9-2006

Parameters Estimation and Bias Correction for Diffusion Processes

Song Xi Chen
Iowa State University, songchen@iastate.edu

Cheng Yong Tang
Peking University

Parameters Estimation and Bias Correction for Diffusion Processes

Abstract

This paper considers parameter estimation for continuous-time diffusion processes which are commonly used to model dynamics of financial securities including interest rates. To understand why the drift parameters are more difficult to estimate than the diffusion parameter as observed in many empirical studies, we develop expansions for the bias and variance of parameter estimators for two mostly employed interest rate processes. A parametric bootstrap procedure is proposed to correct bias in parameter estimation of general diffusion processes. Simulation studies confirm the theoretical findings and show that the bootstrap proposal can effectively reduce both the bias and the mean square error of parameter estimates for both univariate and multivariate processes. The advantages of using more accurate parameter estimators when calculating various option prices in finance are demonstrated by an empirical study on a Fed fund rate data.

Keywords

bias correction, bootstrap, continuous-time models, diffusion processes, jackknife, parameter estimation, Department of Business Statistics and Econometrics, Guanghua School of Management, Peking University

Disciplines

Statistics and Probability

Comments

This preprint was published as Cheng Yong Tang and Song Xi Chen," Parameter estimation and bias correction for diffusion processes", *Journal of Econometrics* (2009): 65-81, doi: 10.1016/j.jeconom.2008.11.001.

Parameters Estimation and Bias Correction [1]
for Diffusion Processes

Song Xi Chen and Cheng Yong Tang

Department of Statistics, Iowa State University, Ames, IA 50011-1210

Emails:songchen@iastate.edu; yongtang@iastate.edu

ABSTRACT

This paper considers parameter estimation for continuous-time diffusion processes which are commonly used to model dynamics of financial securities including interest rates. To understand why the drift parameters are more difficult to estimate than the diffusion parameter as observed in many empirical studies, we develop expansions for the bias and variance of parameter estimators for two mostly employed interest rate processes. A parametric bootstrap procedure is proposed to correct bias in parameter estimation of general diffusion processes. Simulation studies confirm the theoretical findings and show that the bootstrap proposal can effectively reduce both the bias and the mean square error of parameter estimates for both univariate and multivariate processes. The advantages of using more accurate parameter estimators when calculating various option prices in finance are demonstrated by an empirical study on a Fed fund rate data.

Keywords: Bias correction, Bootstrap, Continuous-time models, Diffusion Processes, Jackknife, Parameter estimation.

[1]We thank Professors Yacine Aït-Sahalia and Jianqing Fan for valuable comments and suggestions, and Professor Aït-Sahalia for sending a computing code that allows us to implement his approximate likelihood estimation approach. The project is partially supported by National Science Foundation grants SES 0518904 and DMS 0650938.

1. INTRODUCTION

Diffusion processes have long been used to model stochastic dynamics arising in physics, biology and other natural sciences. One latest surge of interest on these processes comes from molecular biology in modeling the dynamics of proteins as part of an effort to understand how energy transfer and conversion happen within biological cells. Perhaps the most eminent use of these continuous time stochastic processes in the last three decades has been in finance following the works of Merton (1971) and Black and Scholes (1973) which established the foundation of option pricing theory in finance. Since then, there has been phenomenal growth of financial products and instruments powered by these processes as documented in Sundaresan (2000).

Along with this wide range usage of diffusion processes in various fields, there are growing needs and interests in parameter estimation and model testing for diffusion processes, which have been largely encouraged by easily available data in financial applications. See Lo (1988), Bibby and Sørensen (1995), Aït-Sahalia (2002), Aït-Sahalia and Mykland (2003) and Fan (2005) for discussions and overviews.

An important application of diffusion processes is in modeling short-term interest rates, which are fundamental quantities in finance as they define excess asset returns and risk premiums of other assets and their derivative prices. A commonly used family of diffusion processes for the interest rates dynamics is

$$dX_t = \kappa(\alpha - X_t)dt + \sigma(X_t)^\rho dB_t, \tag{1.1}$$

where α, κ, σ and ρ are positive parameters. The linear drift prescribes a mean-reversion of X_t toward the long term mean α at a speed κ. The diffusion function $\sigma(X_t)^\rho$ can accommodate a range of pattens in volatility as X_t gets larger. Important members of this family are the Vasicek model (Vasicek, 1977) with $\rho = 0$ and the CIR model (Cox, Ingersoll, and Ross, 1985) with $\rho = 1/2$. Both Vasicek and CIR models are commonly used in finance due to (i) both have simple and attractive financial interpretations; and (ii) both admit close-form solutions. The latter facilitates explicit calculations of various option prices.

Despite the critical roles played by these interest-rate processes it is well known empirically that estimation of the drift parameters κ and α can incur large bias and/or variability, see for instance Ball and Torous (1996) and Yu and Phillips (2001). This is the case for virtually all the commonly used estimation approaches including the maximum likelihood estimation. The problem exasperates when the process is lack of dynamics which happens when κ is small. Interest rates typically exhibits less amount of changes than stocks, and is typically lack of dynamics. Indeed, as reported in Phillips and Yu (2005) and our simulation study, the maximum likelihood estimator for κ can have more than 200% relative bias even the processes are observed monthly for more than 10 years. This is rather serious as poor qualitative estimates can produce severely biased option prices and serious financial consequences.

In this paper, we first investigate the above empirical phenomena by developing expansions to the bias and variance of estimators for the Vasicek and CIR processes. The bias and variance expansions reveal that, for each process, (i) the bias of the κ estimator is of a larger order of magnitude than the bias of estimators for the other drift parameter α and the diffusion parameter σ^2; (ii) the variances of the estimators for the two drift parameters κ and α are of larger order than that of the estimator of σ^2. These explain why estimation of κ incurs more bias than the other parameters and why the drift parameter (κ and α) estimates are more variable than that of the diffusion parameter σ^2.

We then propose a parametric bootstrap procedure for bias correction in parameter estimation of general diffusion processes. Both theoretical and empirical analysis show that the proposed bias correction effectively reduces the bias without inflating the variance. We demonstrate in numerical simulations that the proposed bootstrap procedure can be combined with a range of parameter estimators including the approximate likelihood estimation of Aït-Sahalia (2002).

The paper is structured as follows. Section 2 outlines parameter estimators used in our analysis. The expansions on the bias and variance of the estimators for Vasicek and CIR processes are presented in Section 3. Section 4 outlines the bootstrap bias correction with justifications. Simulation results are reported in Section 5. Section 6 analyzes a data set of Fed fund rates and we use it to demonstrate (i) the effect of parameter estimation on option prices and (ii) how to carry out bias correction for option prices. All technical details are deferred to the Appendix.

2. PARAMETER ESTIMATION FOR DIFFUSION PROCESSES

2.1 A General Overview

A d-dimensional parametric diffusion process $\{X_t \in R^d; t \geq 0\}$ is defined by the following stochastic differential equation

$$dX_t = \mu(X_t;\theta)dt + \sigma(X_t;\theta)dB_t, \tag{2.1}$$

where θ is a q-dimensional parameter, $\mu(\cdot;\theta) : \mathcal{R}^d \to \mathcal{R}^d$ and $\sigma(\cdot;\theta) = (\sigma_{ij})_{d\times p} > 0 : \mathcal{R}^d \to \mathcal{R}^{d\times p}$ are respectively drift and diffusion functions representing respectively the conditional mean and variance of the infinitesimal change of X_t at time t, and B_t is a p-dimensional Brownian motion. The existence and uniqueness of the process $\{X_t; t \geq 0\}$ satisfying (2.1) and its probability properties are given in Stroock and Varadhan (1979).

A unique feature of statistical inference for diffusion processes is that despite these processes are continuous-time stochastic models, their observations are made only at discrete time points, say at n equally spaced $\{t\delta\}_{t=0}^{n}$. Here δ is the sampling interval and can be made very small that corresponds to high-frequency data. Let $X_0, X_\delta, \cdots X_{n\delta}$ be discrete observations from process (2.1) at equally spaced time points $\{t\delta\}_{t=0}^{n}$ over a time interval $[0,T]$ where $T = n\delta$. To simplify notation, we write these observations as $\{X_t\}_{t=0}^{n}$ by hiding δ whenever doing so does not lead to confusion.

As a diffusion process is Markovian, if its transitional density is known, the maximum likelihood estimation (MLE) is the natural choice for parameter estimation. However, for many diffusion processes, their transitional distributions are not explicitly known which prevents the use of the MLE. In these cases, several methods are available, which include the martingale estimating equation approach by Bibby and Sørensen (1995); the pseudo-Gaussian likelihood approach of Nowman (1997); the Generalized Method of Moments (GMM) estimator of Hansen and Scheinkman (1995); and the approximate likelihood approach of Aït-Sahalia (2002). Aït-Sahalia and Mykland(2003 and 2004) consider likelihood and the GMM based estimation when δ is random and quantify its impacts on parameter estimation.

Nonparametric estimators for the drift and diffusion functions have been also proposed, which include the kernel estimator by Aït-Sahalia (1996) and Stanton (1997), and the semi-parametric estimators of Jiang and Knight (1997). Fan and Zhang (2003) examine the estimators of Stanton (1997) and analyze the effects of high order stochastic expansions on estimation. Bandi and Phillips (2003) consider two stage kernel estimation of the drift and diffusion functions, replacing the strictly stationary assumption with weaker recurrent Markov processes. See Cai and Hong (2003) and Fan (2005) for reviews.

We assume throughout the paper that as $n \to \infty$

$$(i)\,\delta \to 0, \quad (ii)\,T \to \infty \quad \text{and} \quad (iii)\,T\delta^{1/k} \to \infty \quad \text{for some } k > 2. \tag{2.2}$$

The first two assumptions imply that the sampling interval gets finer and the total observation time goes to infinity as $n \to \infty$. The last part of the assumption is used to bound various remainder terms in moment expansions.

2.2 Estimation for Vasicek Process

The Vasicek process satisfies the univariate stochastic differential equation

$$dX_t = \kappa(\alpha - X_t)dt + \sigma dB(t). \tag{2.3}$$

It is the Ornstein-Uhlenbeck process and was proposed by Vasicek (1977) for interest rate dynamics. The conditional distribution of X_t given X_{t-1} is

$$X_t|X_{t-1} \sim N\left(X_{t-1}e^{-\kappa\delta} + \alpha(1 - e^{-\kappa\delta}), \tfrac{1}{2}\sigma^2\kappa^{-1}(1 - e^{-2\kappa\delta})\right)$$

and the stationary distribution of is $N(\alpha, \frac{1}{2}\sigma^2\kappa^{-1})$. The conditional mean and variance of X_t given X_{t-1} are

$$E(X_t|X_{t-1}) = X_{t-1}e^{-\kappa\delta} + \alpha(1-e^{-\kappa\delta}) =: \mu(X_{t-1}) \quad \text{and} \tag{2.4}$$

$$Var(X_t|X_{t-1}) = \tfrac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta}). \tag{2.5}$$

Let $\phi(x)$ be the density function of the standard normal distribution $N(0,1)$. Then, the likelihood function of $\theta = (\kappa, \alpha, \sigma^2)$ is

$$L(\theta) = \phi\left(\sigma^{-1}\sqrt{2\kappa}(X_0-\alpha)\right)\prod_{t=1}^{n}\phi\left(\sigma^{-1}\sqrt{2\kappa(1-e^{-2\kappa\delta})^{-1}}\{X_t-\mu(X_{t-1})\}\right).$$

The maximum likelihood estimators (MLE) are

$$\hat{\kappa} = -\delta^{-1}\log(\hat{\beta}_1), \quad \hat{\alpha} = \hat{\beta}_2 \quad \text{and} \quad \hat{\sigma}^2 = 2\hat{\kappa}\hat{\beta}_3(1-\hat{\beta}_1^2)^{-1} \tag{2.6}$$

where

$$\hat{\beta}_1 = \frac{n^{-1}\sum_{i=1}^n X_iX_{i-1} - n^{-2}\sum_{i=1}^n X_i\sum_{i=1}^n X_{i-1}}{n^{-1}\sum_{i=1}^n X_{i-1}^2 - n^{-2}(\sum_{i=1}^n X_{i-1})^2}, \tag{2.7}$$

$$\hat{\beta}_2 = \frac{n^{-1}\sum_{i=1}^n(X_i-\hat{\beta}_1X_{i-1})}{1-\hat{\beta}_1} \quad \text{and} \tag{2.8}$$

$$\hat{\beta}_3 = n^{-1}\sum_{i=1}^n\{X_i-\hat{\beta}_1X_{i-1}-\hat{\beta}_2(1-\hat{\beta}_1)\}^2. \tag{2.9}$$

The conditional mean and variance (2.4) and (2.5) suggest that the discrete observations $\{X_t\}_{t=0}^n$ follow an AR(1) process with $\beta_1 = e^{-\kappa\delta}$ as the auto-regressive coefficient. As $\beta_1 \to 1$ when $\delta \to 0$, we are having a near unit root situation. Our analysis shows that

$$E(\hat{\beta}_1) = \beta_1 - \frac{4}{n} + \frac{3\kappa\delta}{n} + \frac{7}{n^2\kappa\delta} + o(n^{-2}\delta^{-1} + n^{-1}\delta). \tag{2.10}$$

Here the bias of $\hat{\beta}_1$ is controlled by two forces of asymptotic: δ and n, due to the continuous-time nature of the process. The expansion (2.10) echoes an expansion

$$E(\hat{\beta}_1) = \beta_1 - \frac{1+3\beta_1}{n} + O(n^{-2}) \tag{2.11}$$

given by Marriott and Pope (1954) and Kendall (1954) for discrete-time AR(1) model when δ is treated as fixed.

2.3 Estimation for CIR Process

A CIR (Cox et al., 1985) diffusion process satisfies

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_t}dB(t), \tag{2.12}$$

with $2\kappa\alpha/\sigma^2 > 1$. Let $c = 4\kappa\sigma^{-2}(1-e^{-\kappa\delta})^{-1}$, the transitional distribution of cX_t given X_{t-1} is non-central $\chi^2_\nu(\lambda)$ with the degree of freedom $\nu = 4\kappa\alpha\sigma^{-2}$ and the non-central component $\lambda = cX_{t-1}e^{-\kappa\delta}$.

The conditional mean is the same with (2.4) of the Vasicek process. However, due to the heteroscedasticity in the diffusion function, the conditional variance becomes

$$Var(X_t|X_{t-1}) \;=\; \tfrac{1}{2}\alpha\sigma^2\kappa^{-1}(1-e^{-\kappa\delta})^2 + X_{t-1}\sigma^2\kappa^{-1}(e^{-\kappa\delta} - e^{-2\kappa\delta}). \tag{2.13}$$

Since the non-central χ^2-density function is an infinite series involving central χ^2 densities, explicit expression of the MLEs for $\theta = (\kappa, \alpha, \sigma^2)$ is not available. To gain insight on the parameter estimation, we consider pseudo-likelihood estimators proposed by Nowman (1997), which admit close form expressions. Nowman employed a method of Bergstrom (1984) that approximates the CIR process by

$$dX_t = \kappa(\alpha - X_t)dt + \sigma\sqrt{X_{m\delta}}dB(t) \quad \text{for} \quad t \in [m\delta, (m+1)\delta) \tag{2.14}$$

which discretizes the diffusion function within each $[m\delta, (m+1)\delta)$ by its value at the left end point of the interval while keeping the drift unchanged, instead of discretizing the Brownian motion as in the conventional Euler approximation.

Without confusion in the notation, let $\{X_t\}_{t=0}^n$ be observations from process (2.14). Then, they satisfy the following discrete time series model

$$X_t = e^{-\kappa\delta}X_{t-1} + \alpha(1-e^{-\kappa\delta}) + \eta_t, \tag{2.15}$$

where $E(\eta_t) = 0$, $E(\eta_t\eta_s) = 0$ if $t \neq s$ and $E(\eta_t^2) = \frac{1}{2}\sigma^2\kappa^{-1}(1-e^{-2\kappa\delta})X_{t-1} =: \xi(X_{t-1}, \theta)$. By pretending η_t to be Gaussian distributed, a pseudo log-likelihood

$$\ell(\theta) = -\sum_{t=1}^{n}\left[\tfrac{1}{2}\log\{\xi(X_{t-1},\theta)\} + \tfrac{1}{2}\xi^{-1}(X_{t-1},\theta)\left\{X_t - e^{-\kappa\delta}X_{t-1} - \alpha(1-e^{-\kappa\delta})\right\}\right] \tag{2.16}$$

is obtained which leads to pseudo-MLEs

$$\hat{\kappa} = -\delta^{-1}\log(\hat{\beta}_1), \quad \hat{\alpha} = \hat{\beta}_2 \quad \text{and} \quad \hat{\sigma}^2 = \frac{2\hat{\kappa}\hat{\beta}_3}{1-\hat{\beta}_1^2} \tag{2.17}$$

where

$$\hat{\beta}_1 = \frac{n^{-2}\sum_{t=1}^n X_t \sum_{t=1}^n X_{t-1}^{-1} - n^{-1}\sum_{t=1}^n X_t X_{t-1}^{-1}}{n^{-2}\sum_{t=1}^n X_{t-1}\sum_{t=1}^n X_{t-1}^{-1} - 1}, \tag{2.18}$$

$$\hat{\beta}_2 = \frac{n^{-1}\sum_{t=1}^n X_t X_{t-1}^{-1} - \hat{\beta}_1}{(1-\hat{\beta}_1)n^{-1}\sum_{t=1}^n X_{t-1}^{-1}} \quad \text{and} \tag{2.19}$$

$$\hat{\beta}_3 = n^{-1}\sum_{t=1}^n \left\{X_t - X_{t-1}\hat{\beta}_1 - \hat{\beta}_2(1-\hat{\beta}_1)\right\}^2 X_{t-1}^{-1}. \tag{2.20}$$

We emphasize here that the discretized model (2.14) is used only to produce the estimators. It is the original CIR model (2.12) that is used when we analyze their properties.

3. MAIN RESULTS

We first report our investigation on the MLEs of the Vasicek process.

Theorem 1 *For a stationary Vasicek process and under Condition (2.2),*

$$E(\hat{\kappa}) = \kappa + 4/T - \left\{4\kappa n^{-1} + 7/(\kappa T^2)\right\} + o(n^{-1} + T^{-2}),$$

$$Var(\hat{\kappa}) = 2\kappa/T + o(T^{-1}),$$

$$E(\hat{\alpha}) = \alpha + o(T^{-2}), \quad Var(\hat{\alpha}) = \sigma^2\kappa^{-2}/T + o(T^{-1}),$$

$$E(\hat{\sigma}^2) = \sigma^2 + O(n^{-1}) \quad \textit{and} \quad Var(\hat{\sigma}^2) = 2\sigma^4 n^{-1} + o(n^{-1}).$$

Theorem 1 reveals, first of all, that the leading order bias of $\hat{\kappa}$ is $4/T$, and the leading order relative bias is $4/(\kappa T)$, which gets larger as κ gets smaller (weaker mean-reverting). Secondly, the leading order variance of $\hat{\kappa}$ and $\hat{\alpha}$ are both of $1/T$, which are larger order than $1/n$, the order of $Var(\hat{\sigma}^2)$. Hence, estimation for the two drift parameters are much more variable than $\hat{\sigma}^2$. These confirm the commonly observed empirical bias behavior in κ-estimation as well as larger variability in the drift parameter estimation. The theorem also reveals that despite $\hat{\alpha}$ having a larger order variance, it is almost unbiased. At the same time, contrary to the difficulties in estimating the drift parameters, estimation of σ^2 enjoys both smaller bias and less variability as having been observed in various empirical studies.

The results on the pseudo-MLEs for the CIR process are summarized below

Theorem 2 *For a stationary CIR process, and under Condition (2.2) and $2\kappa\alpha/\sigma^2 \geq 2$,*

$$E(\hat{\kappa}) = \kappa + \left(4 + \frac{2}{\theta_\alpha - 1}\right) T^{-1} + o(T^{-1}), \tag{3.1}$$

$$Var(\hat{\kappa}) = \frac{2\theta_\alpha \kappa}{(\theta_\alpha - 1)^2} \left(4 + \frac{1}{\theta_\alpha - 1}\right) T^{-1} + o(T^{-1}); \tag{3.2}$$

$$E(\hat{\alpha}) = \alpha + \frac{2\alpha}{(\theta_\alpha - 1)\kappa} n^{-1} + o(n^{-1}), \quad Var(\hat{\alpha}) = \frac{2\theta_\alpha^2}{(\theta_\alpha - 1)\theta_\beta^2 \kappa} T^{-1} + o(T^{-1}); \tag{3.3}$$

$$E(\hat{\sigma}^2) = \sigma^2 - \frac{\sigma^2 \kappa \delta}{2(\theta_\alpha - 1)} + O(n^{-1}), \quad Var(\hat{\sigma}^2) = \sigma^4 \left(2 - \frac{1}{\theta_\alpha - 1}\right) n^{-1} + o(n^{-1}), \tag{3.4}$$

where $\theta_\alpha = 2\kappa\alpha/\sigma^2$ and $\theta_\beta = 2\kappa/\sigma^2$.

Theorem 2 reveals similar features to those by Theorem 1 for the Vasicek process. These include (i) the leading order bias of $\hat{\kappa}$ is still T^{-1}; (ii) estimation of κ and α still incurs a larger order variance as compared to the estimation of σ^2. A difference is in the bias of $\hat{\sigma}^2$, which is at the order of δ^{-1}. This can be understood as a result of the piece-wise discretization of the diffusion function used in (2.14). We note that $2\kappa\alpha/\sigma^2 \geq 2$ is needed to ensure terms with $X_i^{-1} X_j^{-1}$ having bounded expectations.

An important message from Theorems 1 and 2 is that it is T, the total observation time, rather than the sample size n, that controls the bias and/or variance in estimation of κ and α. Our analysis is entirely based on each continuous-time process under consideration, and improves the heuristic justification used in Phillips and Yu (2005) which are based on results like (2.11) from discrete-time series. And most importantly, the results in these two theorems nicely explain various empirical results reported in the literature.

4. BOOTSTRAP BIAS CORRECTION

Given the explicit bias expansion in Theorem 1, a simple bias correction for $\hat{\kappa}$ for the Vasicek process is $\hat{\kappa}_1 = \hat{\kappa} - 4/T$. This will remove the leading order bias without altering the variance.

The same may be applied to the CIR process by constructing

$$\hat{\kappa}_1 = \hat{\kappa} - \left(4 + \frac{2}{\hat{\theta}_\alpha - 1}\right) T^{-1}$$

where $\hat{\theta}_\alpha = 2\hat{\kappa}\hat{\alpha}/\hat{\sigma}^2$. The limitation of this approach is that it would not be applicable to other processes unless similar bias expansions are established.

In a significant development, Phillips and Yu (2005) propose a jackknife method to correct bias in parameter estimation of diffusion processes. Their proposal was motivated by the bias expansions (2.11) established for discrete time series. It consists of first dividing the entire sample of n observations into m consecutive non-overlapping blocks of observations of size l such that $n = ml$; and then construct parameter estimators based on each block of observations, say $\hat{\theta}_i$ for the i-th block. The jackknife estimator that corrects bias in an original estimator $\hat{\theta}$ is

$$\hat{\theta}_J = \frac{m}{m-1}\hat{\theta} - \frac{\sum_{i=1}^{m} \hat{\theta}_i}{m^2 - m}.$$

They suggested using $m = 4$ which was shown numerically to produce the best trade-off between bias reduction and variance inflation.

In conventional statistical settings, it is understood (Shao and Tu, 1995) that the jackknife tends to inflate variance more than the bootstrap when both are used for bias correction. Indeed, as shown in our simulations, although using $m = 4$ has reduced the variance of the jackknife estimator as opposed to using $m = 2$, the variance can still be much larger than the original estimator. This may be due to that dividing the data into shorter blocks reduces the observation time which has been shown in Theorems 1 and 2 to be the key force in influencing the variability in the estimation of the drift parameters.

We propose a parametric bootstrap procedure for bias correction. The bootstrap (Efron, 1979) has been shown to be an effective method for bias correction and variance estimation for both independent and dependent observations as summarized in Hall (1992) and Lahiri (2003). Although our analysis was confined to the two specific processes in the previous section, the proposed bootstrap bias correction is applicable to the general multivariate diffusion process (2.1).

Let $\hat{\theta}$ be a mean square consistent estimator of θ. The parametric bootstrap procedure consists of the following steps:

Step 1. Generate a bootstrap sample path $\{X_t^*\}_{t=1}^n$ with the same sampling interval δ from $dX_t = \mu(X_t; \hat{\theta})dt + \sigma(X_t; \hat{\theta})dB_t$;

Step 2. Obtain a new estimator $\hat{\theta}^*$ from the bootstrap sample path by applying the same estimation procedure as $\hat{\theta}$;

Step 3. Repeat Steps 1 and 2 N_B number of times and obtain a set of bootstrap estimates $\hat{\theta}^{*,1}, \cdots, \hat{\theta}^{*,N_B}$.

Let $\bar{\hat{\theta}}^* = N_B^{-1} \sum_{b=1}^{N_B} \hat{\theta}^{*,b}$, the bootstrap bias-corrected estimator is $\hat{\theta}_B = 2\hat{\theta} - \bar{\hat{\theta}}^*$ and the bootstrap estimates for the variance of $\hat{\theta}$ is

$$\widehat{Var}(\hat{\theta}) = N_B^{-1} \sum_{b=1}^{N_B} \left(\hat{\theta}^{*,b} - \bar{\hat{\theta}}^*\right) \left(\hat{\theta}^{*,b} - \bar{\hat{\theta}}^*\right)'.$$

Here A' denotes matrix transpose.

In the above Step 1, we first generate an initial value of X_0^* from the stationary marginal distribution. For a univariate process, the stationary density is known to be

$$\pi_\theta(x) = \frac{\xi(\theta)}{\sigma^2(x, \theta)} \exp\left\{\int_{x_0}^{x} \frac{2\mu(t, \theta)}{\sigma^2(t, \theta)} dt\right\}.$$

If the transitional distribution of $X_{t\delta}$ given $X_{(t-1)\delta}$ is known, we can generate $X_{t\delta}^*$ given $X_{(t-1)\delta}^*$ from that distribution. If the transitional distribution is unknown, we can use the approximate transitional density of Aït-Sahalia(1999). We may also apply the Milstein scheme (Kloeden and Platen, 2000) which is more accurate than the first-order Euler scheme.

The bootstrap bias correction method shares some key features of the jackknife method, for instance it can be applied to a general diffusion process (univariate or multivariate), and for a range of estimators including the MLE, the pseudo-MLE and discretization based estimators. The bootstrap bias correction is justified in the following theorem. Before that, let us introduce some notations.

Let $\theta = (\theta_1, \cdots, \theta_p)^T$ be a vector of parameters of the general diffusion process (2.1), and $\hat{\theta} = (\hat{\theta}_1, \cdots, \hat{\theta}_p)^T$ be a consistent estimator of θ. Write the bootstrap bias corrected estimator

$\hat{\theta}_B = (\hat{\theta}_{B1}, \cdots, \hat{\theta}_{Bp})^T$. For $l = 1, \cdots, p$, let $b_{nl}(\theta) = E(\hat{\theta}_{nl}) - \theta_l$ and $v_{nl}(\theta) = Var(\hat{\theta}_{nl})$ be the bias and variance components of $\hat{\theta}_l$ respectively, and

$$b_{nl}(\theta) = \beta_{nl} b_{nl}^{(0)}(\theta) \quad \text{and} \quad v_{nl}(\theta) = \nu_{nl} v_{nl}^{(0)}(\theta)$$

so that both $|b_{nl}^{(0)}(\theta)|$ and $|v_{nl}^{(0)}(\theta)|$ are uniformly bounded away from ∞ and zero with respect to n and δ. Hence, both β_{nl} and ν_{nl} are the exact orders of magnitude for the bias and variance of $\hat{\theta}_l$ respectively.

Theorem 3 *Suppose that for each $l = 1, \cdots, p$, (i) $\beta_{nl}^2 + \nu_{nl} \to 0$ as both n and $T \to \infty$ and (ii) $b_{nl}^{(0)}(\theta)$ and $v_{nl}^{(0)}(\theta)$, as functions of θ, are twice continuously differentiable within a hypersphere S in R^p that contains the real parameter θ; and (iii) $E\{b_{nl}^{(0)}(\hat{\theta})\}^2 = O(1)$. Then,*

$$E(\hat{\theta}_{Bl}) = \theta_l + o\{b_{nl}(\theta)\} \quad \textit{and} \quad Var(\hat{\theta}_{Bl}) = v_{nl}(\theta) + o\{v_{nl}(\theta)\}. \tag{4.1}$$

The theorem shows that the proposed bootstrap estimator $\hat{\theta}_B$ reduces the bias of the original estimator $\hat{\theta}$ while having the same leading order variance as $\hat{\theta}$. The conditions of the theorem are quite weak, which are no more than the mean square consistent and differentiability of the bias and variance functions near θ. These are satisfied by most of the commonly used estimators, for instance those evaluated in Theorems 1 and 2.

5. SIMULATION STUDIES

We report in this section results from simulation studies which were designed to (i) confirm the theoretical findings of Theorems 1 and 2, (ii) evaluate the performance of the proposed bootstrap bias correction, and (iii) compare the bootstrap proposal with the jackknife bias correction proposed by Phillips and Yu (2005). In the simulation studies, both univariate (Vasicek and CIR processes) and bivariate processes were considered as well as a range of parameter estimators. All the simulation results reported in this section were all based on 5000 simulations and 1000 bootstrap resamples.

5.1 Univariate Processes

To confirm the theoretical results given in Theorems 1 and 2, we simulated two sets of models for both Vasicek and CIR processes. The parameter values used for the Vasicek process were $\theta = (\kappa, \alpha, \sigma^2) = (0.858, 0.0891, 0.00219)$ (Vasicek Model 1) and $(0.215, 0.0891, 0.0005)$ (Vasicek Model 2). For the CIR process, the parameter $\theta = (\kappa, \alpha, \sigma^2) = (0.892, 0.09, 0.033)$ (CIR Model 1) and $(0.223, 0.09, 0.008)$ (CIR Model 2) respectively. Both Vasicek Model 2 and CIR Model 2 have only a quarter of the mean-reverting force of Vasicek Model 1 and CIR 1 respectively. We chose $\delta = 1/12$ that corresponds to monthly observations in annualized term. The sample size n was 120, 300, 500 and 2000. The purpose of trying $n = 2000$ was to confirm the asymptotic bias and variance developed in Theorems 1 and 2. As the transitional distribution of these two processes are known, the simulated sample paths were generated from the known transitional distribution with the initial value X_0 from their known stationary distributions respectively. We only report the simulation results for the CIR models as the pattern of results for the Vasicek simulation was the same.

Tables 1 report the average bias, relative bias (R. Bias), standard deviation (SD) and root mean square error (RMSE) for the two CIR models. We also report in parentheses the asymptotic bias and standard deviation prescribed by expansions in Theorem 2. We observe that the severe bias in κ estimation was very clear, especially when the amount of the mean reverting was weak (Tables 1(b)). At the same time, there was little bias in the estimation of α and the overall quality in estimating σ^2 was very high even for sample size as small as 120. These all confirmed our theoretical findings. We find the difference between the simulated bias and SD and those predicted by the theoretical expansions decreased as n and T were increased, and was very small at $n = 2000$, which was reassuring.

We then applied the bootstrap bias correction to estimation of κ for the Vasicek and CIR models. The jackknife approach proposed by Phillips and Yu (2005) was also performed with $m = 4$. The simulation results are summarized in Tables 2. We see that the bootstrap bias correction effectively reduced the bias without increasing the variance of the estimation much. However for the jackknife bias correction, there was some non-ignorable variance inflation. The bootstrap bias correction had less RMSE than the jackknife bias correction as well as the original estimator.

We also carried out estimation and bootstrap bias correction based on the approximated likelihood estimation of Aït-Sahalia(2002) for the CIR Model 2. This was designed to see if there were significant difference between the approximated MLEs and the pseudo-likelihood estimators of Nowman (1997) which we have analyzed in Theorem 2. The results are reported in Table 3, which were similar to the pseudo-likelihood estimators in Table 1 (b). However, we did see that the use of the approximate likelihood did produce estimates which had slightly smaller bias and standard deviation. Most importantly, the bootstrap bias correction worked well for the approximated likelihood in reducing both the bias and mean square error in κ-estimation.

5.2 Multivariate Processes

To evaluate the general applicability of the proposed bootstrap procedure, we carry out simulations for the following bivariate processes:

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

where

$$X_t = \begin{pmatrix} X_{1t} \\ X_{2t} \end{pmatrix}, \kappa = \begin{pmatrix} \kappa_{11} & 0 \\ \kappa_{21} & \kappa_{22} \end{pmatrix}, \alpha = \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} \quad \text{and} \quad \sigma(X_t) = \begin{pmatrix} \sigma_{11} X_{1t}^{\rho} & 0 \\ 0 & \sigma_{22} X_{2t}^{\rho} \end{pmatrix}$$

with $\rho = 0$ and $1/2$ respectively. Here $\rho = 0$ corresponds a bivariate Ornstein-Uhlenbeck process whose exact transitional density is known to be bivariate Gaussian, whereas $\rho = 1/2$ corresponds to a bivariate extension of the Feller's process. We will report only simulation results for the bivariate Feller's process as those for the bivariate Ornstein-Uhlenbeck were similar.

Unless $\kappa_{21} = 0$, the transitional density of the bivariate Feller's process does not admit an explicit form (Aït-Sahalia and Kimmel, 2002). We consider estimation based on the Euler discretization:

$$X_t - X_{t-1} = \kappa(\alpha - X_{t-1})\delta + \sigma(X_{t-1})\Delta B_t, \tag{5.2}$$

where $\Delta B_t = (B_{1t} - B_{1t-1}, B_{2t} - B_{2t-1})'$ is a discretization of the bivariate Brownian motion. A pseudo-likelihood can be constructed similar to that of (2.16) from the conditional mean and variance structures:

$$E(X_t|X_{t-1}) = \kappa(\alpha - X_{t-1})\delta, \quad \text{and} \quad Var(X_t|X_{t-1}) = \sigma(X_{t-1})\text{diag}(\delta, \delta)\sigma'(X_{t-1}).$$

The approximation (5.2) is subject to a discretization error. However, the pseudo-likelihood estimator is consistent as long as $\delta \to 0$.

We chose $(\kappa_{11}, \kappa_{21}, \kappa_{22}, \alpha_1, \alpha_2, \sigma_{11}, \sigma_{22}) = (0.223, 0.4, 0.9, 0.09, 0.08, 0.008, 0.03)$. The generation of the bivariate diffusion process was via the Milstein's scheme (Kloeden and Platen, 2000). We also pre-runned the process 1000 times before starting the real simulation to make the simulated sample path stationary.

Table 4 summarizes the simulation performance of the psudo-MLEs and the bootstrap bias corrected parameter estimation. We observe that similar to the univariate case as reported in Tables 1 and 2, the estimators of the drift parameters in κ and α had worse performance than the diffusion parameters in σ. It is encouraging to see that the bootstrap worked effectively in reducing both the bias and the mean square errors. The boostrap bias correction did not working as effectively as those for κ_{11} and κ_{22}. However, our simulation for the bivariate Ornstein-Uhlenbeck process showed that the bootstrap work well for all κ coefficient including κ_{21}. Hence, this suggests that it might be due to the use of the pseudo-likelihood that only uses the conditional variance that ignores the dependence between the two marginal processes. We note that there was no need to carry out the bootstrap bias correction for (α_1, α_2) and $(\sigma_{11}, \sigma_{22})$. This is consistent with the recommendations from Theorems 1 and 2.

6. A CASE STUDY AND OPTION PRICING

We analyze a Fed fund interest rate dataset consisting 432 monthly observations from January 1963 to December 1998. This dataset has been analyzed in (Aït-Sahalia, 1999) to demonstrate his approximate likelihood estimation.

In additional to estimate the Vasicek and CIR processes, we computed two option prices driven by these two processes: $P_{t,T}(\theta)$, the price of a zero-coupon-bond at time t that pays \$1 at a maturity time T; and $C_{t,T,S,K}(\theta)$, the price at time t of an European call option with maturity T and a strike price K on a zero-coupon bond maturing at $S > T$. See Vasicek (1977) and Cox et al. (1985) for detailed expressions of these two option prices as functions of parameters of an underlying interest rate process.

We first estimated the parameters of the underlying diffusion processes (Vasicek and CIR) by the maximum likelihood method and carried out the bootstrap bias correction. Then, we calculated the option prices $P_{t,T}(\theta)$ and $C_{t,T,S,K}(\theta)$ based on the estimated parameters of the Vasicek or CIR process with $t = 0$, $T = 1$, $S = 3$ and the initial interest rate at 5%. The face value of the European Call option on a three year discount bond was \$100 with a strike price $K = \$90$.

The parametric bootstrap was used to estimate both the bias of the parameter estimates of the process and the option prices as well as their standard deviations based on 1000 resamples. The bootstrap implementation for the option prices were readily made by extending Steps 2 and 3 in the procedure outlined in Section 4 to include computation of the option prices in each resample. The empirical results are reported in Table 5. It is observed that the bootstrap bias estimates (Estimated Bias) were rather substantial in both $\hat{\kappa}$ and the option price $\hat{C}(0, 1, 3, 90)$. While the large bias in $\hat{\kappa}$ was expected from Theorems 1 and 2, it was rather alarming to see a large under-estimation (more than 10%) in $\hat{C}(0, 1, 3, 90)$. Also, the bootstrap estimate of the standard deviation for both κ and $C(0, 1, 3, 90)$ were quite large too. The large variability in the option price should be taken into consideration and indicates the difficulties in producing accurate estimated prices. The empirical analysis also indicated that the European call option is more affected by the biased parameter estimates for the underlying interest rate process than the zero-coupon bond, which can be understood due to different transformations of the underlying diffusion parameters. We also supplied in parentheses the estimated standard deviation based on the leading order variance terms prescribed by Theorems 1 and 2, which were all comparable with the bootstrap estimates.

7. DISCUSSION

The estimation of the drift parameters in diffusion processes has been known to be challenging when the process is lack of dynamics. Our analyses reported in Theorems 1 and 2 quantify the underlying sources of the challenge for two commonly used interest rate processes. One source of the challenge, apart from being lack of dynamics, is that the accuracy in the estimation of the drift parameters is governed by T, the total amount of time a process is observed, rather than the sample size n. This is different from estimation for discrete time series models where n is the driving force for accuracy. A reason for the proposed parametric bootstrap method working more effectively than the jackknife method is because its re-creation of the full observation length in each resampling that fully utilizes the amount of observation data available and the assumed model for the process.

While we have gained quite complete understanding on parameter estimation for the two popular processes in Theorems 1 and 2, there is a need to understand more on estimation for multivariate processes, in particularly estimation of parameters that control the correlation between components of the process. Another important issue is how to reduce the variability in estimation of the drift parameters and the European call option. We hope that future research will address these issues.

REFERENCES

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

— (1999), "Transition densities for interest rate and other nonlinear diffusions," *Journal of Finance*, 54, 1361–1395.

— (2002), "Maximum likelihood estimation of discretely sample diffusion: A close form approximation approach," *Econometrica*, 70, 223–262.

Aït-Sahalia, Y. and Kimmel, R. (2002), "Estimating affine multifactor term structure models using closed-Form likelihood expansions," *Working Paper.*

Aït-Sahalia, Y. and Mykland, P. (2003), "The Effects of Random and Discrete Sampling When Estimating Continuous-Time Diffusions," *Econometrica*, 71, 483–549.

— (2004), "Estimating diffusions with discretely and possibly Randomly spaced data: a general theory," *Annals of Statistics*, 32, 2186–2222.

Ball, C. and Torous, W. (1996), "Unit roots and the estimation of interest rate dynamics," *Journal of Empirical Finance*, 3, 215–238.

Bandi, F. M. and Phillips, P. C. B. (2003), "Fully nonparametric estimation of Scalar diffusion models," *Econometrica*, 71, 241–283.

Bergstrom, A. R. (1984), "Continuous time stochastic models and issues of aggregation over time," *Handbook of Econometrics (Z. Griliches and M.D. Intriligator Eds) Vol 2.*

Bibby, B. and Sørensen, M. (1995), "Martingale estimating functions for discretely observed diffusion processes," *Bernoulli*, 1, 17–39.

Black, F. and Scholes, M. (1973), "The pricing of options and corporate liabilities," *Journal of Political Economy*, 81, 673–654.

Cai, Z. and Hong, Y. (2003), "Nonparametric methods in continuous-time finance: a selective review," *Recent Advances and Trends in Nonparametric Statistics (M. G. Akritas and D. N. Politis, eds.)*, 283–302.

Cox, J., Ingersoll, J., and Ross, S. (1985), "A theory of the term structure of interest rates," *Econometrica*, 53, 385–407.

Efron, B. (1979), "Bootstrap methods: another look at the jackknife." *Annals of Statistics*, 7, 1–26.

Fan, J. (2005), "A selective overview of nonparametric methods in financial econometrics (with discussion)," *Statistical Science*, 20, 317–357.

Fan, J. and Zhang, C. (2003), "A re-examination of Stanton's diffusion estimations with applications to financial model validation," *Journal of American Statistical Association*, 98, 118–134.

Hall, P. (1992), *The Bootstrap and Edgeworth Expansion*, Springer-Verlag.

Hansen, L. P. and Scheinkman, J. A. (1995), "Back to the future: generating moment implications for continuous-time markov processes," *Econometrica*, 63, 767–804.

Jiang, G. J. and Knight, J. L. (1997), "A nonparametric approach to the estimation of diffusion processes, with an application to a short-Term interest rate model," *Econometric Theory*, 13, 615–645.

Kendall, M. G. (1954), "Note on bias in the estimation of autocorrelation," *Biometrika*, 41, 403–404.

Kloeden, P. and Platen, E. (2000), *Numerical Solution of Stochastic Differential Equations*, Springer.

Lahiri, S. (2003), *Resampling Methods for Dependent data*, Springer-Verlag, Inc.

Lo, A. (1988), "Maximum likelihood estimation of generalized Ito processes with discretely sampled data," *Econometric Theory*, 4, 231–247.

Marriott, F. H. C. and Pope, J. A. (1954), "Bias in the estimation of autocorrelations," *Biometrika*, 41, 390–402.

Merton, R. C. (1971), "Optimum consumption and portfolio rules in a continuous-time model," *Journal of Economic Theory*, 3, 373–413.

Nowman, K. (1997), "Gaussian estimation of single-factor continuous time models of the term structure of interest rates," *Journal of Finance*, 52, 1695–1706.

Phillips, P. C. and Yu, J. (2005), "Jackknifing bond option prices," *Review of Financial Studies*, 18, 707–742.

Sargan, J. (1976), "Econometric estimators and the edgeworth approximation," *Econometrica*, 44, 421–448.

Shao, J. and Tu, D. (1995), *The Jackknife and Bootstrap*, Springer.

Stanton, R. (1997), "A nonparametric model of term structure dynamics and the market price of interest rate risk," *The Journal of Finance*, 52, 1973–2002.

Stroock, D. and Varadhan, S. (1979), *Multidimensional Diffusion Processes*, Springer.

Sundaresan, S. M. (2000), "Continuous-time methods in finance: a review and an assessment," *Journal of Finance*, 55, 1569–1622.

Vasicek, O. (1977), "An equilibrium characterization of the term structure," *Journal of Financial Economics*, 5, 177–186.

Yu, J. and Phillips, P. C. (2001), "A gaussian approach for estimating continuouse time models of short term interest rates," *The Ecnometrics Journal*, 4, 211–225.

APPENDIX: PROOFS

We will only present the proofs of Theorems 2 and 3 here. The proof of Theorem 1 is very similar to that of Theorem 2, but slightly easier as the observed sample path is multivariate normally distributed. Proof of Theorem 1 as well as more a detailed proof of Theorem 2 is available from the authors.

Proof of Theorem 2: We first note two basic facts regarding a sample $\{X_i\}_{i=1}^n$ from a stationary CIR process: (i) for any $i, j \geq 0$, $c \cdot X_i | X_j \sim \chi^2_\nu(\lambda)$ distribution where $\nu = \frac{4\kappa\alpha}{\sigma^2}$, $\lambda = cX_j e^{-|j-i|\kappa\delta}$ and $c = \frac{4\kappa}{\sigma^2(1-e^{-|j-i|\kappa\delta})}$; and (ii) $X_t \sim \Gamma(\theta_\alpha, \theta_\beta)$ where $\theta_\alpha = \frac{2\kappa\alpha}{\sigma^2}$ and $\theta_\beta = \frac{2\kappa}{\sigma^2}$.

Let $F(a_1, a_2; b; Z) = \frac{\Gamma(b)}{\Gamma(a_1)\Gamma(a_2)} \sum_{k=0}^\infty \frac{Z^k}{k!} \frac{\Gamma(a_1+k)\Gamma(a_2+k)}{\Gamma(b+k)}$ be the hypergeometric function. It can be shown using the above facts

$$E(X_i^{-1}) = \sum_{k=0}^\infty \frac{(\lambda/2)^k}{k!} e^{-\lambda/2} \frac{1/2}{\nu/2 + k - 1} \quad \text{and} \tag{A.1}$$

$$E\left(X_i^{-1} X_j^{-1}\right) = E\left\{X_i^{-1} E\left(X_j^{-1} | X_i\right)\right\} = \frac{\theta_\beta^2}{(\theta_\alpha - 1)^2} \cdot F(1, 1; \theta_\alpha, e^{-|j-i|\kappa\delta}). \tag{A.2}$$

The function $F(a_1, a_2, b; Z)$ converges absolutely if $b - a_1 - a_2 \geq 0$, therefore $E(X_i^{-1} X_j^{-1}) < \infty$ if $\theta_\alpha \geq 2$. This is the reason behind assuming $\theta_\alpha \geq 2$ in Theorem 2. Similarly, it can be concluded that

$$E\left(X_{i-1}^{-1} X_i X_{j-1}^{-1}\right) = C_\theta(i, j) S_{i,j} \quad \text{for} \quad i < j \quad \text{and} \tag{A.3}$$

$$E\left(X_{j-1}^{-1} X_{i-1}^{-1} X_i\right) = \frac{\theta_\beta e^{-\kappa\delta}}{\theta_\alpha - 1} + \frac{\theta_\beta^2 \alpha(1 - e^{-\kappa\delta})}{(\theta_\alpha - 1)^2} F(1, 1, \theta_\alpha, e^{-|i-j|\kappa\delta}) \quad \text{for} \quad j < i - 1,$$

where $C_\theta(i, j) = \frac{\theta_\beta \Gamma^2(\theta_\alpha - 1)}{(1-e^{-\kappa\delta})(1-e^{-(j-i-1)\kappa\delta})\Gamma(\theta_\alpha)}$, $S_{i,j} = \sum_{k=0}^\infty \sum_{l=0}^\infty D_1^k D_2^l \frac{\Gamma(\theta_\alpha+l+k+1)}{\Gamma(\theta_\alpha+l)\Gamma(\theta_\alpha+k)}$, $D_1 = e^{-\kappa\delta}/(e^{-\kappa\delta} - 1)$ and $D_2 = e^{-(j-i-1)\kappa\delta}/(e^{-(j-i-1)\kappa\delta} - 1)$;

$$E\left(X_{j-1} X_{i-1}^{-1} X_i\right) = -\frac{\alpha e^{-(i-j)\kappa\delta}}{\theta_\alpha - 1}(1 - e^{-\kappa\delta}) + \mu \quad \text{for} \quad j < i - 1 \quad \text{and} \tag{A.4}$$

$$E\left(X_{i-1}^{-1} X_i X_{j-1}\right) = e^{-(j-i-1)\kappa\delta} C(1 - e^{-\kappa\delta}) + \mu \quad \text{for} \quad j > i,$$

where $\mu = \frac{(\theta_\alpha - e^{-\kappa\delta})\alpha}{\theta_\alpha - 1}$ and $C = \frac{\theta_\alpha}{\theta_\alpha - 1}\left(\frac{\sigma^2}{2\kappa} + \alpha\right)(1 - e^{-\kappa\delta}) + \frac{\sigma^2}{\kappa} e^{-\kappa\delta} + 2\alpha e^{-\kappa\delta} - \alpha(1 + e^{-\kappa\delta}) - \frac{\alpha}{\theta_\alpha - 1}$.

Let $\mu_1 = E(X_t)$, $\mu_2 = E(X_t^{-1})$, $\mu_3 = E(X_t X_{t-1}^{-1})$, $\mu_1' = \mu_1\mu_2 - \mu_3$ and $\mu_2' = \mu_1\mu_2 - 1$. It can be shown that $\mu_1 = \alpha$ and $\mu_2 = \frac{\theta_\beta}{\theta_\alpha - 1}$, $\mu_3 = E\left\{X_{t-1}^{-1} E\left(X_t | X_{t-1}\right)\right\} = \frac{\theta_\alpha - e^{-\kappa\delta}}{\theta_\alpha - 1}$. Then by the definition in (2.18)

$$\hat{\beta}_1 = \frac{\mu_1'}{\mu_2'} + \frac{T_{11}}{\mu_2'} - \frac{\mu_1' T_{21}}{\mu_2'^2} + \frac{T_{12}}{\mu_2'} - \frac{\mu_1' T_{22}}{\mu_2'^2} - \left(\frac{T_{11} T_{21}}{\mu_2'^2} + \frac{\mu_1'^2 T_{21}^2}{\mu_2'^3}\right)\{1 + o_p(1)\}, \tag{A.5}$$

where

$$\begin{aligned}
T_{11} &= \mu_2 n^{-1}\sum_{i=1}^{n}(X_i-\mu_1)+\mu_1 n^{-1}\sum_{j=1}^{n}(X_{j-1}^{-1}-\mu_2)-n^{-1}\sum_{i=1}^{n}(X_iX_{i-1}^{-1}-\mu_3),\\
T_{12} &= n^{-2}\sum_{i=1}^{n}(X_i-\mu_1)\sum_{j=1}^{n}(X_{j-1}^{-1}-\mu_2),\\
T_{21} &= \mu_2 n^{-1}\sum_{i=1}^{n}(X_{i-1}-\mu_1)+\mu_1 n^{-1}\sum_{j=1}^{n}(X_{j-1}^{-1}-\mu_2) \quad \text{and}\\
T_{22} &= n^{-2}\sum_{i=1}^{n}(X_{i-1}-\mu_1)\sum_{j=1}^{n}(X_{j-1}^{-1}-\mu_2).
\end{aligned}$$

It is clear that $E(T_{11})=E(T_{21})=0$. And it can be shown that $E\left\{(X_i-\mu_1)(X_j^{-1}-\mu_2)\right\}=-(\theta_\alpha-1)^{-1}e^{-|j-i|\kappa\delta}$. Then

$$E\left(\frac{T_{12}}{\mu_2'}-\frac{\mu_1'T_{22}}{\mu_2'^2}\right) = -n^{-2}(e^{\kappa\delta}-e^{-\kappa\delta})f_n(\kappa\delta)=-2n^{-1}\left\{1+o(1)\right\}, \tag{A.6}$$

where

$$f_n(\kappa\delta)=\sum_{j>i}e^{-(j-i)\kappa\delta}=n\kappa^{-1}\delta^{-1}-\tfrac{1}{2}n-\kappa^{-2}\delta^{-2}+o(\delta^{-2}+n).$$

The derivations of $E(T_{11}T_{21})$ and $E(T_{21}^2)$ need the following results which can be obtained from (A.1) to (A.4),

$$n^{-2}\frac{\mu_2^2}{\mu_2'^2}E\left\{-\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mu_1)(X_{j-1}-\mu_1)+e^{-\kappa\delta}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}-\mu_1)(X_{j-1}-\mu_1)\right\}$$
$$=-2n^{-1}\left\{1+o(1)\right\}, \tag{A.7}$$

$$n^{-2}\frac{\mu_1^2}{\mu_2'^2}E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}^{-1}-\mu_2)(X_{j-1}^{-1}-\mu_2)\right\}=n^{-1}\frac{2}{(\theta_\alpha-1)\kappa\delta}\left\{1+o(1)\right\}, \tag{A.8}$$

$$n^{-2}\frac{\mu_1}{\mu_2'^2}E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_iX_{i-1}^{-1}-\mu_3)(X_{j-1}^{-1}-\mu_2)\right\}=o(n^{-1}), \tag{A.9}$$

$$n^{-2}\frac{\mu_2}{\mu_2'^2}E\left\{\sum_{i=1}^{n}\sum_{j=1}^{n}(X_iX_{i-1}^{-1}-\mu_3)(X_{j-1}-\mu)\right\}=o(n^{-1}) \quad \text{and} \tag{A.10}$$

$$n^{-2}\frac{\mu_1\mu_2}{\mu_2'^2}E\left\{-\sum_{i=1}^{n}\sum_{j=1}^{n}\left\{(X_i-\mu_1)(X_{j-1}^{-1}-\mu_2)+(X_{i-1}-\mu_1)(X_{j-1}^{-1}-\mu_2)\right\}\right.$$
$$\left.+\,2e^{-\kappa\delta}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}-\mu_1)(X_{j-1}^{-1}-\mu_2)\right\}=4n^{-1}\left\{1+o(1)\right\}. \tag{A.11}$$

We note from (A.5) that

$$
\begin{aligned}
T_{11}T_{21} &= \mu_2^2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mu_1)(X_{j-1}-\mu_1) + \mu_1^2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}^{-1}-\mu_2)(X_{j-1}^{-1}-\mu_2) \\
&+ \mu_1\mu_2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_i-\mu_1)(X_{j-1}^{-1}-\mu_2) + \mu_1\mu_2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}-\mu_1)(X_{j-1}^{-1}-\mu_2) \\
&- \mu_2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_iX_{i-1}^{-1}-\mu_3)(X_{j-1}-\mu_1) - \mu_1 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_iX_{i-1}^{-1}-\mu_3)(X_{j-1}^{-1}-\mu_2)
\end{aligned}
$$

and

$$
\begin{aligned}
T_{21}^2 &= \mu_2^2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}-\mu_1)(X_{j-1}-\mu_1) + \mu_1^2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}^{-1}-\mu_2)(X_{j-1}^{-1}-\mu_2) \\
&+ 2\mu_1\mu_2 n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n}(X_{i-1}-\mu_1)(X_{j-1}^{-1}-\mu_2).
\end{aligned}
$$

Applying the results in (A.7) to (A.11), we have

$$
E\left(\frac{T_{11}T_{21}}{\mu_2'^2} + \frac{\mu_1'^2 T_{21}^2}{\mu_2'^3}\right) = n^{-1}\left(2 + \frac{2}{\theta_\alpha - 1}\right) + o(n^{-1}).
$$

This together with (A.5) and (A.6) then lead to

$$
E(\hat{\beta}_1) = \beta_1 - n^{-1}\left(4 + \frac{2}{\theta_\alpha - 1}\right)\{1 + o(1)\}. \tag{A.12}
$$

To derive $Var(\hat{\beta}_1)$, we take variance operation on both sides of (A.5) so that

$$
Var(\hat{\beta}_1) = \left\{\frac{Var(T_{11})}{\mu_2'^2} + \frac{\mu_1'^2 Var(T_{21})}{\mu_2'^4} - \frac{2\mu_1'}{\mu_2'^3}Cov(T_{11}, T_{21})\right\}\{1 + o(1)\}.
$$

Then we apply (A.7) to (A.11) to yield

$$
Var(\hat{\beta}_1) = n^{-1}\delta\left(4 + \frac{1}{\theta_\alpha - 1}\right) + o(n^{-1}\delta). \tag{A.13}
$$

To establish the bias and variance expansions for $\hat{\kappa}$, we note

$$
\hat{\kappa} = -\frac{1}{\delta}\left[\log(\beta_1) + \frac{(\hat{\beta}_1 - \beta_1)}{\beta_1} - \frac{(\hat{\beta}_1 - \beta_1)^2}{2\beta_1^2} + O_p\{(\hat{\beta}_1 - \beta_1)^3\}\right].
$$

Applying the delta-method, we have from (A.12) and (A.13)

$$\begin{aligned} E(\hat{\kappa}) &= \kappa - \delta^{-1} E\left[\frac{\hat{\beta}_1 - \beta_1}{\beta_1} - \frac{(\hat{\beta}_1 - \beta_1)^2}{2\beta_1^2}\right] + o\left\{E(\hat{\beta}_1 - \beta_1)^2\right\} \\ &= \kappa + \left(4 + \frac{2}{\theta_\alpha - 1}\right) T^{-1} + o(T^{-1}) \quad \text{and} \\ Var(\hat{\kappa}) &= \frac{2\theta_\alpha \kappa}{(\theta_\alpha - 1)^2}\left(4 + \frac{1}{\theta_\alpha - 1}\right) T^{-1} + o(T^{-1}). \end{aligned}$$

These complete proving the part of the theorem regarding $\hat{\kappa}$. The proofs for $\hat{\alpha}$ and $\hat{\sigma}^2$ are almost the same by first carrying out Taylor expansions for the estimators and then applying those intermediate results given from (A.1) to (A.4) and from (A.7) to (A.11).

The proof of Theorem 3 needs the following lemma.

Lemma 1 *Let $\hat{\theta}_n$ be an estimator of θ based on n observations, $b_n(\theta) = E(\hat{\theta}_n) - \theta$ and*

(i) For some integer $N \geq 2$, $E||\hat{\theta}_n - \theta||^N = O(\eta_{n,N})$ where $\eta_{n,N} \to 0$ as $n \to \infty$.

(ii) For some $K \geq 1$, $E\{\phi_n(\hat{\theta})\}^K = O(\xi_{n,K})$ for a sequence of constants $\{\xi_{n,K}\}_{n\geq 1}$.

Then, $E\{\phi_n^k(\hat{\theta}_n)\} - E\{\phi_{nr}^k(\hat{\theta}_n)\} = O(\eta_{n,N}^{r/N} + \xi_{n,K}^{k/K}\eta_{n,N}^{(K-k)/K})$.

Proof: It can be obtained by modifying the proof of Theorem A.2 of Sargan (1976). Noticeably we use $\eta_{n,N}$ and $\xi_{n,K}$ to replace T^{-rR} and T^{λ} respectively in Sargan (1976).

Proof of Theorem 3: Recall $\hat{\theta}_B = \hat{\theta} - (\bar{\hat{\theta}}^* - \hat{\theta})$ where $\bar{\hat{\theta}}^* = N_B^{-1}\sum_{i=1}^n \hat{\theta}_i^*$ and N_B is the replication number of bootstrap resamples. Let χ_n be the σ-field generated by $X_1, \ldots, X_n$. As the bootstrap generates the resamples for the parametric diffusion process, where $\hat{\theta}^*$ are estimations based on the resampled path in the same way as $\hat{\theta}$ based on the original sample, we have

$$E(\hat{\theta}^*|\chi_n) = b_n(\hat{\theta}) \quad \text{and} \quad Var(\hat{\theta}^*|\chi_n) = v_n(\hat{\theta}).$$

First consider the bias of the bootstrap estimator $\hat{\theta}_B$ and note that

$$E(\hat{\theta}_B) = E\left\{E(\hat{\theta}_B|\chi_n)\right\} = E\left[2\hat{\theta}_n - \left\{\hat{\theta}_n + b_n(\hat{\theta})\right\}\right] = \theta + b_n(\theta) - E\{b_n(\hat{\theta})\}.$$

We need to show

$$E\{b_{nl}(\hat{\theta})\} - b_{nl}(\theta) = o\{b_{nl}(\theta)\}. \tag{A.14}$$

Choose $\phi_n(x) = b_{nl}(x)$, $r = 1$, $N = 2$, $k = 1$, $K = 2$, $\eta_{n,N} = O(\nu_{nl})$ and $\xi_{n,K} = 1$ in Lemma 12. Then $E\{b_{nl}^{(0)}(\hat{\theta})\} - b_{nl}^{(0)}(\theta) = O\{(\nu_{nl} + \beta_{nl}^2)^{1/2}\} = o(1)$

which readily leads to (A.14) and the first conclusion of the theorem.

Applying the Lemma in a similar fashion, we have

$$E\{b_{nl}^2(\hat{\theta})\} = b_{nl}^2(\theta) + o(\beta_{nl}^2), \tag{A.15}$$

$$E\{v_{nl}(\hat{\theta})\} = v_{nl}(\theta) + o\{v_{nl}(\theta)\} \tag{A.16}$$

Let us now consider the variance of $\hat{\theta}_B$. Note that

$$\begin{aligned} Var(\hat{\theta}_B) &= Var\left\{E(\hat{\theta}_B|\chi_n)\right\} + E\left\{Var(\hat{\theta}_B|\chi_n)\right\} \\ &= Var\left\{\hat{\theta} - b_n(\hat{\theta})\right\} + E\left\{\frac{1}{N_B}Var(\hat{\theta}^{*,1})\right\} \\ &= Var\left\{\hat{\theta} - b_n(\hat{\theta})\right\} + \frac{1}{N_B}E\left\{v_n(\hat{\theta})\right\} \end{aligned}$$

From (A.16) and by choosing N_B large enough, $N_B^{-1}E\left\{v_n(\hat{\theta})\right\} = o\{v_n(\theta)\}$. Note that (A.14) and (A.15) mean that

$$Var\{b_{nl}(\hat{\theta})\} = Eb_{nl}^2(\hat{\theta}) - E^2\{b_{nl}(\theta)\} = O\{b_{nl}^2(\theta)\} + o\{v_{nl}(\theta)\} = o\{v_{nl}(\theta)\}.$$

This and the Cauthy-Schwartz inequality lead to $|Cov\{\hat{\theta}_{nl}, b_{nl}(\hat{\theta})\}| = o\{v_{nl}(\theta)\}$. Hence $Var\left\{\hat{\theta}_{nl} - b_{nl}(\hat{\theta})\right\} = Var(\hat{\theta}_{nl}) + o\{v_{nl}(\theta)\}$. This establishes the second part of the theorem.

(a) CIR Model 1

		κ	α	σ^2
	True Value	0.892	0.09	0.033
$n = 120$	Bias (A. Bias)	0.464(0.452)	$2.4 \cdot 10^{-4}(4.3 \cdot 10^{-4})$	$8.7 \cdot 10^{-4}(3.5 \cdot 10^{-4})$
	R Bias (%)	52.005	0.270	2.661
	SD (Asy. SD)	0.627(0.497)	0.020(0.021)	0.005(0.004)
	RMSE	0.780	0.020	0.005
$n = 300$	Bias (A. Bias)	0.179(0.180)	$2.2 \cdot 10^{-4}(1.7 \cdot 10^{-4})$	$5.8 \cdot 10^{-4}(3.5 \cdot 10^{-4})$
	R Bias (%)	20.107	0.250	1.778
	SD (Asy. SD)	0.334(0.314)	0.012(0.013)	0.003(0.003)
	RMSE	0.380	0.012	0.003
$n = 500$	Bias (A. Bias)	0.107(0.108)	$6.4 \cdot 10^{-5}(1.0 \cdot 10^{-4})$	$4.9 \cdot 10^{-4}(3.5 \cdot 10^{-4})$
	R Bias (%)	12.037	0.070	1.510
	SD (Asy. SD)	0.247(0.243)	0.009(0.01)	0.002(0.002)
	RMSE	0.269	0.009	0.002
$n = 2000$	Bias (A. Bias)	0.025(0.027)	$3.5 \cdot 10^{-5}(3.0 \cdot 10^{-5})$	$3.5 \cdot 10^{-4}(3.5 \cdot 10^{-4})$
	R Bias (%)	2.805	0.039	1.061
	SD (Asy. SD)	0.112(0.121)	0.005(0.005)	0.001(0.001)
	RMSE	0.115	0.005	0.001

(b) CIR Model 2

		κ	α	σ^2
	True Value	0.223	0.09	0.008
$n = 120$	Bias (A. Bias)	0.509(0.452)	$1.2 \cdot 10^{-3}(1.7 \cdot 10^{-3})$	$1.5 \cdot 10^{-4}(2.9 \cdot 10^{-5})$
	R Bias (%)	228.251	1.343	1.796
	SD (Asy. SD)	0.507(0.242)	0.036(0.042)	0.001(0.001)
	RMSE	0.719	0.036	0.001
$n = 300$	Bias (A. Bias)	0.185(0.180)	$9.2 \cdot 10^{-4}(7.0 \cdot 10^{-4})$	$8.7 \cdot 10^{-5}(2.9 \cdot 10^{-5})$
	R Bias (%)	82.836	1.018	1.062
	SD (Asy. SD)	0.222(0.153)	0.025(0.026)	0.0007(0.0006)
	RMSE	0.289	0.025	0.001
$n = 500$	Bias (A. Bias)	0.108(0.108)	$3.7 \cdot 10^{-4}(4.1 \cdot 10^{-4})$	$5.5 \cdot 10^{-5}(2.9 \cdot 10^{-5})$
	R Bias (%)	48.612	0.408	0.669
	SD (Asy. SD)	0.148(0.118)	0.019(0.02)	0.0005(0.0005)
	RMSE	0.183	0.019	0.001
$n = 2000$	Bias (A. Bias)	0.025(0.027)	$9.2 \cdot 10^{-5}(1.0 \cdot 10^{-4})$	$2.9 \cdot 10^{-5}(2.9 \cdot 10^{-5})$
	R Bias (%)	11.145	0.102 0.346	
	SD (Asy. SD)	0.058(0.060)	0.009(0.01)	$2.6 \cdot 10^{-4}(2.5 \cdot 10^{-4})$
	RMSE	0.063	0.009	$2.6 \cdot 10^{-4}$

Table 1: Bias, Relative bias(R.bias), standard deviation(SD) and the root mean squared error(RMSE) of the pseudo-likelihood estimator for the CIR models; figures inside the parentheses are those predicted by the theoretical expansions in Theorem 2.

(a) Vasicek Model 1 and CIR Model 1

		$\hat{\kappa}$	$\hat{\kappa}_J$	$\hat{\kappa}_B$	$\hat{\kappa}$	$\hat{\kappa}_J$	$\hat{\kappa}_B$
	Bias	0.481	−0.120	0.001	0.464	−0.122	0.002
$n = 120$	R Bias (%)	56.039	14.941	0.118	52.005	13.650	0.178
	SD	0.659	0.767	0.623	0.627	0.730	0.651
	RMSE	0.816	0.778	0.623	0.780	0.739	0.651
	Bias	0.181	−0.026	−0.003	0.179	−0.027	−0.004
$n = 300$	R Bias (%)	21.082	3.070	0.406	20.107	3.094	0.447
	SD	0.329	0.353	0.321	0.334	0.365	0.326
	RMSE	0.375	0.354	0.321	0.380	0.366	0.326
	Bias	0.111	0.005	0.001	0.107	−0.008	0.007
$n = 500$	R Bias (%)	12.880	0.586	0.073	12.037	0.842	0.826
	SD	0.240	0.250	0.235	0.247	0.257	0.245
	RMSE	0.265	0.250	0.235	0.269	0.257	0.245

(b) Vasicek Model 2 and CIR Model 2

		$\hat{\kappa}$	$\hat{\kappa}_J$	$\hat{\kappa}_B$	$\hat{\kappa}$	$\hat{\kappa}_J$	$\hat{\kappa}_B$
	Bias	0.507	−0.112	0.032	0.509	−0.088	0.030
$n = 120$	R Bias (%)	236.344	51.974	14.774	228.251	39.283	13.579
	SD	0.519	0.645	0.510	0.507	0.623	0.501
	RMSE	0.726	0.655	0.511	0.719	0.630	0.502
	Bias	0.191	−0.029	0.002	0.185	−0.032	0.008
$n = 300$	R Bias (%)	88.985	13.465	0.829	82.836	14.428	3.461
	SD	0.221	0.261	0.219	0.222	0.265	0.226
	RMSE	0.292	0.262	0.219	0.289	0.267	0.226
	Bias	0.114	−0.011	0.002	0.108	−0.0161	0.003
$n = 500$	R Bias (%)	53.033	5.230	0.861	48.612	7.209	1.325
	SD	0.150	0.170	0.147	0.148	0.167	0.150
	RMSE	0.189	0.171	0.147	0.183	0.168	0.150

Table 2: Comparisons of bias corrections for the Vasicek and CIR Models, $\hat{\kappa}_J$ and $\hat{\kappa}_B$ are, respectively, the jackknife and bootstrap bias corrected estimators for κ.

CIR model 2

		$\hat{\kappa}$	$\hat{\alpha}$	$\hat{\sigma}^2$	$\hat{\kappa}_J$	$\hat{\kappa}_B$
	True Value	0.223	0.09	0.008	0.223	0.223
$n = 120$	Bias	0.494	0.004	$1 \cdot 10^{-4}$	-0.072	0.035
	R Bias (%)	221.684	4.778	1.507	32.412	15.559
	SD	0.490	0.058	0.001	0.596	0.514
	RMSE	0.696	0.058	0.001	0.601	0.516
$n = 300$	Bias	0.180	0.001	$6 \cdot 10^{-5}$	-0.035	0.013
	R Bias (%)	80.559	1.349	0.700	15.803	5.618
	SD	0.223	0.0262	0.001	0.262	0.234
	RMSE	0.286	0.0262	0.001	0.265	0.234
$n = 500$	Bias	0.1001	$7 \cdot 10^{-4}$	$4 \cdot 10^{-5}$	-0.022	-0.003
	R Bias (%)	45.279	0.834	0.478	9.806	1.493
	SD	0.147	0.019	0.001	0.166	0.151
	RMSE	0.178	0.019	0.001	0.167	0.151

Table 3: Parameters estimation and bias correction for CIR Model 2 based on the approximated likelihood method of Aït-Sahalia (1999).

	Bivariate Feller Process						
$n = 120$	κ_{11}	κ_{21}	κ_{22}	α_1	α_2	σ_1^2	σ_2^2
True Value	0.223	0.4	0.9	0.09	0.08	0.008	0.03
Bias	0.478	0.396	0.531	0.001	-0.0001	-0.0002	-0.0006
	(0.101)	(0.168)	(0.187)	(0.0001)	(-0.001)	(-0.0001)	(-0.0005)
Rbias(%)	214.48	98.948	88.442	1.537	0.147	2.68	2.141
	(45.227)	(41.948)	(31.086)	(0.135)	(1.273)	(1.245)	(1.672)
SD	0.468	0.584	0.561	0.037	0.036	0.001	0.0041
	(0.543)	(0.846)	(0.696)	(0.037)	(0.037)	(0.001)	(0.004)
RMSE	0.669	0.705	0.772	0.037	0.036	0.001	0.0042
	(0.553)	(0.863)	(0.721)	(0.037)	(0.037)	(0.001)	(0.004)
$n = 300$	κ_{11}	κ_{21}	κ_{22}	α_1	α_2	σ_1^2	σ_2^2
Bias	0.174	0.08	0.206	0.002	-0.0009	$-9 \cdot 10^{-5}$	$-7 \cdot 10^{-5}$
	(-0.003)	(0.064)	(-0.006)	(0.002)	(-0.0012)	$(-1 \cdot 10^{-5})$	(-0.0002)
Rbias(%)	78.048	20.048	34.368	2.131	1.12	1.06	0.219
	(1.266)	(15.890)	(1.006)	(1.710)	(1.499)	(0.164)	(0.776)
SD	0.212	0.304	0.303	0.026	0.023	0.0006	0.0027
	(0.208)	(0.297)	(0.288)	(0.026)	(0.023)	(0.0007)	(0.0027)
RMSE	0.275	0.314	0.366	0.026	0.023	0.0007	0.0027
	(0.208)	(0.303)	(0.288)	(0.026)	(0.023)	(0.0007)	(0.0027)
$n = 500$	κ_{11}	κ_{21}	κ_{22}	α_1	α_2	σ_1^2	σ_2^2
Bias	0.102	0.017	0.115	$3 \cdot 10^{-5}$	0.000423	$-5 \cdot 10^{-5}$	0.0002
	(-0.003)	(0.024)	(-0.013)	$(1 \cdot 10^{-5})$	$(-8 \cdot 10^{-5})$	$(-6 \cdot 10^{-6})$	(-0.0001)
Rbias(%)	45.52	4.271	19.087	0.035	0.528	0.656	0.64
	(1.227)	(6.102)	(2.163)	(0.012)	(0.099)	(0.077)	(0.359)
SD	0.148	0.22	0.214	0.021	0.018	0.0005	0.0024
	(0.147)	(0.227)	(0.206)	(0.021)	(0.019)	(0.0005)	(0.0021)
RMSE	0.179	0.22	0.243	0.021	0.018	0.00048	0.0024
	(0.147)	(0.228)	(0.206)	(0.021)	(0.018)	(0.00048)	(0.0023)

Table 4: Bias, Relative bias(R.bias), standard deviation(SD) and the root mean squared error(RMSE) of the pseudo-likelihood estimator for a Bivariate Feller's Process; figures in parentheses are those for the bootstrap bias corrected estimators.

(a) Under Vasicek Process

	$\hat{\kappa}$	$\hat{\alpha}$	$\hat{\sigma}^2$	$\hat{P}$	$\hat{C}$
Estimates	0.261	0.07	0.0005	0.846	3.03
Estimated Bias	0.125	$2 \cdot 10^{-5}$	$2 \cdot 10^{-6}$	-0.004	-0.313
Bootstrap Estimates	0.136	0.07	0.0005	0.852	3.67
$\widehat{SD}(Asy.SD)$	0.17(0.12)	0.015(0.014)	$3.5 \cdot 10^{-5}(3.4 \cdot 10^{-5})$	0.015	1.146

(b) Under CIR Process

	$\hat{\kappa}$	$\hat{\alpha}$	$\hat{\sigma}^2$	$\hat{P}$	$\hat{C}$
Estimates	0.146	0.07	0.0043	0.852	2.64
Estimated Bias	0.127	$8 \cdot 10^{-4}$	$3 \cdot 10^{-5}$	-0.004	-0.294
Bootstrap Estimates	0.018	0.069	0.0043	0.860	3.39
$\widehat{SD}(Asy.SD)$	0.152(0.11)	0.02(0.02)	$3.0 \cdot 10^{-4}(3.0 \cdot 10^{-4})$	0.014	0.996

Table 5: Results for a case study: $\hat{P}$ and $\hat{C}$ are the estimated prices for the discount bond and European call option respectively; Estimated Bias, Bootstrap Estimates and $\widehat{SD}$ are respectively the bootstrap estimate of the bias, the bootstrap bias corrected estimate and the bootstrap estimation of the standard deviation; figures in parentheses are the asymptotic standard deviation (Asy.SD) based on the leading order variance given Theorems 1 and 2.