Finite Sample Comparison of Alternative Estimators of Itô Diffusion Processes

— A Monte Carlo study

by

George J. Jiang
Department of Econometrics
University of Groningen
The Netherlands

and

John L. Knight [1]
Department of Economics
University of Western Ontario
Canada

<Abstract>

In this paper, we consider alternative approaches to the estimation of Itô diffusion processes from discretely sampled observations. Based on Monte Carlo simulation, we investigate the finite sample properties of various estimators and in particular compare the performance of the nonparametric estimators proposed in Jiang and Knight (1997) with common parametric estimators, namely the ML, NLS (or OLS), and GMM estimators. The simulation results show that, with certain large samples over a short sampling period, both the nonparametric diffusion and drift estimators perform reasonably well. However, while all the parametric diffusion estimators perform very well, the parametric drift estimators perform very poorly.

Keywords: Diffusion Process; Parametric Estimation; Nonparametric Estimation; Monte Carlo Simulation

JEL Classification: C14; C15; C22

---

[1] We wish to thank the Chief Editor Domingo Tavella and the anonymous referee for helpful and constructive comments on an earlier draft. We also thank Victor Aguirregabiria, Phelim Boyle, Robin Carter, Yanqin Fan, Reg Kulperger, and Tom Wansbeek for their helpful comments. We would also like to thank seminar participants in the Department of Economics and Department of Statistical and Actuarial Sciences at the University of Western Ontario, the EC^2 Conference in Florence, Italy and the 53rd Econometric Society European Meeting in Berlin, Germany for helpful discussion and comments. The usual disclaimer applies. The second author acknowledges financial support from the Natural Science and Engineering Research Council of Canada (NSERC). Corresponding author: George J. Jiang, Department of Econometrics, University of Groningen, P.O.Box 800, 9700 AV Groningen, The Netherlands. Tel: +31 (50) 363-3711. Fax: +31 (50) 363-3720. e-mail: G.Jiang@ECO.RUG.NL.

1 Introduction

The model specified as the underlying process of the state of asset prices, exchange rates, or spot interest rates $\{X_t, t \geq t_0\}$ in most continuous-time finance literature is a time-homogeneous 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 (or instantaneous mean) and the diffusion (or instantaneous variance) functions of the process. The drift term describes the movement of the process due to time change, while the diffusion term measures the magnitude of the random fluctuations around the drift.

Due to the estimation problem, however, most 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. For instance, in the spot interest rate models, [2] 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 Cox, Ingersoll and Ross (hereafter 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)). Simple specifications allow estimation of the parameters via the use of common parametric estimators, such as ML, NLS (or OLS), or GMM.

The purpose of this paper is to design and perform a simple and small Monte Carlo simulation experiment to investigate the finite sample properties of various estimators of Itô diffusion processes from discretely sampled observations. 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

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

function and drift function estimators proposed in Jiang and Knight (1997) based on relatively large samples over short to medium sampling periods. Our study involves the comparison of nonparametric estimators with common parametric estimators whose implementation relies on either explicit transition density function or exact conditional and/or unconditional moment conditions of the process. Since we wish to minimise sampling error in our Monte Carlo study, we do not examine processes which do not possess an explicit transition density. We have thus excluded processes which could be approximated by the Euler or Milshtein [3] scheme as these may compound the sampling errors and make the results difficult to interpret. Consequently this simulation study will examine estimation performance for the following three basic diffusion models: the Brownian motion with drift process, the Ornstein-Uhlenbeck process, and the Cox-Ingersoll-Ross squared-root process. Our results suggest that, with certain large samples over a short sampling period, [4] both the nonparametric diffusion function estimator and drift function estimator proposed in Jiang and Knight (1997) perform reasonably well. However, while all the parametric diffusion function estimators perform very well, the parametric drift function estimators perform very poorly.

The paper is organised 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 (1) 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.

[3]As the discretization interval goes to zero, it can be shown that the simulated path based on both Euler scheme and Milshtein scheme converges to the true path uniformly in probability on compact sets. The Milshtein scheme has better convergence rate than the Euler scheme for the convergence in $L^p(\Omega)$ and the almost sure convergence (see Talay, 1996).

[4]Intuitively speaking, we need to rely on a sampling path along which the observations are close enough to the points of X_t at which we intend to make inference about the drift and diffusion functions.

2 Estimation of the Diffusion Process from Discretely Sampled Data

Under certain regularity conditions (see e.g. Banon (1978), conditions A1-A4), the stochastic differential equation (SDE) defined in (1) can be applied with Itô stochastic calculus and has a unique solution, moreover the underlying process $\{X_t, t \geq t_0\}$ defined in (1) is a regular strong Markov process with continuous sampling paths and time-stationary transition densities. Thus, the transition density function is the unique fundamental solution of both the Kolmogorov backward equation

$$\frac{1}{2}\sigma^2(y)\frac{\partial^2 p(X_t = x|X_\tau = y)}{\partial y^2} + \mu(y)\frac{\partial p(X_t = x|X_\tau = y)}{\partial y} = -\frac{\partial p(X_t = x|X_\tau = y)}{\partial \tau}$$

and the Kolmogorov forward (or Fokker-Planck) equation (see e.g. Karlin and Taylor (1981))

$$\frac{1}{2}\frac{\partial^2(\sigma^2(x)p(X_t = x|X_\tau = y))}{\partial x^2} - \frac{\partial(\mu(x)p(X_t = x|X_\tau = y))}{\partial x} = \frac{\partial p(X_t = x|X_\tau = y)}{\partial t}$$

where $p(X_t = x|X_\tau = y)$ is the transition density function for $X_t = x$ at time t conditional on $X_\tau = y$ at time τ, and $p(X_\tau = x|X_\tau = y) = \delta(x - y)$ (the Dirac delta function). This implies that the transition density function, i.e. the conditional or dynamic properties, of the underlying Markov process of (1) is fully characterised by its coefficients under the regularity conditions.

Stronger relations can be derived under the condition that the stochastic process is stationary in the strict sense, or equivalently, that there exists a stationary initial probability density $p(X_{t_0})$ such that $p(X_t = x) = \int p(X_t = x|X_{t_0} = u)p(X_{t_0} = u)du = p(X_{t_0} = x)$ for any x in the state space. Under the above condition, the left hand side of the Kolmogorov forward equation becomes zero, multiplying both terms in the right hand side with the marginal density $p(X_\tau)$, then integrating with respect to X_τ, for $-\infty < X_t < +\infty$, one can solve for the solution of $p(X_t)$ as

$$p(X_t) = \frac{A}{\sigma^2(X_t)}exp\{2\int_{X^0}^{X_t}\frac{\mu(u)}{\sigma^2(u)}du\} \tag{2}$$

with the boundary conditions $p(+\infty) = p'(+\infty) = 0$ or $\sigma^2(+\infty) = \sigma^2\prime(+\infty) = 0$, where A is the normalising constant, X^0 is an arbitrary interior point of the state space, i.e., $-\infty < X^0 < +\infty$. This implies further that the marginal density, i.e. the unconditional or static properties, of the underlying Markov process of (1) is fully characterised by its coefficients under the regularity conditions. Similarly, with the boundary condition $p(+\infty) = p'(+\infty) = 0$, we have

$$\mu(X_t) = \frac{1}{2p(X_t)} \frac{d}{dX_t}[\sigma^2(X_t)p(X_t)] \tag{3}$$

or

$$\sigma^2(X_t) = \frac{2}{p(X_t)} \int_{-\infty}^{X_t} \mu(u)p(u)du \tag{4}$$

In other words, with any functional form specification for either the drift or the diffusion term, the other term will be specified given the marginal density function of the diffusion process.

The estimation of stochastic differential equations has been considered in the literature for many years with most of the earlier papers mainly concerned with continuous sampling observations. For instance, the parametric drift and/or diffusion function estimators were proposed by Brown and Hewitt (1975), Lanska (1979), and Kutoyants (1984), and the nonparametric drift and/or diffusion function estimators have been proposed by Banon (1978), Geman (1979), Pham Dinh (1981), and Banon and Nguyen (1981). In most practical situations, however, continuous sampling of the stochastic process is impossible since the characteristic dynamics of the system can be much faster than the sampling rate. The need to estimate the coefficients of a stochastic differential equation from discrete-time observations is the situation arising frequently in practical applications. For instance, when a set of discretely sampled observations is available, the parametric estimation of the drift coefficient has been considered by Dorogovcev (1976), Kasonga (1988), Genon-Catalot (1990) and Laredo (1990) using various estimation methods, and the parametric estimation of diffusion coefficient has been considered by Dacunha-Castelle and Florens-Zmirou (1986) and Donhal (1987) for stationary models, Jacod (1993) for Gaussian models, and Genon-Catalot and Jacod (1993) in a general framework. Lo (1988) proposes a maximum likelihood (ML) estimation method for a jump-diffusion process. Pedersen (1995) suggests an approximate maximum likelihood (AML) parameter estimator for multi-dimensional diffusion processes. Chan, et al (1992) adopted Hansen's (1982) generalised method of moments (GMM) for the parametric estimation. [5] Bibby and Sørensen (1995) propose the estimation of non-linear diffusions based on

[5] It should be noted that any attempts to estimate the model based on simple ”discretization” cannot be recommended, since it may cause severe misspecification for fixed sampling intervals. Lo (1988) and others have provided the examples of inconsistent estimators based on “discretization” of Itô diffusion processes.

martingale estimation functions. Also developed is the alternative “indirect inference” approach to estimating non-linear stochastic differential equations.[6] Particular examples of the “indirect inference” approach include the simulated generalised method of moments proposed by Duffie and Singleton (1993) based on simulated moments of the processes, the quasi maximum likelihood (QML) estimation by Gouriéroux, Monfort and Renault (1993) based on discretized auxiliary models, and the efficient method of moments (EMM) by Gallant and Tauchen (1996) based on the semiparametric (SNP) score generator. Hansen and Scheinkman (1995) formally derived moment conditions for the continuous-time diffusion process based on the infinitesimal generator. Conley, Hansen, Luttmer and Scheinkman (1997) proposed a new estimation method using test functions to discriminate among competing models. When the sample path is discretely observed but the sampling interval is small and goes to zero in the limit, Genon-Catalot, Laredo and Picard (1992) proposed a nonparametric estimator of a time-inhomogeneous diffusion function using a projection method on a wavelets orthonormal basis of $L^2(\mathcal{R})$. Florens-Zmirou (1993) proposed a nonparametric naive estimator of a time-homogeneous diffusion function using the approximations of local time process where the drift term was left unidentified and was restriction-free. Stanton (1997) constructed approximations to both the true drift and diffusion functions.

The nonparametric estimation of the diffusion term suggested by Aït-Sahalia (1996) is a semi-parametric approach to the identification of the strictly stationary diffusion processes. It specifies the drift as a linear mean-reverting function, i.e., $\mu(X_t) = \beta(\alpha - X_t)$. Then based on the Kolmogorov forward equation, $\sigma^2(\cdot)$ can be characterised through the marginal density $p(\cdot)$, and the estimates of α and β. The transition distribution can be used to obtain consistent estimates of α and β through

$$E[X_{t+\rho}|X_t] = \alpha + e^{-\rho\beta}(X_t - \alpha)$$

which is derived from Kolmogorov backward equation. Aït-Sahalia (1996) shows that under certain regularity conditions, the estimator $\hat{\sigma}^2(X_t)$ is pointwise consistent and asymptotically normal. However, this estimation procedure has to rely on the parametric specification of the drift term and works only for the strictly stationary diffusion processes.

[6] See Monfort (1996) for a recent review with focus on misspecified models.

In Jiang and Knight (1997), a nonparametric identification and estimation procedure, based on discretely sampled observations, for a wide range of Itô diffusion processes is proposed. The nonparametric estimator for the diffusion function follows that of Florens-Zmirou (1993). In order to achieve better properties of the estimator and to construct a nonparametric drift function estimator, Jiang and Knight (1997) use the kernel method in place of the naive method (i.e. the indicator function estimator). The technique deals with general Itô diffusion processes and avoids any functional form specification for either the drift function or the diffusion function. Construction of this estimator is based on the local-time properties of the diffusion process and derived from the expansion of the transition density for small changes in time. The proposed estimator of the drift function combines the nonparametric kernel estimator of the diffusion function along with the proposal from Banon (1978) and Banon and Nguyen (1981) who developed a nonparametric drift function estimator based on known diffusion function and continuous sampling observations. It exploits the information contained in the marginal density of the process. Under mild regularity conditions, a nonparametric kernel estimator for the diffusion function $\sigma^2(X_t)$ of the general diffusion process defined in (1) is proposed as:

$$\hat{\sigma}^2(x) = \frac{\sum_{i=1}^{n-1} K(\frac{X_{i\Delta_n}-x}{h_n})[X_{(i+1)\Delta_n} - X_{i\Delta_n}]^2}{\sum_{i=1}^{n} \Delta_n K(\frac{X_{i\Delta_n}-x}{h_n})} \tag{5}$$

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 coefficient function and restriction-free. Thus the nonparametric diffusion function estimator captures the true volatility of the process. It can be shown that as $h_n \to 0, n \to \infty, nh_n \to \infty$, and $nh_n^3 \to 0, \hat{\sigma}^2(x)$ is a pointwise consistent estimator of $\sigma^2(x)$ and asymptotically mixture-normally distributed.[7] The variance of $\hat{\sigma}^2(x)$ can be consistently estimated by $\hat{V}[\hat{\sigma}^2(x)] = \hat{\sigma}^4(x)/\sum_{i=1}^{n} K(\frac{X_{i\Delta_n}-x}{h_n})$.

[7] See Jiang and Knight (1997) for proof of consistency and asymptotic mixture normality for the nonparametric kernel estimator of diffusion function.

Direct estimation of the drift function $\mu(x)$ without any restrictions on the underlying structure of the diffusion process is impossible in general from discretely observed data either over a short time interval (no matter how frequent the sampling is) or with fixed frequency (no matter how long the period spanned). In particular, based on samples with fixed sampling interval one can not distinguish a pair of coefficient functions $(\mu(\cdot), \sigma^2(\cdot))$ from $(c\mu(\cdot), c\sigma^2(\cdot))$ for a constant c without further constraints (see Aït-Sahalia (1996)). It is also well known in the literature that with a discretely observed sample over a short time period, the estimate of the drift function tends to have low precision even though the diffusion function can be precisely estimated when the sampling interval is small (see the example for geometric Brownian motion in this section). Under the condition that the diffusion process is either strictly stationary or has a limiting probability density function, a robust nonparametric estimator of the drift function using the information in the marginal density of the stochastic process is proposed in Jiang and Knight (1997) based on the relationship (3) or: [8]

$$\hat{\mu}(x) = \frac{1}{2}[\frac{d\hat{\sigma}^2(x)}{dx} + \hat{\sigma}^2(x)\frac{\hat{p}'(x)}{\hat{p}(x)}] \tag{6}$$

Further conditions imposed in Banon (1978) (namely A5&A6) ensure that the Kolmogorov forward equation has an unique fundamental solution for the marginal density and that the transition density function converges, as the initial time $t_0 \to -\infty$, to a bounded and continuous limiting density, say $\lim_{h\to+\infty} p(X_t|X_{t-h}) = p(X_t)$, which is the asymptotic marginal density of the process defined in (1) in the steady state. The asymptotic distribution of $\hat{\mu}(x)$ is complicated due to the functional form of its estimator. The variance of $\hat{\mu}(x)$, however, 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 used to derive the standard error of $\hat{\mu}(x)$.

It is noted that in Aït-Sahalia (1996) the same relationship between the diffusion function, the drift function and the marginal density of the process is exploited as in the above estimation procedure. The difference between these two estimation procedures is that instead of specifying a parametric drift function first and then estimating the diffusion function as in Aït-Sahalia (1996),

[8] See Jiang and Knight (1997) for proof of consistency of the estimator in the case of discretely sampled observations.

we estimate a nonparametric diffusion function first and then estimate the nonparametric drift function. It should be also pointed out that our above estimation procedure requires data observed with high frequency. Compared to the semiparametric diffusion function estimator suggested by Aït-Sahalia (1996) in which the asymptotic distribution of the estimator is derived by letting the sample size increase through prolonging the observation period, the nonparametric diffusion function estimator used in this paper requires the sampling frequency to increase in order to obtain asymptotic results. The reason is that when no *a priori* restrictions are imposed on the drift term, the identification of the diffusion term has to rely on the local time properties of the process, i.e. the evolution of the stochastic process for small changes in time. It is noted that in (1) the drift term is of order dt and the diffusion term is of order $\sqrt{dt}$, as $(dW_t)^2 = dt + O((dt)^2)$, i.e., the diffusion term has lower order than the drift term for infinitesimal changes in time. Therefore, the local-time dynamics of the sampling path reflects more of the properties of the diffusion term than those of the drift term, which suggests the possibility of identifying the diffusion term first from high-frequency discrete sampling observations. For the same reason, the drift term can not be estimated precisely from the local-time dynamics of such sampling paths without further constraints. Due to the fact that the observation interval of a discrete sample can never be zero nor can the sample size ever be infinity, there are inevitable biases in the above estimators as a result of discretization.

Thus the general Itô diffusion process defined in (1) can be identified nonparametrically for both the drift and diffusion functions from 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 non-linear diffusion process as defined in (1) without imposing any structural restrictions on the model. In particular, the drift term of the diffusion process (univariate or multivariate) cannot be directly identified on a fixed short time period as we have pointed out, no matter how frequently the observations are sampled, as the Cameron-Martin-Girsanov transformation (see e.g. Øksendal, 1992) can always be applied to give an otherwise unnoticeable change in the drift. The above proposed drift function estimator is based on the estimated diffusion function and marginal density function and exploits the stationarity property of the process. Thus, on the one hand it requires stronger conditions imposed on the diffusion process, on the other hand it overcomes the identification problem. This is consistent with the result that the "aliasing problem" is not present in the stationary univariate diffusions as shown in Hansen and Scheinkman (1995) under very mild technical conditions. Moreover, from (6) we note that estimation of drift function requires $\hat{p}(x) \neq 0$, in other words the estimator $\hat{\mu}(x)$ is valid only at points where the marginal density function can be estimated from the sampling observations.

The observation that the common parametric estimators of the drift parameters perform very poorly over a short sampling period is not new in the literature. 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. 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 $\hat{\mu} \sim N(\mu, \sigma^2/T)$. However, $\sigma^2/T \longrightarrow 0$ only if $T \longrightarrow \infty$. That is, the estimation of the conditional mean requires long sample periods and $\hat{\mu}$ has no efficiency gains even if we increase the sample size by reducing the sampling interval over fixed T. 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 $E[\hat{\sigma}^2] = \sigma^2(1 - \Delta/T)$ and $Var[\hat{\sigma}^2] = \frac{2\sigma^4\Delta}{T}(1 - \Delta/T)$. Our findings in this paper confirm this observation and further reveal that the common parametric estimates of the drift parameters perform poorly even with certain large samples based on which alternative estimator can perform reasonably well.

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 (1997)), while the nonparametric drift function estimator requires stronger conditions (i.e., A1-A8 in Jiang and Knight (1997)), 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. [9] 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 the assumption that, in the economy, there exists a risk-free asset with instantaneous rate of return. However, as Lo and Wang (1995) have shown, alternative drift function specifications can have significant impact on the option prices, 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, and it is shown in Jiang (1997) that the drift function specification also has important economic implications. From this point of view, the drift function estimation is as important as the diffusion function estimation.

[9] 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 certain leverage within a portfolio, the volatility of the prices of underlying assets is the major factor to be considered.

3 Alternative Estimation Methods for Common Diffusion Processes

The commonly used diffusion processes in finance are those 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 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 $Y_t = lnX_t$ 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.

Even though both the Brownian motion with drift process and the geometric Brownian motion process are nonstationary, the first differences of the processes are stationary and one can apply efficient estimators based on the changes of the processes.[10] The maximum likelihood (ML) estimator 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 ML counterparts. However, the aforementioned nonparametric drift function estimator cannot be applied due to the nonstationarity of the level process. Thus the comparison of the performance

[10] We wish to thank an anonymous referee for the suggestion of making this point clearer.

between parametric and nonparametric estimators will have to be constrained only to the diffusion function estimators.

(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 non-parametric 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 (ML) estimation can also be used. Further, as an asymptotic stationary process, the parameters of the Ornstein-Uhlenbeck process can also be estimated using GMM. In this paper, we consider the GMM estimators of α, β and σ^2 for the Ornstein-Uhlenbeck process based on the following four exact conditional and unconditional moments:[11]

$$G_n(\alpha, \beta, \sigma^2) = \frac{1}{n-1}\sum_{i=1}^{n-1} F_i(\alpha, \beta, \sigma^2) \tag{7}$$

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

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

$$E[\epsilon_{i+1}^2 | X_{i\Delta_n}] = \frac{\sigma^2}{2\beta}(1 - e^{-2\beta\Delta_n}) \tag{9}$$

[11]The justification of the choice of specific moments in our GMM estimation is as following. Even though the diffusion processes are in general not normally distributed, it turned out that under regularity conditions the distributions of the process (marginal and transitional densities) are entirely characterised by the first two moments of the process, namely the drift and diffusion functions.

where Δ_n is the sampling interval, and the exact conditional variance of the changes of X_t over a time interval of length Δ_n is given by $E[\epsilon_{i+1}^2|X_{i\Delta_n}] = V[X_{(i+1)\Delta_n}|X_{i\Delta_n}]$ [12]. These moment conditions correspond to transitions of length Δ_n and are not subject to discretization bias [13]. Since these GMM systems are overidentified, we weighted 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 non-linear least squares (NLS) estimators of $(\alpha, \beta, \sigma^2)$ can also be obtained. For the Ornstein-Uhlenbeck process, the NLS estimator is equivalent to the ML estimator, therefore it is not included in the comparison.

(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

[12] 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.

[13] 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 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 (1997).

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 $E[X_t|X_0 = x_0] = x_0 e^{-\beta(t-t_0)} + \alpha(1 - e^{-\beta(t-t_0)})$, $Var[X_t|X_0 = x_0] = x_0(\frac{\sigma^2}{\beta})(e^{-\beta(t-t_0)} - e^{-2\beta(t-t_0)}) + \alpha(\frac{\sigma^2}{2\beta})(1 - e^{-\beta(t-t_0)})^2$. 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 [14]. Similar to the Ornstein-Uhlenbeck process, the Cox-Ingersoll-Ross squared-root process as an asymptotic stationary process can also be estimated using GMM. Similarly, the GMM estimators of α, β and σ^2 are obtained from the following exact conditional and unconditional moment conditions based on its conditional mean and variance, i.e.,

$$E[(X_{(i+1)\Delta_n} - X_{i\Delta_n})|X_{i\Delta_n}] = (1 - e^{-\beta\Delta_n})(\alpha - X_{i\Delta_n}) \tag{10}$$

$$E[\epsilon_{i+1}^2|X_{i\Delta_n}] = \frac{\sigma^2}{\beta}(e^{-\beta\Delta_n} - e^{-2\beta\Delta_n})X_{i\Delta_n} + \alpha(\frac{\sigma^2}{2\beta})(1 - e^{-\beta\Delta_n})^2 \tag{11}$$

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 the conditional mean and variance conditions.

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

[14] The numerical optimisation 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.

Table 1. Diffusion Processes: Stationarity and Alternative Estimators

Diffusions	Stationarity	Estimators of $\sigma^2(\cdot)$	Estimators of $\mu(\cdot)$
BMwD$^{(a)}$	No	NONP$^{(d)}$, ML (or OLS)	ML
O-U$^{(b)}$	Yes $^{(e)}$	NONP, ML, GMM	NONP, ML, 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 $t - t_0 \to +\infty$.

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 (1997). 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.

Since the Itô diffusion process in our study is widely used in finance and in particular in modeling the short-term interest rate process, in our simulation 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 other reason we refer to the interest rate models in our study is that even though both the Brownian motion with drift process and the Ornstein-Uhlenbeck process have been applied to model equity return or currency exchange rate processes, the Cox-Ingersoll-Ross squared-root process is mainly applied to model nominal interest rate process due to its positiveness. As we will see later, the case of high frequency samples over relatively short sampling periods considered in our study are often available in the equity and currency exchange markets but not always available for interest rates, our reference to interest rate models is thus a compromise.

The data generating process (DGP) of each diffusion process is given by its transition density function and based on its Markovian property, the dynamics of the continuous sampling path is explicitly known. For the Ornstein-Uhlenbeck process and the CIR process, the starting values of the DGP are directly drawn from their marginal densities. For the Brownian motion with drift process, 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. 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 $t_0 = -500\Delta_N$ 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 number of replications for each estimator is set to be 1,000 and the sample size of observations in each replication is 1,000 or 5,000 (which is large compared to usual Monte Carlo studies due to the implementation of the nonparametric estimators in our study). The comparison of the performance of different estimators is undertaken at 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). While the estimation of diffusion models in finance based on high frequency data is of particular interest, as pointed out by Merton (1976), in this study we will compare estimators with data mimicing weekly, daily and high frequency over different sampling periods in order to look at the effect of sample size and sampling frequency on the performance of alternative estimators. In particular we consider the following three cases:

Case I: $T = 20$ and $N = 1,000$ to approximate 20 years of weekly data;

Case II: $T = 20$ and $N = 5,000$ to simulate 20 years of daily data. There being approximately 250 trading days per year;

Case III: $T = 1$ and $N = 5,000$ to simulate high frequency data of 20 observations per day for one year.

As we have mentioned, the implementation of the diffusion function estimator $\hat{\sigma}^2(x)$ requires that the observations $\{X_{i\Delta_n}\}$ are close to x and the implementation of the drift function estimator $\hat{\mu}(x)$ in (6) explicitly requires that $\hat{p}(x) \neq 0$, i.e. the marginal density function must be estimated first from the sampling path. This requires that the simulated sampling path provides enough observations that are close to the points of the process at which we intend to make inference about the drift and diffusion functions. Thus to implement the nonparametric estimators over the interval [0.05, 0.10], only the simulated sampling paths which cross the interval [0.06, 0.09] are used in our estimation. That is, we only include in our study the samples which provide sufficient information for the estimation of the diffusion process at relevant levels. [15] The same sampling paths are used for different estimators. [16] 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 minimised, the study of the optimal choice of window-width is not pursued in this paper [17].

[15] More specifically, to infer the behaviour of the short-term interest rate at the level of, say, x%, we need historical information of the short-term interest rate in the neighbourhood of x%. In our study, to infer simultaneously the behaviour of the short-term interest rate over the range of 5% to 10%, we need historical information of the short-term interest rate in a similar range, say, over 6% to 9% as chosen in our simulation.

[16] In this sense, our experiment is designed to accommodate the implementation of the nonparametric estimation procedure and thus all the results should be interpreted as conditional on such restrictions. In practical applications, since we are typically interested in estimating the dynamics of certain processes within a reasonable range of levels or the relevant points of the process, such restrictions are well justified.

[17] 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

Figures 1 to 6 plot the sample means of alternative diffusion function and drift function estimates at different X_t for each process and each case, they give a clear visual impression of how each estimator performs. We note that both the parametric and nonparametric diffusion function estimators perform very well for all processes in all three cases, and the nonparametric drift function estimator also performs reasonably well for both the Ornstein-Uhlenbeck and CIR squared-root process in all three cases. However, while the parametric drift function estimators perform reasonably well in case I and II, i.e. the weekly and daily data over 20 years, they perform very poorly in case III, i.e. the high frequency intra-daily data over relatively short sampling period.

figures 1-6 about here

Tables 2, 3 and 4 report in detail the summary statistics of the sampling distributions for each estimator of the diffusion function, including both nonparametric and parametric estimators, for each process and for each of the three cases. For all three processes in all three cases, 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 nonparametric estimate 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 major observations can be summarised as: (i) as sampling frequency increases from weekly to daily and intra-daily and sample size from 1,000 to 5,000, the performance of all diffusion estimators improves significantly, stressing the importance of high frequency sampling in the estimation of diffusion functions; (ii) however, as sampling frequency increases from daily to intra-daily but with sample size fixed at 5,000, the performance of diffusion function estimators only improves marginally, suggesting that samples with a reasonably high sampling frequency can serve as very good approximations of continuous sampling; (iii) there appears to be not much of a difference

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.

between the sampling distributions of the nonparametric diffusion function estimator and the parametric estimators for all three processes. However, we note that for the CIR squared root process, the parametric estimators are superior at all frequencies. Intuitively, in the nonparametric framework without much restrictions imposed on the diffusion term, a less simplistic diffusion function would present a greater challenge to its nonparametric functional estimator and reduces its level of performance.

Tables 5, 6 and 7 report in detail 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 nonparametric estimator of $\mu(X_t)$, at specified values of X_t, for the Ornstein-Uhlenbeck process and the Cox-Ingersoll-Ross squared-root process. The summary statistics of the parametric estimators of the drift function $\mu(X_t)$ at specified values of X_t are also calculated based on the estimates of α and β in each replication, which are not reported to save space but available upon request. It is clear that (i) for cases I and II, namely the relatively low frequency data (i.e. weekly or daily) over long sampling period (20 years), both the parametric and nonparametric estimators for the drift function perform quite well, and not surprisingly with the parametric estimators often performing much better; (ii) as the sample size increases from 1,000 in case I to 5,000 in case II and the sampling frequency accordingly increases from weekly to daily while keeping the sampling period fixed at 20 years, the performance of drift function estimators only have slight improvement, which is in contrast to the performance of diffusion function estimators, suggesting that the sampling frequency is less relevant to the estimation of drift functions; (iii) most interestingly, in case III, i.e. the case of high frequency intra-daily data over relatively short sampling period (one year), while the nonparametric estimator performs reasonably well, the parametric estimators all perform rather poorly, stressing the importance of long sampling periods in the parametric estimation of drift functions. It is noted that the estimators of α and β both perform poorly, with that of β performs extremely poorly. From the summary statistics of the parametric drift functions estimators, it appears that they are plagued by outliers, due to the poor estimation of α and β and the compounding of their errors in the calculation of $\mu(X_t)$.

Some further analysis of the simulation results of the diffusion function and drift function

Table 2. Brownian Motion with Drift: Simulation Results of the Diffusion Estimators

(1,000 replications, $\sigma^2(X_t) = \sigma^2 = 0.0004$): Case I: $T = 20$, $N = 1,000$ (weekly data); Case II: $T = 20$, $N = 5,000$ (daily data); Case III: $T = 1$, $N = 5,000$ (20 observations per day).

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

Case	X_t	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-10})	M.S.E. (10^{-10})	Skewness (10^{-1})	Kurtosis (10^{-2})	Wilcoxon Test$^{(a)}$ (10^{-1})
I	0.05	3.993	3.997	3.647	3.648	1.158	3.746	6.957 (+)
	0.06	3.993	3.997	3.647	3.648	1.158	3.746	6.958 (+)
	0.07	3.993	3.997	3.647	3.648	1.158	3.746	6.957 (+)
	0.08	3.993	3.997	3.647	3.648	1.158	3.746	6.957 (+)
	0.09	3.993	3.997	3.647	3.648	1.158	3.746	6.957 (+)
	0.10	3.993	3.997	3.647	3.648	1.158	3.746	6.957 (+)
II	0.05	3.999	3.998	0.598	0.599	-0.598	1.093	5.133 (+)
	0.06	3.999	3.998	0.598	0.599	-0.598	1.093	5.138 (+)
	0.07	3.999	3.998	0.598	0.599	-0.598	1.093	5.140 (+)
	0.08	3.999	3.998	0.598	0.599	-0.598	1.093	5.138 (+)
	0.09	3.999	3.998	0.598	0.599	-0.598	1.093	5.137 (+)
	0.10	3.999	3.998	0.598	0.599	-0.598	1.093	5.135 (+)
III	0.05	3.998	3.997	0.588	0.590	-0.587	1.320	0.878 (+)
	0.06	3.998	3.997	0.588	0.590	-0.587	1.320	0.878 (+)
	0.07	3.998	3.997	0.588	0.590	-0.587	1.320	0.878 (+)
	0.08	3.998	3.997	0.588	0.590	-0.587	1.320	0.877 (+)
	0.09	3.998	3.997	0.588	0.590	-0.587	1.320	0.879 (+)
	0.10	3.998	3.997	0.588	0.590	-0.587	1.320	0.879 (+)

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

Case	Method (10^{-4})	Median (10^{-4})	Mean (10^{-10})	Variance (10^{-10})	M.S.E. (10^{-1})	Skewness (10^{-2})	Kurtosis
I	MLE	3.984	3.991	3.350	3.358	1.343	4.149
II	MLE	3.999	3.997	0.596	0.597	-0.586	1.581
III	MLE	4.000	3.998	0.588	0.589	-0.571	1.108

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

Hypothesis H_0:	Case	Test Statistic$^{(b)}$	Critical Value (5%)
Samples of Nonparametric Estimates at Different Values of X_t Have the Same Median	I	1.176	16.92
	II	0.601	16.92
	III	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.

estimators are also reported in Tables 2-7. For instance, for the Cox-Ingersoll-Ross squared-root process in case III (i.e. the high frequency intra-daily data), when compared to the normal density, the sampling distribution of the non-linear least square (NLS) estimator appears slightly skewed to the left, while that of the generalised 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 generalised method of moments (GMM) estimator appears to be slightly more concentrated, and those of the non-linear least square (NLS) estimator and the nonparametric appear 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 analyse 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 nonpara-

Table 3. Ornstein-Uhlenbeck Process: Simulation Results of the Diffusion Estimators

1,000 replications, $\sigma^2(X_t) = \sigma^2 = 0.0004$; Case I: $T = 20$, $N = 1,000$; Case II: $T = 20$, $N = 5,000$; Case III: $T = 1$, $N = 5,000$.

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

Case	X_t	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-10})	M.S.E. (10^{-10})	Skewness (10^{-1})	Kurtosis (10^{-2})	Wilcoxon Test$^{(a)}$
I	0.05	3.998	3.984	6.342	6.368	1.304	2.181	1.747 (+)
	0.06	3.997	3.984	6.339	6.365	1.303	2.501	1.751 (+)
	0.07	3.997	3.984	6.337	6.363	1.303	2.830	1.758 (+)
	0.08	3.996	3.984	6.334	6.361	1.302	3.165	1.768 (+)
	0.09	3.995	3.984	6.333	6.359	1.302	3.508	1.777 (+)
	0.10	3.995	3.984	6.331	6.358	1.302	3.858	1.767 (+)
II	0.05	3.995	3.994	0.892	0.896	0.348	-1.198	1.569 (+)
	0.06	3.995	3.994	0.892	0.895	0.340	-0.357	1.594 (+)
	0.07	3.995	3.994	0.891	0.895	0.331	-0.532	1.521 (+)
	0.08	3.995	3.994	0.891	0.895	0.322	1.460	1.530 (+)
	0.09	3.995	3.994	0.892	0.896	0.312	2.415	1.534 (+)
	0.10	3.995	3.994	0.893	0.897	0.301	3.385	1.523 (+)
III	0.05	3.999	3.997	0.587	0.588	-0.586	1.568	1.437 (+)
	0.06	3.999	3.997	0.587	0.588	-0.584	1.659	1.441 (+)
	0.07	3.999	3.997	0.587	0.588	-0.582	1.752	1.439 (+)
	0.08	3.999	3.997	0.587	0.588	-0.580	1.847	1.437 (+)
	0.09	3.999	3.997	0.587	0.588	-0.577	1.943	1.433 (+)
	0.10	3.999	3.997	0.587	0.588	-0.574	2.040	1.430 (+)

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

Case	Method	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-10})	M.S.E. (10^{-10})	Skewness (10^{-1})	Kurtosis (10^{-2})	Wilcoxon Test$^{(a)}$
I	NLS	3.840	3.985	4.768	4.791	1.058	1.806	
	GMM	3.870	3.991	4.406	4.414	1.006	1.142	-1.934 (+)
II	NLS	3.930	3.994	0.907	0.911	-0.112	-1.073	
	GMM	3.940	3.995	0.762	0.765	-0.026	-0.552	-1.510 (+)
III	NLS	3.944	3.999	0.714	0.714	0.153	1.852	
	GMM	3.942	3.995	0.772	0.779	-0.077	2.602	-1.734 (+)

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

Hypothesis H_0:	Case	Test Statistic^b	Critical Value (5%)
Samples of Nonparametric Estimates at Different Values of X_t have the Same Median	I	6.348	16.92
	II	0.932	16.92
	III	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, $\sigma^2(X_t) = \sigma^2 X_t = 0.007X_t$; Case I: $T = 20$, $N = 1,000$; Case II: $T = 20$, $N = 5,000$; Case III: $T = 1$, $N = 5,000$.

a. Summary Statistics of the Nonparametric Estimator of $\sigma^2(X_t)$ at Different Values of X_t

Case	True Value X_t ($\sigma^2(X_t)10^{-4}$)	Median (10^{-4})	Mean (10^{-4})	Variance (10^{-8})	M.S.E. (10^{-8})	Skewness (10^{-1})	Kurto-sis	Wilcoxon test[(a)]
I	0.05 (3.500)	3.712	3.896	1.570	1.727	8.684	1.308	-8.52 (–)
	0.06 (4.200)	4.331	4.511	1.572	1.669	8.690	1.301	-6.04 (–)
	0.07 (4.900)	5.043	5.229	1.573	1.605	8.706	1.296	-4.66 (–)
	0.08 (5.600)	5.553	5.751	1.573	1.595	8.730	1.294	6.79 (–)
	0.09 (6.300)	6.185	6.377	1.571	1.576	8.756	1.295	8.37 (–)
	0.10 (7.000)	7.314	7.507	1.567	1.874	8.776	1.298	9.79 (–)
II	0.05 (3.500)	3.400	3.497	0.647	0.647	7.876	1.220	-5.27 (–)
	0.06 (4.200)	4.061	4.164	0.629	0.631	8.152	1.267	-3.65 (–)
	0.07 (4.900)	4.840	4.952	0.609	0.612	8.575	1.343	-1.11 (+)
	0.08 (5.600)	5.251	5.385	0.585	0.644	9.082	1.461	3.17 (–)
	0.09 (6.300)	6.248	6.381	0.560	0.567	9.504	1.611	5.16 (–)
	0.10 (7.000)	6.803	6.916	0.539	0.547	9.563	1.753	5.51 (–)
III	0.05 (3.500)	3.864	3.879	0.077	0.221	0.395	-0.437	-5.74 (–)
	0.06 (4.200)	4.320	4.342	0.055	0.075	5.356	0.633	-6.07 (–)
	0.07 (4.900)	4.934	4.958	0.038	0.041	9.373	2.438	-1.37 (+)
	0.08 (5.600)	5.651	5.668	0.056	0.060	8.012	2.495	-1.47 (+)
	0.09 (6.300)	6.307	6.323	0.145	0.145	6.885	1.446	-1.84 (+)
	0.10 (7.000)	7.016	7.083	0.216	0.223	3.469	-4.528	-2.39 (–)

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

Case	Method	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-7})	M.S.E. (10^{-7})	Skewness (10^{-1})	Kurtosis (10^{-2})	Wilcoxon Test[(a)]
I	NLS	6.374	6.745	2.831	3.480	1.014	8.716	2.28 (+)
	GMM	6.655	6.942	1.644	1.678	1.173	8.627	
II	NLS	6.778	6.950	0.691	0.716	0.148	7.401	2.04 (+)
	GMM	6.875	6.997	0.323	0.323	0.447	-2.249	
III	NLS	6.795	6.945	0.577	0.607	-0.778	4.841	1.71 (+)
	GMM	6.856	6.982	0.342	0.345	0.178	-0.187	

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

Hypothesis H_0:	Case	Test Statistic[(b)]	Critical Value (5%)
Sample of Nonparametric Estimates at Different Values of X_t have the Same Median	I	64.65	16.92
	II	10.17	16.92
	III	11.61	16.92

Note: (a) See Table 2 note (a). 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 estimator, and the null hypothesis of the pair-wised Wilcoxon test is against the GMM estimate; (b) See Table 2 note (b).

Table 5. Brownian Motion with Drift: Simulation Results of the Drift Estimators

1,000 replications, $\mu(X_t) = \mu = 0.0055$; Case I: $T = 20, N = 1,000$; Case II: $T = 20, N = 5,000$; Case III: $T = 1, N = 5,000$.

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

Case	Method	Min (10^{-2})	Max (10^{-2})	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-5})	MSE (10^{-5})	Skewness (10^{-2})	Kurtosis (10^{-1})
I	MLE	-0.895	1.758	5.533	5.632	2.176	2.178	-2.905	-1.336
II	MLE	-1.075	1.754	5.398	5.454	2.016	2.017	-6.525	0.344
III	MLE	-6.689	8.149	4.254	4.144	41.32	41.47	-8.398	0.537

metric 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 except for case I of the CIR squared-root process.

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

Remark 1. While it is not surprising to see drift parameter estimates perform badly with samples over short sampling periods, it may be surprising to note the excellent performance of the nonparametric estimators, especially the drift estimator in case III. In response to this observation we make the following points. 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. In fact, even though all the parametric drift function parameter estimators are consistent in our comparison, the asymptotic results are derived under explicitly the condition that the whole sampling period $T \to \infty$ instead of the number of sampling observations $N \to \infty$. [18]

[18]The consistency and asymptotic normality of the drift coefficient estimators are obtained as $\Delta_n \to 0$ and $T = n\Delta_n \to \infty$ with the additional condition that $n\Delta_n^3 \to 0$. Their rate of convergence is $\sqrt{n\Delta_n}$ which means the sampling period T must be quite large, see e.g. Prakasa Rao (1988).

Table 6. Ornstein-Uhlenbeck Process: Simulation Results of the Drift Estimators

1,000 replications, $\mu(X_t) = \beta(\alpha - X_t)$, $\beta = 0.18$, $\alpha = 0.086$; Case I: $T = 20$, $N = 1,000$; Case II: $T = 20$, $N = 5,000$; Case III: $T = 1$, $N = 5,000$.

a. Summary Statistics of the Nonparametric Estimator of $\mu(X_t)$ at Different X_t:

Case	True Value $X_t(\mu(X_t)$ $(10^{-3}))$	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-6})	M.S.E. (10^{-6})	Skewness	Kurtosis
I	0.05 (6.480)	5.695	6.069	2.492	2.661	1.318	1.912
	0.06 (4.680)	4.322	4.517	1.555	1.582	1.068	1.296
	0.07 (2.880)	2.355	2.366	0.945	1.210	0.598	0.223
	0.08 (1.080)	0.764	0.714	0.658	0.792	0.045	-0.864
	0.09 (-0.72)	-0.973	-0.937	0.693	0.741	-0.141	-1.005
	0.10 (-2.52)	-2.299	-2.290	1.054	1.107	-0.316	-0.539
II	0.05 (6.480)	5.302	5.546	1.371	2.243	1.390	3.320
	0.06 (4.680)	4.021	4.184	0.924	1.169	1.103	2.166
	0.07 (2.880)	2.194	2.223	0.645	1.076	0.578	0.558
	0.08 (1.080)	0.817	0.762	0.533	0.634	-0.106	-0.482
	0.09 (-0.72)	-0.569	-0.700	0.589	0.589	-0.597	-0.286
	0.10 (-2.52)	-1.673	-1.861	0.813	1.246	-0.841	0.314
III	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

b. Summary Statistics of the Parametric Estimators of α and β:

Case	Method		Min (10^{-1})	Max	Median (10^{-1})	Mean (10^{-1})	Variance (10^{-2})	MSE (10^{-2})	Skew-ness	Kurto-sis
I	NLS	$\hat{\alpha}$	0.09	1.87	0.708	0.958	0.117	0.118	12.64	176.7
		$\hat{\beta}$	-1.80	1.87	1.017	1.790	1.170	1.171	1.015	0.946
	GMM	$\hat{\alpha}$	0.06	0.52	0.701	0.880	0.122	0.123	4.136	39.59
		$\hat{\beta}$	0.10	1.74	0.802	1.400	0.710	0.901	1.230	1.838
II	NLS	$\hat{\alpha}$	0.06	0.42	0.697	0.867	0.109	0.110	3.228	27.16
		$\hat{\beta}$	-1.28	1.94	1.025	1.852	1.336	1.340	1.216	1.839
	GMM	$\hat{\alpha}$	-0.17	0.25	0.691	0.846	0.068	0.069	2.801	27.18
		$\hat{\beta}$	-0.42	1.54	0.788	1.408	0.737	0.891	1.229	1.744
III	NLS	$\hat{\alpha}$	-2.80	5.60	0.057	0.079	0.803	0.808	8.907	195.8
		$\hat{\beta}$	-2.12	29.1	2.653	5.487	17.059	45.227	1.293	2.654
	GMM	$\hat{\alpha}$	-3.26	7.76	0.058	0.111	0.339	0.340	12.89	195.8
		$\hat{\beta}$	-0.92	24.9	2.514	5.656	19.298	49.289	1.248	1.544

Table 7. Cox-Ingersoll-Ross SR Process: Simulation Results of the Drift Estimators

1,000 replications, $\mu(X_t) = \beta(\alpha - X_t)$, $\beta = 0.23$, $\alpha = 0.076$; Case I: $T = 20$, $N = 1,000$; Case II: $T = 20$, $N = 5,000$; Case III: $T = 1$, $N = 5,000$.

a. Summary Statistics of the Nonparametric Estimator of $\mu(X_t)$ at Different X_t

Case	True Value $X_t(\mu(X_t)(10^{-3}))$	Median (10^{-3})	Mean (10^{-3})	Variance (10^{-6})	M.S.E. (10^{-6})	Skewness (10^{-1})	Kurtosis
I	0.05 (5.980)	4.520	4.616	6.617	8.476	5.062	-0.655
	0.06 (3.680)	3.022	2.976	2.421	2.916	4.446	-0.801
	0.07 (1.380)	0.621	0.759	1.044	1.429	2.122	-1.409
	0.08 (-0.92)	0.308	0.304	1.003	2.501	-0.170	-1.268
	0.09 (-3.22)	-2.895	-2.678	1.650	1.943	-1.367	-0.946
	0.10 (-5.52)	-4.866	-4.642	2.679	3.450	-2.147	-0.705
II	0.05 (5.980)	4.595	5.273	5.816	6.315	7.638	-0.318
	0.06 (3.680)	3.470	3.648	2.116	2.117	6.633	-0.104
	0.07 (1.380)	1.510	1.365	1.127	1.128	-8.454	-1.090
	0.08 (-0.92)	0.664	0.774	1.620	4.491	-7.947	-1.498
	0.09 (-3.22)	-2.773	-2.406	3.028	3.745	0.739	-1.297
	0.10 (-5.52)	-5.107	-4.625	5.293	6.094	0.837	-1.207
III	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	8.466	6.632
	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	-5.488	10.078
	0.10 (-5.52)	-3.394	-3.954	9.871	12.323	-6.403	10.949

b. Summary Statistics of the Parametric Estimators of α and β:

Case	Method		Min (10^{-2})	Max	Median (10^{-1})	Mean (10^{-1})	Variance (10^{-2})	M.S.E. (10^{-2})	Skew-ness	Kurto-sis
I	NLS	$\hat{\alpha}$	0.18	1.66	0.556	0.781	0.091	0.092	2.467	17.46
		$\hat{\beta}$	0.62	1.93	0.866	1.621	1.001	1.468	-1.106	1.648
	GMM	$\hat{\alpha}$	1.73	0.30	0.595	0.765	0.106	0.107	1.768	8.044
		$\hat{\beta}$	2.56	1.65	0.896	1.470	0.640	1.328	-1.116	1.846
II	NLS	$\hat{\alpha}$	0.70	1.86	0.547	1.065	0.432	0.442	6.318	41.35
		$\hat{\beta}$	-9.58	1.93	0.040	1.146	1.055	2.386	-1.343	2.544
	GMM	$\hat{\alpha}$	3.30	0.53	0.599	0.765	0.096	0.097	6.396	75.20
		$\hat{\beta}$	1.32	1.53	0.920	1.481	0.622	1.293	-1.061	1.570
III	NLS	$\hat{\alpha}$	-1.05	6.09	0.056	0.238	1.533	1.559	7.911	124.3
		$\hat{\beta}$	-7.36	39.9	1.884	5.688	30.971	60.76	1.533	4.206
	GMM	$\hat{\alpha}$	-2.06	5.16	0.054	0.081	0.941	0.944	6.904	109.7
		$\hat{\beta}$	-5.49	26.3	1.752	4.355	14.606	31.61	1.568	3.657

Thirdly, as we have pointed out, the simulation experiment is designed to accommodate the implementation of the nonparametric estimation procedure proposed in Jiang and Knight (1997) based on certain restricted sample paths, thus the simulation results should be interpreted as conditional on the restrictions imposed on the sampling paths. 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. Furthermore, 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. It is also interesting to note that estimation of $\hat{\mu}(x)$ only requires a locally well performed marginal density function estimator $\hat{p}(x)$ not necessarily a globally well performed $\hat{p}(x)$ due to its nonparametric specification.

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. As we have mentioned, there have been various methods proposed for the estimation of diffusions based on "indirect inference", and it is shown that (see Gallant and Long (1996)) EMM provides asymptotically efficient estimates of the diffusion parameters. Moreover, the "indirect inference" can reduce the small sample bias existing in standard parametric estimators (see Gouriéroux, Renault and Touzi (1994)). Estimation based on this approach is not included in our simulation study for the reason that the performance of various "indirect inference" estimators depends critically on the choice of the auxiliary model or score generator on the one hand, and they all involve intensive computational burden in each simulation and estimation on the other. Nonetheless, it is worthwhile to investigate their performance in correcting estimation biases of the parametric drift estimators based on samples over short time period.

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 minimise the integrated mean squared error (IMSE) of the estimator [19].

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 the estimator. Whether the above admissible window-width represents the achievable optimal rate of convergence is unknown to us. It is thus clear that

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

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 minimising 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 minimise the numerical IMSE [20].

[20] 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 minimising 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 minimises 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.

5 Conclusion

In this paper, we consider alternative approaches to the estimation of diffusion processes from discretely sampled observations. Through Monte Carlo simulation, we investigate the finite sample properties of different estimators and in particular compare the performance of different estimators based on certain samples generated with three different sampling frequencies over different sampling periods. The simulation results show that both the nonparametric diffusion function and drift function estimators proposed in Jiang and Knight (1997) perform reasonably well. However, with certain high frequency data sampled over relatively short sampling periods, while 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 the nonparametric identification and estimation procedure, proposed for the Itô diffusion process in Jiang and Knight (1997), in which the diffusion function is identified and estimated first 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. As noted, an interesting future research topic would be to investigate the finite sample properties of the "indirect inference" estimators of drift parameters based on short sampling period.

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 Optimisation* 16, 380-95.

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

Boyle, P., M. Broadie, & P. Glasserman, 1996, Monte Carlo methods for security pricing, forthcoming in *Journal of Economics Dynamics and Control.*

Bibby, M. and M. Sørensen, 1995, Martingale estimation function for discretely observed diffusion process, *Bernoulli* 1(1), 17-39.

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

Brown, B.M. & J.I. Hewitt,1975, Asymptotic likelihood theory for diffusion process. *Journal of Applied Probability* 12, 228-238.

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.

Conley, T.G., L.P. Hansen, E.G.J. Luttmer and J.A. Scheinkman, 1997, Short-Term Interest Rates as Subordinated Diffusions, *Review of Financial Studies*, 10(3), 525-577.

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).

Dorogovcev, A. Ja., 1976, The consistency of an Estimate of a Parameter of a Stochastic Differential Equation, *Theoretical Probability and Mathematical Statistics* 10, 73-82.

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.

Gallant, A.R. and J.R. Long, 1997, Estimating stochastic differential equations efficiently by minimum chi-square, *Biometrika* 84, 125-141.

Gallant, A.R. & G. Tauchen, 1996, Which moments to match, *Econometric Theory* 12, 657-681.

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

Genon-Catalot, V., 1990, Maximum contrast estimation for diffusion processes from discrete observations, *Statistics* 21, 99-116.

Genon-Catalot, V. and J. Jacod, 1993, On the estimation of the diffusion coefficient for multi-dimensional diffusion processes, University of Marne-La-Vallée, France.

Genon-Catalot, V., C. Laredo, and D. Picard, 1992, Nonparametric estimation of the Diffusion Coefficient by Wavelets Methods, *Scandinavian Journal of Statistics* 19, 317-335.

Gouriéroux, C., A. Monfort, & E. Renault, 1993, Indirect Inference, *Journal of Applied Econometrics* 8, S85-S118.

Gouriéroux, C., E. Renault, & N. Touzi, 1994, Calibration by simulation for small sample bias correction, Discussion paper CREST.

Hansen, L.P., 1982, Large sample properties of generalised 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).

Jacod, J., 1993, Random Sampling in estimation problems for continuous Gaussian processes with independent increments, *Stochastic Processes and their Applications* 44, 181-204.

Jiang, G.J., 1997, Nonparametric modeling of U.S. term structure of interest rates and implications on the prices of derivative securities, Forthcoming *Journal of Financial and Quantitative Analysis*.

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

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.

Kasonga, R.A., 1988, The consistency of a non-linear least squares estimator from diffusion processes, *Stochastic Processes and their Applications* 30, 263-275.

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

Kutoyants, Y.,1984, *Parameter Estimation for Stochastic Processes*. Heldermann.

Lanska, V.,1979, Minimum contrast estimation in diffusion processes. *Journal of Applied Probability* 16, 65-75.

Laredo, C., 1990, A sufficient condition for asymptotic sufficiency of incomplete observations of a diffusion process, *Annals of Statistics* 18, 1158-1171.

Lo, A.W., 1988, Maximum likelihood estimation of generalised 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.

......, 1976, The impact on option pricing of specification error in the underlying stock price returns, *Journal of Finance* 31, 333-350.

......, 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.

Pham Dinh, T.D.,1981, Non-parametric estimator of the drift coefficient in the diffusion equation. *Math. Operations Statist. Ser Statist.* 12, 61-73.

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

Prakasa Rao, B.L.S., 1988, Statistical inference from sampled data for stochastic processes, *Contemporary Mathematics* 80, 249-284.

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., 1997, A nonparametric model of term structure dynamics and the market price of interest rate risk, *Journal of Finance* 52, 1973-2002.

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.

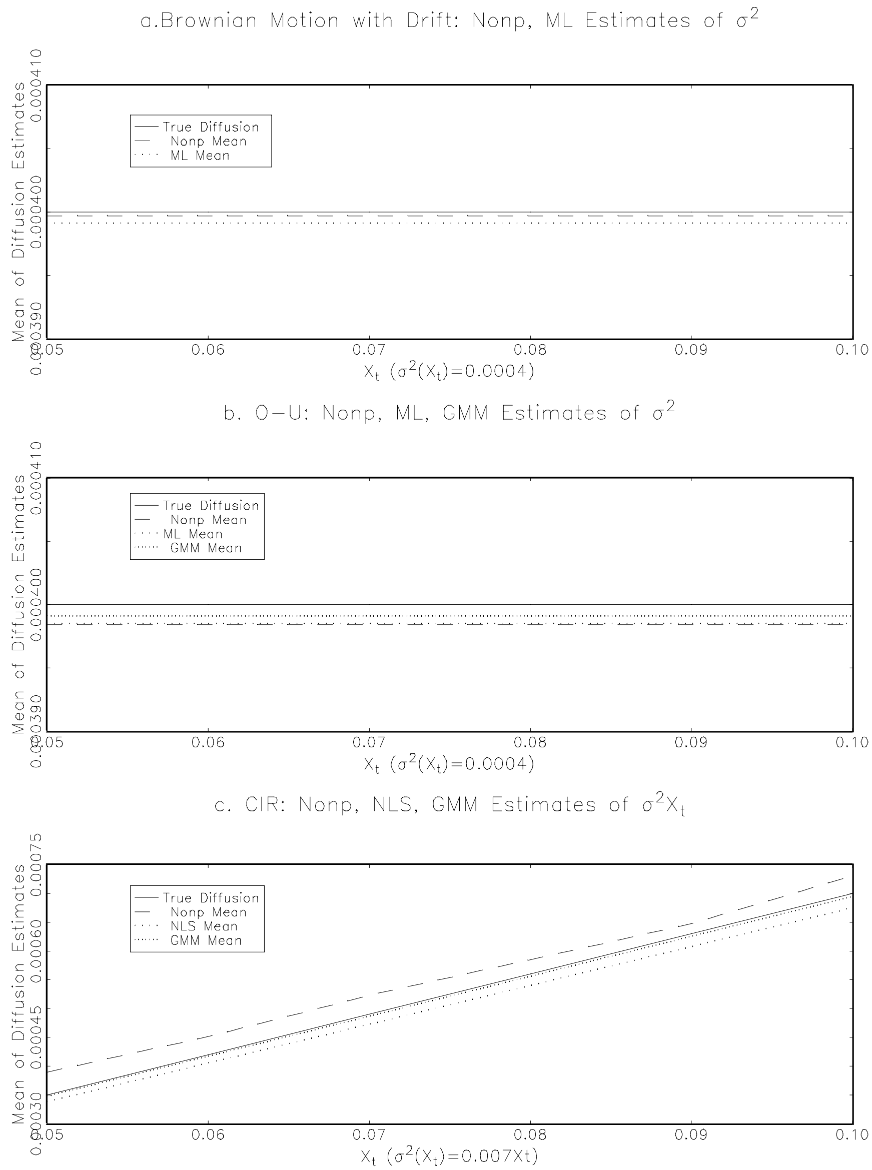

Fig. 1. Case I: Sample Means of Diffusion Estimates at Different X_t
(1,000 replications with sample size $N = 1,000$ and sampling period $T = 20$)

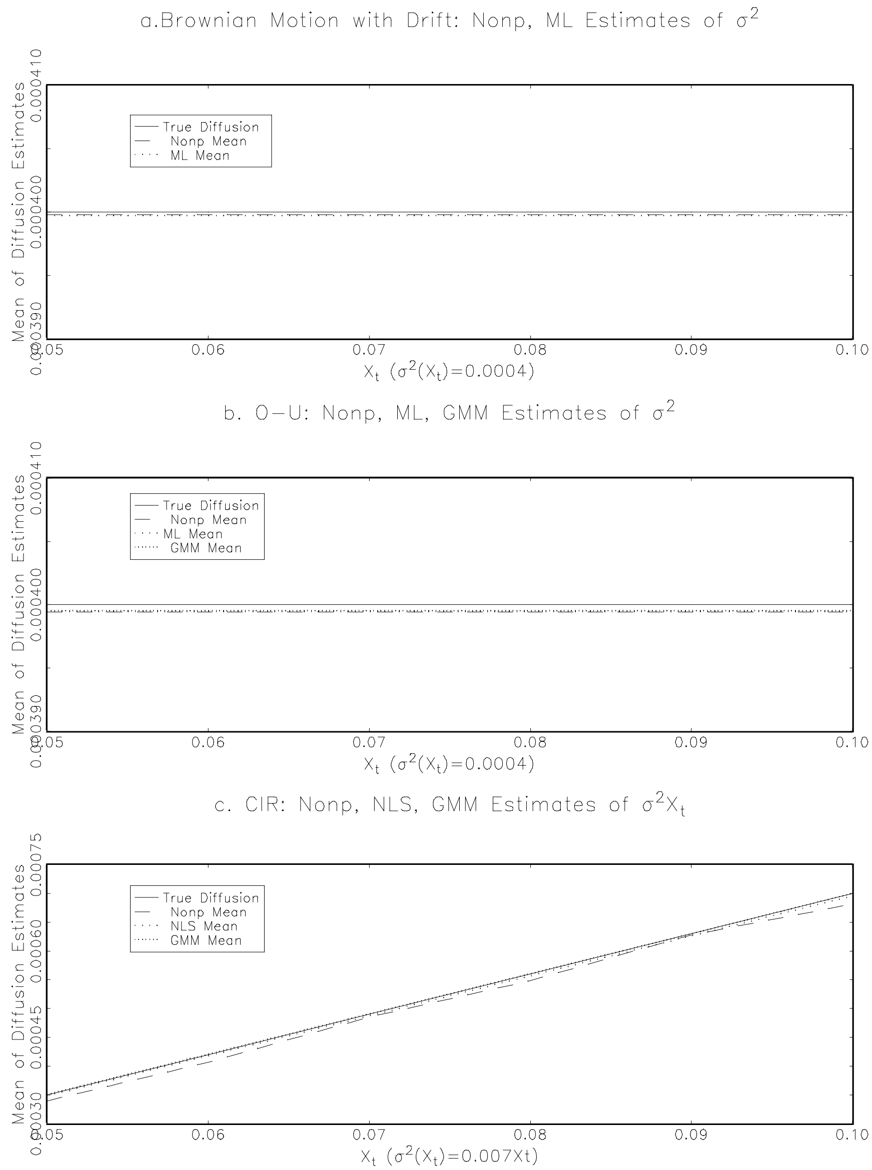

Fig. 2. Case II: Sample Means of Diffusion Estimates at Different X_t
(1,000 replications with sample size $N = 5,000$ and sampling period $T = 20$)

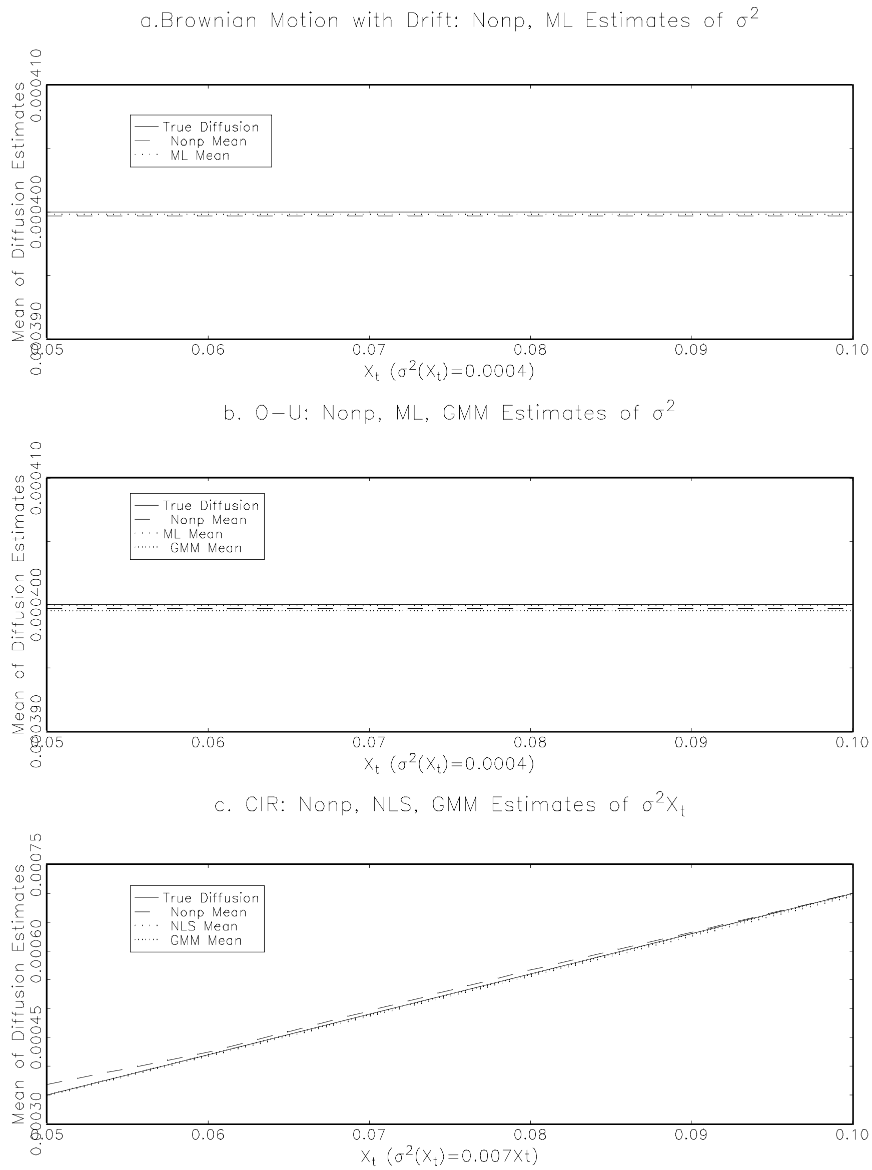

Fig. 3. Case III: Sample Means of Diffusion Estimates at Different X_t
(1,000 replications with sample size $N = 5,000$ and sampling period $T = 1$)

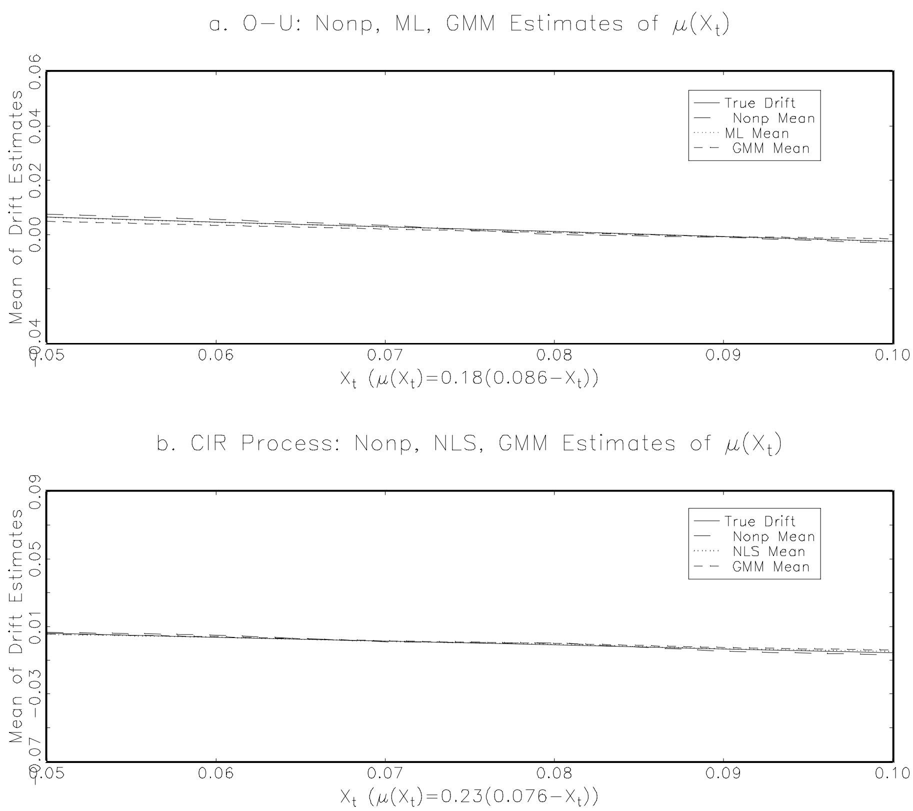

Fig. 4. Case I: Sample Means of Drift Estimates at Different X_t
(1,000 replications with sample size $N = 1,000$ and sampling period $T = 20$)

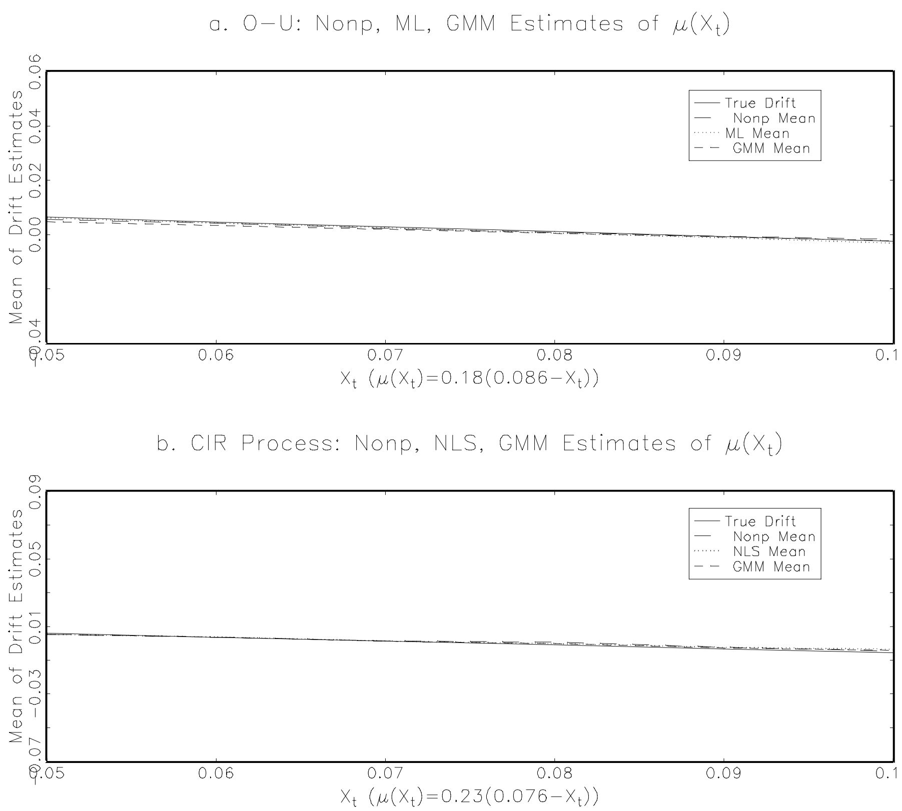

Fig. 5. Case II: Sample Means of Drift Estimates at Different X_t
(1,000 replications with sample size $N = 5,000$ and sampling period $T = 20$)

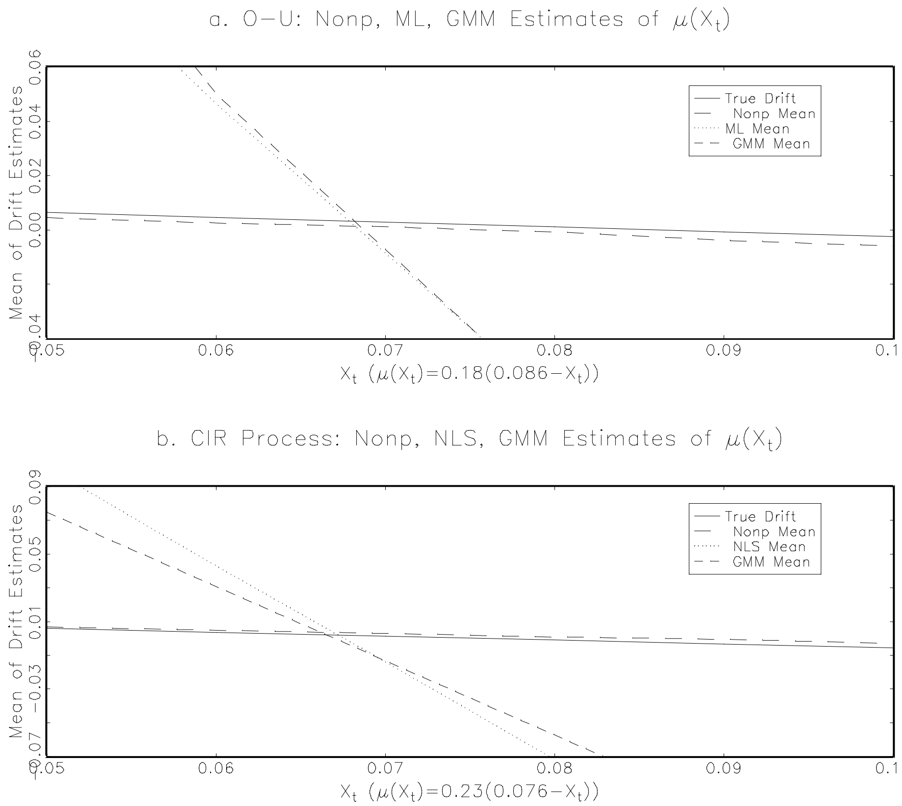

Fig. 6. Case III: Sample Means of Drift Estimates at Different X_t
(1,000 replications with sample size $N = 5,000$ and sampling period $T = 1$)