Parametric versus Nonparametric Estimation of Diffusion Processes — A Monte Carlo Comparison

George J. Jiang [1] and John L. Knight [2]

May, 1997

Abstract

In this paper, a Monte Carlo simulation is performed to investigate the finite sample properties of various estimators, based on discretely sampled observations, of the continuous-time Itô diffusion process. The simulation study aims to compare the performance of the nonparametric estimators proposed in Jiang and Knight (1996) with common parametric estimators based on those diffusion processes which have explicit transition density functions. The simulation results show that, with a large sample over a short sampling period, although all the parametric diffusion estimators perform very well, the parametric drift estimators perform very poorly. However, both the nonparametric diffusion and drift estimators perform reasonably well.

Keywords: Diffusion Process; Nonparametric Estimation; Monte Carlo Simulation

1 Department of Econometrics, University of Groningen, P.O. Box 800, 9700 AV Groningen, The Netherlands, e-mail: G.Jiang@eco.rug.nl

2 Department of Economics, University of Western Ontario, Canada.

1. Introduction

The purpose of this paper is to design and perform a small Monte Carlo simulation experiment to investigate the finite sample properties of various estimation methods, based on discretely sampled observations, of a continuous-time Itô diffusion process represented by the following stochastic differential equation (SDE):

$$dX_t = \mu(X_t)dt + \sigma(X_t)dW_t \tag{1}$$

with initial condition

$$X_{t_0} = X$$

where $\{W_t, t \geq t_0\}$ is a standard Brownian motion process or a Wiener process. The functions $\mu(\cdot)$ and $\sigma^2(\cdot)$ are respectively the drift function (or instantaneous mean) and the diffusion function (or instantaneous variance) of the process.

In particular, the simulation study aims to investigate the performance of those common parametric diffusion function and drift function estimators, namely the MLE, NLS (or OLS), and GMM estimators, as well as the nonparametric diffusion function and drift function estimators proposed in Jiang and Knight (1996). It is noted that all these estimators are developed when only discretely sampled data of the continuous-time diffusion process are available. The continuous record of observation of the process between the sampling points is unobservable. Therefore the data generating process (DGP) in our simulation study requires that the explicit transition density functions of the diffusion processes are known, in order that the realizations of the process can be observed at discrete time along the exact continuous sampling path.

For this reason, our simulation study is based on those diffusion processes which have explicit transition densities, i.e., the Brownian motion with drift process, the Ornstein-Uhlenbeck process, and the Cox-Ingersoll-Ross squared-root process. The simulation study provides us with some surprising yet interesting results, namely that, with a large sample over a short sampling period, even though all the parametric diffusion function estimators perform very well, the parametric drift function estimators can perform very poorly. However, both the nonparametric diffusion function estimator and drift function estimator proposed in Jiang and Knight (1996) perform well.

The paper is organized as follows. Section 2 reviews the common parametric estimators and the nonparametric diffusion function and drift function estimators proposed for the Itô diffusion process defined in (**??**) when only observations of the process at discrete time are available. Section 3 details the transition density functions and the applicable common parametric diffusion function and drift function estimators as well as the nonparametric diffusion function and drift function estimators for each of the diffusion processes on which our simulation study is based, namely the Brownian motion with drift process, the Ornstein-Uhlenbeck process, and the Cox-Ingersoll-Ross squared-root process. Section 4 outlines the experimental design of the Monte Carlo simulation and analyses the results. A brief conclusion is contained in Section 5.

2. Estimation of the Diffusion Process from Discretely Sampled Data

Estimation of the Itô diffusion process or stochastic differential equation (SDE) defined in (**??**) has been considered in the literature for many years, with most of the papers being concerned with estimating the drift and diffusion functions from continuously sampled data. Unfortunately, in practice, more often than not we can only obtain data in discrete time since the dynamics of the process can be much faster than the sampling rate. With discretely sampled observations from the continuous sampling path, identification and estimation of the continuous-time Itô diffusion process proves to be much more complicated and difficult. The first parametric estimator of the coefficients of a stationary diffusion process from discretely sampled observations is the ML estimator proposed by Dacunha-Castelle and Florens-Zmirou (1986). Other parametric estimators include the ML estimators derived by Lo (1988) for more general jump-diffusion processes, the method of moments based on simulated sampling paths from given parameter values proposed by Duffie and Singleton (1993), the purely theoretic approximate maximum likelihood (AML) estimator proposed by Pedersen (1995), as well as the nonlinear least square (NLS) or ordinary least square (OLS), or most commonly Hansen's (1982) generalized method of moments (GMM) based on the "discretized" model. Recently, Hansen and Scheinkman (1995) also derived moment conditions based on the infinitesimal generator. The first nonparametric diffusion function estimator is proposed by Florens-Zmirou (1993) which imposes no restriction on either the drift term or diffusion term, but her procedure leaves the drift term unidentified and the diffusion function estimator can not be used for the construction of the drift function estimator. Stanton (1996) develops approximations to the true drift and diffusion functions and estimates these approximations nonparametrically. Aït-Sahalia (1996) proposes a nonparametric diffusion function estimator based on the linear mean-reverting drift function for the strictly stationary diffusion processes. We will notice later in this paper that the parameter estimators of the linear mean reverting drift function are not robust in that they are extremely sensitive to the sampling path (and/or the discrete observations along the sampling path) and consequently perform very poorly even with a large sample over a short sampling period.

Diffusion processes as defined in (**??**) are widely used in the finance literature to model the dynamics of certain financial variables, e.g., the stock prices, the exchange rates, and the term structure of interest rates [1]. Due to the estimation problem, however, all the diffusion models in the finance literature have to rely on parametric or semi-parametric specifications for the drift and diffusion functions in order to implement available estimation methods based on discretely observed data. The diffusion function is usually specified as a power function of the stochastic process, i.e., $\sigma(X_t) = \sigma X_t^{\gamma}$ ($\gamma = 0$ for Merton (1973) and Vasicek (1977); $\gamma = 1/2$ for CIR (1985); $\gamma = 1$ for Dothan (1978) and Brennan and Schwartz (1977, 1979, 1980); $\gamma = 3/2$ for CIR (1980)). The drift function is typically specified as either a constant $\mu(X_t) = \mu$ (as in Merton (1973), Dothan (1978), and CIR (1980)) or a linear mean reverting function $\mu(X_t) = \beta(\alpha - X_t)$, $\beta > 0$ (as in Vasicek (1977), CIR (1985), Brennan and Schwartz (1977, 1979, 1980), and Aït-Sahalia (1996)). Such specifications allow estimation of the parameters via the use of common parametric estimators, such as MLE, NLS (or OLS), or GMM. The discussion and empirical

---

1 See Chan, et al (1992) for a review of various parametric spot interest rate models.

results in Jiang and Knight (1996), however, show that both parametric and semi-parametric specifications impose very strong and unrealistic assumptions on the underlying process of the model.

In Jiang and Knight (1996), a nonparametric identification and estimation procedure, based on discretely sampled observations, for a wide range of Itô diffusion processes is proposed. Under mild regularity conditions, a nonparametric kernel estimator for the diffusion function $\sigma^2(X_t)$ of the general diffusion process defined in (**??**) is proposed as:

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n-1} K(\frac{X_{i\Delta_n} - x}{h_n})[X_{(i+1)\Delta_n} - X_{i\Delta_n}]^2}{\sum_{i=1}^{n} \Delta_n K(\frac{X_{i\Delta_n} - x}{h_n})} \tag{2}$$

based on observing X_t at $\{t = t_1, t_2, ..., t_n\}$ in the time interval [0, T], with $T \geq T_0 > 0$ where T_0 is a positive constant, $\{X_t = X_{\Delta_n}, X_{2\Delta_n}, ..., X_{n\Delta_n}\}$ are n equispaced observations at $\{t_1 = \Delta_n, t_2 = 2\Delta_n, ..., t_n = n\Delta_n\}$ with $\Delta_n = T/n$, and $K(\cdot)$ is a kernel density function satisfying regularity conditions. The nonparametric diffusion function estimator is developed without imposing any functional form restrictions on either the drift term or diffusion term with the drift term $\mu(\cdot)$ being a nuisance parameter and restriction-free. Thus the nonparametric diffusion function estimator captures the true volatility of the process. The variance of $\hat{\sigma}^2(x)$ can be consistently estimated by $\hat{V}[\hat{\sigma}^2(x)] = \hat{\sigma}^4(x)/\sum_{i=1}^{n} K(\frac{X_{i\Delta_n} - x}{h_n})$. Under a further condition that there exists a limiting probability density function for the process or that the process is stationary in the strict sense, based on the equation derived from the Kolmogorov forward equation, a consistent nonparametric drift function estimator is proposed as:

$$\hat{\mu}(x) = \frac{1}{2}[\frac{d\hat{\sigma}^2(x)}{dx} + \hat{\sigma}^2(x)\frac{\hat{p}'(x)}{\hat{p}(x)}] \tag{3}$$

where $\hat{\sigma}^2(x)$ is the nonparametric diffusion function estimator in (2), $\hat{p}(x)$ is a consistent kernel estimator of the marginal density function of the process. This is the first nonparametric estimator proposed in the literature for the drift function based on discretely sampled observations. The variance of $\hat{\mu}(x)$ can be obtained using the δ-method conditional on either $\hat{p}(x)$ or $\hat{\sigma}^2(x)$, or otherwise unconditionally if the covariance of $\hat{p}(x)$ and $\hat{\sigma}^2(x)$ is known. In practice, a bootstrapping technique could be applied to derive the standard error of $\hat{\mu}(x)$.

Thus the general Itô diffusion process defined in (**??**) can be identified nonparametrically for both the drift and diffusion functions based on discretely observed data. The fact that the identification and estimation of the drift function requires stronger conditions than the diffusion function is similar to the so-called “aliasing problem" for a system of linear stochastic differential equations (SDE), as discussed in Phillips (1973) and Hansen and Sargent (1983). Phillips (1973) points out that, unless there are sufficient *a priori* restrictions on the parameters of a system of linear stochastic differential equations, we cannot distinguish between structures generating cycles whose frequencies differ by integer multiples of the reciprocal of the observation period. Similarly, it is impossible to identify a nonlinear diffusion process as defined in (**??**) without imposing any structural restrictions on the model. Especially, the drift term of the diffusion process (univariate or multivariate) cannot be directly identified on a fixed time interval, no matter how frequently the observations are sampled, as the Cameron-Martin-Girsanov transformation (see e.g. Øksendall, 1992) can always be applied to give an otherwise unnoticeable change in the drift. The above proposed estimator is based on the estimated diffusion function and marginal density function and exploits the stationarity property of the process.

As some authors (see e.g. Merton, 1980) have already observed, even though the diffusion term of a stochastic process can be estimated very precisely when the sampling interval is small, the estimates of the drift term tend to have low precision. Our findings in this paper not only confirm this observation but also further reveal that the parametric estimates of the common drift function specifications can perform very poorly even with large samples of data, no matter how frequently the observations are sampled over a short sampling period. The following simple example can help to illustrate the problem. Suppose that the log return of a stock price follows a Brownian motion with drift process, i.e., $dlnX_t = \mu dt + \sigma dW_t$, where μ and σ are constants. The ML estimator of μ from equispaced discretely sampled observations $\{X_{t_1=0}, X_{t_2}, \cdot, \cdot, \cdot, X_{t_n=T}\}$, with $\Delta = X_{t_i} - X_{t_{i-1}}$, is the average of log-returns, i.e., $\hat{\mu} = (1/T)\sum_{i=1}^{n} ln(X_{t_i}/X_{t_{i-1}})$, or $\hat{\mu} = (lnX_T - lnX_0)/T$. It is easy to verify that $\hat{\mu}$ is a consistent estimator of μ as $\hat{\mu}|X_0 \sim N(\mu, \sigma^2/T)$. However, it is also very easy to see that, for any finite sample of observations, $\hat{\mu}$ is very sensitive to the first and last observations of the sample and is actually determined only by these two values. Thus, if we have a sample of, say, 5,000 observations, it is only the first and last observations that matter for the estimate of μ. Moreover, $\hat{\mu}$ has no efficiency gains even if we increase the sample size by reducing the sampling interval over fixed T. Thus, it is not hard to see that the estimate of μ will not be robust in that it will be very sensitive to the sampling path and/or the discrete observations along the sampling path. On the other hand, when the sampling interval is small, the ML diffusion function estimator, $\hat{\sigma}^2 = \sum_{i=1}^{n}(ln(X_{t_i}/X_{t_{i-1}}) - \hat{\mu}\Delta)^2/T$, performs very well, regardless of the poor performance of the drift function estimator as $E[\hat{\sigma}^2|\hat{\mu}] = \sigma^2 + (\mu - \hat{\mu})^2\Delta$ and $Var[\hat{\sigma}^2|\hat{\mu}] = \frac{2\sigma^2\Delta}{T}[(\mu - \hat{\mu})^2\Delta + \sigma^2]$.

Our simulation analysis will focus on both the diffusion function estimator and the drift function estimator. Of course, the nonparametric diffusion function estimator requires only mild regularity conditions (i.e., A1-A6 in Jiang and Knight, 1996), while the nonparametric drift function estimator requires stronger conditions (i.e., A1-A8 in Jiang and Knight, 1996), i.e., the stochastic process must be at least asymptotically stationary in the strict sense. This excludes the application of the proposed nonparametric drift function estimator to such processes as Brownian motion with drift and geometric Brownian motion. It is noted that, in the finance literature, drift function estimation has received much less attention than the diffusion function estimation. One reason is that the diffusion function, as the second moment and the measurement of instantaneous volatility of the stochastic process, is of more interest in modeling the movements of interest rates, asset prices, or exchange rates. For instance, the volatility of the riskless interest rate is one of the key determinants of the value of contingent claims and one of the key factors determining optimal portfolio hedging strategies for risk-averse investors. Therefore, to predict the movements of derivative security prices, to hedge an investment portfolio, or to create a certain leverage within a portfolio, the volatility of the prices of underlying assets is the major factor to be considered. Another and maybe more direct reason is that, in the famous Black-

Scholes option pricing formula, the prices of derivative securities are affected by the price of underlying assets only through its instantaneous volatility, i.e. the diffusion function. The drift function does not appear in the option pricing formula at all due to an assumption that, in the economy, there exists a risk-free asset with nonstochastic rate of return. However, as Lo and Wang (1995) point out, predictability of an asset's return is typically induced by the drift and will affect the prices of options on that asset, even though the drift term does not enter the option pricing formula. Moreover, in models with stochastic spot interest rates, both the diffusion function and drift function will enter the derivative security pricing formulation. Therefore the prices of derivative securities in these cases are explicitly affected by the price of underlying assets through not only the diffusion function but also drift function. From this point of view, the drift function estimation is as important as the diffusion function estimation.

3. Diffusion Processes: Explicit Transition Density Functions and Common Estimators

As we have mentioned, our aim is to investigate the finite sample properties and the performance of common parametric estimators and the nonparametric diffusion function and drift function estimators proposed for the continuous-time Itô diffusion process when only discretely sampled observations of the continuous sampling path over a short time period are available. Therefore the data generating process (DGP) for the sampling path in our simulation has to be continuous in time, while realizations along the continuous sampling path are observed only at discrete time. For this reason, we have to focus our simulation study on models which have explicit transition density functions, namely the Brownian motion with drift process, the Ornstein-Uhlenbeck process, and the Cox-Ingersoll-Ross squared-root process. For stationary diffusion processes, the functional forms of the transition density functions corresponding to specifications which are essentially different from the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross squared-root process are not known explicitly. As Wong (1964) shows, one can only construct a stationary continuous-time Markov process with known explicit transition density function from a linear functional specification for the drift function $\mu(\cdot)$ and a quadratic function specification for the diffusion function $\sigma^2(\cdot)$. In this section, we will detail the transition density function and the applicable common parametric and nonparametric diffusion function and drift function estimators for each of the above processes.

(a) The Brownian Motion with Drift Process: The Brownian motion with drift process, $dX_t = \mu dt + \sigma dW_t$ where both μ and σ are constants, has a normal transition density function given by $f(X_t = x, t; X_{t_0} = x_0, t_0) = \frac{1}{\sqrt{2\pi\sigma^2(t-t_0)}} exp\{-\frac{(x-x_0-\mu(t-t_0))^2}{2\sigma^2(t-t_0)}\}$, with its marginal density function varying over time. Since the geometric Brownian motion process, $dX_t = \mu X_t dt + \sigma X_t dW_t$, implies that $Y_t = ln X_t$ follows a Brownian motion with drift process, i.e., , $dY_t = (\mu - \sigma^2/2)dt + \sigma dW_t$, we do not consider separately the geometric Brownian motion. Both the Brownian motion with drift process and the geometric Brownian motion process are nonstationary.

Since the Brownian motion with drift process is neither strictly stationary nor has a limiting probability density function, the aforementioned nonparametric drift function estimator cannot be applied. Thus the comparison of the performance between parametric and nonparametric estimators will have to be constrained only to the diffusion function estimators. The maximum likelihood estimator (MLE) can of course be used to estimate both the drift and diffusion for the Brownian motion with drift process as the transition density function and hence the joint probability density function has an explicit form. It can be shown that the OLS estimators of μ and σ^2 based on the conditional mean and variance conditions (or the first and second moment conditions) are identical to their MLE counterparts. However, GMM estimation is not applicable to the Brownian motion with drift process due to its nonstationarity.

(b) The Ornstein-Uhlenbeck Process: The Ornstein-Uhlenbeck process, $dX_t = \beta(\alpha - X_t)dt + \sigma dW_t$ where α, β, and σ are constants, also has a normal transition density function given by $f(X_t = x, t; X_{t_0} = x_0, t_0) = \frac{1}{\sqrt{2\pi s^2(t)}} exp\{-\frac{(x-\alpha-(x_0-\alpha)e^{-\beta(t-t_0)})^2}{2s^2(t)}\}$, where $s^2(t) = \frac{\sigma^2}{2\beta}[1 - e^{-2\beta(t-t_0)}]$. If the process does display the property of mean reversion ($\beta > 0$), then as $t_0 \to -\infty$ or $t - t_0 \to +\infty$, the marginal density of the stochastic process is invariant to time, i.e., the Ornstein-Uhlenbeck process is stationary in the strict sense in the steady state.

The Ornstein-Uhlenbeck process has a limiting probability density function, thus both the nonparametric diffusion function estimator and drift function estimator can be applied. Since the transition density function and hence the joint probability density function of the Ornstein-Uhlenbeck process also has an explicit functional form, the maximum likelihood estimation (MLE) can also be used. Further, as an asymptotic stationary process, the parameters of Ornstein-Uhlenbeck process can also be estimated using GMM, based on the exact conditional moment conditions. That is, the GMM estimates of α, β and σ^2 for the Ornstein-Uhlenbeck process can be obtained based on the following four exact conditional and unconditional moments:

$$G_n(\alpha, \beta, \sigma^2) = \frac{1}{n-1} \sum_{i=1}^{n-1} F_i(\alpha, \beta, \sigma^2) \tag{4}$$

with

$$F_i(\alpha, \beta, \sigma^2) = \begin{bmatrix} \epsilon_{i+1} \\ \epsilon_{i+1} X_{i\Delta_n} \\ \epsilon_{i+1}^2 - E[\epsilon_{i+1}^2 | X_{i\Delta_n}] \\ (\epsilon_{i+1}^2 - E[\epsilon_{i+1}^2 | X_{i\Delta_n}]) X_{i\Delta_n} \end{bmatrix}$$

where $\epsilon_{i+1} = (X_{(i+1)\Delta_n} - X_{i\Delta_n}) - E[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) | X_{i\Delta_n}]$ and

$$E[(X_{(i+1)\Delta_n} - X_{i\Delta_n}) | X_{i\Delta_n}] = (1 - e^{-\beta\Delta_n})(\alpha - X_{i\Delta_n}) \tag{5}$$

$$E[\epsilon_{i+1}^2 | X_{i\Delta_n}] = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\Delta_n}) \tag{6}$$

where Δ_n is the sampling interval, and the exact conditional variance of the changes of X_t over time interval of length Δ_n is given by $E[\epsilon_{i+1}^2 | X_{i\Delta_n}] = V[X_{(i+1)\Delta_n} | X_{i\Delta_n}]$ [2]. These moment conditions correspond to transitions of length Δ_n and are not subject to discretization bias [3]. Since these GMM systems are overidentified, we weighted the

---

2 To obtain the conditional mean and variance of the diffusion process, one can solve for the transition density functions from the Kolmogorov backward equation $\partial f(X_t, t; X_{t_0}, t_0)/\partial t = \mu(X_t, \theta)\partial f(X_t, t; X_{t_0}, t_0)/\partial X_t + \frac{1}{2}\sigma^2(X_t, \sigma^2)\partial^2 f(X_t, t; X_{t_0}, t_0)/\partial X_t^2$, and then calculate the exact conditional mean and variance.

3 It is noted that in most financial economics literature, using GMM to estimate the parameters of the diffusion processes consists in first discretizing the continuous-time model, then based on the discrete-time model deriving the moment conditions. The GMM approach in this case no longer requires that the distribution of the changes of X_t be normal. Actually the discrete-time model specifies that the instantaneous variance of the residual is proportional to the length of sampling interval, i.e., $E[\epsilon_{i+1}^2] = \sigma^2(X_{t\Delta_n})\Delta_n$. Therefore the asymptotic justification for the GMM procedure requires only that the

criterion optimally (see Hansen (1982)). The positive-definite symmetric weighting matrix is chosen such that the GMM estimator of $(\alpha, \beta, \sigma^2)$ has the smallest asymptotic covariance matrix. With the above conditional mean and variance conditions and the property of independent increments of the process, the nonlinear least squares (NLS) estimators of $(\alpha, \beta, \sigma^2)$ can also be obtained.

(c) The Cox-Ingersoll-Ross Squared-Root Process: The Cox-Ingersoll-Ross (CIR) squared-root (SR) process, $dX_t = \beta(\alpha - X_t)dt + \sigma X_t^{1/2} dW_t$ where α, β, and σ are constants, has the transition density function given by $f(X_t = x, t; X_{t_0} = x_0, t_0) = ce^{-u-v}(\frac{v}{u})^{q/2} I_q(2(uv)^{1/2})$ with X_t taking nonnegative values, where $c = \frac{2\beta}{\sigma^2(1-e^{-\beta(t-t_0)})}$, $u = cx_0 e^{-\beta(t-t_0)}$, $v = cx$, $q = \frac{2\beta\alpha}{\sigma^2} - 1$, and $I_q(\cdot)$ is the modified Bessel function of the first kind of order q. The transition distribution function is a noncentral chi-square, $\chi^2[2cx; 2q+2, 2u]$, with $2q+2$ degrees of freedom and parameter of noncentrality $2u$ proportional to the current level of the stochastic process. The conditional expected value and variance of X_t is given by $E[X_t | X_0 = x_0] = x_0 e^{-\beta(t-t_0)} + \alpha(1 -$

---

distribution of interest rate changes be stationary and ergodic and that the relevant expectations exist. The moment conditions used in the literature are as follows:

$$F_i(\alpha, \beta, \sigma^2) = \begin{bmatrix} \epsilon_{i+1} \\ \epsilon_{i+1} X_{i\Delta_n} \\ \epsilon_{i+1}^2 - \sigma^2 \Delta_n \\ (\epsilon_{i+1}^2 - \sigma^2 \Delta_n) X_{i\Delta_n} \end{bmatrix}$$

for the Vasicek model, and

$$F_i(\alpha, \beta, \sigma^2) = \begin{bmatrix} \epsilon_{i+1} \\ \epsilon_{i+1} X_{i\Delta_n} \\ \epsilon_{i+1}^2 - \sigma^2 X_{i\Delta_n} \Delta_n \\ (\epsilon_{i+1}^2 - \sigma^2 X_{i\Delta_n} \Delta_n) X_{i\Delta_n} \end{bmatrix}$$

for the Cox-Ingersoll-Ross squared-root model, with $\epsilon_{i+1} = X_{(i+1)\Delta_n} - X_{i\Delta_n} - \beta(\alpha - X_{i\Delta_n})\Delta_n$ where $\Delta_n = t_{i+1} - t_i = T/N$ due to equal sampling interval. It is clear that these moment conditions are different from those derived from the continuous-time model. The misspecification and inconsistency caused by "discretization" is discussed in Jiang and Knight (1996).

$e^{-\beta(t-t_0)})$, $Var[X_t|X_0 = x_0] = x_0(\frac{\sigma^2}{\beta})(e^{-\beta(t-t_0)} - e^{-2\beta(t-t_0)}) + \alpha(\frac{\sigma^2}{2\beta})(1 - e^{-\beta(t-t_0)})^2$. If the process displays the property of mean reversion ($\beta > 0$), then as $t_0 \to -\infty$ or $t - t_0 \to +\infty$, its marginal density function will approach a gamma probability density function, i.e. $f(X_t = x, t) = \frac{\omega^\nu}{\Gamma(\nu)} x^{\nu-1} e^{-\omega x}$ where $\omega = 2\beta/\sigma^2$ and $\nu = 2\alpha\beta/\sigma^2$, with mean α and variance $\sigma^2\alpha/2\beta$. That is, the Cox-Ingersoll-Ross squared-root process is also stationary in the steady state.

As the Cox-Ingersoll-Ross squared-root process has a limiting probability density function, both the nonparametric diffusion function and drift function estimators can be applied. However, the ML estimator is not performed for the Cox-Ingersoll-Ross squared-root process due to the complexity of the Bessel function [4]. Similar to the Ornstein-Uhlenbeck process, the Cox-Ingersoll-Ross squared-root process as an asymptotic stationary process can also be estimated using GMM. The GMM estimates of α, β and σ^2 can be obtained from the exact conditional and unconditional moment conditions based on its conditional mean and variance, i.e.,

$$E[(X_{(i+1)\Delta_n} - X_{i\Delta_n})|X_{i\Delta_n}] = (1 - e^{-\beta\Delta_n})(\alpha - X_{i\Delta_n}) \tag{7}$$

$$E[\epsilon_{i+1}^2|X_{i\Delta_n}] = \frac{\sigma^2}{\beta}(e^{-\beta\Delta_n} - e^{-2\beta\Delta_n})X_{i\Delta_n} + \alpha(\frac{\sigma^2}{2\beta})(1 - e^{-\beta\Delta_n})^2 \tag{8}$$

The estimation procedure is exactly the same as for the Ornstein-Uhlenbeck process. For the same reason, the NLS estimators of $(\alpha, \beta, \sigma^2)$ can also be obtained based on

---

4 The numerical optimization procedure, such as subroutine E04UCF of the NAG library could be used to perform the ML estimation of the CIR process. However, it involves the evaluation of the modified Bessel function of the first kind of order q, $I_q(x)$, over different intervals. When both x and q are small numbers, $I_q(x)$ has to be evaluated using, e.g. the backward recursion in Section 19.4.2 of Luke (1977). For large values of x and/or q, the asymptotic expansion, e.g. in Olver (1965) has to be used. Therefore the computation would be too intensive for a Monte Carlo study.

the conditional mean and variance conditions.

Insert Table 1 around here

Table 1 summaries the stationarity, in the asymptotic sense, of each process and identifies the common parametric and nonparametric estimators of the diffusion and drift functions for each process. The Monte Carlo comparison will be based on the simulation results of these estimators.

4. Monte Carlo Study: Parametric versus Nonparametric Estimators

The aim of the Monte Carlo study is twofold. Firstly, to examine the finite sample properties of the nonparametric diffusion function and drift function estimators developed in Jiang and Knight (1996). Secondly, to undertake a detailed comparison of the nonparametric estimator with common parametric estimators. As the nonparametric diffusion function and drift function estimators are both functions of X_t, the Monte Carlo analysis is based on a sample of estimates for each parametric and nonparametric estimator at given values of X_t. In each replication, one set of discrete observations along the continuous sampling path of a known diffusion process is generated and based on this sample different estimators are applied. With the sample of estimates for each estimator, we investigate its finite sample properties and compare their performance based on their respective sampling distributions. The number of replications for each estimator is set to be 1,000 and/or 5,000 and the sample size of observations in each replication is 5,000 and/or 10,000. The comparison of the performance of different estimators is first undertaken at one single point of X_t (i.e, $X_t = 0.07$), and then extended to different points in an interval of X_t (i.e., $X_t \in [0.05, 0.10]$ or 5% to 10% of interest rate levels).

The data generating process (DGP) of each diffusion process is given by its transition probability density function and based on its Markovian property, the dynamics of the continuous sampling path is explicitly known. The discrete sampling observations along the continuous sampling path are observed over equispaced intervals with sampling interval Δ_N. In all the simulations, the discrete observations of sample size N from the sampling path are recorded over a time period from $t = -500\Delta_N$ to $T = N\Delta_N$ with sampling interval $\Delta_N = T/N$. We discard the first 500 observations to eliminate any start-up effects.

The values for the parameters of different processes are set to be approximately equal to those of the corresponding interest rate models estimated in Chan, et al (1992) using the American monthly Treasury bill yield data from June 1964 to December 1989, i.e.,

(a) $\mu = 0.0055, \sigma^2 = 0.0004$ for the Brownian motion with drift process;

(b) $\alpha = 0.086, \beta = 0.18, \sigma^2 = 0.0004$ for the Ornstein-Uhlenbeck process; and

(c) $\alpha = 0.076, \beta = 0.23, \sigma^2 = 0.007$ for the Cox-Ingersoll-Ross squared-root process.

The models are, consequently, close to the true term structure of interest rate models. The starting value of the DGP is set to be equal to the average interest rate level of the data set in Chan, et al (1992), i.e. 0.067.

As the relative performance of the parametric estimators and the nonparametric estimators are similar for different sample sizes and different number of replications, the simulation results are reported for only one sample size (i.e. 5,000), one replication number (i.e. 1,000), and also one window-width for the nonparametric estimators. The study is obviously not comprehensive. As we will mention later, our choice of window-width is based on the numerical criteria that the integrated mean squared error (IMSE) is minimized, the study of the optimal choice of window-width is not pursued in this paper [5].

Figures 1 and 2 plot the sample means of different diffusion function and drift function estimates at different X_t for each process, which gives a clear visual impression of how each estimator performs. It is clear that while the parametric diffusion function estimators perform very well for all processes, the parametric drift function estimators all perform poorly. However, both the nonparametric diffusion function estimator and

---

5 In the nonparametric estimation procedure, both the kernel diffusion function estimator and the kernel marginal density function estimator and its first derivative, which are used in estimating the drift function, involve the choice of kernel functions and optimal window-width. The regularity conditions of the kernel function of order r for both diffusion function and marginal density estimation are as follows:
(i) The kernel $K(\cdot)$ is symmetric about zero, continuously differentiable to order r on R, belongs to $L^2(R)$, and $\int_{-\infty}^{+\infty} K(x)dx = 1$;
(ii) $K(\cdot)$ is of order r: $\int_{-\infty}^{+\infty} x^i K(x)dx = 0$, $i = 1, ..., r-1$, and $\int_{-\infty}^{+\infty} x^r K(x)dx \neq 0$, $\int_{-\infty}^{+\infty} |x|^r |K(x)| dx < \infty$.
The regularity conditions for the admissible window-width are as follows: as the sample size $n \to \infty$, and the sampling interval $\Delta_n \to 0$,
(i) $h_n \to 0, nh_n \to \infty$, and $nh_n^{r+1} \to 0$ for the diffusion function estimation;
(ii) $h_n \to 0, nh_n \to \infty$, and $nh_n^{2r+1} \to 0$ for the marginal density function estimation; and
(iii) $h_n \to 0, nh_n^3 \to \infty$, and $nh_n^{2r+1} \to 0$ for the first derivative of the marginal density function estimation.
The above conditions ensure that for all cases, the bias in the estimator is asymptotically negligible and at the same time the variance of the estimator goes to zero as sample size increases to infinite. The conditions are stronger than usual nonparametric estimation due to the correlation among the data and the requirement for estimating the derivative of marginal density function.

drift function estimator perform reasonably well.

Tables 2, 3 and 4 report the summary statistics of the sampling distributions for each estimator of the diffusion function, including both nonparametric and parametric estimators, for each process. For all three processes, the nonparametric as well as the parametric diffusion function estimators perform extremely well, as measured by the sample median, mean, variance, and mean squared error (MSE). The plots of the empirical cumulative density function (ECDF) based on the sample of diffusion function estimates for each estimator in Figures 3, 4 and 5 further show the characteristics of the sampling distributions for the estimators of σ^2, at a given value of X_t ($X_t = 0.07$), for all three processes. The sampling distributions are all highly concentrated around the true value of σ^2. The nonparametric estimates of σ^2 is obtained from $\hat{\sigma}^2 = \hat{\sigma}^2(X_t)$ for the Brownian motion with drift process and the Ornstein-Uhlenbeck process and $\hat{\sigma}^2 = \hat{\sigma}^2(X_t)/X_t$ for the Cox-Ingersoll-Ross squared-root process. The empirical cumulative density function is simply based on the sample of estimates of σ^2 from the 1,000 replications, i.e., $ECDF(x) = \frac{\sum_{i=1}^{M} 1(\hat{\sigma}_i^2 \leq x)}{M}$, where $x \in (-\infty, +\infty)$, $M = 1,000$ and $\hat{\sigma}_i^2$ is the estimate of σ^2 in the ith replication, $i = 1, 2, ..., M$. Visually there is not much of a difference between the sampling distributions of the nonparametric diffusion function estimator and other parametric estimators for all three processes.

Tables 5, 6 and 7 report the summary statistics of the sampling distribution for the parametric estimator of μ for the Brownian motion with drift process, the parametric estimators of (α, β) for the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross squared-root process, as well as the parametric and nonparametric estimator of $\mu(X_t)$, at given value of X_t ($X_t = 0.07$), for the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross process. It is clear that, for all three processes, the parametric estimators of the drift function parameters all perform poorly, especially estimators of β, with its estimates spreading over a wide range of intervals as indicated by the maximum and minimum values of the samples of the estimates. However, the nonparametric drift function estimator for both the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross squared-root process performs reasonably well, measured by the sample median, mean, variance, and mean squared error (MSE). Similarly, the plots of the empirical cumulative distribution function (ECDF) based on the sample of drift function estimates in Figures 6 and 7 show the sampling distributions of the drift function parameter estimators for each process as well as the parametric and nonparametric drift function estimator, at a given value of X_t ($X_t = 0.07$), for the Ornstein-Uhlenbeck process and Cox-Ingersoll-Ross squared-root process. The sampling distributions of the parameter estimators of μ, α, and β overlap a wide range of support. In contrast, the sampling distribution of the nonparametric estimator of $\mu(X_t)$, at $X_t = 0.07$, is highly concentrated around its true value. There is clearly a big difference between the sampling distributions of the nonparametric drift function estimator and the parametric drift function estimators.

Insert Tables 2-7 and Figures 1-10 around here

In the aforementioned comparison we noted that the performance of the nonparametric diffusion function and drift function estimators were based on fixing values of X_t, i.e., $X_t = 0.07$. All the simulation results and inferences thus could be interpreted as conditional on $X_t = 0.07$. We extended the Monte Carlo simulation study of the nonparametric estimators from one single point ($X_t = 0.07$) to an interval of X_t ($X_t \in [0.05, 0.10]$). These simulation results of the nonparametric diffusion function and drift function estimators are reported in Tables 2-7 as well. Tables 2-7 report the summary statistics of the sampling distributions of the nonparametric diffusion function and drift function estimators at different values of X_t. The results show that the nonparametric diffusion function and drift function estimators perform reasonably well at different values of the whole interval of $X_t (X_t \in [0.05, 0.1]$, or 5% to 10% of interest rate level). Figures 8, 9 and 10 plot the empirical cumulative distribution functions (ECDF) of the nonparametric and parametric diffusion function and drift function estimators at different values of X_t, the sampling distributions further show that the performance of the nonparametric estimators over the interval is quite reasonable.

Some further analysis of the simulation results of the diffusion function and drift function estimators are also reported in Tables 2-7. For instance, for the Cox-Ingersoll-Ross squared-root process, when compared to the normal density, the sampling distribution of the nonlinear least square (NLS) estimator appears slightly skewed to the left, while that of the generalized method of moments (GMM) estimator and the nonparametric estimator, at different values of X_t, appear slightly skewed to the right. The sampling distribution of the generalized method of moments (GMM) estimator appears to be slightly more concentrated, and that of the nonlinear least square (NLS) estimator and the nonparametric appears to be less concentrated, while that of the nonparametric estimator shows no consistent sign over different values of X_t. The Wilcoxon matched-pairs signed-ranks test is employed here to analyze pair-wise the differences of the absolute bias between different estimators based on the sample of estimates. The Wilcoxon matched-pairs signed-rank test is employed for its robustness against the violation of the normality assumption. This is basically a test of H_0: the median of the population of the differences between two random variables is zero, against either H_1: the median of the population of the differences between two random variables is positive (or non-negative) or negative (or non-positive). The indicator in the brackets beside the statistics denotes whether the null hypothesis is not rejected (+) or rejected (–) at the 5 % significance level.

An extension of the multi-sample median test is also conducted for the samples of the nonparametric estimates at different points. The null hypothesis of the test is H_0: all n populations have the same median, against its alternative H_1: at least one population has a median different from the others. The results of this test are reported in Tables 2-4 as well, which indicate that the null hypothesis is not rejected for all three processes at 5 % significance level.

The following are a few remarks on the above Monte Carlo simulation study:

Remark 1. The fact that the nonparametric diffusion function estimator performs as well as the parametric diffusion function estimators and the nonparametric drift function estimator outperforms all the parametric drift function estimators, including the maximum likelihood estimator (MLE), might seem surprising. First of all, it should be noted that the Monte Carlo simulation study is designed to investigate only the finite sample properties of the estimators and by no means explores the asymptotic properties of the estimators. Therefore all the simulation results and analysis are only valid for the finite samples. Secondly, the issue which we are dealing with here is more or less an identification problem rather than an estimation problem. Poor performance of the parametric drift function estimators simply imply that the drift term of a diffusion process cannot be *directly* identified from the discretely sampled data over a short sampling period, no matter how large the sample. Hence any attempt to estimate the drift function parameters based on such a sample, without using extra information, is doomed to fail. Comparing carefully the ML, NLS, and GMM estimators with the nonparametric estimators, we can see that the ML, NLS, and GMM estimators employ only the information contained in the transition density functions of the diffusion process, while the nonparametric estimator employs the information contained in both the transition density function and the marginal density function. It is through the marginal density function that we establish the relationship between $\mu(X_t)$ and $\sigma(X_t)$ given in equation (3), and based on this there is a unique drift function corresponding to a given diffusion function and marginal density function. Further, from an estimation point of view, since the kernel density function estimator performs very well for a large sample of observations, the performance of the nonparametric drift function estimator is thus mainly determined by the nonparametric diffusion function estimator. In other words, if the nonparametric diffusion function estimator performs well, then the nonparametric drift function estimator will also perform well. However, in the case of parametric drift function and diffusion function estimators, since all the moment conditions are explicitly expressed in terms of time t and the level of the diffusion process X_t, there is no way to incorporate the information contained in the limiting steady state of the process into the estimation. As a result, as our experiments indicate, the drift function parameter estimates are not robust in that they are extremely sensitive to the sampling path of the diffusion process and/or the discrete observations along the sampling path. Therefore, even though all the parametric diffusion function estimators perform very well, the parametric drift function estimators can perform very poorly.

Remark 2. Since the parametric estimators of the linear mean-reverting drift function perform very poorly with finite samples, the semiparametric identification and estimation approach proposed by Aït-Sahalia (1996) is not specifically included in our Monte Carlo simulation analysis as, in his approach, the diffusion function is estimated using the estimates of the linear mean-reverting drift function parameters.

Remark 3. The derivation of the moment conditions of the diffusion process based on the infinitesimal generator, as proposed by Hansen and Scheinkman (1995), is not necessary since the exact moment conditions of all the diffusion processes in our simulation study can be solved from the Kolmogorov backward equations.

Remark 4. It is worthwhile pointing out that the choice of the values of $(\alpha, \beta, \sigma^2)$ does not affect the simulation results and hence the performance of the nonparametric diffusion function and drift function estimators.

Remark 5. Choice of the Kernel: In the nonparametric estimation of both the diffusion function and the marginal density function, which is used in the estimation of the drift function, we have to choose the kernel functions. The kernel we chose for both the nonparametric diffusion function estimator and the marginal density function estimator is the standard Gaussian density, $K(x) = \frac{1}{\sqrt{2\pi}} exp\{-\frac{x^2}{2}\}$, which is continuously differentiable of any order.

Remark 6. Choice of the Smoothing Parameter for the Kernel Marginal Density Function Estimator: the actual window-width or smoothing parameter for the kernel marginal density function estimator and its first derivative is set as $h_n = c_n n^{-1/5}$ where $c_n = c/ln(n)$ and c is chosen to minimize the integrated mean squared error (IMSE) of the estimator [6].

Remark 7. Choice of the Smoothing Parameter for the Kernel Diffusion Function Estimator: in order to achieve convergence in distribution and consistency of the nonparametric diffusion function estimator $\hat{\sigma}^2(x)$, the window-width or smoothing parameter h_n is required to converge to zero faster than in the case of nonparametric density estimation, that is, not only $h_n \to 0, nh_n \to \infty, nh_n^5 \to 0$ as $\Delta_n \to 0$, but also $nh_n^3 \to 0$ as $\Delta_n \to 0$. Therefore the actual window-width is chosen as $h_n = c_n n^{-1/3}$, where $c_n = c/ln(n)$. As in the case of nonparametric kernel density estimation, implementation of the nonparametric kernel diffusion function estimator also requires that we deal with the problem of selecting the window-width or smoothing parameter h_n. Our experiments show that the nonparametric kernel diffusion function estimator $\hat{\sigma}^2(\cdot)$ is sensitive to the choice of the value of h_n in that different values of h_n generate different standard deviations for the sampling distribution of

6 This window-width choice was also used in Aït-Sahalia (1996) for marginal density function estimation.

the estimator. Whether the above admissible window-width represents the achievable optimal rate of convergence is unknown to us. It is thus clear that further research is required concerning the optimal choice of the window-width h_n. This research is beyond the scope of this paper and hence not pursued in this study. However, it is noted that both the nonparametric kernel density estimator and the nonparametric kernel diffusion function estimator can be regarded as a weighted averaging scheme in which the role of the window-width is to determine the span of the sampling of points and therefore the relative weights over different points, given the kernel function. In the case of the nonparametric kernel diffusion function estimator, it is not hard to see that a wider window-width means the estimate is an average with significant weights over a larger number of points and, hence, tends to have a smaller variance. On the other hand, with wider window-width, the fluctuating movements of the diffusion function over an interval might be averaged out and the estimate tends to have increased bias. The trade-off for the choice of the value of h_n is as follows. For a larger value of h_n, the estimated diffusion function $\hat{\sigma}^2(x)$ tends to be smoother in x, therefore the estimates tend to have higher bias but smaller variance, and vice versa. Moreover, given the criteria for the bias and variance of the estimates, a lower h_n tends to be a stronger requirement than a higher h_n for the sampling density of discrete observations. Therefore the choice of window-width involves the delicate task of balancing the two components: the variance on the one hand, and the bias on the other. This trade-off leads to the choice of minimizing the integrated mean squared error (IMSE) as a natural criteria for the optimal window-width selection. Therefore, in our simulation the coefficient c of the window-width sequence is also chosen to minimize the numerical IMSE [7].

5. Conclusion

From the Monte Carlo simulation results reported in the previous section we note that, based on a large sample of discrete observations over a short sampling period, both the nonparametric diffusion function and drift function estimators proposed in Jiang and Knight (1996) perform reasonably well. However, even though the parametric estimators of the diffusion function parameters all perform very well, none of the parametric estimators of the drift function parameters performs satisfactorily. This fact further suggests that, with the same data set, the identifications for the drift function and the diffusion function are not necessarily mutually dependent and the identification of the diffusion function is less troublesome than that of the drift function. In other words, a correct identification of the diffusion function does not necessarily rely on a correct identification of the drift function. This finding justifies

---

7 Unfortunately, we normally do not have knowledge of the true underlying process that generates the data and the minimum IMSE criteria are not available if one wants to estimate the diffusion function. The various ways to get around this problem for the nonparametric functional estimation, such as the cross-validation approach, the plug-in approach, the smoothed bootstrap approach, etc. (see surveys by Delgado and Robinson (1992) and Jones, Marron, and Sheather (1996)), can used here as well. Similar research can be pursued for the automatic selection of the optimal window-width of the nonparametric diffusion function estimator. Since the nonparametric diffusion function estimator has a relatively strong requirement for the smoothness of the diffusion function $\sigma^2(X_t)$, a very small h_n is not desirable. In practice, a reference value for h_n can be determined as following from the minimizing the variance with fixed amount of bias. Since the consistent estimator of the variance of $\hat{\sigma}^2(x)$ is given by $\frac{nh_n\hat{\sigma}^4(x)}{L_1(x)}$, let α be the allowable relative error of the estimate at, say, 95% confidence level in terms of the percentage of the true diffusion $\sigma^2(x)$, we can set the width of the 95 % confidence band such that $1.96(\frac{nh_n\hat{\sigma}^4(x)}{\hat{L_1}(x)})^{1/2}+\sigma^2(x) \leq (1+\alpha)\sigma^2(x)$ or $-1.96(\frac{nh_n\hat{\sigma}^4(x)}{\hat{L_1}(x)})^{1/2}+\sigma^2(x) \geq (1-\alpha)\sigma^2(x)$, where $\hat{L_1}(x)$ is estimated by setting h_n to be the optimal window-width which minimizes the integrated mean square error of $\hat{L_1}(x)$. As $\hat{\sigma}^2(x) \to \sigma^2(x)$ in probability, therefore we have $h_n \leq \frac{\hat{L_1}(x)}{n}(\frac{\alpha}{1.96})^2$. Thus the value of h_n can be calculated from $h_n^* = sup\{h_n|h_n \leq \frac{\hat{L_1}(x)}{n}(\frac{\alpha}{1.96})^2, \forall x \in I\}$, where I represents the inference area.

the nonparametric identification and estimation procedure, proposed for the Itô diffusion process in Jiang and Knight (1996), in which the diffusion function is identified and estimated without imposing any *a priori* restrictions on either the drift term or the diffusion term. The Monte Carlo simulation results also suggest that, for both stationary and non-stationary processes, the nonparametric diffusion function estimator captures the true volatility.

References:

Aït-Sahalia, Y., 1996, Nonparametric pricing of interest rate derivative securities, *Econometrica* 64, No.3, 527-60.

Banon, G., 1978, Nonparametric identification for diffusion processes, *S.I.A.M. Journal of Control and Optimization* 16, 380-95.

Banon, G. and H.T. Nguyen, 1981, Recursive estimation in diffusion models, *S.I.A.M. Journal of Control and Optimization* 19, 676-85.

Black, F. and M. Scholes, 1973, The pricing of options and corporate liabilities, *Journal of Political Economy* 81, 637-54.

Brennan, M.J. and E.S. Schwartz, 1977, Saving bonds, retractable bonds, and callable bonds, *Journal of Financial Economics* 3, 67-88.

......, 1979, A continuous time approach to the pricing of bonds, *Journal of Banking and Finance,* 3, 133-55.

......, 1980, Analyzing convertible bonds, *Journal of Financial and Quantitative Analysis*, 15, 907-29.

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* Vol.XLVII, No.3, 1209-27.

Cox, J.C., J.E. Ingersoll and S.A. Ross, 1980, An analysis of variable rate loan contracts, *Journal of Finance,* Vol.XXXV, No.2, 389-403.

......, 1985, A theory of the term structure of interest rates, *Econometrica* 53, No.2, 385-407.

Cox, J.C. and S.A. Ross, 1976, The valuation of options for alternative stochastic processes, *Journal of Financial Economics* 3, 145-66.

Dacunha-Castelle, D. and D. Florens-Zmirou, 1986, Estimation of the coefficient of a diffusion from discrete observations, *Stochastics* 19, 263-84.

Daniel, W.W., 1978, *Applied nonparametric statistics* (Houghton Mifflin Company).

Delgado, M.A. and P.M. Robinson, 1992, Nonparametric and semiparametric methods for economic research, *Journal of Economic Surveys* Vol.6, No.3, 201-249.

Donhal, G., 1987, On estimating the diffusion coefficient, *Journal of Applied Probability* 24, 105-114.

Doob, J.L., 1953, *Stochastic processes* (John Wiley & Sons).

Dothan, L.U., 1978, On the term structure of interest rates, *Journal of Financial Economics* 6, 59-69.

Duffie,D. and K.J. Singleton, 1993, Simulated moments estimation of Markov models for asset prices, *Econometrica* 61, 929-52.

Feller, W., 1951, Two singular diffusion problems, *Annals of Mathematics* 54: 173-82.

Florens-Zmirou, D., 1993, On estimating the diffusion coefficient from discrete observations, *Journal of Applied Probability* 30, 790-804.

Geman, S.A., 1979, On a common sense estimator for the drift of a diffusion, *Pattern Analysis* #79, Division of Applied Math., Brown University, Providence, Rhode Island.

Hansen, L.P., 1982, Large sample properties of generalized method of moments estimators, *Econometrica* 50, 1029-54.

Hansen, L.P. & T.J. Sargent, 1983, The dimensionality of the aliasing problem, *Econometrica* 51, 377-88.

Hansen, L.P. and J.A. Scheinkman, 1995, Back to the future: generating moment implications for continuous-time Markov processes, *Econometrica* 63(4), 767-804.

Hendry, D.F., 1984, Monte Carlo experimentation in econometrics, in: Z. Griliches and M.D. Intriligator, eds., *Handbook of econometrics,* Volume II. (North-Holland, Amsterdam).

Ingersoll, J.E.Jr., 1987, *Theory of financial decision making* (Rowman & Littlefield).

Jiang, G.J. and J.L. Knight, 1996, A nonparametric approach to the estimation of diffusion process – with an application to a short-term interest rate model, forthcoming *Econometric Theory*.

Jones, M.C., J.S. Marron, and S.J. Sheather, 1996, A brief survey of bandwidth selection for density estimation, *Journal of the American Statistical Association* Vol.91, No.433, 401-407.

Kumar, T.K. and Markman, J.M., 1975, Estimation of probability density function: A Monte Carlo comparison of parametric and nonparametric methods, mimeo.

Lo, A.W., 1988, Maximum likelihood estimation of generalized Itô processes with discretely sampled data, *Econometric Theory* 4, 231-247.

Lo, A.W. and J. Wang, 1995, Implementing option pricing models when asset returns are predictable, *Journal of Finance* Vol. L, No.1, 87-129.

Luke, Y., 1977, *Algorithms for the computation of mathematical functions* (Academic Press, New York, NY).

Merton, R.C., 1973, Theory of rational option pricing, *Bell Journal of Economics and Management Science* 4, 141-83.

......, 1980, On estimating the expected return on the market: an exploratory investigation, *Journal of Financial Economics* 8, 323-361.

Øksendal, B., 1992, *Stochastic differential equations* (Springer-Verlag, New York).

Olver, F.W.J., 1965, Bessel functions of integer order, in: M.A. Abramovitz and I.A.Stegun, eds., *Handbook of Mathematical Functions* (Dover, New York, NY).

Pedersen, A.R., 1995, Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes, *Bernoulli* 1(3), 257-279.

Phillips, P.C.B., 1973, The problem of identification in finite parameter continuous-time model, *Journal of Econometrics* 1, 351-62.

Prohorov, Yu. V. and Yu. A. Rozanov, 1969, *Probability theory*, translated by K. Krickeberg and H. Urmitzer (Springer-Verlag).

Robinson, P.M., 1983, Nonparametric estimators for time series, *Journal of Time Series Analysis* 4, 185-207.

Rosenblatt, M., 1971, *Markov processes, structure and asymptotic behaviour* (Springe-Verlag, New York).

Silverman, B.W., 1986, *Density estimation for statistical and data analysis* (Chapman-Hall, London).

Stanton, R., 1996, A nonparametric model of term structure dynamics and the market price of interest rate risk, Working paper, Haas School of Business, U.C. Berkeley.

Vasicek, O., 1977, An equilibrium characterization of the term structure, *Journal of Financial Economics* 5, 177-88.

Wong, E., 1964, The construction of a class of stationary Markov processes, in: *Stochastic Processes in Mathematical Physics and Engineering*, Proceedings of Symposia in Applied Mathematics, Vol. XVI.

Table 1
Diffusion Processes: Stationarity and Common Estimators

Diffusions	Stationarity	Estimators of $\sigma^2(\cdot)$	Estimators of $\mu(\cdot)$
BMwD$^{(a)}$	No	NONP$^{(d)}$, MLE (or OLS)	MLE
O-U$^{(b)}$	Yes $^{(e)}$	NONP, MLE, NLS, GMM	NONP, MLE, NLS, GMM
CIR SR$^{(c)}$	Yes$^{(e)}$	NONP, NLS, GMM	NONP, NLS, GMM

Note: (a)– the Brownian motion with drift process;

(b)– the Ornstein-Uhlenbeck process;

(c)– the Cox-Ingersoll-Ross squared-root process;

(d)– the Nonparametric estimator;

(e)– the process is strictly stationary in the steady state as the initial time $t_0 \to -\infty$ or $t - t_0 \to +\infty$.

Table 2

Brownian Motion with Drift: Simulation Results of the Diffusion Estimators

(1,000 replications with sample size in each replication=5,000)

($\sigma^2(X_t) = \sigma^2 = 0.0004$):

a. Summary Statistics of the Sampling Distributions of the Nonparametric Estimator of σ^2 at Different Values of X_t:

X_t	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-11})	M.S.E. (10^{-11})	Skewness (10^{-2})	Kurtosis (10^{-2})	Wilcoxon Test[(a)] (10^{-2})
0.05	3.998	3.997	5.886	5.895	-5.872	1.320	8.779 (+)
0.06	3.998	3.997	5.886	5.895	-5.872	1.320	8.779 (+)
0.07	3.998	3.997	5.886	5.895	-5.872	1.320	8.779 (+)
0.08	3.998	3.997	5.886	5.895	-5.872	1.320	8.768 (+)
0.09	3.998	3.997	5.886	5.895	-5.872	1.320	8.790 (+)
0.10	3.998	3.997	5.886	5.895	-5.872	1.320	8.790 (+)

b. Summary Statistics of the Sampling Distribution of the Parametric Estimators of σ^2:

Method	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-11})	M.S.E. (10^{-11})	Skewness (10^{-2})	Kurtosis (10^{-2})
MLE	4.000	3.998	5.877	5.887	-5.708	1.108

c. Median Test for Multi-Sample of Nonparametric Estimates at Different Values of X_t:

Hypothesis H_0:	Test Statistic[(b)]	Critical Value (5%)
Samples of Nonparametric Estimates at Different Values of X_t Have the Same Median	0.600	16.92

Note: (a) The null hypothesis of the pair-wised Wilcoxon test is H_0: the median of the samples of the absolute bias of the nonparametric estimates at certain value of X_t is not greater than that of the maximum likelihood (ML) estimates. In the large sample case (say $n > 25$), an approximate statistic of the Wilcoxon matched-pairs signed-ranks test is $z = \frac{T-[n(n-1)]/4}{\sqrt{n(n+1)(2n+1)/24}}$ which follows a standard normal distribution, where n is the sample size, T is the sum of the positive ranks of the difference between two samples based on the null hypothesis. The test statistic reported in the table is corresponding to the above null hypothesis H_0 at 5 % significance level. "+" in the brackets denotes the hypothesis is not rejected, "–" denotes the hypothesis is rejected.

(b) The extension of the multi-sample median test statistic is given by $\chi^2 = \sum_{i=1}^{2}\sum_{j=1}^{n}[\frac{(O_{ij}-E_{ij})^2}{E_{ij}}]$ which follows the χ^2 distribution with degrees of freedom $(n-1)$, where n is the number of independent samples to be tested, E_{ij} is the combined sample median of all the samples, O_{1j} is the number of observations in the jth sample which are less than the combined sample median, and O_{2j} is the number of observations in the jth sample which are greater than the combined sample median.

Table 3
Ornstein-Uhlenbeck Process: Simulation Results of the Diffusion Estimators
(1,000 replications with sample size in each replication=5,000)
($\sigma^2(X_t) = \sigma^2 = 0.0004$)

a. Summary Statistics of the Sampling Distributions of the Nonparametric Estimator of σ^2 at Different Values of X_t:

X_t	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-11})	M.S.E. (10^{-11})	Skewness (10^{-2})	Kurtosis (10^{-2})	Wilcoxon Test$^{(a)}$
0.05	3.999	3.997	5.873	5.883	-5.863	1.568	1.437 (+)
0.06	3.999	3.997	5.872	5.882	-5.845	1.659	1.441 (+)
0.07	3.999	3.997	5.871	5.881	-5.824	1.752	1.439 (+)
0.08	3.999	3.997	5.871	5.881	-5.800	1.847	1.437 (+)
0.09	3.999	3.997	5.871	5.881	-5.774	1.943	1.433 (+)
0.10	3.999	3.997	5.871	5.881	-5.745	2.040	1.430 (+)

b. Summary Statistics of the Sampling Distributions of the Parametric Estimators of σ^2:

Method	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-11})	M.S.E. (10^{-11})	Skewness	Kurtosis	Wilcoxon Test$^{(a)}$
MLE	3.944	3.999	7.141	7.144	1.536	1.852	
NLS	3.935	3.992	9.692	9.760	-0.509	1.382	-3.548 (+)
GMM	3.942	3.995	7.725	7.792	-0.773	2.602	-17.34 (+)

c. Median Test for Multi-Sample of Nonparametric Estimates at Different Values of X_t

Hypothesis H_0:	Test Statistic^b	Critical Value (5%)
Samples of Nonparametric Estimates at Different Values of X_t have the Same Median	1.800	16.92

Note: (a) See Table 2 note (a); (b) See Table 2 note (b).

Table 4
Cox-Ingersoll-Ross SR Process: Simulation Results of the Diffusion Estimators
(1,000 replications with sample size in each replication=5,000)
($\sigma^2(X_t) = \sigma^2 X_t = 0.007X_t$)

a. Summary Statistics of the Sampling Distributions of the Nonparametric Estimator of σ^2 at Different Values of X_t

X_t ($\sigma^2(X_t)10^{-4}$)	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-10})	M.S.E. (10^{-10})	Skewness 10^{-1}	Kurtosis 10^{-1}	Wilcoxon test$^{(a)}$
0.05 (3.500)	3.864	3.879	7.740	22.111	0.395	-4.377	-5.742 (–)
0.06 (4.200)	4.320	4.342	5.519	7.525	5.356	6.333	-6.070 (–)
0.07 (4.900)	4.934	4.958	3.819	4.154	9.373	24.38	-1.372 (+)
0.08 (5.600)	5.651	5.668	5.566	6.030	8.012	24.95	-1.471 (+)
0.09 (6.300)	6.307	6.323	14.496	14.548	6.885	14.46	-1.842 (+)
0.10 (7.000)	7.016	7.083	21.643	22.337	3.469	-45.28	-2.393 (–)

b. Summary Statistics of the Sampling Distributions of the Parametric Estimators of σ^2:

Method	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-8})	M.S.E. (10^{-8})	Skewness (10^{-1})	Kurtosis (10^{-1})	Wilcoxon Test$^{(a)}$
NLS	6.795	6.945	5.774	6.074	-7.787	48.414	1.71 (+)
GMM	6.856	6.982	3.421	3.454	1.777	-0.187	

c. Median Test for Multi-Sample of Nonparametric Estimates at Different Values of X_t

Hypothesis H_0:	Test Statistic$^{(b)}$	Critical Value (5%)
Sample of Nonparametric Estimates at Different Values of X_t have the Same Median	11.61	16.92

Note: (a) See Table 2 note (a) for the explanation of the test statistic. For CIR process, the Wilcoxon test is based on the sampling distribution of $\hat{\sigma}^2$, which is calculated from $\hat{\sigma}^2 = \hat{\sigma}^2(X_t)/X_t$ for the nonparametric diffusion function estimator, and the null hypothesis of the pair-wised Wilcoxon test is against the sample of GMM estimates. (b) See Table 2 note (b).

Table 5
Brownian Motion with Drift: Simulation Results of the Drift Estimators
(1,000 replications with sample size in each replication =5,000)
($\mu(X_t) = \mu = 0.0055$)

Summary Statistics of the Sampling Distribution of the Parametric Estimators of μ:

Method	Para-meters	Min (10^{-2})	Max (10^{-2})	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-4})	MSE (10^{-4})	Skewness (10^{-2})	Kurtosis (10^{-2})
MLE	$\hat{\mu}$	-6.689	8.149	4.254	4.144	4.132	4.147	-8.398	5.371

Table 6
Ornstein-Uhlenbeck Process: Simulation Results of the Drift Estimators
(1,000 replications with sample size in each replication =5,000)
($\mu(X_t) = \beta(\alpha - X_t)$, $\beta = 0.18$, $\alpha = 0.086$)

a. Summary Statistics of the Sampling Distributions of the Parametric Estimators of α and β:

Method	Para-meters	Min	Max	Median	Mean	Variance	MSE	Skew-ness	Kurto-sis
MLE	$\hat{\alpha}$	-2.801	5.606	0.057	0.079	8.056 10^{-2}	8.059 10^{-2}	8.907	195.86
	$\hat{\beta}$	-2.119	29.176	2.653	5.487	17.059	45.227	1.293	2.654
NLS	$\hat{\alpha}$	-2.695	7.509	0.059	0.112	0.219	0.220	10.523	144.34
	$\hat{\beta}$	-3.079	24.667	3.707	7.205	22.825	72.171	0.779	0.345
GMM	$\hat{\alpha}$	-3.266	7.765	0.058	0.111	0.339	0.340	12.896	195.83
	$\hat{\beta}$	-0.926	24.941	2.514	5.656	19.298	49.289	1.248	1.544

b. Summary Statistics of the Sampling Distributions of the Estimators of $\mu(X_t)$ at Different X_t:

Method	$X_t(\mu(X_t)\ (10^{-3}))$	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-6})	M.S.E. (10^{-6})	Skewness (10^{-1})	Kurtosis (10^{-1})
NONP	0.05 (6.480)	4.720	4.570	1.120	4.773	-7.419	5.893
	0.06 (4.680)	2.567	2.479	1.057	5.996	-4.902	4.745
	0.07 (2.880)	1.426	1.135	1.025	4.084	-2.013	4.976
	0.08 (1.080)	-0.073	-0.075	1.026	2.360	1.068	6.882
	0.09 (-0.72)	-2.886	-2.895	1.070	5.801	4.129	8.316
	0.10 (-2.52)	-5.052	-4.975	1.134	7.161	6.915	9.631
MLE	0.05 (6.480)	80.478	101.466	1225.2	2127.4	14.265	34.901
	0.06 (4.680)	39.283	46.423	794.81	969.06	8.786	27.046
	0.07 (2.880)	-3.761	-8.619	707.62	720.85	-3.746	25.367
	0.08 (1.080)	-46.08	-63.66	963.64	1382.8	-12.198	32.474
	0.09 (-0.72)	-91.66	-118.70	1562.8	2954.90	-13.130	32.299
	0.10 (-2.52)	-139.03	-173.74	2505.3	5437.18	-13.814	29.775
NLS	0.05 (6.480)	113.20	136.29	1651.0	3336.1	9.401	11.282
	0.06 (4.680)	55.865	64.246	1068.5	1423.3	5.152	11.484
	0.07 (2.880)	-0.179	-7.801	942.44	953.84	-4.474	13.208
	0.08 (1.080)	-60.532	-79.847	1272.9	1927.8	-10.093	14.574
	0.09 (-0.72)	-123.16	-151.89	2059.9	4345.2	-10.288	11.799
	0.10 (-2.52)	-191.02	-223.94	3303.4	8206.0	-9.524	8.863
GMM	0.05 (6.480)	85.490	107.309	1271.1	2287.8	13.017	24.482
	0.06 (4.680)	40.933	50.072	795.92	1001.9	8.115	22.363
	0.07 (2.880)	-0.265	-7.166	709.21	719.30	-3.649	19.676
	0.08 (1.080)	-45.063	-64.403	1010.98	1439.79	-11.633	22.081
	0.09 (-0.72)	-89.312	-121.64	1701.25	3163.44	-13.444	22.050
	0.10 (-2.52)	-138.47	-178.88	2780.02	5890.25	-13.561	20.844

Table 7
Cox-Ingersoll-Ross SR Process: Simulation Results of the Drift Estimators
(1,000 replications with sample size in each replication =5,000)
($\mu(X_t) = \beta(\alpha - X_t), \beta = 0.23, \alpha = 0.076$)

a. Summary Statistics of the Sampling Distributions of the Parametric Estimators of α and β:

Method	Para-meters	Min	Max	Median	Mean	Variance	M.S.E.	Skew-ness	Kurtosis
NLS	$\hat{\alpha}$	-1.058	6.090	0.056	0.238	1.533	1.559	7.911	124.38
	$\hat{\beta}$	-7.363	39.906	1.884	5.688	30.971	60.760	1.533	4.206
GMM	$\hat{\alpha}$	-2.065	5.164	0.054	0.081	$9.430\ 10^{-2}$	$9.433\ 10^{-2}$	6.904	109.77
	$\hat{\beta}$	-5.492	26.294	1.752	4.355	14.606	31.614	1.568	3.657

b. Summary Statistics of the Sampling Distributions of the Estimators of $\mu(X_t)$ at Different X_t

Method	$X_t(\mu(X_t)(10^{-3}))$	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-6})	M.S.E. (10^{-6})	Skewness (10^{-1})	Kurtosis (10^{-1})
NONP	0.05 (5.980)	6.149	6.734	8.968	9.537	9.397	8.241
	0.06 (3.680)	4.854	5.218	4.378	6.744	12.466	25.632
	0.07 (1.380)	2.887	2.933	2.282	4.694	4.206	4.164
	0.08 (-0.92)	0.940	-0.913	4.473	7.833	-3.163	9.548
	0.09 (-3.22)	-1.472	-1.550	8.199	10.988	-9.488	20.078
	0.10 (-5.52)	-3.394	-3.954	9.871	12.323	-6.403	10.949
NLS	0.05 (5.980)	72.505	101.25	1735.9	2643.5	20.103	95.22
	0.06 (3.680)	28.161	43.470	1035.7	1193.9	18.568	93.07
	0.07 (1.380)	-11.550	-14.31	948.29	927.91	1.297	45.40
	0.08 (-0.92)	-49.03	-72.09	1473.8	1980.4	-13.805	39.06
	0.09 (-3.22)	-94.00	-129.87	2612.3	4216.3	-17.454	50.64
	0.10 (-5.52)	-140.00	-187.65	4363.7	7680.8	-17.700	53.66
GMM	0.05 (5.980)	56.920	74.948	856.05	1331.7	13.116	25.96
	0.06 (3.680)	22.228	30.879	567.52	641.50	8.924	25.46
	0.07 (1.380)	-12.375	-13.191	570.54	591.77	-3.276	23.16
	0.08 (-0.92)	-42.935	-57.260	865.09	1182.6	-13.026	35.56
	0.09 (-3.22)	-77.483	-101.33	1451.2	2413.7	-16.349	44.34
	0.10 (-5.52)	-112.61	-145.39	2328.8	4285.4	-17.136	46.78

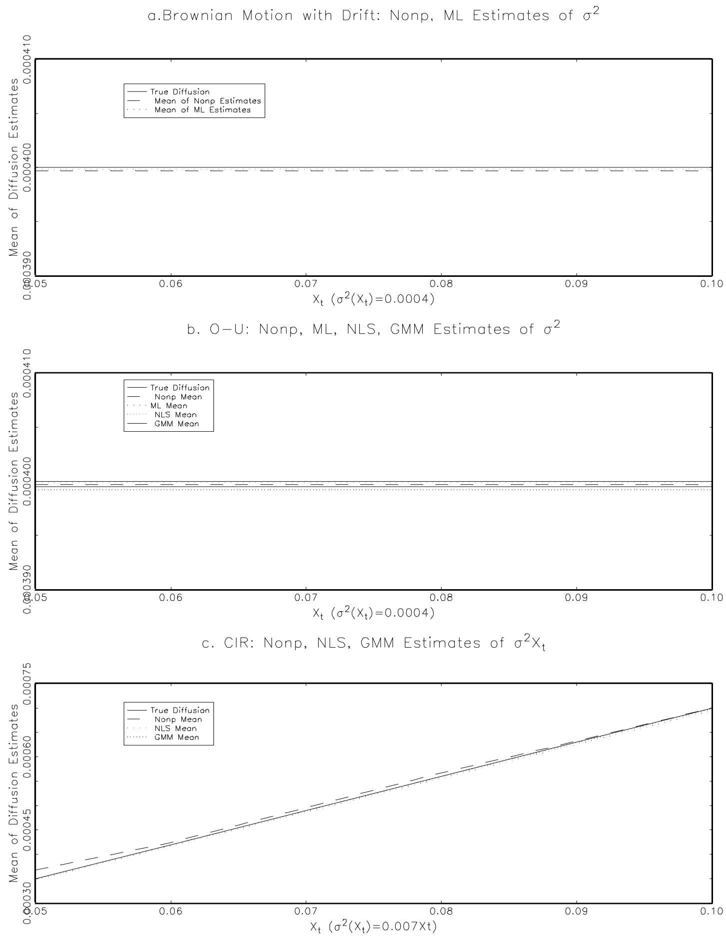

Fig. 1. Sample Means of Diffusion Estimates at Different X_t
(1,000 replications with sample size in each replication=5,000)

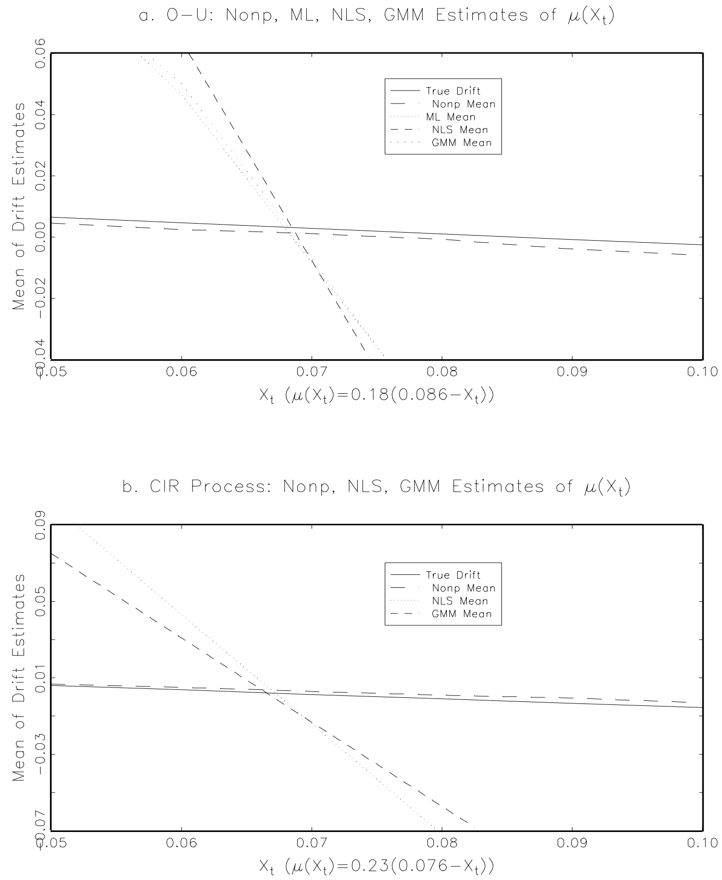

Fig. 2. Sample Means of Drift Estimates at Different X_t
(1,000 replications with sample size in each replication=5,000)

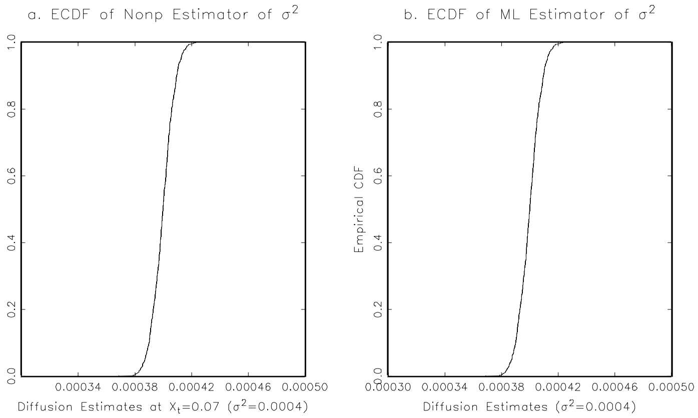

Fig. 3. Brownian Motion with Drift: Empirical CDF of Diffusion Estimators
(1,000 replications with sample size in each replication=5,000)

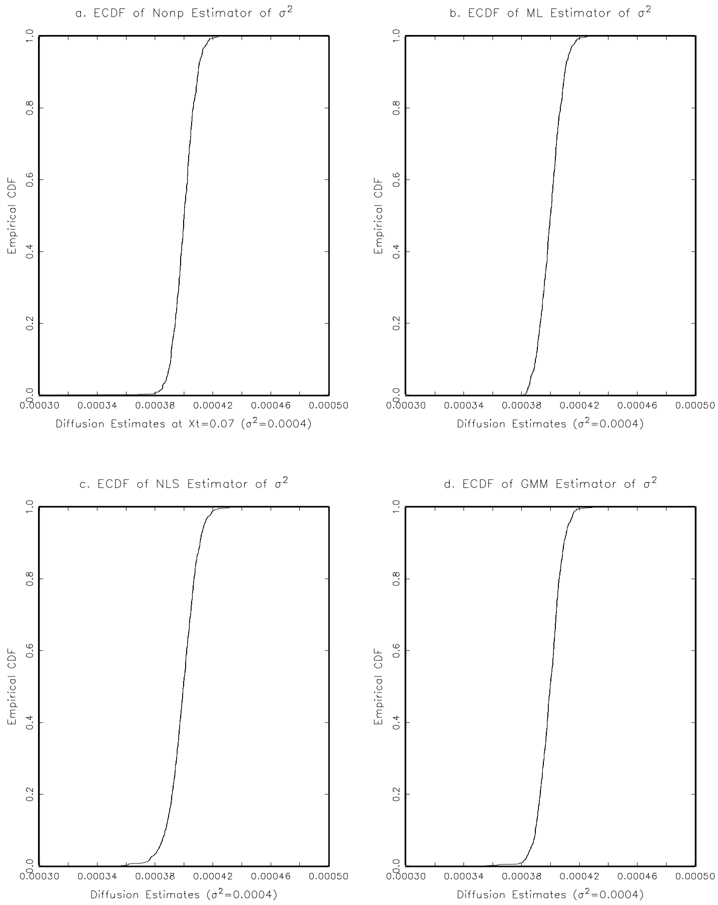

Fig. 4. O-U Process: Empirical CDF of Diffusion Estimators
(1,000 replications with sample size in each replication=5,000)

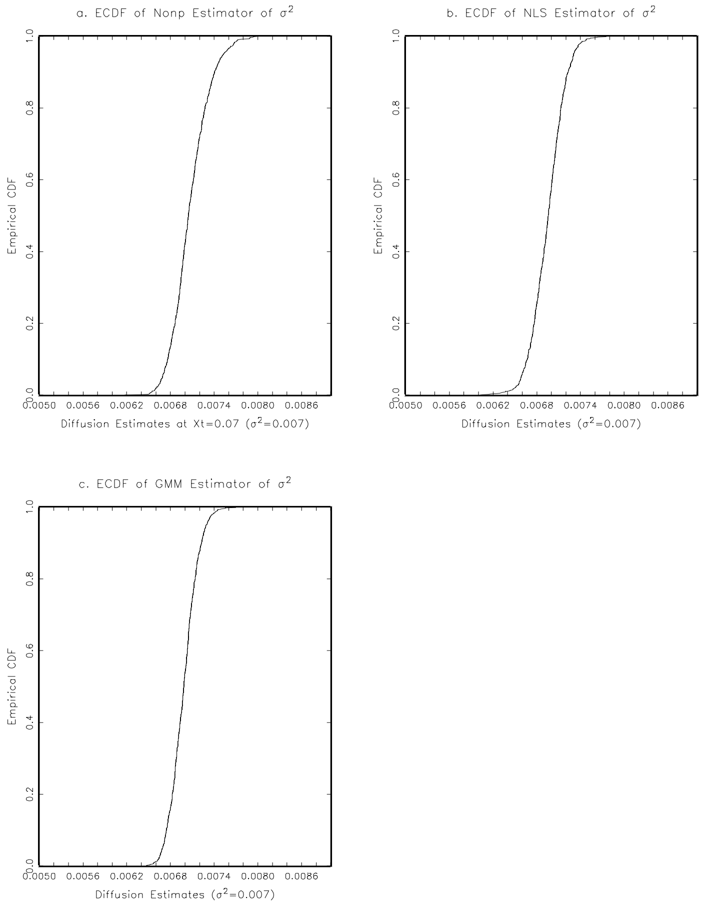

Fig. 5. CIR Process: Empirical CDF of Diffusion Estimators
(1,000 replications with sample size in each replication=5,000)

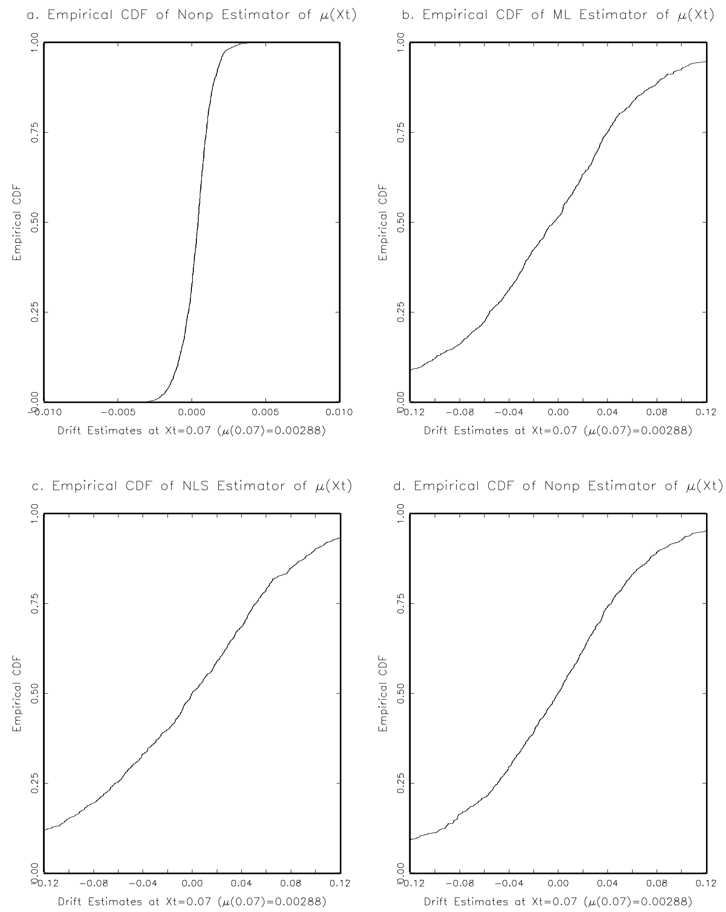

Fig. 6. O-U Process: ECDF of Drift Estimators at $X_t = 0.07$
(1,000 replications with sample size in each replication=5,000)

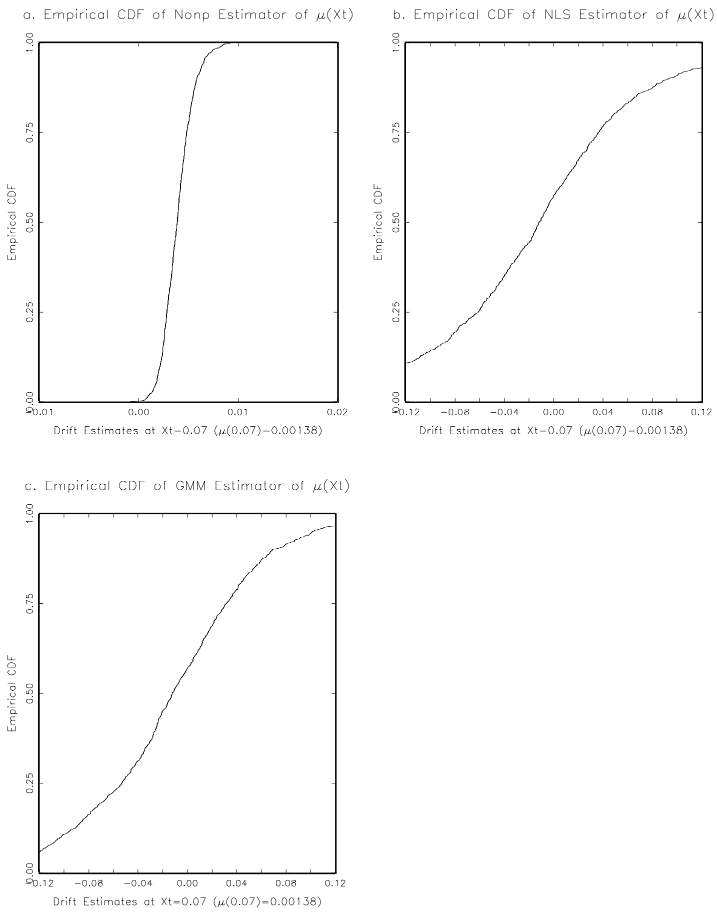

Fig. 7. CIR Process: ECDF of Drift Estimators at $X_t = 0.07$
(1,000 replications with sample size in each replication=5,000)

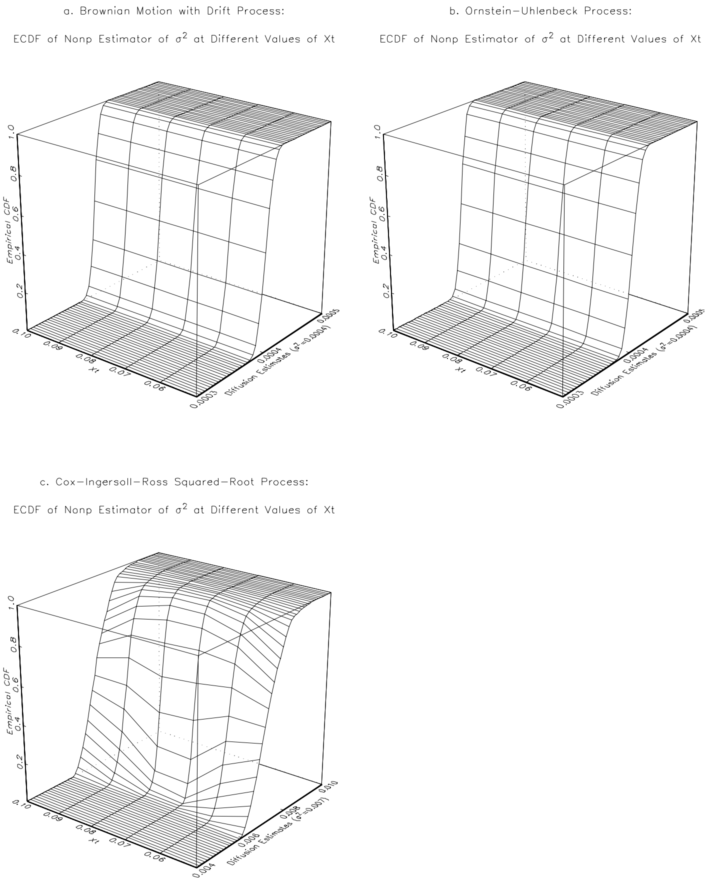

Fig. 8. ECDF of Nonp Diffusion Estimators at Different Values of X_t
(1,000 replications with sample size in each replication=5,000)

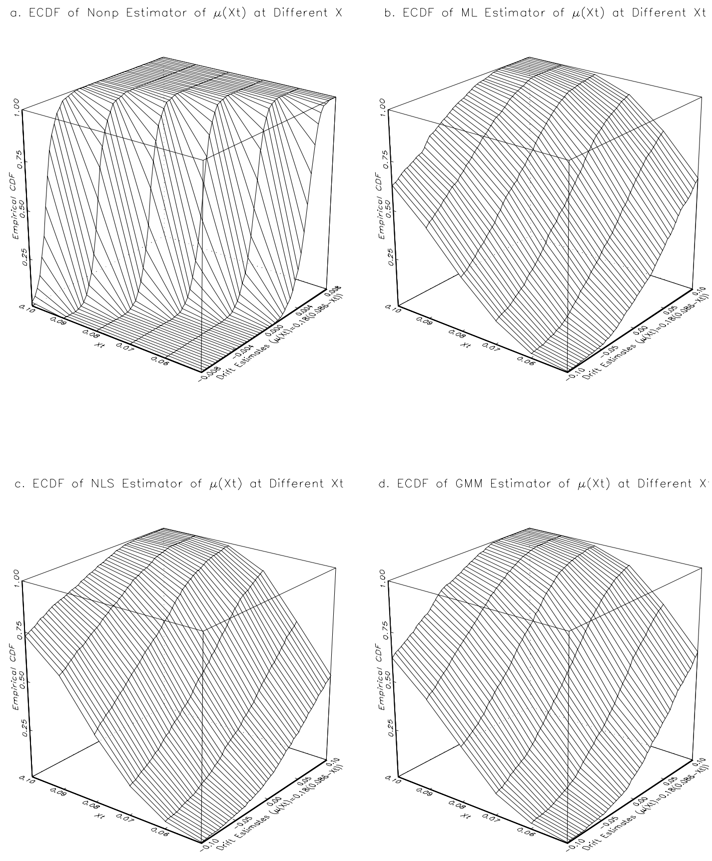

Fig. 9. O-U Process: ECDE of Drift Estimators at Different Values of X_t
(1,000 replications with sample size in each replication=5,000)

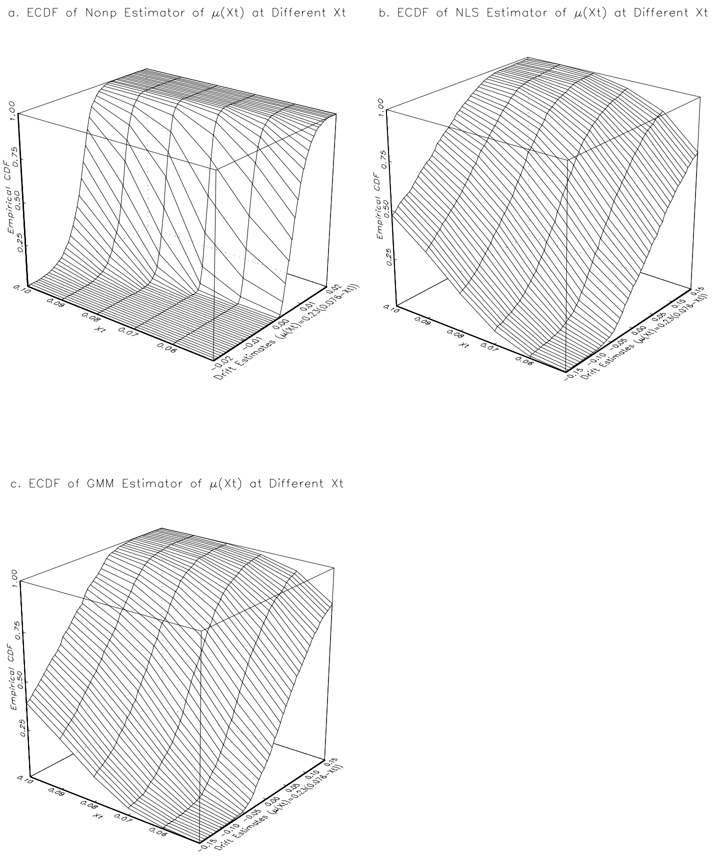

Fig. 10. CIR Process: ECDE of Drift Estimators at Different Values of X_t
(1,000 replications with sample size in each replication=5,000)