline{x}$. The constant $\xi(\theta)$ is applied so that the marginal density

integrates to one. The idea of Aït-Sahalia was to check if the true density of the diffusion process is the same with the parametric density given in (2.3). A weight L_2 discrepancy measure between the true density $f(\cdot)$ and the parametric density $f(\cdot, \theta)$ is

$$
\begin{aligned}
M &\equiv \min_{\theta \in \Theta} \int_{\underline{x}}^{\overline{x}} (f(u, \theta) - f(u))^2 f(u) du & (2.4) \\
&= \min_{\theta \in \Theta} E[(f(X, \theta) - f(X))^2]. & (2.5)
\end{aligned}
$$

In fact, this is the integrated squared difference between the true and parametric density weighted by $f(\cdot)$. From the measure of distance, it is clear that under the null hypothesis M is small, while M is large under the alternative hypothesis.

Aït-Sahalia (1996a) applied the nonparametric kernel estimator to replace the true marginal density. The parametric and nonparametric density estimators should be quite the same under H_0. Under H_1, the parametric density estimator would deviate from the nonparametric estimator. In his test, he used the standard kernel estimator:

$$
\hat{f}(x) = \frac{1}{N} \sum_{t=1}^{N} \frac{1}{h} K(\frac{x - X_t}{h}), \tag{2.6}
$$

where N is the number of observations, h is called the bandwidth and $K(\cdot)$ is a function which is commonly a symmetric probability density and satisfies :

$$
\int_{R} K(x) dx = 1, \tag{2.7}
$$

$$
\int_{R} x K(x) dx = 0, \tag{2.8}
$$

$$
\int_{R} x^2 K(x) dx = \sigma_k^2, \tag{2.9}
$$

where σ_k^2 is a positive constant. Table 2.1 lists some of the common kernels used in literature on nonparametric kernel estimators.

Kernel	K(u)
Gaussian	$\dfrac{1}{\sqrt{2\pi}}e^{-\frac{u^2}{2}}$
Epanechnikov	$\dfrac{3}{4}(1-u^2)I(u)$
Biweight	$\dfrac{15}{16}(1-u^2)^2 I(u)$

Table 2.1: Common used Kernels ($I(\cdot)$ signifies the indicator function)

Aït-Sahalia (1996a) applied Gaussian kernel in his empirical studies. To estimate the marginal density, we choose the bandwidth such that $h \to 0$, $lim_{N\to\infty}Nh = \infty$ and $lim_{N\to\infty}Nh^{4.5} = 0$.

Finally, the test statistic proposed by Aït-Sahalia (1996a) is

$$\hat{M} \equiv Nh \min_{\theta\in\Theta} \frac{1}{N}\sum_{t=1}^{N}(f(X_t,\theta) - \hat{f}(X_t))^2, \qquad (2.10)$$

where Nh is a normalizing constant. Aït-Sahalia (1996a) estimated θ, say $\hat{\theta}_M$ that minimizes the distance between the densities with the same bandwidth, i.e,

$$\hat{\theta}_M \equiv arg \min_{\theta\in\Theta} \frac{1}{N}\sum_{t=1}^{N}(f(X_t,\theta) - \hat{f}(X_t))^2. \qquad (2.11)$$

2.2.2 Distribution Of The Test Statistic

Aït-Sahalia (1996a) used the asymptotic distribution of the kernel density estimate to derive the asymptotic distribution of the test statistic $\hat{M}$. He showed under the

conditions $lim_{N\to\infty}Nh = \infty$, $h \to 0$ and $lim_{N\to\infty}Nh^{4.5} = 0$, the test statistic $\hat{M}$ is distributed as

$$h^{-1/2}\{\hat{M} - E_M\} \xrightarrow{D} N(0, V_M), \tag{2.12}$$

where

$$E_M \equiv (\int_{-\infty}^{+\infty} K^2(x)dx)(\int_{\underline{x}}^{\overline{x}} f^2(x)dx), \tag{2.13}$$

$$V_M \equiv 2(\int_{-\infty}^{+\infty}\{\int_{-\infty}^{+\infty} K(u)K(u+x)du\}^2dx)(\int_{\underline{x}}^{\overline{x}} f^4(x)dx), \tag{2.14}$$

where $\underline{x}$ and $\overline{x}$ are the lowest and highest realizations of X_t in the data.

Therefore, the procedure of the test at level α is to

$$reject \quad H_0 : when \quad \hat{M} \geq \hat{c}(\alpha) \equiv \hat{E}_M + h^{1/2}z_{1-\alpha}/\hat{V}_M^{1/2}, \tag{2.15}$$

where $\hat{E}_M$ and $\hat{V}_M$ are the estimators of E_M and V_M. The estimators are the plugged-in types and have the expressions:

$$\hat{E}_M \equiv (\int_{-\infty}^{+\infty} K^2(x)dx)(\frac{1}{N}\sum_{t=1}^{N} \hat{f}(X_t)), \tag{2.16}$$

$$\hat{V}_M \equiv 2(\int_{-\infty}^{+\infty}\{\int_{-\infty}^{+\infty} K(u)K(u+x)du\}^2dx)(\frac{1}{N}\sum_{t=1}^{N} \hat{f}^3(X_t)). \tag{2.17}$$

2.3 Pritsker's Study

Using the test statistic above, Aït-Sahalia (1996a) only did empirical test for diffusion models on a data set and did not do simulation studies. Pritsker (1998) carried out simulation study on Aït-Sahalia's (1996a) test. As a result, he found that the empirical size of Aït-Sahalia's test is poor.

If the marginal density of the diffusion model is complicated (what's more is that many marginal densities of the diffusion models have no close form), studying the finite sample properties of the test of the diffusion model is a challenge work. It is well-known that the marginal density of the Vasicek (1977) model is Gaussian, which is the most used statistical distribution and well-developed in theory. Therefore, Pritsker selected the Vasicek (1977) model which is the most tractable to study Aït-Sahalia's (1996a) test.

Now we turn to know more details on the properties of the Vasicek (1977) model. The Vasicek (1977) model has the form:

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{2.18}$$

where the parameters κ and σ are restricted to be positive, and the value of α is finite.

Under the diffusion process in Equation (2.19), X has a normal marginal density.

$$f(x|\kappa, \alpha, \sigma) = \frac{1}{\sqrt{2\pi V_E}} e^{-0.5(\frac{x-\alpha}{\sqrt{V_E}})^2}, \tag{2.19}$$

where $V_E = \dfrac{\sigma^2}{2\kappa}$.

From equation (2.19), it is clear that the marginal density of X is a normal density with the unconditional mean α and variance $\dfrac{\sigma^2}{2\kappa}$. The rate of mean reversion becomes slowly when we lower the value of κ. Therefore, the parameter κ determines the persistence of the diffusion process.

In order to quantify the effect of κ on persistence, Pritsker fixed the marginal distribution but varied the persistence of the diffusion process. He changed the

value of σ^2 and κ in the same proportion, in this case the persistence of the process varied but the marginal density is not changed. The parameters in the baseline model Pritsker selected are $\kappa = 0.85837, \alpha = 0.089102$ and $\sigma^2 = 0.0021854$. These parameters were from Aït-Sahalia (1996b), which were obtained by applying the GMM based on the seven-day Eurodollar deposit rate between June 1, 1973 and February 25, 1995 from Bank of American. Pritsker also considered models in which the baseline κ and σ^2 are doubled, quadrupled, halved and quartered. Table 2.2 lists the corresponding models which are labeled model -2, model -1, model 0, model 1 and model 2. Although models toward the top of the table are less persistence, all models have the same marginal distribution.

	Parameters		
Model	κ	α	σ^2
-2	3.433480	0.089102	0.008742
-1	1.716740	0.089102	0.004371
0	0.858370	0.089102	0.002185
1	0.429185	0.089102	0.001093
2	0.214592	0.089102	0.000546

Table 2.2: Models considered by Pritsker (1998)

Pritsker (1998) performed 500 Monte Carlo simulations for each of the Vasicek (1977) model. In each simulation, he generated 22 years of daily data which gave

a total of 5500 observations. The bandwidth applied was the optimal bandwidth which minimized the Mean Integrated Squared Error (MISE) of the nonparametric kernel density estimate (More details about the bandwidth selection refer to Prisker (1998)). To compute the test statistic of Aït-Sahalia (1996a), he generated the following consistent estimates of M, V_M and E_M:

$$\hat{M} \quad \equiv \quad \min_{\theta \in \Theta} Nh \int_{\underline{x}}^{\overline{x}} [f(\hat{\theta}, u) - \hat{f}(u)]^2 \hat{f}(u) du, \tag{2.20}$$

$$\hat{E}_M \quad \equiv \quad (\int_{-\infty}^{+\infty} K^2(x) dx)(\int_{\underline{x}}^{\overline{x}} \hat{f}^2(u) du), \tag{2.21}$$

$$\hat{V}_M \quad \equiv \quad 2(\int_{-\infty}^{+\infty} \{\int_{-\infty}^{+\infty} K(u)K(u+x) du\}^2 dx)(\int_{\underline{x}}^{\overline{x}} \hat{f}^4(u) du), \tag{2.22}$$

where $\underline{x}$ and $\overline{x}$ are the highest and lowest realization in the data. The difference of these consistent estimators between Pritsker (1998) and Aït-Sahalia (1996a) is that Aït-Sahalia calculated these estimators by Riemann sum while Pritsker used Riemann Integral.

Using asymptotic critical values, Pritsker (1998) got the empirical rejection frequencies which showed in Table 2.3. In the case the Vasicek model 0, the empirical rejection frequeny is about 50% at the 5% confidence level. The rejection rates increase from model -2 to model -1 but they decrease from model -1 to model 2 rapidly. For the Vasicek model 2 which has the highest persistence, the empirical rejection frequeny is only 21% at the 5% confidence level.

Pritsker (1998) also showed the finite sample properties of kernel density estimates of the marginal distribution when interest rates are generated from the Vasicek model. He derived analytic expressions of finite sample bias, variance,

covariance and MISE for the nonparametric kernel estimator. He found that the optimal choice of bandwidth depends on the persistence of the process but not on the frequency with which the process was sampled. After comparing the finite sample and asymptotic properties of kernel density estimators of the marginal distribution for the Vasicek model, he maintained that the asymptotic approximation understated the finite sample magnitudes of the bias, variance, covariance and correlation of the kernel density estimator. In particular, he found that to obtain a reasonable agreement between the empirical size and the nominal size required about 2755 years of data.

Model	Rej.freq(5%)	Optimal Bandwidth
-2	45.60%	0.0140979
-1	57.40%	0.0175509
0	51.60%	0.0217661
1	40.80%	0.0268048
2	21.00%	0.0325055

Table 2.3: Empirical rejection frequencies using asymptotic critical values at 5% level, extracted from Pritsker(1998).

Chapter 3

Goodness-of-fit Test

3.1   Introduction

From the early chapters, we are aware that the misspecification for the diffusion process may be produced when a parametric model is used in a study. Therefore, goodness-of-fit tests arise aiming at testing the validity of the parametric model. The purpose of this chapter is to apply a version of Owen's (1988, 1990) empirical likelihood to formulate a test procedure on the specification of the stationary density of a diffusion model.

The null and alternative hypotheses we considered are:

$$H_0 \quad : \quad f(\cdot, \theta) = f(\cdot) \qquad for \qquad some \qquad \theta \in \Theta,$$

$$H_1 \quad : \quad f(\cdot, \theta) \neq f(\cdot) \qquad for \qquad all \qquad \theta \in \Theta, \qquad (3.1)$$

where Θ is a compact parameter space.

We take the opportunity to reformulate Aït-Sahalia's (1996a) test statistic via a version of the empirical likelihood. The test statistic Aït-Sahalia (1996a) proposed was directly based on the difference between the parametric density and the nonparametric kernel density estimator which brings undersmoothing. Our test statistic avoids undersmoothing as we carry out a local linear smoothing of the parametric density implied by the diffusion model under consideration.

We use a bootstrap procedure to profile the finite sample distribution of the test statistic in order to remove part of the problem appeared in Aït-Sahalia's (1996a) test. It is well known that both the bootstrap and the full empirical likelihood are computing intensive methods. Fortunately, we note that one version of empirical likelihood, the least squares empirical likelihood, can be computed efficiently. This least squares empirical likelihood was introduced by Brown and Chen (1998) and has a simpler form in one-dimension than the full empirical likelihood. It avoids maximizing a nonlinear function, and hence makes the computation of the test statistic straightforward. At the same time, this least squares empirical likelihood has a high level of approximation to the full empirical likelihood under some mild conditions. The difference between the full empirical likelihood and the least squares empirical likelihood based test statistic is just a smaller order, as indicated in Brown and Chen (2003). Therefore, we propose the test statistic based on the least square empirical likelihood to make the computation more efficient.

In this chapter, we introduce the empirical likelihood in Section 3.2 for the case of the mean parameter first. Then we extend the full empirical likelihood and the

least square empirical likelihood for the stationary density of the diffusion model as well. The least squares empirical likelihood based goodness-of-fit test and some of its properties is presented in Section 3.3.

3.2 Empirical Likelihood

3.2.1 The Full Empirical Likelihood

The conception of empirical likelihood is presented for the case of the mean parameter first. Then the details on the empirical likelihood for the stationary density of the diffusion model are described.

The early idea of empirical likelihood ratio appeared in Thomas and Grunke-meier (1975), who used a nonparametric likelihood ratio to construct confidence intervals for survival probabilities. It was Owen (1988) who extended the idea and proposed using empirical likelihood ratio to form confidence intervals for the mean parameter. Like other nonparametric statistical methods, the empirical likelihood is applied to data without assuming that they come from a known family of distribution. Other nonparametric inferences include the jackknife and the bootstrap. These nonparametric methods give confidence intervals and tests with validity not depending on strong distributional assumptions. Among these, the empirical like-lihood is known to be effective in certain aspects of inference as summarized in Owen (2001).

Let $X_1, X_2, \cdots, X_N$ be independent random vectors in R^p, with a common

distribution F. Then the empirical distribution function $\hat{F}$ is

$$\hat{F}(x) = N^{-1} \sum_{t=1}^{N} I(X_t \leq x),$$

where $I(\cdot)$ is the indicator function. Assume that what we are interested in is the

mean of the population, say $\theta = \theta(F)$. Let $p_1, p_2, \cdots, p_N$ be nonnegative probability

weight allocated to the sample. The empirical weighted distribution function is

$$\hat{F}_p(x) = \sum_{t=1}^{N} p_t I(X_t \leq x).$$

Then

$$\theta(p) = \int x d\hat{F}_p(x) = \sum_{t=1}^{N} p_t X_t$$

is the mean based on the distribution $\hat{F}_p$. The empirical likelihood of θ, evaluated

at $\theta = \theta_0$ is

$$L(\theta_0) = sup\{\prod_{t=1}^{N} p_t | \theta(p) = \theta_0, \sum_{t=1}^{N} p_t = 1\}. \tag{3.2}$$

If we only keep the natural constraint $\sum_{t=1}^{N} p_t(x) = 1$, after applying the basic

inequality, we have

$$\prod_{t=1}^{N} p_t \leq (\frac{1}{N} \sum_{t=1}^{N} p_t)^{1/N} = (\frac{1}{N})^{1/N}.$$

Since the equality holds if and only if $p_1 = p_2 = \cdots = p_N = \frac{1}{N}$. Therefore, the

maximum empirical likelihood is

$$L(\hat{\theta}) = N^{-N},$$

where the maximum empirical likelihood estimator is $\hat{\theta} = \bar{X} = 1/N \sum_{t=1}^{N} X_t$. The

empirical log-likelihood ratio $\ell(\theta_0)$ is

$$-2log\{L(\theta_0)/L(\hat{\theta})\} = -2inf\{\sum_{t=1}^{N} log(Np_t) | \theta(p) = \theta_0, \sum_{t=1}^{N} p_t = 1\}. \tag{3.3}$$

Introducing the Lagrange multiplier λ and γ, let

$$G = \sum_{t=1}^{N} logNp_t + \gamma(1 - \sum_{t=1}^{N} p_t) + N\lambda \sum_{t=1}^{N} p_t(X_t - \theta_0).$$

Setting to zero the partial derivative of G with respect to p_t gives

$$\frac{\partial G}{\partial p_t} = \frac{1}{p_t} - \gamma + N\lambda(X_t - \theta_0) = 0.$$

Applying the restriction $\sum_{t=1}^{N} p_t X_t = \theta_0$,

$$0 = \sum_{t=1}^{N} p_t \frac{\partial G}{\partial p_t} = N - \gamma.$$

So $\gamma = N$. Therefore we may write

$$p_t(x) = \frac{1}{N}\{1 + \lambda(X_t - \theta_0)\}^{-1}, t = 1, \cdots, N, \tag{3.4}$$

where $\lambda(x)$ is the root of

$$\sum_{t=1}^{N} \frac{X_t}{1 + \lambda(x)(X_t - \theta_0)} = 0. \tag{3.5}$$

Finally, we get the log empirical likelihood ratio

$$
\begin{aligned}
-2log\{L(\theta_0)/L(\hat{\theta})\} &= -2\{\sum_{t=1}^{N} log(Np_t)|\theta(p) = \theta_0, \sum_{t=1}^{N} p_t = 1\} \tag{3.6}\\
&= 2\sum_{t=1}^{N} log\{1 + \lambda(X_t - \theta_0)\}. \tag{3.7}
\end{aligned}
$$

Now we turn to the empirical likelihood for the stationary density of the diffusion model which is our interest of this thesis. For the diffusion model (1.1), we observe the process X_t at dates $\{t\Delta|t = 0, 1, \cdots, N\}$, where $\Delta > 0$ is generally small, but fixed, for example $\Delta = 1/250$(daily) and $\Delta = 1/12$(monthly). Let

$K_h(\cdot) = h^{-1}K(\cdot/h)$, then the standard kernel density estimator of $f(x)$ can be

expressed $\hat{f}(x) = \dfrac{1}{N}\sum_{t=1}^{N}K_h(x - X_t)$.

Let

$$\tilde{f}(x, \hat{\theta}) = \sum_{t=1}^{N}w_t(x)f(X_t, \hat{\theta}) \tag{3.8}$$

be the kernel smoothed density of the parametric density $f(x, \hat{\theta})$ by using the

same kernel and bandwidth. Here $\hat{\theta}$ is a consistent estimator of θ and $w_t(x) =$

$\dfrac{1}{N}K_h(x - X_t)\dfrac{s_2(x) - s_1(x)(x - X_t)}{s_2(x)s_0(x) - s_1^2(x)}$ is the local weight, where $s_r(x) = \dfrac{1}{N}\sum_{s=1}^{N}K_h(x -$

$X_s)(x - X_s)^r$ for $r = 0, 1, 2$.

In Chapter 2, we have already known that the test statistic proposed by Aït-

Sahalia (1996a) was based directly on the difference between the parametric density

of the diffusion model $f(x, \hat{\theta})$ and the nonparametric kernel density estimator $\hat{f}(x)$.

While the test statistic we considered is based on the difference between $\tilde{f}(x, \hat{\theta})$

and $\hat{f}(x)$. By doing this, the issue of bias associated with the nonparametric fit is

canceled so as to avoid undersmoothing. To appreciate this point, we note that if

$\hat{\theta}$ is a $\sqrt{N}$-consistent estimator of θ, then it may be shown from some algebra that

$$E\{\tilde{f}(x, \hat{\theta}) - \tilde{f}(x, \theta)\}^2 = O(\dfrac{1}{N}).$$

It follows a standard derivatation in kernel density estimator, for instance that

given in Silverman (1986), where $f(x)$ is the real density:

$$E[\hat{f}(x) - f(x)] = \dfrac{1}{2}h^2\sigma_k^2 f''(x) + o(h^2)$$

and

$$E[\tilde{f}(x, \hat{\theta}) - f(x)] = \dfrac{1}{2}h^2\sigma_k^2 f''(x) + o(h^2)$$

provided that the first three derivation of $f(x)$ exist, where $\sigma_k^2 = \int x^2 K(x)dx$, and

they are the same in the first term.

This implies that as $N \to \infty$

$$E^2[\hat{f}(x) - \tilde{f}(x,\hat{\theta})] = o(h^4). \tag{3.9}$$

From standard results in kernel estimator, the mean square error of $\hat{f}(x)$ is

$$
\begin{aligned}
MSE\{\hat{f}(x)\} &= E\{\hat{f}(x) - f(x)\}^2 \\
&= \frac{1}{4}h^4 f''^2(x)\sigma_k^4 + \frac{f(x)R(k)}{Nh} + o(h^4) + O(N^{-1}), \tag{3.10}
\end{aligned}
$$

where $R(K) = \int K^2(u)du$ which is $< \infty$.

Then the optimal local bandwidth that minimizes the leading term (first two terms)

of MSE is

$$h^* = (\frac{f(x)R(K)}{f''^2(x)\sigma_k^4})^{1/5}N^{-1/5}.$$

Finally, we get the optimal mean square error

$$MSE^*\{\hat{f}(x)\} = \frac{5}{4}\{f(x)R(K)\}^{4/5}\{f''(x)\sigma_k^2\}^{3/5}N^{-4/5}.$$

On the other hand, Aït-Sahalia's (1996a) test statistic was based on $\hat{f}(x) - f(x,\hat{\theta})$,

which measures directly the difference $\hat{f}(x)$ and $f(x,\hat{\theta})$. It can be shown that under

H_0,

$$E^2[\hat{f}(x) - f(x,\hat{\theta})] = O(h^4). \tag{3.11}$$

This means that it has the same order as the variance of $\hat{f}(x)$ if h is chosen to be

$O(N^{-1/5})$. Thus, to obtain an asymptotically normal distribution with zero mean,

h has to be smaller order than $N^{-1/5}$. This implies undersmoothing. By contrast, it can be seen from (3.9) that the use of the difference $\hat{f}(x) - \tilde{f}(x,\hat{\theta})$ can avoid undersmoothing. In other words, one can still use h at order of $N^{-1/5}$ and means that we also can use the Cross-Validation method to choose h.

In the following, we formulate the empirical likelihood ratio for the marginal density. At an arbitrary $x \in S$ where S is a compact set, let $p_t(x)$ be nonnegative numbers representing weights allocated to X_t. The empirical likelihood for $\tilde{f}(x,\hat{\theta})$ is

$$L\{\tilde{f}(x,\hat{\theta})\} = \max \prod_{t=1}^{N} p_t(x) \tag{3.12}$$

subject to $\sum_{t=1}^{N} p_t(x) = 1$ and $\sum_{t=1}^{N} p_t(x)Q_t(x) = 0$, where $Q_t(x) = [K_h(x - X_t) - \tilde{f}(x,\hat{\theta})]$. The idea of the empirical likelihood is to find the optimal $p_t(x)$ at each X_t in order to maximize $\prod_{t=1}^{N} p_t(x)$ under the two restrictions.

We apply the method of Lagrange multipliers to work out the optimal problem with restrictions (see Owen (2000)). Introducing the Lagrange multiplier $\lambda(x)$ and $\gamma(x)$, we suppose

$$G = \sum_{t=1}^{N} log p_t(x) - N\lambda(x) \sum_{t=1}^{N} p_t(x)Q_t(x) + \gamma(x)\{\sum_{t=1}^{N} p_t(x) - 1\}.$$

Setting to zero the partial derivative of G with respect to $p_t(x)$ gives

$$\frac{\partial G}{\partial p_t} = \frac{1}{p_t} - N\lambda(x)Q_t(x) + \gamma(x) = 0.$$

Applying the restriction $\sum_{t=1}^{N} p_t(x)Q_t(x) = 0$ and $\sum_{t=1}^{N} p_t(x) = 1$,

$$0 = \sum_{t=1}^{N} p_t \frac{\partial G}{\partial p_t} = N + \gamma(x).$$

So $\gamma = -N$. Therefore we may write

$$p_t(x) = \frac{1}{N}\{1 + \lambda(x)Q_t(x)\}^{-1}, t = 1, \cdots, N, \tag{3.13}$$

where $\lambda(x)$ is the root of

$$\sum_{t=1}^{N} \frac{Q_t(x)}{1 + \lambda(x)Q_t(x)} = 0. \tag{3.14}$$

The case where $p_t(x) = \dfrac{1}{N}$ corresponds to the conventional kernel density estimate.

Finally, we get the log empirical likelihood ratio for the marginal density

$$
\begin{aligned}
\ell\{\tilde{f}(x,\hat{\theta})\} &= -2log[L\{\tilde{f}(x,\hat{\theta})\}N^N] \\
&= 2\sum_{t=1}^{N} log[1 + \lambda(x)\{K_h(x - Xt) - \tilde{f}(x,\hat{\theta})\}].
\end{aligned}
\tag{3.15}
$$

Clearly the computation of $\ell\{\tilde{f}(x,\hat{\theta})\}$ involves solving $\lambda(x)$ as a root of a non-linear equation (3.14). People use the conjugation gradient method which requires derivative calculations and one-dimensional sub-minimization, which is quite computation intensive. This is on top of the fact that we need to evaluate $\ell\{\tilde{f}(x,\hat{\theta})\}$ at many x points when formulating the empirical likelihood test statistic.

3.2.2 The Least Squares Empirical Likelihood

To overcome the computational difficulty of the empirical likelihood, Brown and Chen (1998) proposed a "least-squares" version of the empirical likelihood.

The empirical likelihood actually maximizes such function $\sum_t log(Np_t)$ whereas the least squares empirical likelihood maximizes the function $-\sum_t (Np_t(x) - 1)^2$ under some restriction. It is also called the Euclidean likelihood (Owen 2001).

Brown and Chen (1998) showed that the least squares empirical likelihood curves followed those of the full empirical likelihood closely under some mild conditions. In particular, the least squares empirical likelihood has a close form and this character makes its computation straightforward.

We provide here the details of the method in a general setting following Brown and Chen (1998) because the least squares empirical likelihood for the marginal density is based on this theory. We assume the dimension of the parameter θ (which has a true value θ_0) is p. Let $Z_1(\theta), Z_2(\theta), \cdots, Z_N(\theta)$ be k dimensional independent but not necessarily identically distributed random vectors and $E\{Z_i(\theta_0)\} = 0, i = 1, \cdots, N$. The least squares empirical likelihood for θ is defined as

$$lsl(\theta) = min \sum_{t=1}^{N} (Np_t(x) - 1)^2, \tag{3.16}$$

subject to $\sum_{t=1}^{N} p_t(x) = 1$ and $\sum_{t=1}^{N} p_t(x)Z_t(\theta) = 0$.

Actually $lsl(\theta) = N^2 min \sum_t p_t^2 - 2N min \sum_t p_t + N = N^2 min \sum_t p_t^2 - N$. Let $M(\theta) = min \sum_t p_t^2$, then we just should compute $M(\theta)$ directly.

Applying Lagrange multipliers $\alpha = (\alpha_1, \cdots, \alpha_p)^T$, the objective function is

$$G = \sum_t p_t^2 + \alpha_0 \sum_t p_t + \alpha^T \sum_t p_t Z_t(\theta).$$

Setting to zero the partial derivative of G with respect to p_t gives

$$\frac{\partial G}{\partial p_t} = 2p_t + \alpha_0 + \sum_j \alpha_j Z_{tj}(\theta) = 0.$$

Therefore, we get

$$p_t = -\frac{1}{2}\{\alpha_0 + \sum_t \alpha_j Z_{tj}(\theta)\}. \tag{3.17}$$

Let $\alpha^T = (\alpha_0, \alpha_1, \cdots, \alpha_k)$, from the structural constraints we write

$$(10\cdots0)^T = -\frac{1}{2}\begin{pmatrix} N & V^T \\ V & R \end{pmatrix}\alpha$$

where $V^T = (V_1, V_2, \cdots, V_k)$, $V_j = \sum_t Z_{tj}(\theta)$ and $R = (R_{jj'})_{k\times k}$, $R_{jj'} = \sum_t Z_{tj}(\theta)Z_{tj'}(\theta)$.

Then we get the optimal p_t which is

$$p_t = N^{-1} + N^{-1}(N^{-1}V - Z_t(\theta))^T H^{-1}V, \tag{3.18}$$

where $H = R - N^{-1}VV^T$. Therefore, the least squares empirical likelihood for the

mean parameter is

$$lsl(\theta) = V^T H^{-1}V. \tag{3.19}$$

Now we turn to our interest, the marginal density. At an arbitrary $x \in S$,

let $p_t(x)$ be nonnegative numbers representing weights allocated to X_t. The least

squares empirical likelihood for the marginal density is

$$lsl\{\tilde{f}(x,\hat{\theta})\} = min \sum_{t=1}^{N}(Np_t(x) - 1)^2, \tag{3.20}$$

subject to $\sum_{t=1}^{N} p_t(x) = 1$ and $\sum_{t=1}^{N} p_t(x)Q_t(x) = 0$, where $Q_t(x) = [K_h(x - X_t) - \tilde{f}(x,\hat{\theta})]$.

Let $V = \sum_t Q_t(x)$, $R = \sum_t Q_t^2(x)$ and $H = R - N^{-1}V^2$, plugging (3.18) we

have

$$\begin{aligned} p_t &= N^{-1} + N^{-1}(N^{-1}V - Q_t)H^{-1}V \\ &= N^{-1}H^{-1}\{N^{-1}V^2 - Q_tV + H\} \\ &= N^{-1}H^{-1}\{R - VQ_t\}, \end{aligned} \tag{3.21}$$

and the least squares empirical likelihood for the marginal density is

$$
\begin{aligned}
lsl\{\tilde{f}(x,\hat{\theta})\} &= N^2 min \sum_t p_t^2 - n = \frac{V^2}{H} \\
&= (\sum_t Q_t)^2 \{\sum_t Q_t^2 - N^{-1}(\sum_t Q_t)^2\}^{-1} \\
&= \{\sum_t Q_t^2 (\sum_t Q_t)^{-2} - N^{-1}\}^{-1}.
\end{aligned}
\tag{3.22}
$$

Compared with the full empirical likelihood, the least squares empirical like-lihood for the marginal density needs only two simple statistics $\sum_t Q_t(x)$ and $\sum_t Q_t^2(x)$ while computation of the full empirical likelihood is more complicated.

3.3   Goodness-of-fit Test

Based on the full empirical likelihood and the least squares empirical likelihood for the marginal density given in Section 3.2, we define the full empirical likelihood and least squares empirical likelihood test statistics as

$$
\begin{aligned}
\hat{N}(h) &= \int \ell\{\tilde{f}(x,\hat{\theta})\}\pi(x)dx, \\
\hat{N}_{LS}(h) &= \int lsl\{\tilde{f}(x,\hat{\theta})\}\pi(x)dx,
\end{aligned}
\tag{3.23}
$$

where $\pi(x)$ is a probability weight function satisfying $\int \pi(x)dx = 1$ and $\int \pi^2(x)dx < \infty$, for example simple function.

Let $\gamma(x)$ be a random process with $x \in S$. Denote $\gamma(x) = \tilde{o}_p(\delta_n)$ for the fact that $\sup_{x \in S} |\gamma(x)| = o_p(\delta_n)$ for a sequence δ_n. Using the technique proposed by Chen (1996), one can develop the expansion for the log EL ratio for the marginal density

as

$$
\begin{aligned}
\ell\{\tilde{f}(x,\hat{\theta})\} &= -2log[L\{\tilde{f}(x,\hat{\theta})\}N^N] \\
&= (Nh)\frac{(\hat{f}(x) - \tilde{f}(x,\theta))^2}{R(K)f(x)} + \tilde{o}_p\{(Nh)^{-1/2}log(N)\}. \qquad (3.24)
\end{aligned}
$$

Hence, the test statistic for the full empirical likelihood for the marginal density is

$$
\begin{aligned}
\hat{N}(h) &= \int \ell\{\tilde{f}(x,\hat{\theta})\}\pi(x)dx \\
&= (Nh)\int \frac{(\hat{f}(x) - \tilde{f}(x,\theta))^2}{R(K)f(x)}\pi(x)dx + o_p\{(Nh)^{-1/2}log(N)\}. \quad (3.25)
\end{aligned}
$$

Brown and Chen (1998) pointed out that both the full empirical likelihood and the least squares empirical likelihood have the same first order term. Therefore, we have

$$
\ell(\tilde{f}(x,\hat{\theta})) = lsl(\tilde{f}(x,\hat{\theta})) + \tilde{o}_p((Nh)^{-1/2}logN). \qquad (3.26)
$$

More details refer to Brown and Chen (1998).

Hence

$$
\int \ell(\tilde{f}(x,\hat{\theta}))\pi(x)dx = \int lsl(\tilde{f}(x,\hat{\theta}))\pi(x)dx + o_p((Nh)^{-1/2}logN). \qquad (3.27)
$$

The test statistic of the least squares empirical likelihood for the marginal density is

$$
\begin{aligned}
\hat{N}_{LS}(h) &= \int lsl\{\tilde{f}(x,\hat{\theta})\}\pi(x)dx \\
&= \hat{N}(h) + o_p\{(Nh)^{-1/2}log(N)\}. \qquad (3.28)
\end{aligned}
$$

It is clear that the full empirical likelihood test statistic $\hat{N}(h)$ and the least squares empirical likelihood test statistic $\hat{N}_{LS}(h)$ are same in the first order. However, the

computation of the least squares empirical likelihood test statistic is more efficient

than that of the full empirical likelihood test statistic. Therefore, we will use it for

our test of the marginal density in this thesis.

Let the standard test statistic be $\hat{L} = \dfrac{\hat{N}_{LS}(h) - 1}{\sigma_h}$, where $\sigma_h^2 = 2hC(K, \pi)$ and

$C(K, \pi) = R^{-2}(K)K^{(4)}(0)\int \pi^2(x)dx$. Then under some assumptions for instance

these given in Chen (1996) and H_0 in (3.1), we have

$$\hat{L} = \frac{\hat{N}_{LS}(h) - 1}{\sigma_h} \to_D N(0, 1) \tag{3.29}$$

as $N \to \infty$.

In the following, we discuss how to get a critical value for the test statistic

based on the least squares empirical likelihood. The exact α-level critical value,

$l_\alpha(0 < \alpha < 1)$ is the $1 - \alpha$ quantile of the exact finite-sample distribution of the

test statistic. However, l_α can not be evaluated in practice because the distribution

of the test statistic is unknown. We get an asymptotic α-level critical value, say

l_α^*, by the bootstrap. The bootstrap procedure is:

1. Use the data set $\{X_t; t = 1, 2, \cdots, N\}$ to estimate θ by $\hat{\theta} = \text{argmax}_\theta L(\theta; \Delta)$,

where

$$L(\theta; \Delta) = \frac{1}{N}\sum_{t=1}^{N} log\{p_\theta(X_{t+1}|X_t, \Delta)\} \tag{3.30}$$

is the likelihood under H_0. Denote the resulting estimate by $\hat{\theta}$.

2. Compute the test statistic $\hat{N}_{LS}(h)$ for a given h.

3. Generate a bootstrap resample $\{X_t^*; t = 1, 2, \cdots, N\}$ from the transition

density $p_{\hat{\theta}}(X_{t+1}^*|X_t^*, \Delta)$ with X_0^* generated from $f(x, \hat{\theta})$. Use the new data set

$\{X_t^*; t = 1, 2, \cdots, N\}$ and the function $L(\hat{\theta}, \Delta) = \dfrac{1}{N} \sum\limits_{t=1}^{N} log\{p_{\hat{\theta}}(X_{t+1}^* | X_t^*, \Delta)\}$ to re-estimate $\hat{\theta}$. Denote the resulting estimate by $\hat{\theta}^*$. Compute the statistic $\hat{N}_{LS}^*(h)$ that is obtained by replacing X_t and $\hat{\theta}$ with X_t^* and $\hat{\theta}^*$.

4. Repeat the above steps B times for example $B = 300$ and produce B versions of $\hat{N}_{LS}^{*1}, \cdots, \hat{N}_{LS}^{*m}, \cdots, \hat{N}_{LS}^{*B}$ for $m = 1, 2, \cdots, B$. Use the B values of $\hat{N}_{LS}^*(h)$ to construct their empirical bootstrap distribution function, that is , $F(u) = \dfrac{1}{B} I(\hat{N}_{LS}^* \leq u)$. Use the Ordered statistic, we have $\hat{N}_{LS}^{*(1)}, \leq, \cdots, \leq, \hat{N}_{LS}^{*(B)}$. Hence the asymptotic critical value is $l_\alpha^* = \hat{N}_{LS}^{*(T)}$ where $T = N(1 - \alpha)$.

In fact, under some assumptions and H_0 in (3.1) it may be showed

$$\lim_{N \to \infty} P(\hat{N}_{LS}^*(h) > l_\alpha^*) = \alpha. \tag{3.31}$$

The main result on the behavior of the test statistic $\hat{N}_{LS}$ under H_0 is that l_α^* is an asymptotically correct α-level critical value under the null hypothesis.

Chapter 4

Simulation Studies

4.1   Introduction

In this chapter we report results from simulation studies designed to evaluate the performance of the proposed empirical likelihood goodness-of-fit test for a diffusion process. We also compare our test with the test proposed by Aït-Sahalia (1996a).

In Section 4.2, we discuss the details on the simulation procedure including some practical issues such as the parameters estimation, bandwidth selection, initial value and the generation of a diffusion process. Many diffusion models have been developed so far. Similar to Pritsker (1998), we only focus on the simplest and the most important model, the Vasicek (1977) model, in this thesis. We discuss the computation of the test statistic and how to obtain the critical value for the test statistic. The simulation results including the empirical size and power of the test

for both the IID case and the diffusion models are presented in Section 4.3. Finally,

we reevaluate the performance of the test proposed by Aït-Sahalia (1996a).

4.2 Simulation Procedure

Under the null hypothesis H_0 in (3.1), the conditional likelihood of θ based on

the observed data $\{X_t\}_{t=1}^N$ is

$$L(\theta; \Delta) = \frac{1}{N} \sum_{t=1}^N \log \{p_\theta(X_{t+1}|X_t; \Delta)\}, \quad (4.1)$$

in which $p_\theta(\cdot|\cdot, \Delta)$ is transition density specified by H_0. Hence, the maximum

likelihood estimator of θ is $\hat{\theta} = \arg\max_\theta L(\theta; \Delta)$.

In this thesis, the simulation is focused on the Vasicek (1977) model which has

the form

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \quad (4.2)$$

where the parameters κ and σ are restricted to be positive, and value of α is finite.

Under the diffusion process (4.2), the marginal and transition densities of the

diffusion process are Gaussian. The marginal density of X is

$$f(x|\kappa, \alpha, \sigma) = \frac{1}{\sqrt{2\pi V_E}} exp^{-0.5(\frac{x-\alpha}{\sqrt{V_E}})^2}, \quad (4.3)$$

where $V_E = \dfrac{\sigma^2}{2\kappa}$. The transition density of X is

$$p(X_{t+1}, |X_t, \Delta, \kappa, \alpha, \sigma) = \frac{1}{\sqrt{2\pi V(X_{t+1}|X_t)}} exp^{-0.5(\frac{X_{t+1}-\mu(X_{t+1}|X_t)}{\sqrt{V(X_{t+1}|X_t)}})^2}, \quad (4.4)$$

where $\mu(X_{t+1}|X_t) = \alpha + (X_t - \alpha)e^{-\kappa\Delta}$ and $V(X_{t+1}|X_t) = V_E(1 - e^{-2\kappa\Delta})$.

Following the formula (4.1), the conditional likelihood of θ on the observed Vasicek process $\{X_t\}_{t=1}^N$ is

$$L(\theta; \Delta) = \frac{1}{N} \sum_{t=1}^N \{-\frac{1}{2} log(2\pi V(X_{t+1}|X_t)) - \frac{1}{2} \frac{(X_{t+1} - \mu(X_{t+1}|X_t))^2}{V(X_{t+1}|X_t))}\}. \qquad (4.5)$$

The maximum likelihood estimator of θ, say $\hat{\theta}$, can be obtained by maximizing (4.5).

In the simulation study, we use the parameters which are applied in the simulation study of Pritsker (1998). To be consistent, we also call these Vasicek models as model -2, model -1, model 0, model 1 and model 2 which all have the same marginal density but different levels of dependence. Model -2 has the least persistent and model 2 has the most persistent.

Now we turn to bandwidth selection in the simulation. The choice of bandwidth is important to the kernel density estimate and the test statistic under consideration. Small values of bandwidth make the estimate look "wiggly" and show spurious features, whereas too big values of bandwidth lead to too much smoothing and may not reveal structural features for the observations. In general, a bandwidth should be chosen to minimize the Integrated Squared Error (ISE) or the Mean Integrated Squared Error (MISE). There are a number of bandwidth selection methods which have been proposed by researchers over the years, for example the reference to a standard distribution approach, the Cross-Validation and the Plug-in Method. Berwin (1993) gave a review on bandwidth selection in kernel density estimation. Our interest in the simulation is the Vasicek model whose marginal density is Nor-

mal, it is favorable for us to employ the reference to a normal distribution approach

for bandwidth selection. Based on the Mean Integrated Squared Error, the optimal

global bandwidth is

$$h^* = \{\frac{R(K)}{\sigma_k^4 R(f^{(2)})}\}^{\frac{1}{5}} N^{-\frac{1}{5}} \tag{4.6}$$

where $R(K) = \int K^2(t)dt$, $\sigma_k^2 = \int K(t)t^2 dt$ (see Chapter 3 for details) and N is the

sample size.

Usually the term $R(f^{(2)})$ is unknown in the expression. The reference to a

normal distribution approach replaces the unknown density function f in (4.6) by

a normal density function, which matches the empirical mean and variance of the

data.

If we use Gaussian kernel $K(u) = \frac{1}{2\pi}e^{-\frac{u^2}{2}}$, the reference to a normal distribution

approach yields the optimal bandwidth

$$h^* = 1.06\hat{\sigma} N^{-\frac{1}{5}},$$

where $\hat{\sigma}^2$ is the sample variance and N is the sample size. In our simulation, we

employ the Biweight kernel $K(u) = \frac{15}{16}(1-u^2)I(u)$ where $I(\cdot)$ signifies the indicator

function and get the optimal bandwidth

$$h^* = 2.78\hat{\sigma} N^{-\frac{1}{5}}.$$

Table 4.1 lists the optimal bandwidth for a variety of sample sizes considered in

the simulation. We would like to highlight that bandwidths for IID are same as

those for dependent observations generated from a diffusion model as long as IID

and dependent observations have the same marginal density. This is due to the so called "prewhitening" effect by a bandwidth in the kernel smoothing of dependent data. The effect of dependence is only felt in the second order.

	Optimal	Bandwidth					
Sample size	n=100	n=120	n=200	n=250	n=500	n=1000	n=2000
h^*	0.0398	0.0384	0.0347	0.0332	0.0289	0.0251	0.219

Table 4.1: Optimal bandwidth corresponding different sample size

To simulate a diffusion process, the first step is generating the starting value X_0. As mentioned above, our interest is the Vasicek model where the exact marginal distribution is Normal. Therefore, we simulate X_0 simply from the Normal stationary distribution.

After generating the initial value X_0, we can generate a diffusion process. As the transition distribution of X_{t+1} given X_t is available from the transition density $p(X_{t+1}|X_t, \Delta)$ under H_0, we can simulate X_{t+1} from the transition distribution given X_t, whiles X_1 is simulated based on X_0 given above. For the Vasicek model, the transition density follows a conditional normal density where the mean is $\alpha + (X_t - \alpha)e^{-\kappa\Delta}$ and the variance is $V_E(1 - e^{-2\kappa\Delta})$.

To profile the finite sample distribution of the test statistic, we employ the bootstrap procedure which is known to be time-consuming. If we choose $\Delta = \dfrac{1}{250}$ (daily), the calculation of the test statistic will take long time. To improve the computing efficiency of the test statistic, we choose another reasonable interval

$\Delta = \dfrac{1}{12}$ (monthly) in simulation study.

Since the support of the density function $f(\cdot)$ may not be compact, we choose

the weight function $\pi(\cdot)$ to be compactly supported to truncate out the tail regions

of the marginal density, in particular we may use

$$\pi(x) = \begin{cases} (R_2 - R_1)^{-1} & \text{if } x \in [R_1, R_2], \\ 0 & \text{otherwise,} \end{cases}$$

where $0 \leq R_1 < R_2$ for some constant R_1 and R_2, which should be chosen properly

so that the two tail regions $(0, R_1)$ and (R_2, ∞) cover around 10% of data. In the

simulation, we use the Biweight kernel function.

Let $\{t_l\}_{l=1}^{Q}$ be equally spaced points within $[R_1, R_2]$. At each fixed points $t_l, l =$

$1, \cdots, Q$, the likelihood goodness-of-fit is $lsl\{\tilde{f}(t_l, \hat{\theta}\}$. Then a discretization of the

test statistic for a bandwidth h is $\hat{N}_{LS}(h) = \dfrac{1}{Q} \sum_{l=1}^{Q} lsl\{\tilde{f}(t_l, \hat{\theta})\}$.

Lastly, we find critical value l_{α}^{*} following the bootstrap procedure which is al-

ready completely described in Chapter 3.

4.3 Simulation Result

4.3.1 Simulation Result For IID Case

Before we start to evaluate the performance of the proposed empirical likeli-

hood goodness-of-fit test for diffusion models, we first consider the test for IID case

to make sure that the method we proposed works for IID. We generate X which fol-

lows a Normal distribution with mean 0.089102 and variance 0.001273052, and also

has the same marginal density as dependent observations generated from diffusion

models which are considered in the later simulation. We apply MLE to estimate

the mean and variance parameters from the IID. The process of computation of

test statistics and critical values is the same as that of diffusion models which had

already discussed in the early section.

To estimate the empirical size of the test for IID case, we performed 500 sim-

ulations on 19 spaced bandwidths ranging from 0.003 to 0.048. The range of

bandwidths includes the optimal bandwidth given by Table 4.1 and offers a wide

range of smoothness. In order to learn the trend with increased sample size, we

consider three different sample sizes which are 100, 200 and 500 respectively. Ta-

ble 4.2 lists the size of the bootstrap based least squares empirical likelihood test

for IID case for a set of bandwidth values and their sample sizes. Figure 4.1 is

a graphical illustration of Table 4.2 where h^* is the optimal bandwidth given in

Table 4.1 and is indicated by the vertical line. It is obviously that the empirical

rejection frequencies become more stable around 0.05 with increased sample size.

In the case the sample size is 100, the empirical size first increases with increased

bandwidth. When the bandwidth equals 0.009, the empirical size reaches 0.04.

After that, the empirical size is decreasing with increased bandwidth. The per-

formance of our test is improved when the sample size is doubled. The empirical

size remains steady around 0.05 but it decreases rapidly with the bandwidth in-

creasing after bandwidth equals 0.04. When the sample size is as large as n=500,

the empirical size rates are steadily around 0.05 for a wide range of bandwidths.

Therefore, the empirical likelihood goodness-of-fit test we proposed has reasonable empirical rejection frequencies for IID case when the critical value is generated via the bootstrap.

bandwidth	Sample Size			bandwidth	Sample Size		
	100	200	500		100	200	500
0.003	0.08	0.04	0.054	0.03	0.056	0.046	0.052
0.006	0.068	0.038	0.05	0.032	0.054	0.05	0.052
0.009	0.04	0.052	0.054	0.034	0.05	0.048	0.052
0.012	0.052	0.056	0.05	0.036	0.048	0.048	0.052
0.015	0.056	0.05	0.048	0.038	0.046	0.05	0.056
0.018	0.06	0.05	0.048	0.04	0.044	0.048	0.058
0.021	0.07	0.052	0.056	0.042	0.044	0.04	0.054
0.024	0.064	0.052	0.058	0.044	0.04	0.034	0.058
0.027	0.058	0.05	0.056	0.046	0.04	0.036	0.054
				0.048	0.034	0.032	0.046

Table 4.2: Size of the bootstrap based LSEL Test for IID for a set of bandwidth values and their sample sizes of 100, 200 and 500
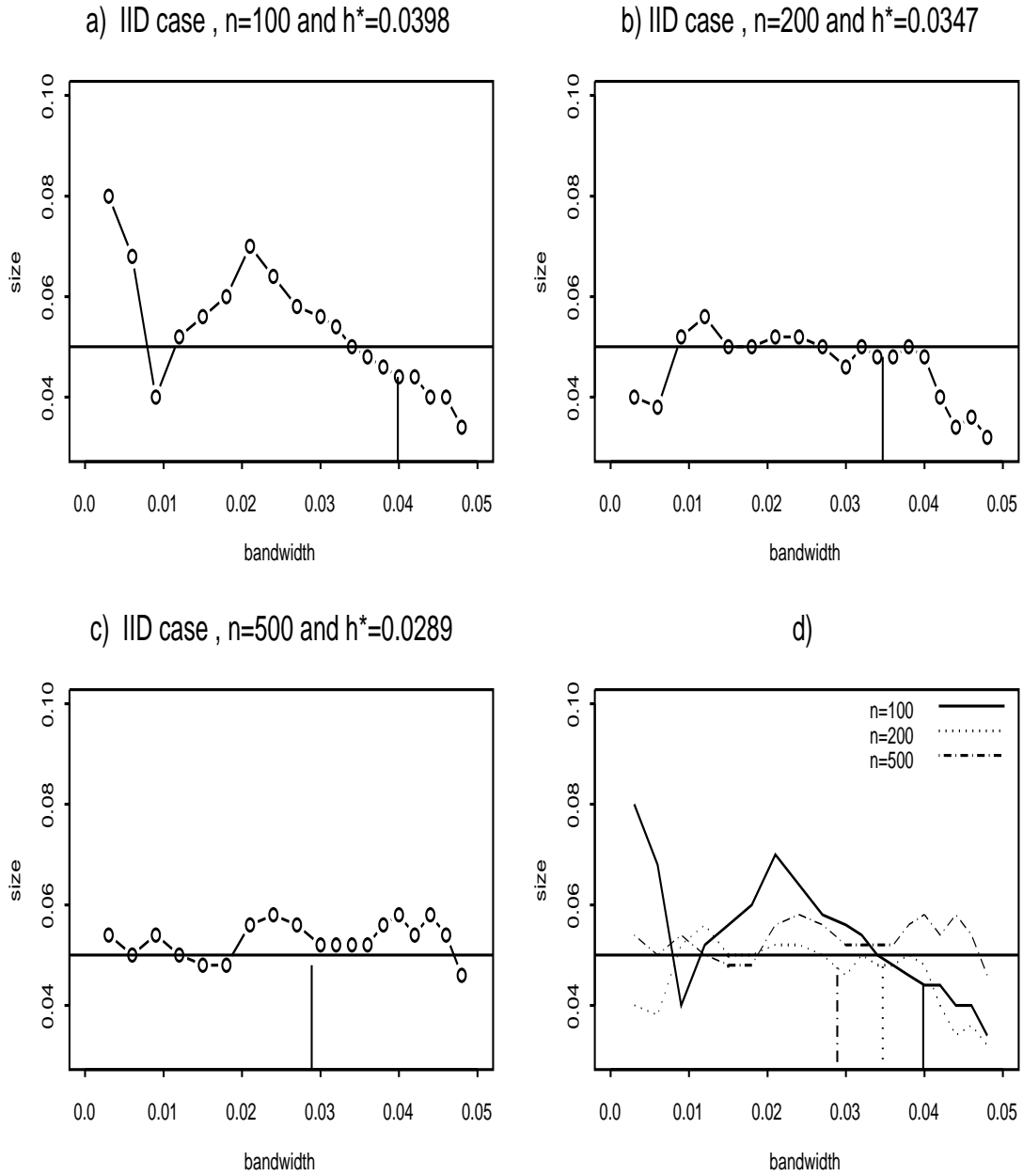
Figure 4.1: Graphical illustrations of Table 4.2, where h* are the optimal bandwidths given in Table 4.1 and are indicated by vertical lines.

4.3.2 Simulation Result For Diffusion Processes

We then carry out simulations on our empirical likelihood goodness-of-fit test for the marginal density for each of the Vasicek models, from model -2 to model 2 on 10 equally spaced bandwidths ranging from 0.005 to 0.05. This range of bandwidths includes the optimal bandwidth given in Table 4.1 and offers a wide range of smoothness. On the whole, the empirical likelihood goodness-of-fit test we proposed has reasonable empirical rejection rates for diffusion models when the critical value is generated by the bootstrap and the performance of our test is much better than that of Pritsker (1998). The performance of the test improves with increased sample size. On the other hand, these tests for the Vasicek models with low persistence (model -2 and model -1) have better performance than those with high persistence (model 2 and model 1). In the case the Vasicek model -2 which has the least persistence, the test has reasonable size even when the sample size is as small as n=120 (about ten years). The empirical size is about 0.05 when the bandwidth changes from 0.005 to 0.03 and it decreases sharply to 0.006 when the larger bandwidth is applied. The empirical size is about 0.016 when the bandwidth equals 0.04 which is near the optimal bandwidth. The trends for sample sizes n=250 (about 20 years), 500 (about 40 years) and 1000 (about 80 years) are similar with that of sample size n=120 but the empirical sizes are steady around 0.05. The range of bandwidth where the test has reasonable sizes also extends and has reasonable size around the optimal bandwidth. In the case the Vasicek model

2 which has the highest persistence in five models, the empirical size is decreasing with increased bandwidth. When the bandwidth is 0.005, the empirical size is as large as 0.164 which is worse comparing with the result of the Vasicek model -2. Only when the bandwidth is around 0.025, the test has a reasonable size. With the sample sizes increasing, the performance of tests improves and has reasonable sizes around the optimal bandwidth.

bandwidth	Size			
model-2	n=120	n=250	n=500	n=1000
0.005	0.044	0.064	0.044	0.034
0.01	0.044	0.078	0.052	0.058
0.015	0.058	0.086	0.048	0.064
0.02	0.06	0.076	0.048	0.06
0.025	0.06	0.072	0.052	0.06
0.03	0.042	0.066	0.046	0.064
0.035	0.03	0.066	0.044	0.058
0.04	0.016	0.052	0.044	0.046
0.045	0.01	0.038	0.036	0.034
0.05	0.006	0.022	0.024	0.028
h^*	0.02	0.072	0.048	

Table 4.3: Size of the bootstrap based LSEL Test for the Vasicek model -2 for a set of bandwidth values and their sample sizes of 120, 250, 500 and 1000

bandwidth	Size			
model-1	n=120	n=250	n=500	n=1000
0.005	0.04	0.062	0.062	0.048
0.01	0.046	0.068	0.058	0.052
0.015	0.046	0.072	0.06	0.048
0.02	0.056	0.062	0.06	0.05
0.025	0.058	0.064	0.062	0.048
0.03	0.036	0.06	0.062	0.054
0.035	0.018	0.048	0.058	0.052
0.04	0.012	0.03	0.052	0.048
0.045	0.004	0.016	0.044	0.042
0.05	0.002	0.008	0.028	0.020
h^*	0.012	0.044	0.066	0.052

Table 4.4: Size of the bootstrap based LSEL Test for the Vasicek model -1 for a set of bandwidth values and their sample sizes of 120, 250, 500 and 1000

bandwidth	Size			
model0	n=120	n=250	n=500	n=1000
0.005	0.06	0.072	0.074	0.068
0.01	0.074	0.07	0.082	0.064
0.015	0.066	0.068	0.072	0.068
0.02	0.062	0.064	0.068	0.070
0.025	0.052	0.064	0.072	0.074
0.03	0.036	0.054	0.064	0.074
0.035	0.014	0.03	0.048	0.068
0.04	0.006	0.01	0.036	0.062
0.045	0.004	0.004	0.03	0.046
0.05	0.002	0	0.016	0.03
h^*	0.008	0.038	0.062	0.074

Table 4.5: Size of the bootstrap based LSEL Test for the Vasicek model 0 for a set of bandwidth values and their sample sizes of 120, 250, 500 and 1000

bandwidth	Size				
model1	n=120	n=250	n=500	n=1000	n=2000
0.005	0.092	0.084	0.064	0.054	0.054
0.01	0.106	0.068	0.062	0.052	0.052
0.015	0.102	0.076	0.06	0.06	0.05
0.02	0.106	0.076	0.058	0.062	0.058
0.025	0.074	0.066	0.046	0.056	0.06
0.03	0.038	0.028	0.024	0.05	0.064
0.035	0.014	0.014	0.016	0.04	0.062
0.04	0.008	0.004	0.004	0.028	0.056
0.045	0.002	0.002	0.004	0.016	0.034
0.05	0	0	0.004	0.012	0.014
h*	0.008	0.02	0.03	0.06	0.052

Table 4.6: Size of the bootstrap based LSEL Test for the Vasicek model 1 for a set of bandwidth values and their sample sizes of 120, 250, 500, 1000 and 2000

bandwidth	Size				
model2	n=120	n=250	n=500	n=1000	n=2000
0.005	0.164	0.104	0.078	0.05	0.042
0.01	0.158	0.084	0.088	0.046	0.042
0.015	0.152	0.084	0.076	0.056	0.04
0.02	0.092	0.07	0.06	0.064	0.042
0.025	0.042	0.038	0.036	0.038	0.032
0.03	0.022	0.012	0.01	0.022	0.026
0.035	0.012	0.004	0.004	0.008	0.022
0.04	0.002	0.004	0.004	0	0.016
0.045	0.002	0.002	0	0	0.006
0.05	0.002	0	0	0	0.002
h^*	0.006	0.01	0.016	0.038	0.038

Table 4.7: Size of the bootstrap based LSEL Test for the Vasicek model 2 for a set of bandwidth values and their sample sizes of 120, 250, 500, 1000 and 2000
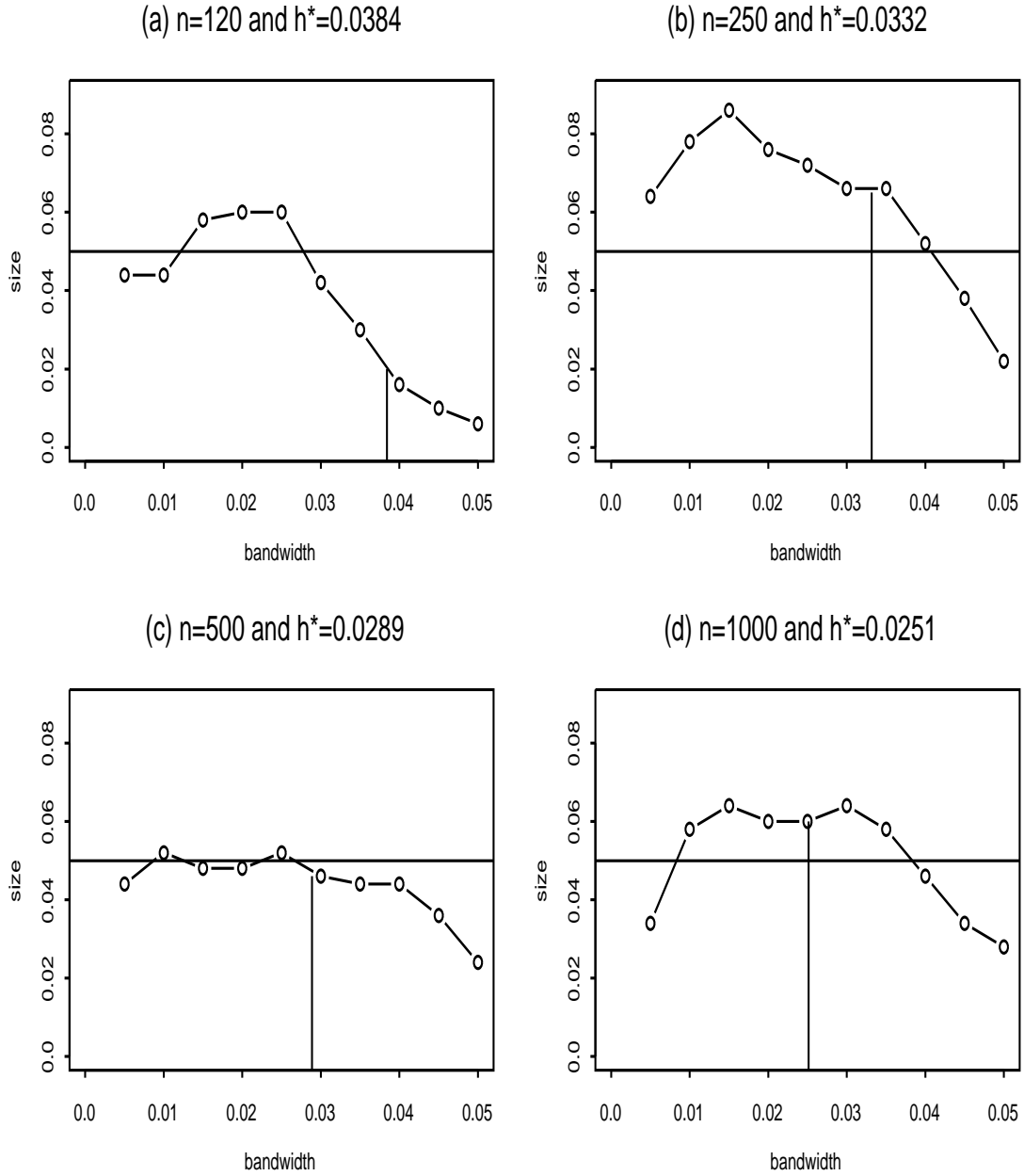
Figure 4.2: Graphical illustrations of Table 4.3 for the Vasicek model -2, where h* are the optimal bandwidth given in Table 4.1 and are indicated by the vertical lines.

Figure 4.3: Graphical illustrations of Table 4.4 for the Vasicek model -1, where h* are the optimal bandwidth given in Table 4.1 and are indicated by the vertical lines.

Figure 4.4: Graphical illustrations of Table 4.5 for the Vasicek model 0, where h*
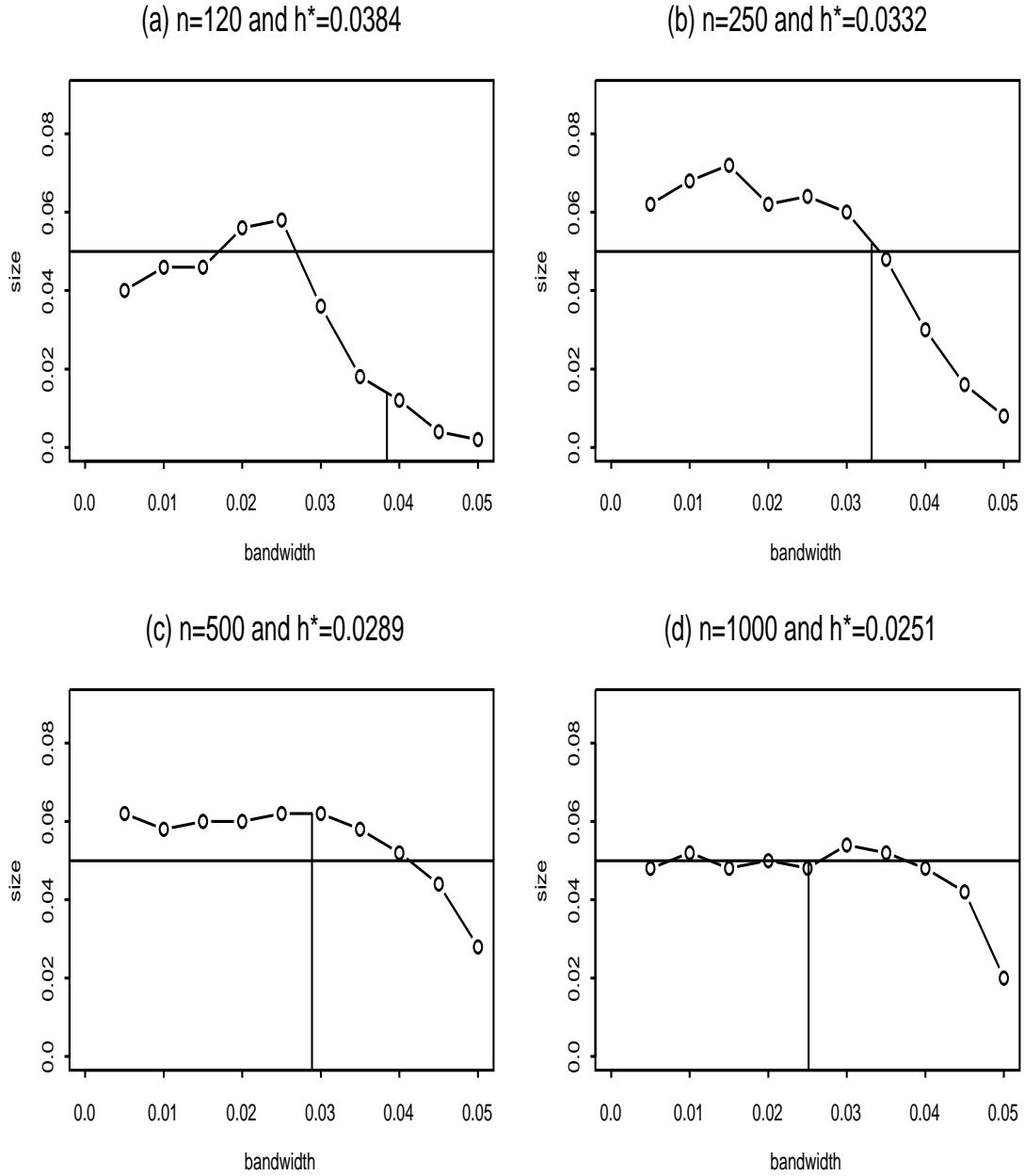are the optimal bandwidth given in Table 4.1 and are indicated by the vertical
lines.

Figure 4.5: Graphical illustrations of Table 4.6 for the Vasicek model 1, where h* are the optimal bandwidth given in Table 4.1 and are indicated by the vertical lines.

Figure 4.6: Graphical illustrations of Table 4.7 for the Vasicek model 2, where h*
are the optimal bandwidth given in Table 4.1 and are indicated by the vertical
lines.

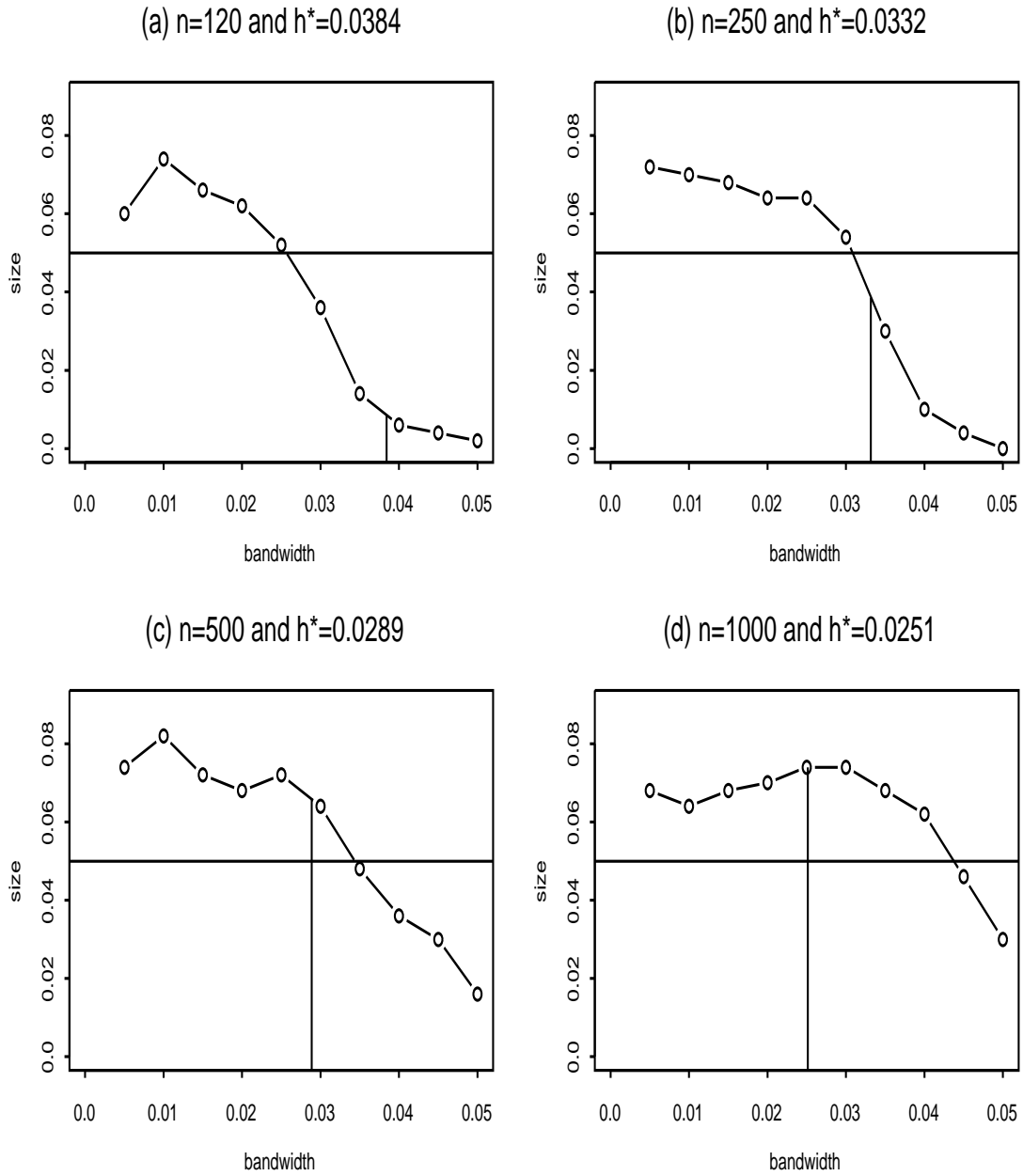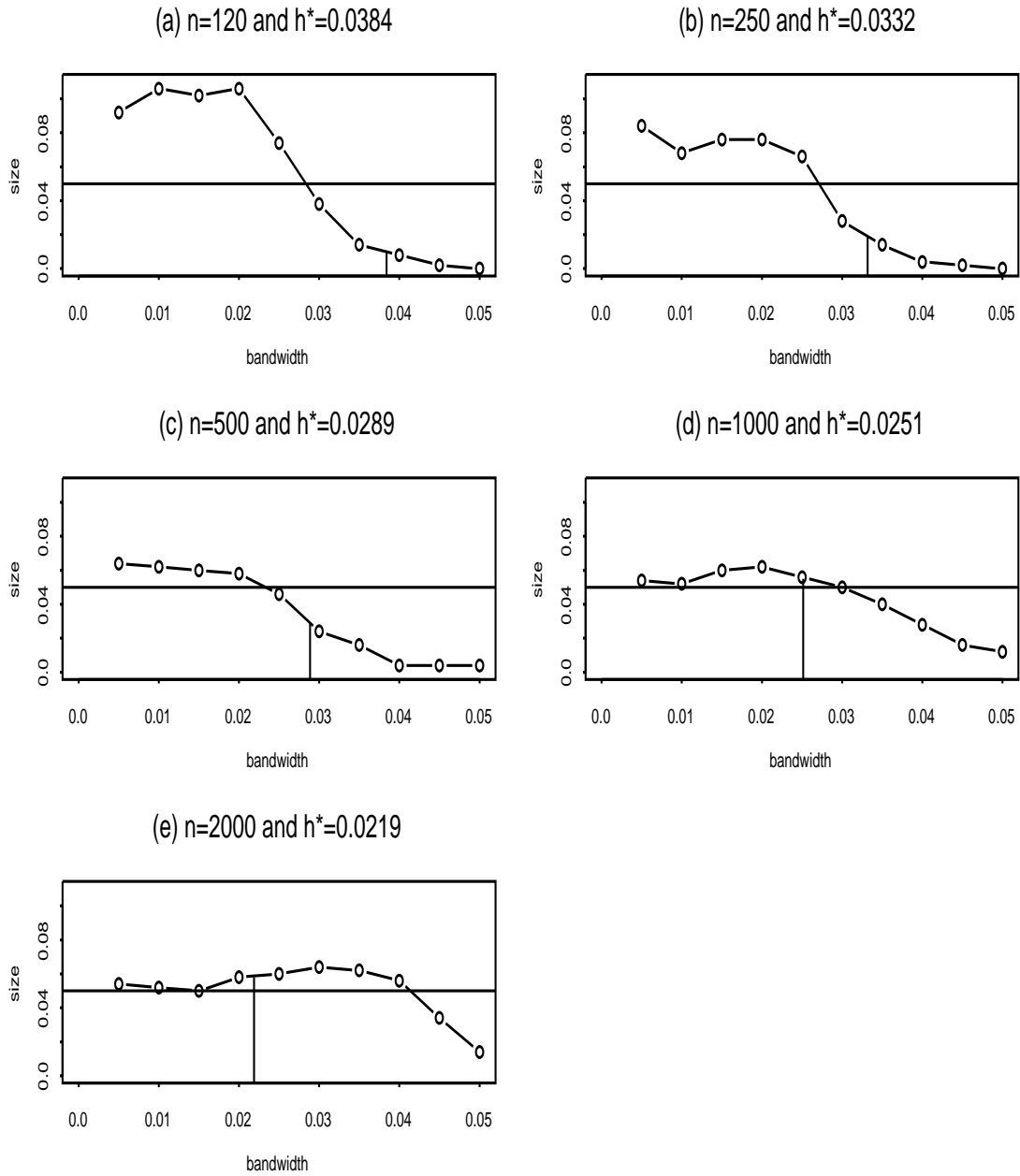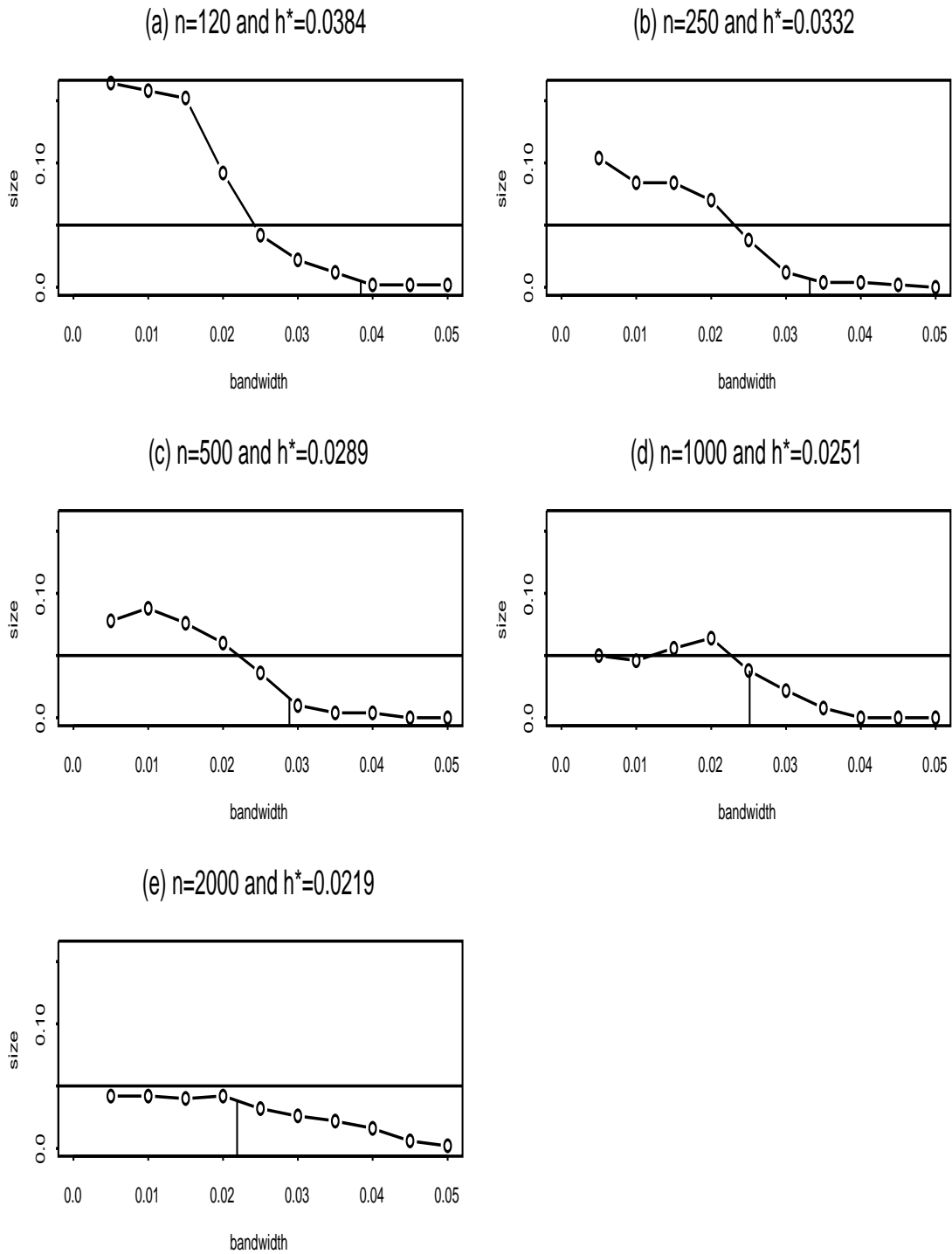To investigate the power of the test, we simulate data from the following Cox-Ingersoll-Ross (CIR, 1985) Model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \tag{4.11}$$

where κ, α, σ are all positive.

The marginal density of CIR is a gamma distribution. It is

$$f(x|\kappa, \alpha, \sigma) = \frac{w^v}{\Gamma(v)}x^{v-1}e^{-wx},$$

where $w = 2\kappa/\sigma^2$ and $v = 2\kappa\alpha/\sigma^2$.

The transition density of CIR is

$$p_\theta(X_{t+1}|X_t; \Delta) = ce^{-u-v}(v/u)^{q/2}I_q(2(uv)^{1/2}), \tag{4.12}$$

where $c = 2\kappa/(\sigma^2\{1 - e^{-\kappa\Delta}\})$, $u = cX_t e^{\kappa\Delta}$, $v = cX_{t+1}$ and I_q is the modified Bessel function of the first kind of order $q = \dfrac{2\kappa\alpha}{\sigma^2} - 1$. Therefore, we can generate the CIR process via its transition density.

In our simulation, we select the same parameters as the Vasicek model 0 in empirical size study: $(\kappa, \alpha, \sigma^2) = (0.85837, 0.089102, 0.002185)$. The procedure of simulation is similar to the empirical study just described for the Vasicek model. Table 4.9 shows the empirical rejection frequencies when the critical value is from the bootstrap for the Vasicek model. The power of the empirical likelihood test fairly equal 1 when these bandwidths are larger than 0.02.

bandwidth	Size		
CIR model	n=120	n=250	n=500
0.005	0.104	0.164	0.15
0.01	0.384	0.224	0.314
0.015	0.782	0.828	0.914
0.02	0.98	0.992	1
0.025	1	1	1
0.03	1	1	1
0.035	1	1	1
0.04	1	1	1
0.045	1	1	1
0.05	1	1	1

Table 4.8: Power of the bootstrap based LSEL Test for the CIR model for a set of bandwidth values and their sample sizes of 120, 250, 500

4.4 Comparing With Early Study

4.4.1 Pritsker's Studies

In this part, we simulate the test statistic proposed by Aït-Sahalia(1996a) again. Similar to Pritsker (1998) simulation study, we also perform 500 Monte Carlo simulations for each parameterization of the Vasicek model. For each simulation we generated 22 years of daily data for a total 5500 observations. We estimated $\hat{f}(x)$ using the standard kernel density estimation with a Gaussian kernel function. X changes from -0.07 to 0.25 and the range of X is 0.32 covered about all generated data. Table 4.9 lists the empirical rejecting frequencies of the test proposed by Aït-Sahalia (1996a). The result is similar with the Pritsker study. In Pritsker's study, he got the highest rejection frequency for model -1. However, we get the highest rejection frequency for model 0. This may due to some small difference in the simulation.

4.4.2 Simulation On Aït-Sahalia(1996a)'s Test

Pritsker (1998) used 1.645 as the asymptotic value of the test statistic at the confidence levels of 5%. We reformulate Aït-Sahalia's (1996a) test corresponding to our design. We perform 500 Monte Carlo simulations for each of the Vasicek model. In each simulation we generated observations about 10 years, 20 years and 40 years respectively for a total of sample size is 120, 250 and 500. Figure 4.7 presents the size of the Aït-Sahalia (1996a) test for the Vasicek models for a set

Model	Rej.freq (5%)	Optimal Bandwidth
-2	46.40%	0.0140979
-1	48.40%	0.0175509
0	52.50%	0.0217661
1	45.80%	0.0268048
2	32.60%	0.0325055

Table 4.9: Empirical rejection frequencies using asymptotic critical values at 5% level from Normal distribution.

of bandwidths. The performance of the empirical size is again very poor. In case of the Vasicek model -2 which has the least persistence, the empirical size is less than 0.02. It decreases with increased bandwidth. The performance is improved in model -1 and model 0.
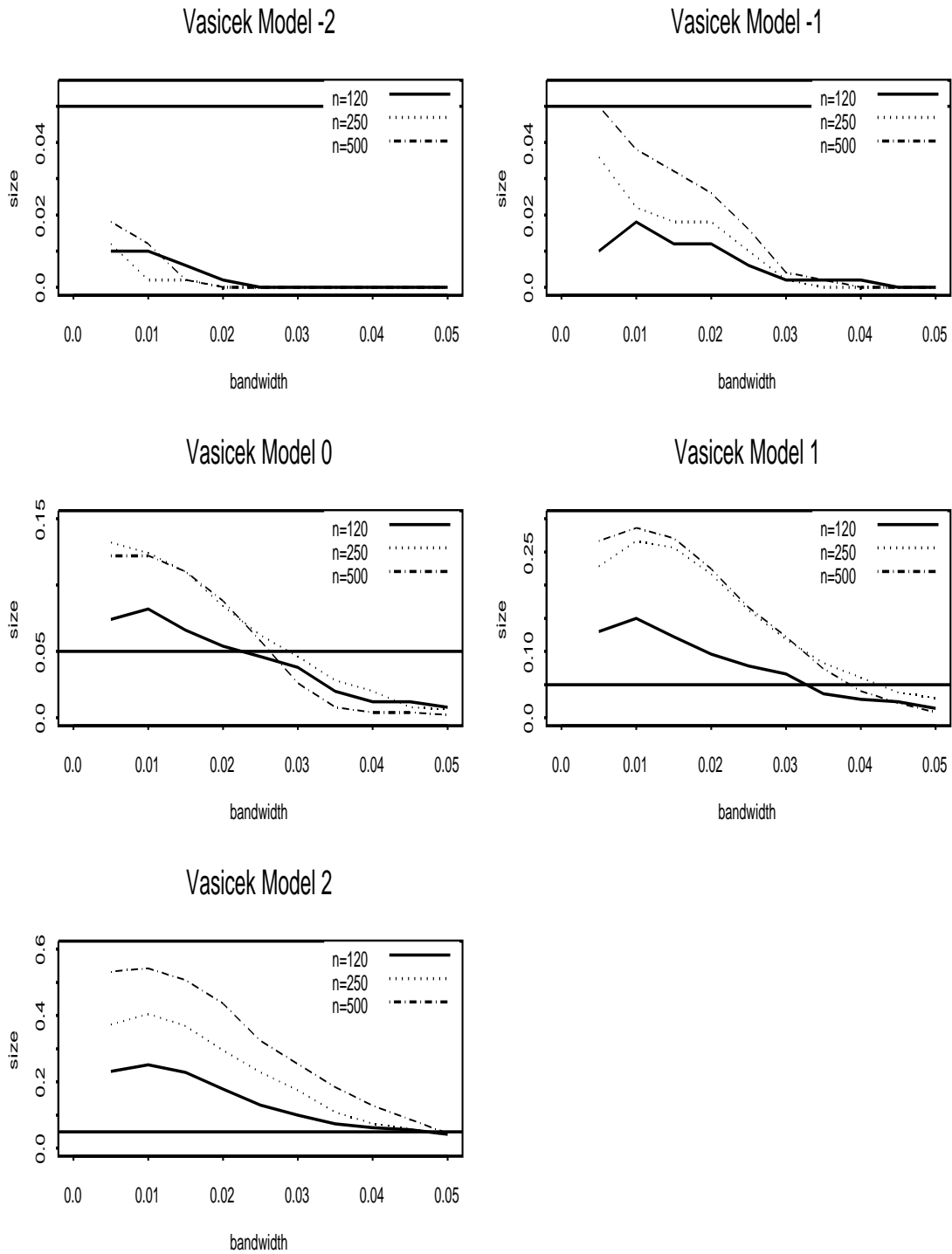
Figure 4.7: Size of Aït-Sahalia(1996a) Test for the Vasicek models for a set of bandwidth values and their sample sizes of 120, 250, 500

Chapter 5

Case Study

In early chapters, we have proposed a version of empirical likelihood goodness-of-fit test and have carried out simulation study on its empirical performance. We know that there are many existing models which are applied to capture the dynamics of the spot interest rate. It is an important work to evaluate these parametric models for the spot interest rate. Therefore in this chapter, we apply the least squares empirical likelihood specification test to evaluate five important diffusion models which are widely used to model the dynamics of the interest rate in the literature.

5.1   The Data

The interest spot rate data used here are the monthly Fed Fund Rates between January 1963 and December 1998 with a total of N=432 observed rates. The source
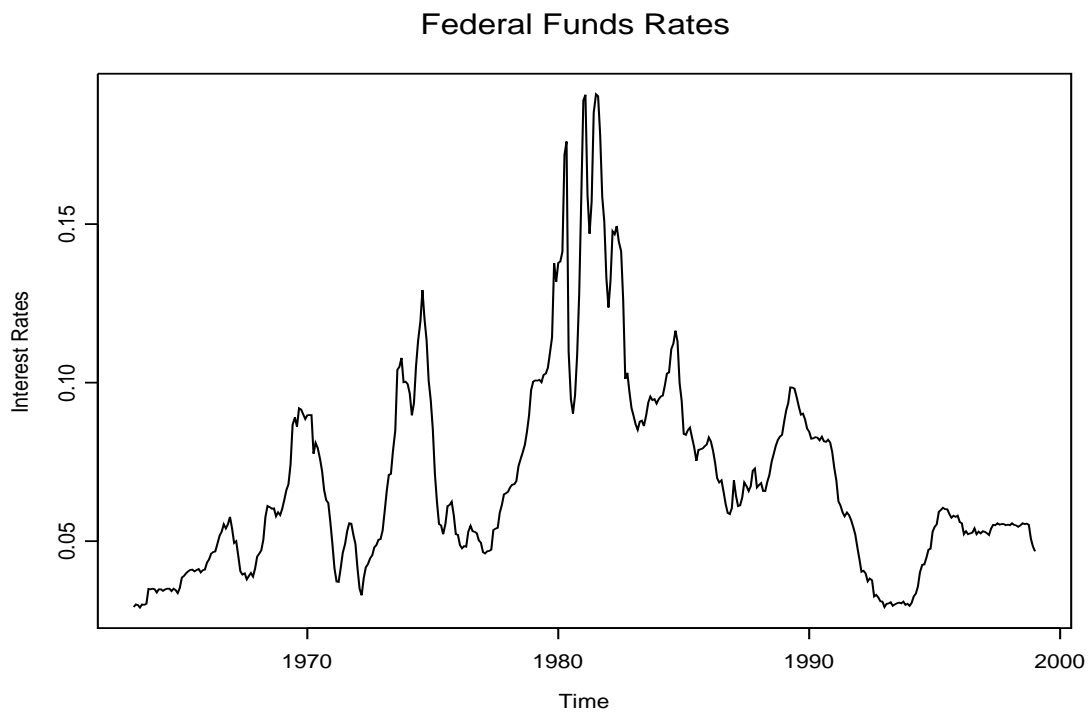
Figure 5.1: The Federal Fund Rate Series between January 1963 and December 1998.

for the data is H-15 Federal Reserve Statistical Release. The raw interest rate series

are displayed in Figure 5.1.

5.2   Early Study

Aït-Sahalia (1999) used the monthly Fed Fund Rates data to carry out the maximum likelihood estimation of parameters based on either the exact or the approximate transition density functions for the following five diffusion models.

1) Vasicek (1977) Model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB_t, \tag{5.1}$$

where the parameters κ and σ are restricted to be positive and the value of α is finite. In the Vasicek model, the volatility of the spot rate process is constant and its mean term structure is a linear function. It is generally thought that the constant diffusion structure is too simple to capture the real variability of the interest rate process.

2) Cox, Ingersoll and Ross (CIR, 1985) Model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB_t, \tag{5.2}$$

where the parameters κ, α and σ are all positive. It also contains the linear drift function but improves the constant diffusion function to the linear structure which may describe the higher variation of the interest rate.

3) Ahn and Gao's (1999) Inverse CIR Model

$$dX_t = X_t\{\kappa - (\sigma^2 - \kappa\alpha)X_t\}dt + \sigma X_t^{3/2}dB_t. \tag{5.3}$$

If X_t follows the CIR Model, $1/X_t$ satisfies the above process. Therefore, it is

called inversion of the CIR Model. In this model, it is clear that the parameter of

diffusion also affects the parameter of the drift.

4) Constant Elasticity of Volatility (CEV) Model

$$dX_t = \kappa(\alpha - X_t)dt + \sigma X_t^\rho dB_t, \tag{5.4}$$

where $\rho > 1/2$. This model is proposed by Chan,et al. (1992) and it relaxes

the diffusion function to a general power function while still keeps the linear drift

structure.

5) Aït-Sahalia (1996a) Nonlinear Drift Model (NDM)

$$dX_t = (\alpha_{-1}X_t^{-1} + \alpha_0 + \alpha_1 X_t + \alpha_2 X_t^2)dt + \sigma X_t^{3/2}dB_t. \tag{5.5}$$

It is well known for improving the general linear drift function to a quadratic form

and the diffusion function is regarded as a scale of $X_t^{3/2}$.

First, we measure the goodness-of-fit for these five models for the interest data.

In this thesis, the Biweight kernel $K(u) = \dfrac{15}{16}(1-u^2)I(u)$, where $I(\cdot)$ is the indicator

function on $[-1, 1]$, has been employed in all the numerical studies. We still apply

the reference to a normal distribution approach to select the bandwidth. This

method gives the optimal bandwidth $h = 0.0264$ with the sample size N=432.

In Figure 5.2-5.6, we plot the nonparametric kernel estimates of the marginal

density $\hat{f}(x)$, the parametric marginal density $f(x, \hat{\theta})$ and the smoothed parametric

density $\tilde{f}(x, \hat{\theta})$ with three different bandwidths, where $\hat{\theta}$ are maximum likelihood

estimates given in Table VI of Aït-Sahalia (1999). In the figures, $R1$ and $R2$,

which are indicated by the vertical lines, are 0.031 and 0.138 respectively. Each

of two tail regions $(0, R1)$ and $(R2, \infty)$ cover around 5% of the Federal Fund Rate data. The effect of smoothing on the parametric density is prominent especially for model (5.3)-(5.6). It reduced the discrepancies between the nonparametric kernel estimates and the parametric estimates for models (5.3)-(5.5). For the Vasicek model (5.1), it is clear that the nonparametric kernel estimate of the marginal density does not agree well with the smoothed parametric specifications in the range of $[R1, R2]$. With the increased bandwidth, the discrepancies of these two estimates do not change much. As will be reported shortly, this is strongly supported by the testing results. For the CIR model (5.2), the performance is better than the Vasicek model. In Figure 5.3, the nonparametric kernel estimates of the marginal density agree reasonably well with the smoothed parametric specifications in the range of $[0.10, 0.20]$. Also, the discrepancies between these two estimates in other regions are also smaller than that of model (5.3) and model (5.4). In the case of model (5.3) and (5.4), the situations are similar. The nonparametric kernel estimates of the marginal density agree reasonably well with the smoothed parametric specifications in the range of $[0.10, 0.2]$ but have large discrepancies in other range. For Aït-Sahalia (1996a) nonlinear drift model (5.5), the nonparametric kernel estimates of the marginal density agree well with the smoothed parametric specifications almost in the whole range while the nonparametric kernel estimates do not fit well with the parametric specifications. On the whole, the discrepancies between the nonparametric kernel estimates and the smoothed parametric estimates become smaller than that between the nonparametric kernel estimates and the parametric

estimates. Secondly, among these five models, the behavior of Aït-Sahalia (1996a) nonlinear drift model is the best one and the Vasicek model may be improper for mimicing the dynamics of the interest rate.
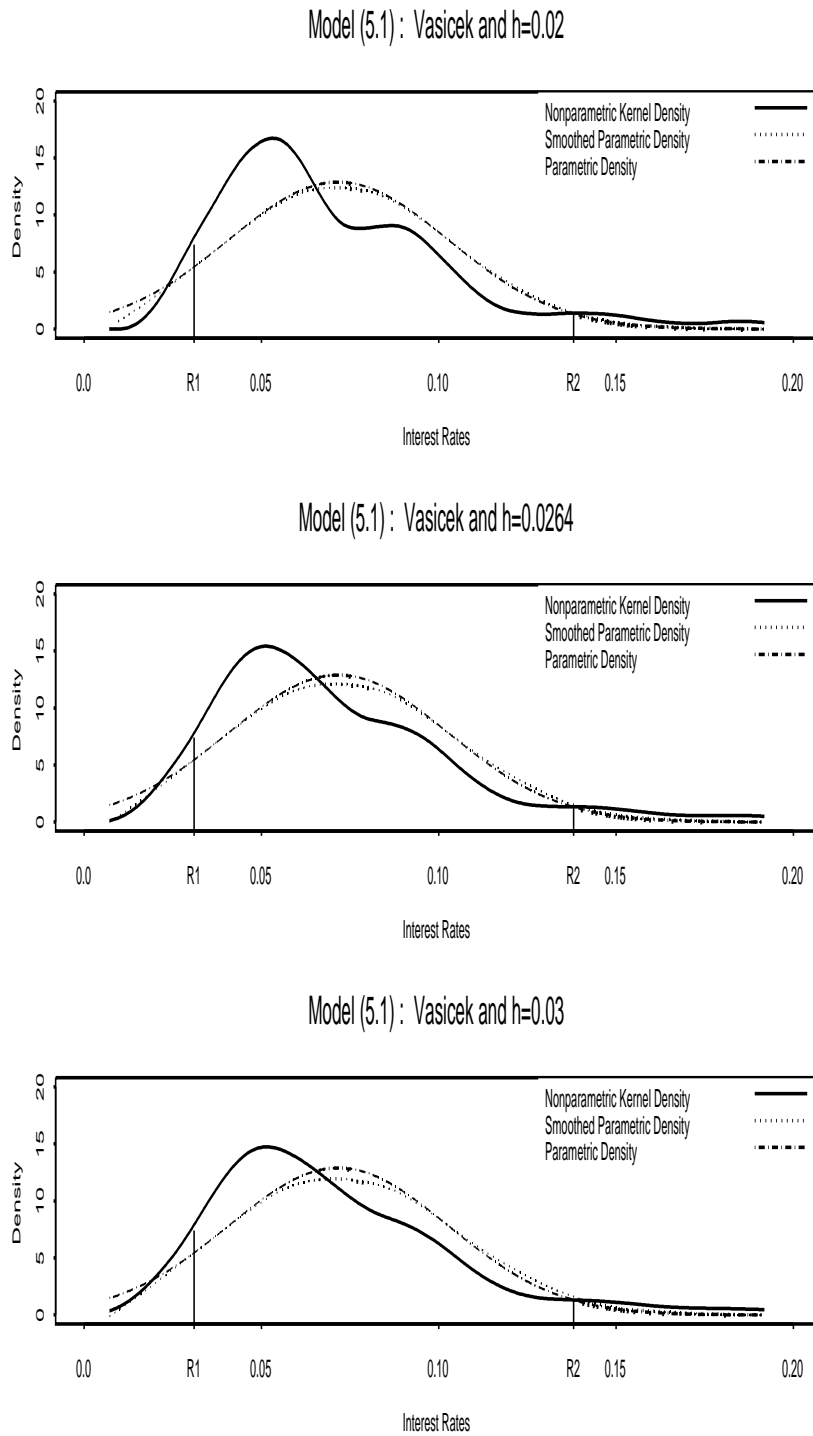
Figure 5.2: Nonparametric kernel estimates, parametric and smoothed parametric estimates of the marginal density for the Federal Fund Rate Data and R1=0.031, R2=0.138.

Figure 5.3: Nonparametric kernel estimates, parametric and smoothed parametric estimates of the marginal density for the Federal Fund Rate Data and R1=0.031, R2=0.138.

Figure 5.4: Nonparametric kernel estimates, parametric and smoothed parametric estimates of the marginal density for the Federal Fund Rate Data and R1=0.031, R2=0.138.

Model (5.4) : CEV and h=0.02

Model (5.4) : CEV and h=0.0264

Model (5.4) : CEV and h=0.03

Figure 5.5: Nonparametric kernel estimates, parametric and smoothed parametric estimates of the marginal density for the Federal Fund Rate Data and R1=0.031, R2=0.138.
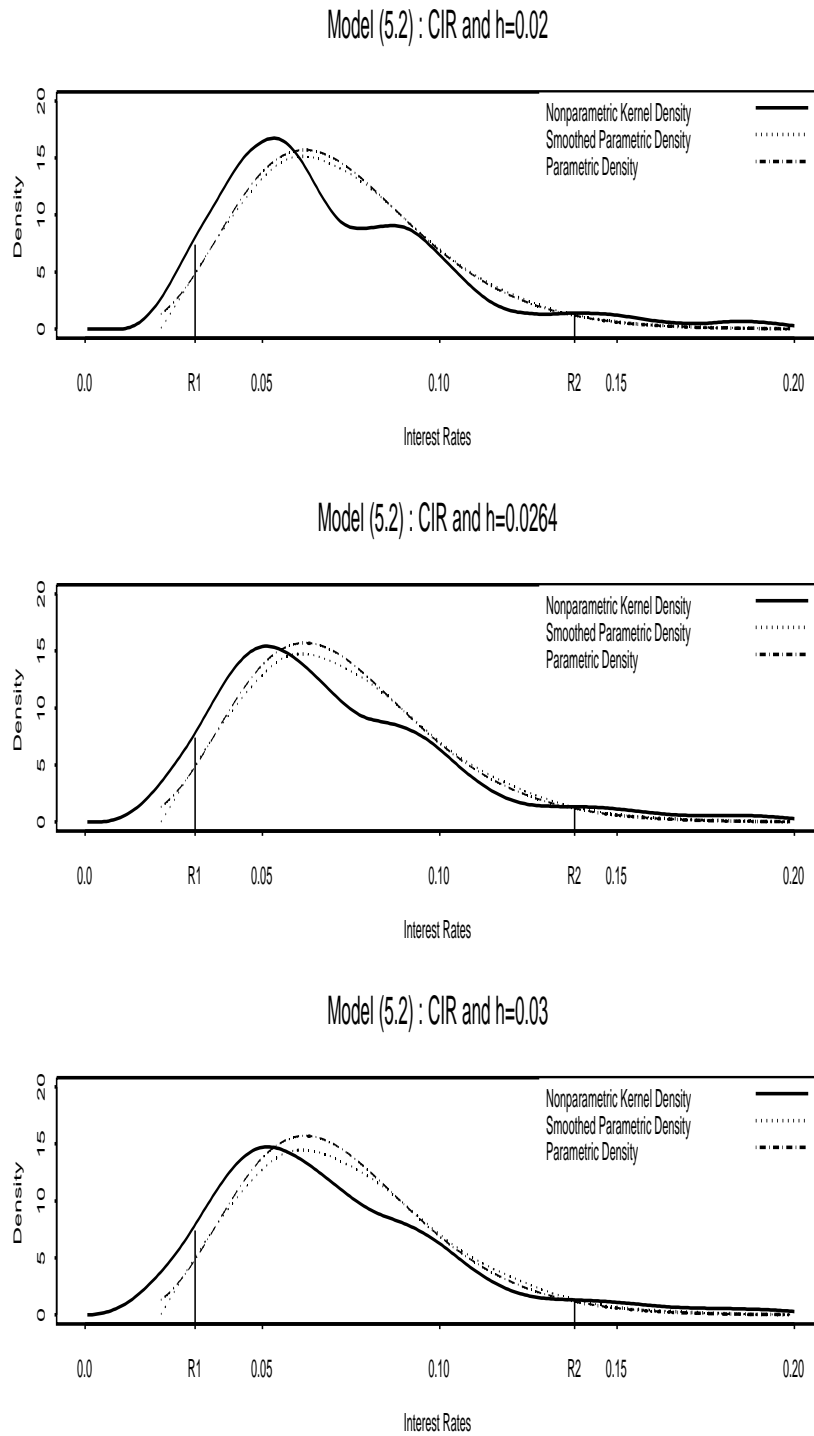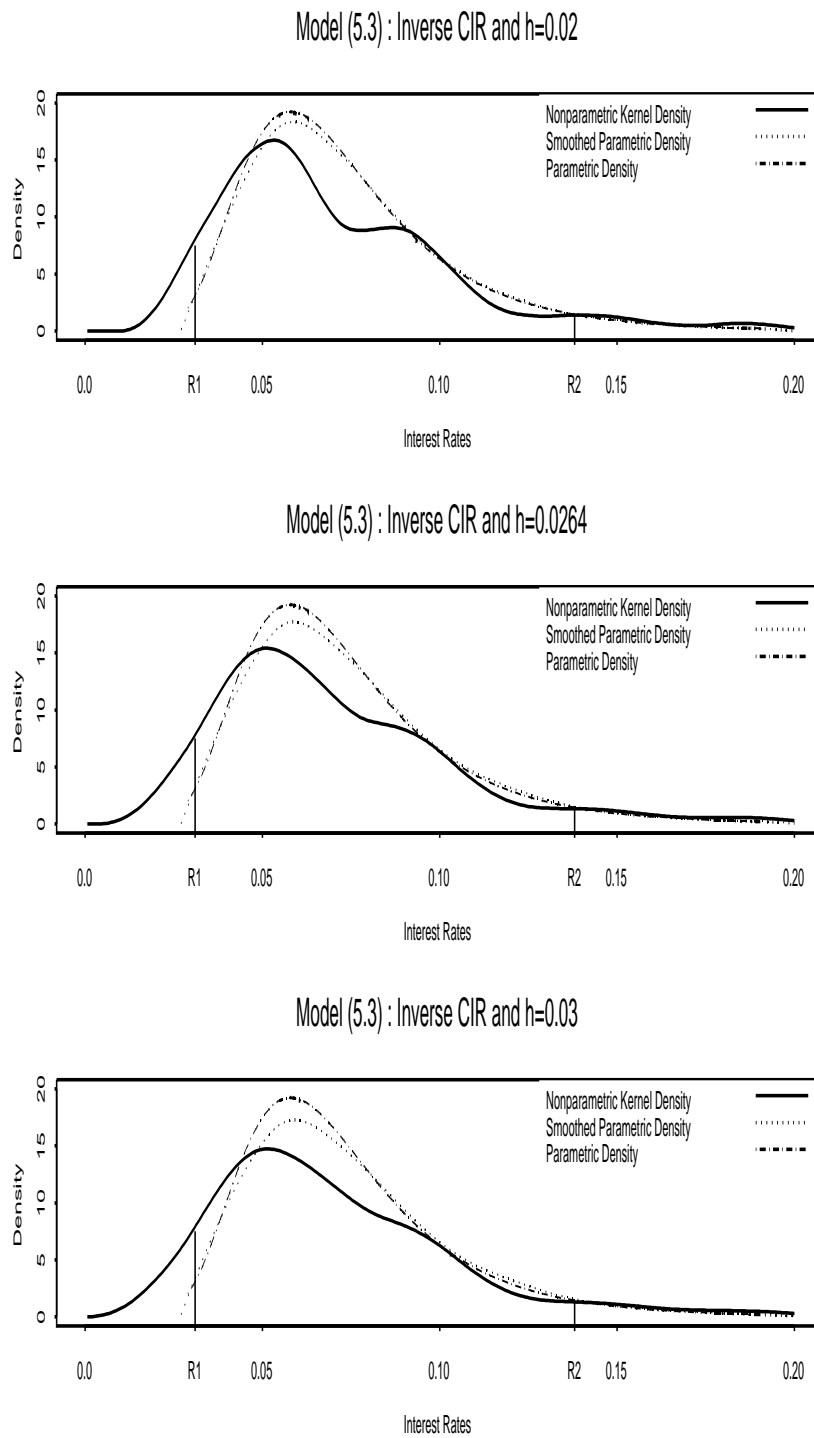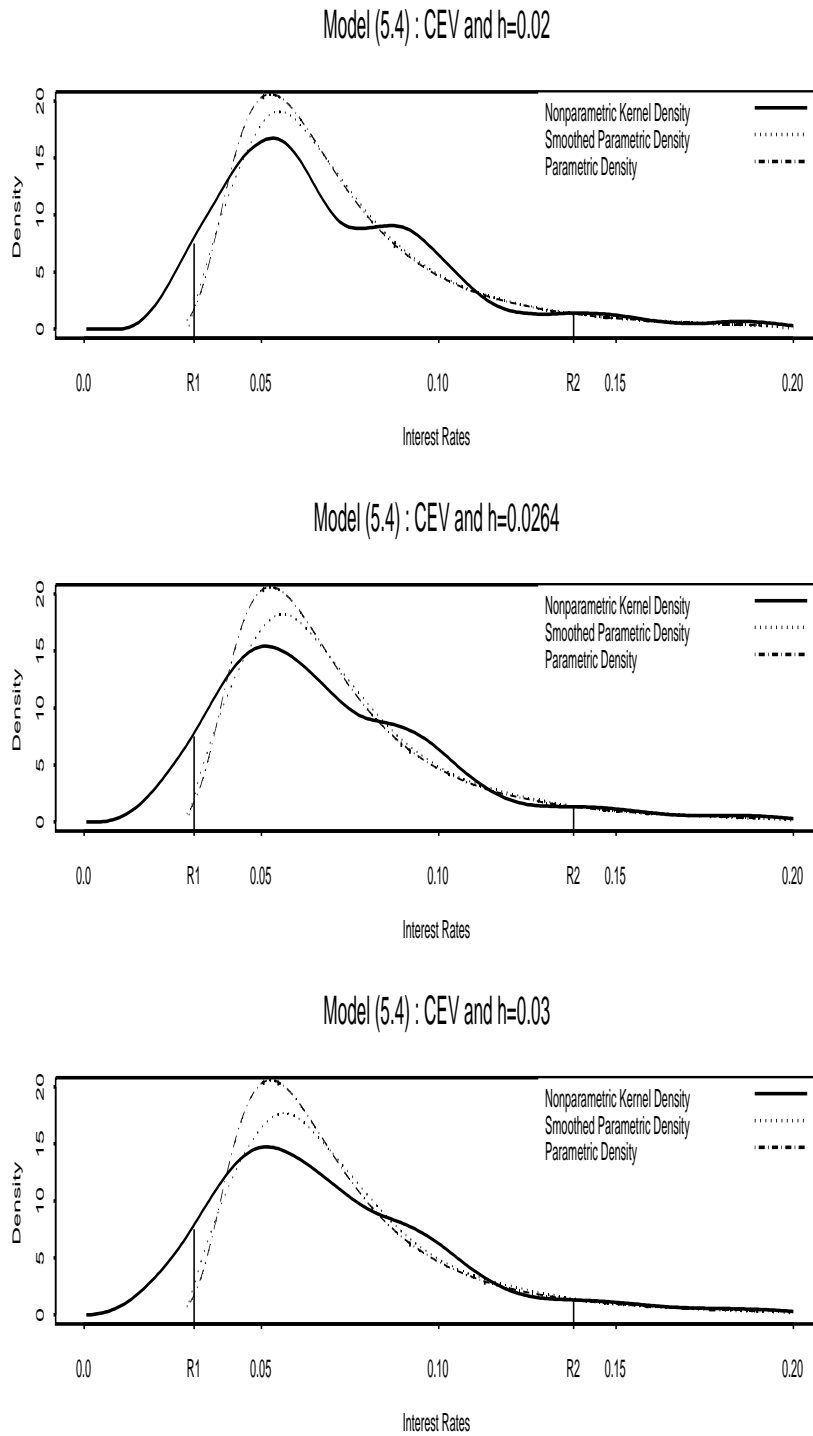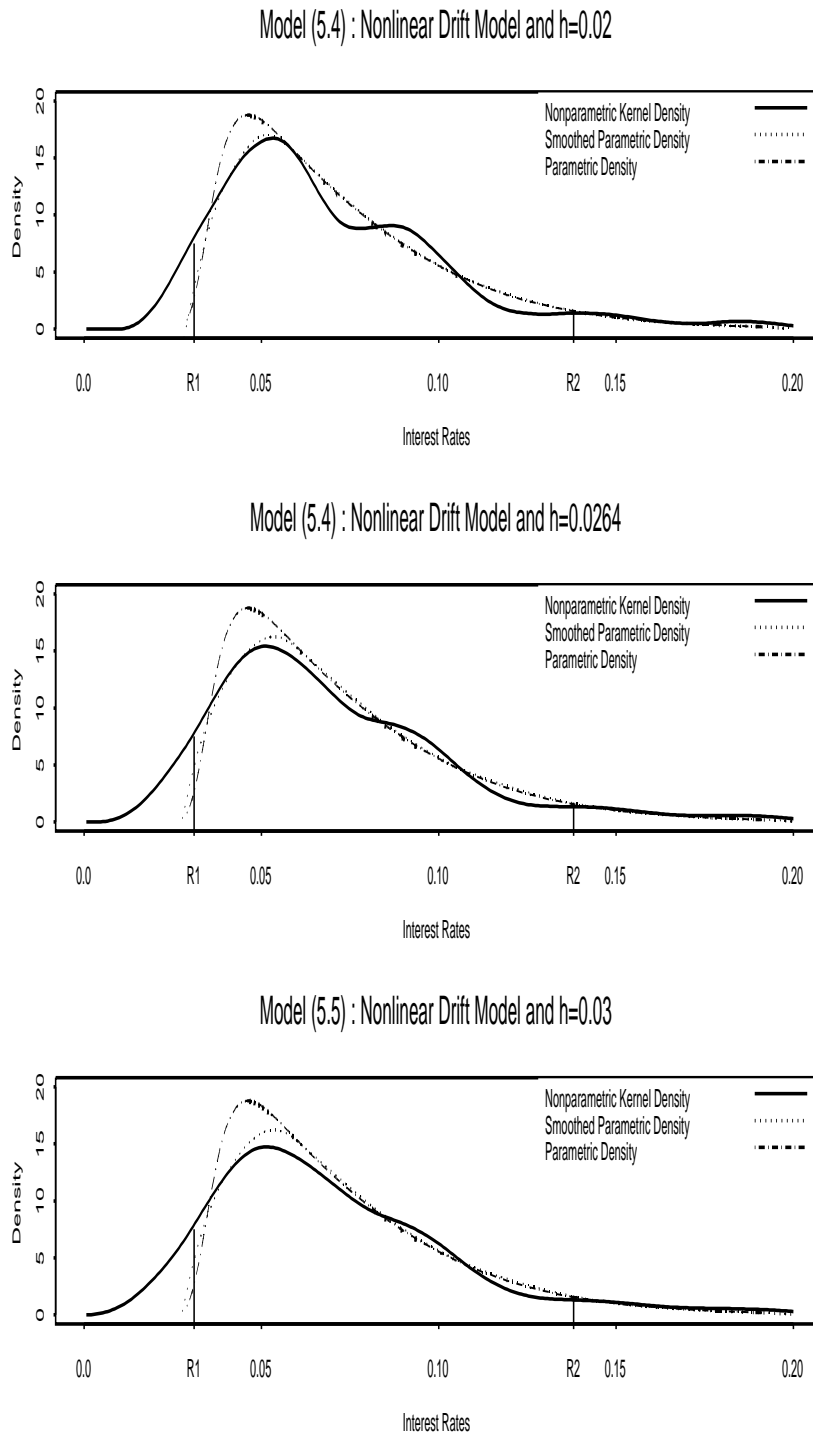
Figure 5.6: Nonparametric kernel estimates, parametric and smoothed parametric estimates of the marginal density for the Federal Fund Rate Data and R1=0.031, R2=0.138.

5.3   Test

We carry out the least squares empirical likelihood goodness-of-fit test for the marginal density for five diffusion model with 10 equally spaced bandwidths ranging from 0.005 to 0.05 and one optimal bandwidth. The optimal bandwidth is included in this range of bandwidths and this range offers a wide range of smoothness. The weight function is $\pi(x) = I(R1 < x < R2) = I(0.031 < x < 0.138)$ which implies a constant weight in the range that contains about 90% of the Federal Fund Rate data.

Table 5.1 contains the p-value of the test for the Vasicek and the CIR model. It is observed that for the Vasicek model, while the bandwidth changes from 0.005 to 0.04, the p-value is steadily around 0.1. The p-value is 0.10 when the optimal bandwidth 0.0264 is applied. In this test, we get much larger p-value than those early empirical studies which almost strongly reject the Vasicek model. Therefore, it may be the first one that shows we can not strongly reject the Vasicek model for the spot interest rate.

The p-values of the test for the CIR model are much larger than those of the Vasicek model. When the bandwidth is 0.0264, the p-value of the test already reaches 0.496. The p-value of the test still keeps increasing with increased bandwidth. From the early measurement of the goodness-of-fit, we know the reasonable agreement between the nonparametric kernel estimates and the smoothed parametric estimates, which may justify the large p-value.

Table 5.2 lists the p-value of the test for the inverse CIR model and CEV model. The p-value of the test is increasing with increased bandwidth. For the inverse CIR model, the p-values of the test are smaller than those of the CIR model but larger than those of Vasicek model. When the optimal bandwidth 0.264 is applied, the p-value of the test reaches 0.424, litter smaller than that of the CIR model which is 0.46. For the CEV model, the p-values of the test are larger than those of the CIR model. The p-value of the test is 0.894 when the optimal bandwidth 0.264 is used.

Table 5.3 lists the p-values of the test for the nonlinear drift model. The p-values of the test are the largest than those of other four diffusion models. The p-value of the test reaches 0.942 when the bandwidth is 0.264. From the measurement of the goodness-of-fit in early section, we know the behavior of the nonlinear drift model is the best one, which may justify the largest p-value.

Furthermore, testing of the marginal density is not conclusive for the specification of diffusion models as pointed out in Aï-Sahalia (1996a). The transition density describes the short-run time-series behavior to the diffusion process so it captures the full dynamics of the diffusion process. Whereas the marginal density of the process describes the long-run behavior of the diffusion processes. Therefore, further specification study on the transition density is required. In the test, these results show that we may not strongly reject the Vasicek model and the nonlinear drift model may be the most satisfying model for the interest rate.

We also compute the standard test statistic $\hat{L} = \dfrac{\hat{N}_{LS}(h) - 1}{\sigma_h}$ where $\sigma_h^2 = 2hC(K, \pi)$ and $C(K, \pi) = R^{-2}(0)K^{(4)}(0) \int \pi^2(x)dx$ which is asymptotically standard Normal distribution under some assumptions. We observe that the p-values for the Vasicek model, the CIR Model, the inverse CIR model and the CEV model are all almost 0. For the nonlinear drift model, the p-value is 0.0003 when the bandwidth is 0.264. The p-values for the nonlinear drift model are also very small. It means we would reject all these models if we applied the asymptotic normal distribution. This was unfortunately a test similar to that proposed in Aï-Sahalia (1996a) and studied in Pritsker (1998). These very contrasting p-values indicate that we have to excercise cares when we carry out the specification test for the diffusion models. They also highlight the danger of using a test based on the asymptotically normality.

	Vasicek Model (5.1)			CIR Model (5.2)		
Bandwidth	Test Statistic	P-V[1]	P-V[2]	Test Statistic	P-V[1]	P-V[2]
0.005	6.56	0.128	$0_{(10.10)}$	5.25	0.312	$0_{(7.72)}$
0.01	13.88	0.114	$0_{(16.54)}$	9.08	0.336	$0_{(10.37)}$
0.015	19.02	0.112	$0_{(18.90)}$	11.00	0.346	$0_{(10.49)}$
0.02	22.97	0.11	$0_{(19.95)}$	11.90	0.374	$0_{(9.90)}$
0.025	26.01	0.102	$0_{(20.32)}$	12.23	0.46	$0_{(9.12)}$
0.0264	26.70	0.10	$0_{(20.32)}$	12.24	0.496	$0_{(8.88)}$
0.03	28.17	0.098	$0_{(20.15)}$	12.11	0.60	$0_{(8.24)}$
0.035	29.65	0.092	$0_{(19.67)}$	11.73	0.698	$0_{(7.37)}$
0.04	30.65	0.124	$0_{(19.05)}$	11.39	0.792	$0_{(6.67)}$
0.045	30.66	0.186	$0_{(18.32)}$	11.48	0.826	$0_{(6.35)}$
0.05	31.26	0.334	$0_{(17.49)}$	12.68	0.866	$0_{(6.71)}$

Table 5.1: Test statistics and P-values (P-V[1]) of Vasicek Model and CIR Model of the empirical tests for the marginal density for the Fed fund rate data, and P-values (P-V[2]) when the asymptotic normal distribution is applied and the corresponding standard test statistics show in brackets.

Bandwidth	INVCIR Model (5.3)			CEV Model (5.4)		
	Test Statistic	P-V^1	P-V^2	Test Statistic	P-V^1	P-V^2
0.005	6.86	0.294	$0_{(10.64)}$	4.38	0.453	$0_{(6.15)}$
0.01	11.97	0.312	$0_{(14.08)}$	6.74	0.515	$0_{(7.37)}$
0.015	15.05	0.330	$0_{(14.74)}$	7.54	0.671	$0_{(6.86)}$
0.02	17.35	0.334	$0_{(14.85)}$	8.08	0.79	$0_{(6.43)}$
0.025	19.52	0.41	$0_{(15.04)}$	8.77	0.88	$0_{(6.31)}$
0.0264	20.11	0.424	$0_{(15.11)}$	8.97	0.894	$0_{(6.3)}$
0.03	21.53	0.53	$0_{(15.22)}$	9.43	0.916	$0_{(6.25)}$
0.035	23.23	0.628	$0_{(15.26)}$	9.98	0.95	$0_{(6.16)}$
0.04	24.85	0.726	$0_{(15.32)}$	10.78	0.962	$0_{(6.28)}$
0.045	27.19	0.816	$0_{(15.86)}$	12.84	0.966	$0_{(7.17)}$
0.05	31.54	0.852	$0_{(17.54)}$	17.72	0.952	$0_{(9.61)}$

Table 5.2: Test statistics and P-values (P-V^1) of Inverse CIR Model and CEV Model of the empirical tests for the marginal density for the Fed fund rate data, and P-values (P-V^2) when the asymptotic normal distribution is applied and the corresponding standard test statistics show in brackets.

	NDM Model (5.5)		
Bandwidth	Test Statistic	P-V^1	P-V^2
0.005	3.60	0.552	$0_{(4.72)}$
0.01	6.13	0.552	$0_{(6.59)}$
0.015	6.42	0.718	$0_{(5.68)}$
0.02	6.04	0.86	$0_{(4.58)}$
0.025	5.57	0.93	$0.0001_{(3.71)}$
0.0264	5.42	0.942	$0.0003_{(3.5)}$
0.03	5.02	0.966	$0.0014_{(3.00)}$
0.035	4.58	0.986	$0.0065_{(2.46)}$
0.04	4.67	0.992	$0.0086_{(2.35)}$
0.045	5.96	0.992	$0.0013_{(3.00)}$
0.05	9.60	0.984	$0_{(4.94)}$

Table 5.3: Test statistics and P-values (P-V^1) of Nonlinear Drift Model of the empirical tests for the marginal density for the Fed fund rate data, and P-values (P-V^2) when the asymptotic normal distribution is applied and the corresponding standard test statistics show in brackets.

Bibliography

[1] Ahn, D.H., and B. Gao, (1999), A Parametric Nonlinear Model of Term Structure Dynamics, *Review of Financial Studies*, 12, 721-762.

[2] Aït-Sahalia, Y., (1996a), Testing continuous-time models of the spot interest rate, *Review of Financial Studies*, 9, 385-426.

[3] Aït-Sahalia, Y., (1996b), Nonparametric Pricing of Interest Rate Derivative Securities, *Econometrica*, 64, 527-560.

[4] Aït-Sahalia, Y., (1999), Transition Densities for Interest Rate and Other Nonlinear Diffusions, *Journal of Finance*, 54, 1361-1395.

[5] Aït-Sahalia, Y., P. Bickel, and T. Stoker, (2001), Goodness-of-fit Tests for Regression using Kernel Methods, *Journal of Econometrics*, 105, 363-412.

[6] Black, F., M. Scholes, (1973), The Pricing of Options and Corporate Liabilities, *Journal of Political Economy*, 3, 133-155.

[7] Brennan, M. J., E. S. Schwartz, (1979), A Continuous-Time Approach to the Pricing of Bonds, *Journal of Banking and Finance*, 3, 133-155.

[8] Brown, B. M., Chen, S.X., (1998), Combined and Least Squares Empirical Likelihood, *Ann. Inst. Statist. Math*, 50, 697-714.

[9] Brown, S. J., P. H. Dybvig, (1986), The Empirical Implications of the Cox,, Ingersoll, Ross Theory of the Term Structure of Interest Rates, *Journal of Finance*, 41, 617-630.

[10] Chapman, D. A. and Pearson, N. D., (2000), Is the Short Rate Drift Actually Nonlinear, *Journal of Finance*, 55, 355-388.

[11] Cox, J. C., Ingersoll, J. E. and Ross, S. A., (1985), A theory of term structure of interest rates, *Econometrica*, 53, 385-407.

[12] Chan, K. C., G. A. Karolyi, F. A. Longstaff, and A. B. Sanders, (1992), An Empirical Comparison of Alternative Models of the Short-Term Interest Rate, *Journal of Finance*, 47, 1209-1227.

[13] Chapman, D. A. and Pearson, N. D., (2000), Is the Short Rate Drift Actually Nonlinear, *Journal of Finance*, 55, 355-388.

[14] Chen, R. and L. Scott, (1993), Maximum Likelihood Estimation for a Multi-Factor Equilibrium Model of the Term Structure of Interest Rates, *Journal of Fixed Income*, 4, 14-31.

[15] Chen, S. X., (1996), Empirical likelihood for nonparametric density estimation, *Biometrika*, 83, 329-341.

[16] Chen, S. X., Härdle,W. and Li, M., (2003), An Empirical likelihood Goodness-of-Fit Test for Time Series, *Biometrika*, 83, 329-341.

[17] Constantinides, G. M., (1992), A Theory of the Nominal Term Structure of Interest Rates, *Review of Financial Studies*, 5, 531-552.

[18] Corrade, V., and Swanson, N., (2001), Bootstrap Specification Tests with Dependent Observations and Parameter Estimation Error, working paper, Department of Economics, Univeristy of Exeter, UK.

[19] Courtadon, G., (1982), The Pricing of Options on Default-free Bonds, *Journal of Financial and Quantitative Analysis*, 17, 75-100.

[20] Dacunha-Castelle, D., and D. Florens-Zmirou, (1986), Estimation of the Coefficient of a Diffusion from Discrete Observations, *Stochastics*, 19, 263-284.

[21] Dohnal, G. (1987), On estimating the Diffusion Coefficient, *Journal of Applied Probability*, 24, 105-114.

[22] Duffie, D., and R. Kan, (1993), A Yield Factor Model of Interest Rates, *Stanford University Mimeo*.

[23] Florens-Zmirou, D., (1993), On Estimating the Diffusion Coefficient from Discrete Observations, *Journal of Applied Probability*, 30, 790-804.

[24] Gao, J., and M. King, (2001), Estimation and Model Specification Testing in Nonparametirc and Semiparametric Time Series Econometric Models, work-

ing paper, School of Mathematics and Statistics, The University of Western Australia, Perth, Australia.

[25] Genon-Catalot, V. J. Jacod, (1993), On the Estimation of the Diffusion Co-efficient for Multidimensional Diffusion Processes, Working paper, Univerity of Marne-La-Vallée, France.

[26] Gibbons, M. R., K. Ramaswamy, (1993), A Test of the Cox, Ingersoll and Ross Model of the Term Structure, *Review of Financial Studies*, 6, 619-658.

[27] Gibson, Rajna and Eduardo S. Schwartz, (1990), Stochastic convenience yield and the pricing of oil contingent claims, *Journal of Finance*, 45, 959-976.

[28] Hansen, L. P., (1982), Large Sample Properties of Generalized Mehtod of Moments Estimators, *Econometrica*, 50, 1029-1054.

[29] Hansen, L. P., and J. A. Scheinkman, (1995), Back to the Future: Generating Moment Implications for Continuous Time Markov Processes, *Econometrica*, 63, 767-804.

[30] Hong, Y. M. and Li, H. T., (2001), Nonparametric Specification Testing for Continuous-Time Models with Application to Spot Interest Rates, Working Paper, Cornell University.

[31] Horowits, J. L. and Spokoiny, V. G., (2001), An adaptive, rate-optimal test of a parametric mean-regression model against a nonparametric alternative, *Econometrica*, 69, 599-631.

[32] Jamshidian and Farshid, (1989), An exact bond option formula, *Journal of Finance*,44, 205-209.

[33] Karlin S., Taylor H. M., (1981), *A second course in stochastic processes*, Academic Press Inc. California.

[34] Kasonga, R. A., (1988), The Consistency of A Non-linear Least Squares Estimator from Diffusion Processes *Stochastic Processes and Their applications*, 30, 263-275.

[35] Lo, Andrew W., (1988), Maximum Likelihood Estimation of Generalized Ito Processes with Discretely Sampled Data, *Econometric Theory*, 4, 231-247.

[36] Longstaff, Francis A., (1990), The valustion of options on yields, *Journal of Financial Economics*, 26, 97-122.

[37] Marsh, T. A., and E. R. Rosenfeld, (1983), Stochastic Processes for Interest Rates and Equilibirium Bond Prices, *Journal of Finance*, 38, 635-646.

[38] Merton, R. C., (1969), A Golden Golden-Rule for the Welfare-Maximization in an Economy with a Varying Population Growth Rate, *Western Economic Journal*, 4, December 1969b.

[39] Merton, R. C., (1973), Theory of Rational Option Pricing, *Bell Journal of Economics*, 4,141-183.

[40] Øksendal, Bent, (1985), Stochastic Differential Equations: an Introduction with Applications, third edition, Springer-Verlag, New York.

[41] Owen, A., (1988), Empirical likelihood ratio confidence intervals for a single functional, *Biometrika*, 75,237-249.

[42] Owen, A., (1990), Empirical likelihood ratio confidence regions, *Ann. Statist*, 18,90-120.

[43] Owen, A., (2001), Empirical likelihood, *Chapman and Hall/CRC*, Boca Raton.

[44] Pritsker M., (1998), Nonparametric Density Estimation and Tests of Continuous Time Interest Rate Models, *The Review of Financial Studies*, 11,449-487.

[45] Pearson, Neil D. and Tong-Sheng Sun, (1994), Exploiting the Conditional Density in Estimating the Term Structure: An application to the Cox, Ingersoll, and Ross Model, *Journal of Finance*, 49,1279-1304.

[46] Ramaswamy and Sundaresan, (1986), The valuation of floating-rate instruments, *Journal of Financial Economics* , 17, 251-272.

[47] Silverman, B. W., (1986), *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, New York.

[48] Stanton, R., (1997), A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk, *Journal of Finance*, 52, 1973-202.

[49] Sundaresan, S. M., (2000), Continuous-Time Methods in Finance: A Review and an Assessment, *The Journal of Finance* , 4, 1569-1622.

[50] Tan, J. and Yao Q., (1998), Efficient esimtaion of conditional variance functions in stochastic regression, *Biometrika*, 85, 645-660.

[51] Thomas, D. R. and Grunkemeier, G. L., (1975), Confidence interval estimation of survival probabilities for censored data, *J. Am. Statist. Assoc.*, 70, 865-871.

[52] Vasicek, Oldrich (1977), An equilibrium characterization of the term structure,*Journal of Financial Economics*, 5,177-188.