COMPARATIVE STUDY OF ESTIMATION METHODS FOR CONTINUOUS TIME STOCHASTIC PROCESSES

BY ISAO SHOJI AND TOHRU OZAKI

University of Tsukuba, Ibaraki and Institute of Statistical Mathematics, Tokyo

First version received April 1996

Abstract. In this paper we investigate the finite sample performances of five estimation methods for a continuous-time stochastic process from discrete observations. Applying these methods to two examples of stochastic differential equations, one with linear drift and state-dependent diffusion coefficients and the other with nonlinear drift and constant diffusion coefficients, Monte Carlo experiments are carried out to evaluate the finite sample performance of each method. The Monte Carlo results indicate that the differences between the methods are large when the discrete-time interval is large. In addition, these differences are noticeable in estimations of the diffusion coefficients.

Keywords. Stochastic differential equation; discretization; maximum likelihood estimation; generalized method of moments.

1. INTRODUCTION

To describe the time evolution of dynamic phenomena, it is often convenient to use a continuous-time stochastic process as a statistical model of dynamics. When making a statistical inference based on a continuous-time model, we must first estimate the parameters of the model from real data which are almost always sampled with a regular time interval. Hence, without an appropriate method for linking a continuous-time model and discrete-time data, it is difficult to estimate a continuous-time model. Many statisticians have paid much attention to estimating a continuous-time model from discrete-time data. Specifically, studies on estimation of a continuous-time model are rooted in Phillips (1959), Bergstrom (1976) and Robinson (1976), where continuous-time linear models are considered. Recently, Singer (1993) proposed an alternative estimation method for continuous-time linear models, and Hansen and Scheinkman (1995) proposed a new estimation method applicable not only to linear continuous-time models but also to nonlinear models. Although various methods to estimate a continuous-time model from discrete-time data have been proposed, there has been little study on a comparison of the estimation performance of each method. On evaluating estimation performance, we may first recall the asymptotic properties of the estimator derived from each estimation method. For example,
Yoshida (1992) and Prakasa Rao (1983) investigate the asymptotic properties of the maximum likelihood estimator based on the likelihood ratio and the least squares estimator, respectively. Time series data, however, are often of finite length; we should therefore pay attention to the finite sample performance from a practical point of view. In fact, practitioners have much interest in the performance for the case where both the time interval and the number of data are finite. A finite interval and finite observations cannot meet the requirements for asymptotic theory.

The aim of this paper is to compare the finite sample performance of the methods popularly used in some fields—engineering, mathematics and econometrics. The methods considered in this paper are classified into two categories, the maximum likelihood method and the generalized method of moments (GMM) developed by Hansen (1982).

When estimating parameters by the maximum likelihood method, one approach is to use first a discretization method to link a continuous-time stochastic process with the discrete-time data. A continuous-time stochastic process can be easily discretized by applying a conventional discretization method used for deterministic differential equations to the stochastic differential equation of the process. Furthermore, once a process is discretized, the transition probability distribution of the discretized process may be obtained immediately. For example, when applying a conventional method, say the Euler method, to the stochastic differential equation, the transition probability distribution of the discretized process is normal. Thus, by using this normality of the discretized process together with the Markov property, the log-likelihood can be easily obtained. The other discretization methods recently developed by Ozaki (1985) and Shoji and Ozaki (1994) also have the property that the transition probability distribution is normal, and so the log-likelihood can be easily obtained.

From another viewpoint, Kutoyants (1984) and Yoshida (1992) propose an alternative approach based on the maximum likelihood method in which parameters are estimated by maximizing the likelihood ratio instead of the likelihood itself. It is a key point that this method uses the exact likelihood ratio. Of course, since the likelihood ratio is written in a continuous-time framework, the quantities comprising the likelihood ratio must be approximated by a discretization method to link the continuous-time process with discrete observations. Even though using the discrete approximation, Yoshida (1992) shows that the maximum likelihood estimator has consistency and asymptotic normality as the time interval goes to zero and the number of data goes to infinity. Comparing the above two approaches, maximum likelihood estimation based on the discrete approximate process and maximum likelihood estimation based on the discrete approximate likelihood ratio, the former method uses the exact likelihood of a discretized process and the latter uses the approximate likelihood ratio in the sense that the original quantities are replaced by the quantities derived from a discretized process. However, interestingly, it can be shown that the estimator derived from the Euler method and the estimator

© Blackwell Publishers Ltd 1997
derived from the maximum likelihood ratio are equivalent when the diffusion coefficient is constant (see Shoji, 1995).

A GMM estimator is obtained by minimizing a criterion function of sample moments which are derived from orthogonality conditions (see Hansen, 1982). The GMM has an advantage over the maximum likelihood method, particularly in practical situations, since no prior knowledge about the transition distribution of a stochastic process is assumed. In addition, the GMM is applicable to a wide class of econometric models, for example, whose disturbances show heteroskedasticity and autocorrelation. Thus, the GMM is widely used in the literature related to the estimation of linear and nonlinear econometric models.

When studying the finite sample performance of an estimator, we can choose the analytical approach. However, it is often the case that we have analytical difficulty in deriving the finite sample properties of an estimator. In this paper, we compare the performance of each estimation method through Monte Carlo simulation.

Considering the fact that real data exhibit many kinds of behavior with contributions from many different frequencies, the discrete-time sampling interval is also an important aspect in evaluating the performance of an estimator; it is important to check whether an estimator remains effective as the time interval becomes large, because it is probable that the discretization approximation becomes worse as the time interval becomes large. In this paper, we also examine the influence of time interval on estimation performance.

The organization of the paper is as follows. Section 2 presents five estimation methods, four representing maximum likelihood approaches and one the GMM approach. In Section 3, applying these methods to two examples of stochastic differential equations, one with linear drift and state-dependent diffusion coefficients and the other with nonlinear drift and constant diffusion coefficients, Monte Carlo experiments are carried out and the performances of the five methods are compared. Some conclusions are presented in Section 4.

2. ESTIMATION METHODS FOR A CONTINUOUS-TIME PROCESS

In this section we describe how the parameters of a continuous-time stochastic process are estimated by the alternative methods. We focus on a one-dimensional continuous-time stochastic process which is characterized by a one-dimensional stochastic differential equation (SDE). First, we consider the following somewhat general SDE in the sense that the diffusion term is dependent on the state of the process:

\[ dx_t = f(x_t) \, dt + g(x_t) \, dB_t \]  

where \( f(\cdot) \) and \( g(\cdot) \) are twice continuously differentiable functions of the process \( x_t \), and \( B_t \) is a standard Brownian motion. This SDE can be transformed into an SDE with constant diffusion term by Ito's formula. Let \( \phi(\cdot) \) be a twice continuously differentiable function of \( x \) such that

© Blackwell Publishers Ltd 1997
\[ \phi'(x) g(x) = \sigma \]  \hspace{1cm} (2)

where \( \sigma \) is constant. From Ito’s formula, the new process \( y_t \equiv \phi(x_t) \) satisfies the SDE

\[ dy_t = \left( \frac{g^2}{2} \phi'' + f \phi' \right) dt + \phi' g \ dB_t \]

\[ = \left( \frac{g^2}{2} \phi'' + f \phi' \right) dt + \sigma \ dB_t. \hspace{1cm} (3) \]

Thus, from Ito’s formula, we have only to consider the SDE

\[ dx_t = f(x_t) \ dt + \sigma \ dB_t. \hspace{1cm} (4) \]

Hereafter, we describe how to use the five estimation methods to estimate the parameters of the above SDE.

2.1. The Euler method

The Euler method assumes \( f(x_t) \) of (4) to be piecewise constant; e.g. \( f(x_t) \) is constant on a small time interval \([t, t + \Delta t]\). This assumption is commonly used to solve a differential equation numerically. On the basis of this assumption, the SDE (4) is solved from

\[ x_{t+\Delta t} - x_t = f(x_t) \Delta t + \sigma (B_{t+\Delta t} - B_t). \hspace{1cm} (5) \]

Since \( B_{t+\Delta t} - B_t \) follows the normal distribution with mean zero and variance \( \Delta t \), the logarithm of the joint density function \( p(x_{t_0}, \ldots, x_{t_N}) \) for \( N + 1 \) discrete observations \( x_{t_0}, \ldots, x_{t_N} \) is given by

\[ \log \{ p(x_{t_1}, \ldots, x_{t_N}) \} = \sum_{i=1}^{N} \log \{ p(x_{t_i} | x_{t_{i-1}}) \} + \log \{ p(x_{t_0}) \} \]

\[ = -\frac{1}{2} \sum_{i=1}^{N} \left[ \frac{(x_{t_i} - x_{t_{i-1}} - f(x_{t_{i-1}}))^2}{\sigma^2 \Delta t} + \log(2\pi\sigma^2 \Delta t) \right] \]

\[ + \log \{ p(x_{t_0}) \}. \hspace{1cm} (6) \]

The first equality is derived from the Markov property of \( x_t \). The parameters of (4) are those values which maximize the log-likelihood of (6).

2.2. The local linearization method

The local linearization method was developed by Ozaki (1985). Intuitively, it assumes \( f(x_t) \) to be locally linear with respect to the process \( x_t \). In order to derive a discretized process, (4) is first considered in a deterministic context, i.e. the following differential equation is considered.

© Blackwell Publishers Ltd 1997
\[
\frac{dx_t}{dt} = f(x_t).
\] (7)

By using an appropriate approximation on the small time interval \([t, t + \Delta t]\), the above differential equation is solved from

\[
x_{t+\Delta t} = x_t + \frac{f(x_t)}{f'(x_t)} \left[ \exp \left\{ f'(x_t) \Delta t \right\} - 1 \right]
\] (8)

(see Ozaki, 1992).

Also, using the local linearization and setting \(f(x_s) = K_t x_s\) for \(s \in [t, t + \Delta t]\) in (4), we obtain a linear SDE which can easily be solved from

\[
x_{t+\Delta t} = \exp(K_t \Delta t) x_t + \int_t^{t+\Delta t} \exp \left\{ K_t (t + \Delta t - s) \right\} dB_s.
\] (9)

From (9), we see that the conditional expectation with respect to \(t\), \(E_t(x_{t+\Delta t}) = \exp(K_t \Delta t) x_t\), and comparing with (8) we obtain

\[
K_t = \frac{1}{\Delta t} \log \left( 1 + \frac{f(x_t)}{x_t f'(x_t)} \left[ \exp \left\{ f'(x_t) \Delta t \right\} - 1 \right] \right).
\] (10)

Since \(\int_t^{t+\Delta t} \exp \left\{ K_t (t + \Delta t - s) \right\} dB_s\) follows the normal distribution with mean zero and variance

\[
V_t = \sigma^2 \frac{\exp(2K_t \Delta t) - 1}{2K_t}
\]

the log-likelihood is

\[
\log \left\{ p(x_{t_0}, \ldots, x_{t_N}) \right\} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{x_{t_i} - \exp(K_{t_{i-1}} \Delta t)^2}{V_{t_{i-1}}} + \log(2\pi V_{t_{i-1}}) \right\}
\]

\[+ \log \{ p(x_{t_0}) \}
\]

\[
K_t = \frac{1}{\Delta t} \log \left( 1 + \frac{f(x_t)}{x_t f'(x_t)} \left[ \exp \left\{ f'(x_t) \Delta t \right\} - 1 \right] \right)
\]

\[
V_t = \sigma^2 \frac{\exp(2K_t \Delta t) - 1}{2K_t}.
\] (11)

The local linearization method is also applicable to an SDE such as (1) by using a data transformation \(\phi(\cdot)\) which satisfies (2), since an SDE of \(y_t \equiv \phi(x_t)\) has a constant diffusion coefficient and \(y_t\) can be discretized as above. We should note that this data transformation requires a little change in the log-likelihood as follows. The joint density function \(p(x_{t_0}, \ldots, x_{t_N})\) is

\[
p(y_{t_0}, \ldots, y_{t_N}) \left| \frac{\partial(y_{t_0}, \ldots, y_{t_N})}{\partial(x_{t_0}, \ldots, x_{t_N})} \right|
\]

© Blackwell Publishers Ltd 1997
where $|\frac{\partial (\ldots)}{\partial (\ldots)}|$ is a Jacobian satisfying
\[
\left| \frac{\partial (y_{t_0}, \ldots, y_{t_N})}{\partial (x_{t_0}, \ldots, x_{t_N})} \right| = \prod_{j=0}^{N} \left| \frac{d\phi}{dx} (x_{t_j}) \right|.
\]

Therefore,
\[
\log \{ p(x_{t_0}, \ldots, x_{t_N}) \} = \log \{ p(y_{t_0}, \ldots, y_{t_N}) \} + \sum_{j=0}^{N} \log \left\{ \left| \frac{d\phi}{dx} (x_{t_j}) \right| \right\}
\]  (12)

where $\log \{ p(y_{t_0}, \ldots, y_{t_N}) \}$ has the same form as (11).

2.3. The new local linearization method

The new local linearization method developed by Shoji and Ozaki (1994) is based on the motivation of the original local linearization method. That is, $f(x_t)$ is assumed to be locally linear with respect to $x_t$. Since we are interested in the local behavior of $f(x_t)$, the differential of $f(x_t)$ should be considered. This differential can be clarified by Ito’s formula which gives
\[
df = f'(x_u) dx_u + \frac{\sigma^2}{2} f''(x_u) du.
\]  (13)

Here, $f'(x_u)$ and $f''(x_u)$ are assumed to be equal to $f'(x_t)$ and $f''(x_t)$ respectively for $u \in [t, t + \Delta t)$. Using the local constancy of $f'(x_u)$ and $f''(x_u)$, (13) is replaced by
\[
df = a \ dx_u + b \ du
\]
where $a$ and $b$ are local constants with $a = f'(x_t)$ and $b = (\sigma^2 / 2) f''(x_t)$. Thus we obtain
\[
f(x_u) = f(x_t) + f'(x_t)(x_u - x_t) + \frac{\sigma^2}{2} f''(x_t)(u - t)
\]  (14)

where $u \in [t, t + \Delta t)$. Using this linear approximation of $f(\cdot)$ in (4), the following SDE is obtained.
\[
dx_u = (L_t x_u + M_t u + N_t) \ du + \sigma \ dB_u
\]  (15)

where
\[
L_t = f'(x_t)
\]
\[
M_t = \frac{\sigma^2}{2} f''(x_t)
\]
\[
N_t = f(x_t) - f'(x_t)x_t - \frac{\sigma^2}{2} f''(x_t)t.
\]

Solving this linear SDE, we get the discretized process as

© Blackwell Publishers Ltd 1997
\[ x_{t+\Delta t} = x_t + \frac{f(x_t)}{L_t} \{ \exp (L_t \Delta t) - 1 \} + \frac{M_t}{L_t^2} \{ \{ \exp (L_t \Delta t) - 1 \} - L_t \Delta t \} + \sigma \int_0^{t+\Delta t} \exp \{ L_t (t + \Delta t - u) \} dB_u \]

\[ L_t = f'(x_t) \]

\[ M_t = \frac{\sigma^2}{2} f''(x_t). \]

Since \( \int_0^{t+\Delta t} \exp \{ L_t (t + \Delta t - u) \} dB_u \) follows a normal distribution with mean zero and variance

\[ V_t = \sigma^2 \frac{\exp (2L_t \Delta t) - 1}{2L_t} \]

the log-likelihood is given by

\[ \log \{ p(x_{t_0}, \ldots, x_{t_N}) \} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{(x_{t_i} - E_{t_{i-1}})^2}{V_{t_{i-1}}} + \log (2\pi V_{t_{i-1}}) \right\} + \log \{ p(x_{t_0}) \} \]

\[ E_t = x_t + \frac{f(x_t)}{L_t} \{ \exp (L_t \Delta t) - 1 \} + \frac{M_t}{L_t^2} \{ \{ \exp (L_t \Delta t) - 1 \} - L_t \Delta t \} \]

\[ V_t = \sigma^2 \frac{\exp (2L_t \Delta t) - 1}{2L_t} \]

\[ L_t = f'(x_t) \]

\[ M_t = \frac{\sigma^2}{2} f''(x_t). \]

Like the original local linearization method, the new local linearization method is also applicable to an SDE such as (1). By using the same derivation as the original local linearization method, the log-likelihood can be easily obtained.

2.4. The likelihood ratio maximum likelihood method

On the assumption that continuous observations are available, the parameters of \( f(\cdot) \) can be estimated by maximizing the likelihood ratio

\[ \exp \left[ \int_{t}^{T} f(x_{u}) \sigma^2 du - \frac{1}{2} \int_{t}^{T} \left( \frac{f(x_{u})}{\sigma} \right)^2 du \right]. \]
Here, on this assumption, note that the variance $\sigma^2$ is explicitly obtained from the definition of quadratic variation (see Kutoyants, 1984; Yoshida, 1992). When only discrete observations are available, the above likelihood ratio is replaced by

$$
\exp \left[ \sum_{i=1}^{N} \frac{f(x_{t_{i-1}})}{\sigma^2} (x_{t_i} - x_{t_{i-1}}) - \frac{\Delta t}{2} \sum_{i=1}^{N} \left\{ \frac{f(x_{t_{i-1}})}{\sigma} \right\}^2 \right].
$$

(20)

Since $\sigma$ is usually unknown in practical situations, it must be estimated from the discrete observations. As is shown by Florens-Zmirou (1989) and Yoshida (1992), the consistent estimate $\hat{\sigma}^2$ of $\sigma^2$ is given by

$$
\hat{\sigma}^2 = \frac{1}{N\Delta t} \sum_{i=1}^{N} (x_{t_i} - x_{t_{i-1}})^2.
$$

(21)

Therefore, the likelihood ratio of (4) is given by

$$
\exp \left[ \sum_{i=1}^{N} \frac{f(x_{t_{i-1}})}{\hat{\sigma}^2} (x_{t_i} - x_{t_{i-1}}) - \frac{\Delta t}{2} \sum_{i=1}^{N} \left\{ \frac{f(x_{t_{i-1}})}{\hat{\sigma}} \right\}^2 \right]
$$

(22)

where

$$
\hat{\sigma}^2 = \frac{1}{N\Delta t} \sum_{i=1}^{N} (x_{t_i} - x_{t_{i-1}})^2.
$$

2.5. Generalized method of moments

When estimating parameters from a discrete time series by the GMM the following procedure is widely used (see for example Chan et al., 1992). First, (4) is discretized by a conventional method, say the Euler method, which gives

$$
x_{t_i} = x_{t_{i-1}} + f(x_{t_{i-1}})\Delta t + \sigma (B_{t_i} - B_{t_{i-1}}).
$$

Next we construct orthogonality conditions based on the first and second moments in terms of conditional expectations. From the above discretized process, the first and second moments are

$$
E_{t_{i-1}}(x_{t_i}) = x_{t_{i-1}} + f(x_{t_{i-1}})\Delta t
$$

$$
E_{t_{i-1}} \left[ \{x_{t_i} - x_{t_{i-1}} - f(x_{t_{i-1}})\Delta t\}^2 \right] = \sigma^2 \Delta t.
$$

Moving the right-hand sides of the equalities to the left, we get

$$
E_{t_{i-1}} \{x_{t_i} - x_{t_{i-1}} - f(x_{t_{i-1}})\Delta t\} = 0
$$

$$
E_{t_{i-1}} \{x_{t_i} - x_{t_{i-1}} - f(x_{t_{i-1}})\Delta t\}^2 - \sigma^2 \Delta t = 0.
$$

(24)

Here, the two arguments of the expectation operator $E_{t_{i-1}}(\cdot)$ are designated $\epsilon_{1, t_i}$ and $\epsilon_{2, t_i}$. When using one instrumental variable, say $x_{t_i}$, the orthogonality conditions are defined as

$$
E_{t_{i-1}}(y_{t_i}) = 0
$$

(25)
where

\[ y_{t_i} = (\varepsilon_{1,t_i}, \varepsilon_{2,t_i}, x_{t_i-1,1,t_i}, x_{t_i-1,2,t_i})'. \]

Taking the sample mean of \( y_{t_i} \), i.e.

\[ z_N = \frac{1}{N} \sum_{i=1}^{N} y_{t_i} \]

the following objective function is obtained:

\[ z_N' W_N z_N \]  \hspace{1cm} (26) \]

where \( W_N \) is a consistent covariance matrix which yields efficient parameter estimates (see Hansen, 1982). The model parameters are those values which minimize the above objective function.

3. MONTE CARLO EXPERIMENTS

Simulation methods are used to evaluate the finite sample performances of the estimation methods described in the previous section: the new local linearization method (NLL), the original local linearization method (LL), the Euler method (Euler), the likelihood ratio maximum likelihood method (LR), and the generalized method of moments (GMM).

We consider the following two examples of SDEs, one with linear drift and state-dependent diffusion coefficients and the other with nonlinear drift and constant diffusion coefficients:

(i) \( dx_t = (1 - x_t) dt + x_t dB_t; \)
(ii) \( dx_t = -x_t^3 dt + dB_t; \)

It should be noted that since aliasing is inevitable in the estimation of a continuous-time stochastic process from discrete observations, an estimation method may show different performances depending on the discrete-time sampling interval. To check the influence of changes in discrete-time sampling interval, experiments are carried out for rather large time intervals as well as small time intervals: \( \Delta t = 0.05, 0.1, 0.15 \) and \( 0.2 \).

3.1. Method of experiments

In the following experiments, we set the number of sample paths, \( N \), to 1000. The number of sample points, \( n \), varied depending upon the discrete time interval: \( n = 3600, 1800, 1200 \) and \( 900 \) for \( \Delta t = 0.05, 0.1, 0.15 \) and \( 0.2 \), respectively.

To obtain sample paths from the above examples, data points were generated by the Euler method using a small time interval \( \Delta t_g = 0.005 \). This method is widely used in numerical studies of SDEs. In addition, with a small time interval, the Euler method presents a reasonable approximation of a true sample
path. Even if there is some effect of using this method, it favors the performance of the Euler method but does not favor that of the other estimation methods. Applying the Euler method to the first example, we need to pay attention to the diffusion term, which is not constant but state dependent. When the diffusion term is state dependent, it is possible that the sample path will explode numerically because the variance becomes large. A large variance may be brought about once the process $x_t$ reaches a large value. Thus, in order to avoid numerical explosion, we transformed the process $x_t$ into the process $y_t = \log(x_t)$ whose SDE has a constant diffusion coefficient and we then applied the Euler method to this SDE. Of course, the sample path for $x_t$ can be easily obtained by inverse transformation of $y_t$.

After generating the data points, we have only to pick up every tenth, twentieth, thirtieth and fortieth point from the total points to get a sample path for $\Delta t = 0.05, 0.1, 0.15$ and $0.2$, respectively. For each sample path, we applied the five alternative estimation methods to the following general models for the first and second examples, respectively:

$$dx_t = (\alpha + \beta x_t) \, dt + \sigma x_t \, dB_t$$

$$dx_t = \alpha x_t^3 \, dt + \sigma \, dB_t.$$  

Here, for the first example, the true values of the parameters $(\alpha, \beta, \sigma^2)$ are $(1, -1, 1)$, and for the second example, those of $(\alpha, \sigma^2)$ are $(-1, 1)$. When estimating the first example by maximum likelihood methods whose objective functions are not the likelihood ratio, we used the following log-likelihood:

$$\log \{ p(x_{t_1}, \ldots, x_{t_N}) \} = \log \{ p(y_{t_1}, \ldots, y_{t_N}) \} + \sum_{i=1}^{N} \log \left\{ \left| \frac{dy}{dx}(x_{t_i}) \right| \right\}$$

where $y = \log(x)$. Furthermore, because of numerical error in the LL method in the estimation of the first example, we used a modified LL method (mLL) for LL, whose likelihood is equivalent to the LL method except that the local variance of the mLL method is the same as that of the NLL method.

After 1000 repetitions of this procedure, we constructed the distribution of the estimators and evaluated the performance of the five alternative methods by comparing (i) bias, (ii) variance and (iii) mean squared errors of the estimate (MSEs).

3.2. Results of experiments

The results for the first example are presented in Tables I–III. The distributions for estimates of $\alpha$, $\beta$ and $\sigma^2$ are represented in Figures 1, 2 and 3, respectively.

Looking at Figure 1, the NLL estimator of $\alpha$ basically dominates the other estimators for large $\Delta t$. By contrast, the GMM estimator is found to be inefficient for all $\Delta t$; the distribution of the GMM estimator is relatively flat compared with the other estimators. In addition, the estimators other than the NLL estimator tend to underestimate $\alpha$. This is also confirmed from the

© Blackwell Publishers Ltd 1997
Figure 1. Estimation of $\alpha$ in the simulation of $dx = (\alpha + \beta x)dt + \sigma x dB$: estimate interval (a) 0.05, (b) 0.1, (c) 0.15 and (d) 0.2.

statistics for the estimators in Table I. The NLL estimator shows basically a small bias and MSE. The small bias and MSE are noticeable when $\Delta t$ is large. The GMM estimator shows a relatively large MSE. The statistics for bias show that the estimators other than the NLL estimator underestimate $\alpha$. Here, focusing on the Euler estimator, it shows a small variance for all $\Delta t$; however, it shows large bias and MSE.

© Blackwell Publishers Ltd 1997
TABLE I

<table>
<thead>
<tr>
<th>Δt</th>
<th>Method</th>
<th>Bias</th>
<th>Variance</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>NLL</td>
<td>0.0140</td>
<td>0.0079</td>
<td>0.0080</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.0497</td>
<td>0.0085</td>
<td>0.0109</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.0312</td>
<td>0.0067</td>
<td>0.0077</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>-0.0057</td>
<td>0.0076</td>
<td>0.0076</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.0974</td>
<td>0.0787</td>
<td>0.0881</td>
</tr>
<tr>
<td>0.1</td>
<td>NLL</td>
<td>0.0091</td>
<td>0.0085</td>
<td>0.0086</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.1113</td>
<td>0.0101</td>
<td>0.0225</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.0757</td>
<td>0.0064</td>
<td>0.0121</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>-0.0309</td>
<td>0.0082</td>
<td>0.0092</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.0013</td>
<td>0.0933</td>
<td>0.0932</td>
</tr>
<tr>
<td>0.15</td>
<td>NLL</td>
<td>0.0014</td>
<td>0.0093</td>
<td>0.0093</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.1656</td>
<td>0.0112</td>
<td>0.0387</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.1154</td>
<td>0.0060</td>
<td>0.0193</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>-0.0538</td>
<td>0.0085</td>
<td>0.0113</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.0790</td>
<td>0.0849</td>
<td>0.0911</td>
</tr>
<tr>
<td>0.2</td>
<td>NLL</td>
<td>-0.0109</td>
<td>0.0094</td>
<td>0.0095</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.2178</td>
<td>0.0121</td>
<td>0.0595</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.1532</td>
<td>0.0055</td>
<td>0.0289</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>-0.0767</td>
<td>0.0088</td>
<td>0.0146</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.1307</td>
<td>0.0651</td>
<td>0.0821</td>
</tr>
</tbody>
</table>

From Figure 2 it can be seen that the NLL estimator of β basically performs better than the other estimators for large Δt and that the GMM estimator is inefficient for all Δt. This is confirmed from the statistics in Table II. In contrast with the estimation of α, the estimators other than the NLL estimator tend to overestimate β.

In comparison with the estimation of α and β, there is a remarkable difference between the alternative estimators of σ². It can easily be seen from Figure 3 that the NLL estimator is much better than the other estimators for all Δt. By contrast, the GMM estimator is inefficient with large bias. The statistics for bias and MSE in Table III clearly show the dominance of the NLL estimator over the other estimates. In particular, the NLL estimator performs well even for large Δt; however, the mLL, Euler and GMM estimators considerably underestimate σ², and the LR estimator considerably overestimates it.

The results for the second example are presented in Tables IV and V, and distributions of the estimates of α and σ² are represented in Figures 4 and 5, respectively.

Looking at Figure 4, it can be seen that all the estimators show large bias as Δt increases. The NLL estimator basically performs better than the other estimators. The Euler and LR estimators show a similar performance. The LL estimator is inefficient with large bias. Unlike the first example, the GMM estimator shows reasonable performance. This is also confirmed from the statistics for the estimators in Table IV. In particular, the bias and MSE of the LL estimator are quite large.

© Blackwell Publishers Ltd 1997
Figure 2. Estimation of $\beta$ in the simulation of $dx = (\alpha + \beta x) dt + \sigma x dB$. estimate interval (a) 0.05, (b) 0.1, (c) 0.15 and (d) 0.2.

It can be clearly seen from Figure 5 that the NLL estimator of $\sigma^2$ performs considerably better than the other estimators for all $\Delta t$. The estimators other than the NLL estimator tend to underestimate $\sigma^2$ as $\Delta t$ increases. In contrast with the estimation of $\alpha$, the LL and LR estimators show a similar performance and so do the Euler and GMM estimators. This is confirmed from the statistics in Table V. For all $\Delta t$, the NLL estimator shows the smallest bias and MSE of

© Blackwell Publishers Ltd 1997
<table>
<thead>
<tr>
<th>$\Delta t$</th>
<th>Method</th>
<th>Bias</th>
<th>Variance</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>NLL</td>
<td>-0.0177</td>
<td>0.0229</td>
<td>0.0232</td>
</tr>
<tr>
<td></td>
<td>mL</td>
<td>0.0532</td>
<td>0.0260</td>
<td>0.0288</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.0148</td>
<td>0.0204</td>
<td>0.0206</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>-0.0001</td>
<td>0.0219</td>
<td>0.0219</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.1074</td>
<td>0.1076</td>
<td>0.1190</td>
</tr>
<tr>
<td>0.1</td>
<td>NLL</td>
<td>-0.0122</td>
<td>0.0237</td>
<td>0.0239</td>
</tr>
<tr>
<td></td>
<td>mL</td>
<td>0.1231</td>
<td>0.0305</td>
<td>0.0456</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.0501</td>
<td>0.0193</td>
<td>0.0218</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.0261</td>
<td>0.0225</td>
<td>0.0232</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.0084</td>
<td>0.1218</td>
<td>0.1217</td>
</tr>
<tr>
<td>0.15</td>
<td>NLL</td>
<td>-0.0039</td>
<td>0.0243</td>
<td>0.0243</td>
</tr>
<tr>
<td></td>
<td>mL</td>
<td>0.1847</td>
<td>0.0336</td>
<td>0.0677</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.0820</td>
<td>0.0175</td>
<td>0.0243</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.0493</td>
<td>0.0215</td>
<td>0.0239</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.0624</td>
<td>0.1108</td>
<td>0.1145</td>
</tr>
<tr>
<td>0.2</td>
<td>NLL</td>
<td>0.0091</td>
<td>0.0240</td>
<td>0.0240</td>
</tr>
<tr>
<td></td>
<td>mL</td>
<td>0.2449</td>
<td>0.0364</td>
<td>0.0963</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.1136</td>
<td>0.0159</td>
<td>0.0288</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.0724</td>
<td>0.0215</td>
<td>0.0267</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.1056</td>
<td>0.0862</td>
<td>0.0973</td>
</tr>
</tbody>
</table>

all the estimators. The bias and MSE of the LL estimator are almost the same as those of the LR estimator and those of the Euler estimator are almost the same as those of the GMM estimator. It is not surprising that the LL and LR methods show the same performance since the two estimators are equivalent. See Shoji (1995).

4. CONCLUSION

We have investigated the finite sample performances of five estimation methods for a continuous-time stochastic process using simulation methods. Applying these methods to two examples of SDEs, one with linear drift and state-dependent diffusion coefficients and the other with nonlinear drift and constant diffusion coefficients, Monte Carlo experiments were carried out to compare the performance of the estimation methods and to evaluate the influence of the sampling time interval on estimation.

In the simulation for the first example, i.e. an SDE with linear drift and state-dependent diffusion coefficients, the Monte Carlo results indicate that the differences between the methods except for the GMM are not large when the discrete-time interval is small. When the time interval is large, however, the local linearization method (NLL) developed by Shoji and Ozaki (1994) performs much better than the other methods. In particular, in the estimation of

© Blackwell Publishers Ltd 1997
the diffusion coefficients, the dominance of the NLL method is noticeable and the other methods show a large bias, particularly for large $\Delta t$. The GMM is somewhat inferior to the other methods.

In the simulation for the second example, i.e. an SDE with nonlinear drift and constant diffusion coefficients, the Monte Carlo results are similar to those for the first example. In the estimation of the diffusion coefficients, the NLL
### TABLE III

**Bias, Variance and MSE of Estimates of \( \sigma^2 \) in the Simulation of \( dx = (\alpha + \beta x) dt + \sigma x dB \) with \( \alpha = 1, \beta = -1 \) and \( \sigma = 1 \)**

<table>
<thead>
<tr>
<th>( \Delta t )</th>
<th>Method</th>
<th>Bias</th>
<th>Variance</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>NLL</td>
<td>0.0083</td>
<td>0.0006</td>
<td>0.0007</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.0164</td>
<td>0.0006</td>
<td>0.0009</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.0635</td>
<td>0.0005</td>
<td>0.0045</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.0558</td>
<td>0.0009</td>
<td>0.0040</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.0190</td>
<td>0.0033</td>
<td>0.0036</td>
</tr>
<tr>
<td>0.1</td>
<td>NLL</td>
<td>0.0084</td>
<td>0.0013</td>
<td>0.0014</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.0395</td>
<td>0.0014</td>
<td>0.0029</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.1261</td>
<td>0.0009</td>
<td>0.0168</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.1000</td>
<td>0.0026</td>
<td>0.0126</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.0564</td>
<td>0.0081</td>
<td>0.0113</td>
</tr>
<tr>
<td>0.15</td>
<td>NLL</td>
<td>0.0073</td>
<td>0.0022</td>
<td>0.0022</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.0622</td>
<td>0.0021</td>
<td>0.0060</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.1822</td>
<td>0.0013</td>
<td>0.0345</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.1404</td>
<td>0.0050</td>
<td>0.0247</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.1115</td>
<td>0.0138</td>
<td>0.0262</td>
</tr>
<tr>
<td>0.2</td>
<td>NLL</td>
<td>0.0053</td>
<td>0.0029</td>
<td>0.0030</td>
</tr>
<tr>
<td></td>
<td>mLL</td>
<td>-0.0839</td>
<td>0.0029</td>
<td>0.0099</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>-0.2319</td>
<td>0.0015</td>
<td>0.0553</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.1783</td>
<td>0.0089</td>
<td>0.0407</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>-0.1785</td>
<td>0.0210</td>
<td>0.0528</td>
</tr>
</tbody>
</table>

### TABLE IV

**Bias, Variance and MSE of Estimates of \( \alpha \) in the Simulation of \( dx = \alpha x^3 \, dt + \sigma dB \) with \( \alpha = -1 \) and \( \sigma = 1 \)**

<table>
<thead>
<tr>
<th>( \Delta t )</th>
<th>Method</th>
<th>Bias</th>
<th>Variance</th>
<th>MSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.05</td>
<td>NLL</td>
<td>-0.0013</td>
<td>0.0087</td>
<td>0.0087</td>
</tr>
<tr>
<td></td>
<td>LL</td>
<td>-0.0989</td>
<td>0.0121</td>
<td>0.0218</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.0561</td>
<td>0.0077</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.0561</td>
<td>0.0077</td>
<td>0.0108</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.0219</td>
<td>0.0106</td>
<td>0.0111</td>
</tr>
<tr>
<td>0.1</td>
<td>NLL</td>
<td>0.0264</td>
<td>0.0084</td>
<td>0.0090</td>
</tr>
<tr>
<td></td>
<td>LL</td>
<td>-0.2094</td>
<td>0.0192</td>
<td>0.0631</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.1176</td>
<td>0.0072</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.1177</td>
<td>0.0072</td>
<td>0.0210</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.0547</td>
<td>0.0102</td>
<td>0.0132</td>
</tr>
<tr>
<td>0.15</td>
<td>NLL</td>
<td>0.0666</td>
<td>0.0085</td>
<td>0.0130</td>
</tr>
<tr>
<td></td>
<td>LL</td>
<td>-0.3589</td>
<td>0.0354</td>
<td>0.1642</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.1698</td>
<td>0.0070</td>
<td>0.0358</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.1699</td>
<td>0.0069</td>
<td>0.0358</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.0839</td>
<td>0.0102</td>
<td>0.0172</td>
</tr>
<tr>
<td>0.2</td>
<td>NLL</td>
<td>0.1145</td>
<td>0.0078</td>
<td>0.0209</td>
</tr>
<tr>
<td></td>
<td>LL</td>
<td>-0.5610</td>
<td>0.0658</td>
<td>0.3804</td>
</tr>
<tr>
<td></td>
<td>Euler</td>
<td>0.2134</td>
<td>0.0063</td>
<td>0.0519</td>
</tr>
<tr>
<td></td>
<td>LR</td>
<td>0.2132</td>
<td>0.0063</td>
<td>0.0517</td>
</tr>
<tr>
<td></td>
<td>GMM</td>
<td>0.1130</td>
<td>0.0094</td>
<td>0.0221</td>
</tr>
</tbody>
</table>

© Blackwell Publishers Ltd 1997
method performs much better than the other methods, particularly when the discrete-time interval is large. By contrast, the other methods show a large bias. Unlike the first example, the GMM shows reasonable performance and the original local linearization method is somewhat inferior to the other methods.

The NLL estimator shows considerably better performance in both experiments. This may result from the efficiency of the discretization based on the NLL method.

In this paper, we focus on one-dimensional SDEs. However, because of the complexity of practical problems, it is often necessary to use high dimensional SDEs in order to model dynamic phenomena more efficiently. Fortunately, all the estimation methods presented in this paper are applicable to high dimensional SDEs. (The NLL method for high dimensional cases is presented by Shoji and Ozaki (1995).) The performance comparison for high dimensional cases is an interesting topic for future work.

APPENDIX

To clarify the difference between objective functions used to estimate parameters in the Monte Carlo studies, this Appendix presents the objective functions of the five estimation methods for each example.

For the first example, to estimate $\alpha$, $\beta$ and $\sigma$ of $dx = (\alpha + \beta x) \, dt + \sigma x \, dB$, the following functions are optimized.
Figure 4. Estimation of $\alpha$ in the simulation of $dx = \alpha x^3 \, dt + \sigma \, dB$: estimate interval (a) 0.05, (b) 0.1, (c) 0.15 and (d) 0.2.

(1) The log-likelihood function for the Euler method:

$$
\log \left\{ p(x_0, \ldots, x_N) \right\} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{(y_i - E_{i-1})^2}{V_{i-1}} + \log (2\pi V_{i-1}) \right\} + \log \left\{ p(y_0) \right\} \\
+ \sum_{i=0}^{N} \log \left( \frac{1}{x_i} \right)
$$

© Blackwell Publishers Ltd 1997
Figure 5. Estimation of $\sigma^2$ in the simulation of $dx = ax^3 \, dt + \sigma \, dB$: estimate interval (a) 0.05, (b) 0.1, (c) 0.15 and (d) 0.2.

where

$$y_i = \log(x_i)$$

$$E_i = \left\{ \alpha \exp(-y_i) + \beta - \frac{\sigma^2}{2} \right\} \Delta t$$

$$V_i = \sigma^2 \Delta t.$$
(2) The log-likelihood function for the mLL method:
\[
\log \{ p(x_0, \ldots, x_N) \} = -\frac{1}{2} \sum_{i=1}^{N} \left( \frac{(y_i - E_{i-1})^2}{V_{i-1}} + \log (2\pi V_{i-1}) \right) + \log \{ p(y_0) \} + \sum_{i=0}^{N} \log \left( \frac{1}{x_i} \right)
\]
where
\[
y_i = \log (x_i)
\]
\[
E_i = y_i - \left\{ 1 + \frac{1}{\alpha} \left( \beta - \frac{\sigma^2}{2} \right) \exp (y_i) \right\} \left[ \exp \{-\alpha \exp (-y_i)\Delta t\} - 1 \right]
\]
\[
V_i = \sigma^2 \frac{\exp \{-2\alpha \exp (-y_i)\Delta t\} - 1}{-2\alpha \exp (-y_i)}.
\]

(3) The log-likelihood function for the NLL method:
\[
\log \{ p(x_0, \ldots, x_N) \} = -\frac{1}{2} \sum_{i=1}^{N} \left( \frac{(y_i - E_{i-1})^2}{V_{i-1}} + \log (2\pi V_{i-1}) \right) + \log \{ p(y_0) \} + \sum_{i=0}^{N} \log \left( \frac{1}{x_i} \right)
\]
where
\[
y_i = \log (x_i)
\]
\[
E_i = y_i - \left\{ 1 + \frac{1}{\alpha} \left( \beta - \frac{\sigma^2}{2} \right) \exp (y_i) \right\} \left[ \exp \{-\alpha \exp (-y_i)\Delta t\} - 1 \right] + \frac{\sigma^2 \exp (y_i)}{2\alpha} \left[ \exp \{-\alpha \exp (-y_i)\Delta t\} - 1 + \alpha \exp (-y_i)\Delta t \right]
\]
\[
V_i = \sigma^2 \frac{\exp \{-2\alpha \exp (-y_i)\Delta t\} - 1}{-2\alpha \exp (-y_i)}.
\]

(4) The objective function for the LR method: the estimate \( \hat{\sigma}^2 \) of \( \sigma^2 \) is given as
\[
\hat{\sigma}^2 = \frac{1}{N\Delta t} \sum_{i=1}^{N} \frac{(x_i - x_{i-1} - (\alpha + \beta x_{i-1})\Delta t)^2}{x_{i-1}^2}
\]
To estimate \( \alpha \) and \( \beta \), the following likelihood ratio is maximized.
\[
\exp \left\{ \sum_{i=1}^{N} \frac{\alpha + \beta x_{i-1}}{\hat{\sigma}^2 x_{i-1}^2} (x_i - x_{i-1}) - \frac{\Delta t}{2} \sum_{i=1}^{N} \frac{(\alpha + \beta x_{i-1})^2}{\hat{\sigma}^2 x_{i-1}^2} \right\}.
\]

(5) The objective function for the GMM: to estimate \( \alpha \), \( \beta \) and \( \sigma \), the following objective function is minimized.
\[
z_N^t W_N z_N
\]

© Blackwell Publishers Ltd 1997
where
\[
z_N = \frac{1}{N} \sum_{i=1}^{N} y_i
\]
\[
y_i = \left(\begin{array}{c}
x_i - x_{i-1} - (\alpha + \beta x_{i-1})\Delta t \\
{x_i - x_{i-1} - (\alpha + \beta x_{i-1})\Delta t}^2 - \sigma^2 x_{i-1}^2\Delta t \\
x_{i-1} - x_{i-1} - (\alpha + \beta x_{i-1})\Delta t \\
x_{i-1} \{x_i - x_{i-1} - (\alpha + \beta x_{i-1})\Delta t\}^2 - \sigma^2 x_{i-1}^2\Delta t
\end{array}\right)
\]

and \( W_N \) is a consistent covariance matrix defined by Hansen (1982).

For the second example, to estimate \( \sigma \) of \( \dot{x} = \alpha x^3 dt + \sigma dB \), the following functions are optimized.

1. The log-likelihood function for the Euler method:
\[
\log \{ p(x_0, \ldots, x_N) \} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{(x_i - E_i)^2}{V_i} + \log (2\pi V_i) \right\} + \log \{ p(x_0) \}
\]
where
\[
E_i = \alpha x_i^3 \Delta t \\
V_i = \sigma^2 \Delta t.
\]

2. The log-likelihood function for the LL method:
\[
\log \{ p(x_0, \ldots, x_N) \} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{(x_i - E_i)^2}{V_i} + \log (2\pi V_i) \right\} + \log \{ p(x_0) \}
\]
where
\[
E_i = x_i + \frac{x_i}{3} \{ \exp (3\alpha x_i^2 \Delta t) - 1 \} \\
V_i = \sigma^2 \frac{\exp (2K_i \Delta t) - 1}{2K_i} \\
K_i = 1 + \frac{1}{3} \{ \exp (3\alpha x_i^2 \Delta t) - 1 \}.
\]

3. The log-likelihood function for the NLL method:
\[
\log \{ p(x_0, \ldots, x_N) \} = -\frac{1}{2} \sum_{i=1}^{N} \left\{ \frac{(x_i - E_i)^2}{V_i} + \log (2\pi V_i) \right\} + \log \{ p(x_0) \}
\]
where
\[
E_i = x_i + \frac{x_i}{3} \{ \exp (3\alpha x_i^2 \Delta t) - 1 \} + \frac{\sigma^2}{3\alpha^2} \{ \exp (3\alpha x_i^2 \Delta t) - 1 - 3\alpha x_i^2 \Delta t \} \\
V_i = \sigma^2 \frac{\exp (6\alpha x_i^2 \Delta t) - 1}{6\alpha x_i^2}.
\]

4. The objective function for the LR method: first,
\[
\hat{\sigma}^2 = \frac{1}{N\Delta t} \sum_{i=1}^{N} (x_i - x_{i-1} - \alpha x_{i-1}^3 \Delta t)^2.
\]

© Blackwell Publishers Ltd 1997
To estimate $\alpha$ and $\beta$, the following likelihood ratio is maximized.

$$
\exp \left\{ \sum_{i=1}^{N} \frac{\alpha x_{i-1}^3}{\sigma^2} (x_i - x_{i-1}) - \frac{\Delta t}{2} \sum_{i=1}^{N} \left( \frac{\alpha x_{i-1}^3}{\sigma} \right)^2 \right\}.
$$

(5) The objective function for the GMM: to estimate $\alpha$ and $\sigma$, the following objective function is minimized.

$$
z_N' W_N z_N
$$

where

$$
z_N = \frac{1}{N} \sum_{i=1}^{N} y_i
$$

$$
y_i = \begin{pmatrix}
 x_i - x_{i-1} - \alpha x_{i-1}^3 \Delta t \\
 (x_i - x_{i-1} - \alpha x_{i-1}^3 \Delta t)^2 - \sigma^2 \Delta t \\
 x_{i-1}(x_i - x_{i-1} - \alpha x_{i-1}^3 \Delta t) \\
 x_{i-1} \{ (x_i - x_{i-1} - \alpha x_{i-1}^3 \Delta t)^2 - \sigma^2 \Delta t \}
\end{pmatrix}.
$$

REFERENCES


© Blackwell Publishers Ltd 1997