Bernoulli **1(1/2)**, 1995, 017–039

Martingale estimation functions for discretely observed diffusion processes

BO MARTIN BIBBY and MICHAEL SØRENSEN*

Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus, DK-8000 Aarhus C, Denmark

We consider three different martingale estimating functions based on discrete-time observations of a diffusion process. One is the discretized continuous-time score function adjusted by its compensator. The other two emerge naturally when optimality properties of the first are considered. Subject to natural regularity conditions, we show that all three martingale estimating functions result in consistent and asymptotically normally distributed estimators when the underlying diffusion is ergodic. Practical problems with implementing the estimation procedures are discussed through simulation studies of three specific examples. These studies also show that our estimators have good properties even for moderate sample sizes and that they are a considerable improvement compared with the estimator based on the unadjusted discretized continuous-time likelihood function, which can be seriously biased.

Keywords: discrete-time sampling; inference for diffusion processes; optimality; quasi-likelihood; simulation; stochastic differential equation

1. Introduction

Ideally, parametric inference concerning diffusion processes based on discrete-time observations should be based on the likelihood function L. Under weak conditions the maximum likelihood estimator has the usual good properties (Dacunha-Castelle and Florens-Zmirou 1986). However, the likelihood function for discrete observations is a product of transition densities, and these are only known in special cases. One way around this problem is to find good approximations to L. This approach has been pursued in Pedersen (1995a; 1995b).

The likelihood theory for continuously observed diffusions is well studied (see, for example, Taraskin 1974; Brown and Hewitt 1975; Liptser and Shiryayev 1977; Ibragimov and Has'minskii 1981; Kutoyants 1984; and Sørensen 1991). Inference from discrete-time observations can be based on an approximation, $\tilde{L}$, to the continuous-time likelihood function, obtained by replacing Lebesgue integrals and Itô integrals by Riemann–Itô sums (see Section 2). This approach works well when the observation times are closely spaced, as has been demonstrated by several authors (see, for example, Le Breton 1976; Prakasa Rao 1988; Florens-Zmirou 1989; Genon-Catalot 1990; Yoshida 1992; and Kloeden *et al.* 1992). However, the estimator obtained in this way is not consistent when the time between observations is bounded away from zero (Florens-Zmirou 1989),

* To whom correspondence should be addressed.

1350-4864

and when the time between observations is not small it can be strongly biased (see Pedersen 1995a; and Section 4 below).

In the present paper these defects of estimators based on $\tilde{L}$ are avoided by constructing a martingale estimating function from $\tilde{L}$. One cannot expect to obtain a consistent estimator from $\tilde{L}$ since the corresponding pseudo-score function $\dot{\tilde{\ell}}$ is biased. A dot denotes differentiation with respect to the parameter, and $\tilde{\ell} = \log \tilde{L}$. When the time between observations is bounded away from zero, then so is the bias of $\dot{\tilde{\ell}}$. Our approach is to compensate $\dot{\tilde{\ell}}$ so that a martingale $\tilde{G}$ is obtained. This is done in Section 2. Here we also find, within a natural class of martingale estimating functions, the optimal estimating function G^* *sensu* Godambe and Heyde (1987). It turns out that $\tilde{G}$ can be viewed as an approximation to G^*. We also derive another approximation to G^*. Good approximations to G^* are important in practice since in many cases it is difficult to find the roots of G^* because of problems of numerical instability. The estimating functions can also be used when the diffusion coefficient depends on the parameter.

In Section 3 we prove that the estimators based on $\tilde{G}$ and G^* are consistent and asymptotically normal. In fact, we give a result for a general class of martingale estimating equations based on discretely observed ergodic diffusions. Also questions of asymptotic efficiency are discussed. In Section 4 we study three examples in detail, including simulation studies of the behaviour of the estimators.

2. Derivation of the estimating functions

We will consider one-dimensional diffusion processes defined by the following class of stochastic differential equations:

$$dX_t = b(X_t;\theta)\,dt + \sigma(X_t;\theta)\,dW_t, \qquad X_0 = x_0. \tag{2.1}$$

Here, then, the drift and the diffusion coefficient do not depend on the time t; however, the method we shall discuss extends straightforwardly to the time-inhomogeneous case. We assume that (2.1) has a unique solution for all θ in some open subset, Θ, of the real line. The assumption that X and θ are one-dimensional is only made to simplify the exposition. We give results for the multi-dimensional case later. The function σ is assumed to be positive. Furthermore, the functions b and σ are supposed to be known and twice continuously differentiable with respect to both arguments. The parameter θ is to be estimated from discrete equidistant observations of $\{X_t\}$, $X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ say. It is no real limitation that we only consider equidistant observations as the general case can be treated in exactly the same way.

If σ does not depend on θ, then under some additional conditions (Liptser and Shiryayev 1977, Theorem 7.19) the measures corresponding to continuous observation of the solution of (2.1) for different values of θ are equivalent. The most important condition is that the integrals in (2.2) should exist. The continuous-time log-likelihood function is

$$\ell_t(\theta) = \int_0^t \frac{b(X_s;\theta)}{\sigma^2(X_s)}\,dX_s - \frac{1}{2}\int_0^t \frac{b^2(X_s;\theta)}{\sigma^2(X_s)}\,ds. \tag{2.2}$$

Using an Itô sum and a Riemann sum to approximate the integrals in (2.2) and differentiating with respect to θ, we get an approximate score function of the form

$$\dot{\tilde{\ell}}_n(\theta) = \sum_{i=1}^{n} \frac{\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta})}(X_{i\Delta} - X_{(i-1)\Delta}) - \Delta\sum_{i=1}^{n} \frac{b(X_{(i-1)\Delta};\theta)\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta})},$$

where the dot denotes differentiation with respect to θ. This approximate score function can also be derived using the approximation that, conditionally on the past, the increment $X_{i\Delta} - X_{(i-1)\Delta}$ is normally distributed with mean $b(X_{(i-1)\Delta};\theta)\Delta$ and variance $\sigma^2(X_{(i-1)\Delta})\Delta$. If σ does depend on θ, we use the same estimating function, but now with σ depending on θ:

$$\dot{\tilde{\ell}}_n(\theta) = \sum_{i=1}^{n} \frac{\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta};\theta)}(X_{i\Delta} - X_{(i-1)\Delta}) - \Delta\sum_{i=1}^{n} \frac{b(X_{(i-1)\Delta};\theta)\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta};\theta)}. \tag{2.3}$$

As mentioned in Section 1 this estimating function is biased, and the idea is now to adjust (2.3) by subtracting its compensator in order to get a zero-mean P_θ-martingale with respect to the filtration defined by $\mathcal{F}_i = \sigma(X_\Delta, \ldots, X_{i\Delta})$, $i = 1, 2, \ldots$. Using the notation

$$F(x;\theta) = \mathrm{E}_\theta(X_\Delta \mid X_0 = x), \tag{2.4}$$

the compensator is

$$\sum_{i=1}^{n} \mathrm{E}_\theta\{\dot{\tilde{\ell}}_i(\theta) - \dot{\tilde{\ell}}_{i-1}(\theta) \mid \mathcal{F}_{i-1}\} = \sum_{i=1}^{n} \frac{\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta};\theta)}\{F(X_{(i-1)\Delta};\theta) - X_{(i-1)\Delta}\}$$
$$- \Delta\sum_{i=1}^{n} \frac{b(X_{(i-1)\Delta};\theta)\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta};\theta)}.$$

This means that we get the following estimating function which is a zero-mean P_θ-martingale,

$$\tilde{G}_n(\theta) = \sum_{i=1}^{n} \frac{\dot{b}(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta};\theta)}\{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)\}. \tag{2.5}$$

The fact that $\tilde{G}_n(\theta)$ is a zero-mean P_θ-martingale estimating function does not depend on whether σ is a function of θ or not. This gives some justification for using (2.5) also in cases where σ depends on θ. A stronger justification will be given below.

In order to explore the properties of $\tilde{G}$ we consider the class of zero-mean P_θ-martingale estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^{n} g_{i-1}(\theta)\{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)\}, \tag{2.6}$$

where g_{i-1} is $\mathcal{F}_{i-1}$-measurable and a continuously differentiable function of θ, $i = 1, \ldots, n$. We are interested in finding the optimal estimating function within the class given by (2.6) in the sense of giving the smallest asymptotic confidence interval around θ and yielding an estimator with the smallest asymptotic dispersion (see Godambe and Heyde 1987). Straightforward calculations show that the quadratic characteristic of $G(\theta)$ is

$$\langle G(\theta)\rangle_n = \sum_{i=1}^{n} g_{i-1}^2(\theta)\phi(X_{(i-1)\Delta};\theta),$$

where

$$\phi(X_{(i-1)\Delta};\theta) = \mathrm{E}_\theta\{(X_{i\Delta} - F(X_{(i-1)\Delta};\theta))^2 \mid X_{(i-1)\Delta}\}, \qquad i = 1, \ldots, n, \tag{2.7}$$

and that

$$\dot{G}_n(\theta) = \sum_{i=1}^{n} \dot{g}_{i-1}(\theta)\{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)\} - \sum_{i=1}^{n} g_{i-1}(\theta)\dot{F}(X_{(i-1)\Delta};\theta), \tag{2.8}$$

where F is assumed to be a differentiable function of θ.

The first term on the right-hand side of (2.8) is a martingale, so the compensator of $\dot{G}(\theta)$ is given by

$$\begin{aligned}\bar{G}_n(\theta) &= \sum_{i=1}^{n} \mathrm{E}\{\dot{G}_i(\theta) - \dot{G}_{i-1}(\theta) \mid \mathcal{F}_{i-1}\} \\ &= -\sum_{i=1}^{n} g_{i-1}(\theta)\dot{F}(X_{(i-1)\Delta};\theta).\end{aligned}$$

According to Heyde (1988), an estimating function $G_n^*(\theta)$ of the type (2.6) is optimal within the class (2.6) if and only if

$$\frac{\langle G(\theta), G^*(\theta)\rangle_n}{\bar{G}_n(\theta)} = \frac{\langle G^*(\theta)\rangle_n}{\bar{G}_n^*(\theta)}, \tag{2.9}$$

for all n and all G satisfying (2.6). Now, in our case criterion (2.9) becomes

$$\frac{\sum_{i=1}^{n} g_{i-1}(\theta)g_{i-1}^*(\theta)\phi(X_{(i-1)\Delta};\theta)}{-\sum_{i=1}^{n} g_{i-1}(\theta)\dot{F}(X_{(i-1)\Delta};\theta)} = \frac{\sum_{i=1}^{n} g_{i-1}^*(\theta)^2\phi(X_{(i-1)\Delta};\theta)}{-\sum_{i=1}^{n} g_{i-1}^*(\theta)\dot{F}(X_{(i-1)\Delta};\theta)},$$

which is clearly the case if and only if $g_{i-1}^*(\theta) = \alpha(\theta)\dot{F}(X_{(i-1)\Delta};\theta)/\phi(X_{(i-1)\Delta};\theta)$, where $\alpha(\theta)$ is a non-random function of θ. We can choose $\alpha = 1$. The optimal estimating function in the asymptotic sense in the class (2.6) is thus

$$G_n^*(\theta) = \sum_{i=1}^{n} \frac{\dot{F}(X_{(i-1)\Delta};\theta)}{\phi(X_{(i-1)\Delta};\theta)}\{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)\}. \tag{2.10}$$

Note that the right-hand side of (2.9) is non-random when $G_n^*(\theta)$ is of the form (2.10). This means that $G_n^*(\theta)$ is also optimal in what Godambe and Heyde (1987) call the fixed sample sense, that is $G_n^*(\theta)$ is in some sense closest to the score function based on the usually unknown exact likelihood function within the class (2.6) (see Godambe and Heyde 1987).

It is almost never the case that $\dot{b}(x;\theta)/\sigma^2(x;\theta) = \dot{F}(x;\theta)/\phi(x;\theta)$, so $\tilde{G}_n(\theta)$ is rarely optimal. However, for small Δ we have the expansion (Florens-Zmirou 1989)

$$F(x;\theta) = x + \Delta b(x;\theta) + \tfrac{1}{2}\Delta^2\{b(x;\theta)b'(x;\theta) + \tfrac{1}{2}\sigma^2(x;\theta)b''(x;\theta)\} + O(\Delta^3), \tag{2.11}$$

and

$$\phi(x;\theta) = \Delta\sigma^2(x;\theta) + \Delta^2[b(x;\theta)\sigma(x;\theta)\sigma'(x;\theta) + \tfrac{1}{2}\sigma^2(x;\theta)\{2b'(x;\theta) + \sigma'(x;\theta)^2 + \sigma(x;\theta)\sigma''(x;\theta)\}] + O(\Delta^3), \tag{2.12}$$

where the prime denotes differentiation with respect to x. Hence $\dot{F}(x;\theta) = \dot{b}(x;\theta)\Delta + O(\Delta^2)$ and $\phi(x;\theta) = \sigma^2(x;\theta)\Delta + O(\Delta^2)$, so that for small Δ we have that $\tilde{G}_n(\theta)$ is approximately optimal.

As we shall indicate later (see Section 4), it is in some cases not easy to determine $\dot{F}(x;\theta)$ satisfactorily due to numerical problems. In those instances we might turn to the expansion result (2.11) once again and substitute the terms up to order $O(\Delta^2)$ or higher for $\dot{F}(x;\theta)$. This gives a third estimating function which we shall denote by $G^{\dagger}$, given by

$$G_n^{\dagger}(\theta) = \sum_{i=1}^{n}(\dot{b}(X_{(i-1)\Delta};\theta)\Delta + \tfrac{1}{2}\Delta^2[\dot{b}(X_{(i-1)\Delta};\theta)b'(X_{(i-1)\Delta};\theta) + b(X_{(i-1)\Delta};\theta)\dot{b}'(X_{(i-1)\Delta};\theta) + \tfrac{1}{2}\{\dot{\sigma}^2(X_{(i-1)\Delta};\theta)b''(X_{(i-1)\Delta};\theta) + \sigma^2(X_{(i-1)\Delta};\theta)\dot{b}''(X_{(i-1)\Delta};\theta)\}])\frac{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)}{\phi(X_{(i-1)\Delta};\theta)}. \tag{2.13}$$

Starting from the continuous-time likelihood function we have now found expressions for three different zero-mean P_θ-martingale estimating functions. One of them, namely G^*, is optimal in the class (2.6) in both senses of Godambe and Heyde (1987) and the other two, $\tilde{G}$ and $G^{\dagger}$, are the first- and second-order approximations in Δ of G^*.

In the multidimensional case we let θ be k-dimensional and $\{X_t\}$ d-dimensional. The drift is also d-dimensional while the diffusion coefficient σ is assumed to be a $(d \times m)$-dimensional matrix such that $\sigma\sigma^{\mathrm{T}}$ is positive definite. Here m is the dimension of the Wiener process, and T denotes transposition. The $(k \times 1)$-dimensional estimating functions are

$$\tilde{G}_n(\theta) = \sum_{i=1}^{n}\dot{b}(X_{(i-1)\Delta};\theta)^{\mathrm{T}}\{\sigma(X_{(i-1)\Delta};\theta)\sigma(X_{(i-1)\Delta};\theta)^{\mathrm{T}}\}^{-1}\{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)\} \tag{2.14}$$

$$G_n^*(\theta) = \sum_{i=1}^{n}\dot{F}(X_{(i-1)\Delta};\theta)^{\mathrm{T}}\phi(X_{(i-1)\Delta};\theta)^{-1}\{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)\}, \tag{2.15}$$

where $\dot{b}$ and $\dot{F}$ are $(d \times k)$-dimensional matrices of partial derivatives with respect to the components of θ, and where we have assumed that ϕ is positive definite.

Example 2.1 *The Ornstein–Uhlenbeck process*
The Ornstein–Uhlenbeck process is the solution to the stochastic differential equation

$$\mathrm{d}X_t = \theta X_t\,\mathrm{d}t + \sigma\,\mathrm{d}W_t, \qquad X_0 = x_0,$$

that is $b(x;\theta) = \theta x$ and $\sigma(x;\theta) \equiv \sigma$ is assumed to be known. It is well known that in this case the transition probability is normal with mean $F(x;\theta) = x\,\mathrm{e}^{\theta\Delta}$ and variance $\phi(\theta) = \sigma^2(\mathrm{e}^{2\theta\Delta} - 1)/2\theta$. This means that we get the following explicit formula for $\tilde{G}$:

$$\tilde{G}_n(\theta) = \frac{1}{\sigma^2}\sum_{i=1}^{n}X_{(i-1)\Delta}(X_{i\Delta} - X_{(i-1)\Delta}\,\mathrm{e}^{\theta\Delta}),$$

giving an estimator for θ of the form

$$\tilde{\theta}_n = \frac{1}{\Delta}\log\frac{\sum_{i=1}^{n} X_{(i-1)\Delta}X_{i\Delta}}{\sum_{i=1}^{n} X_{(i-1)\Delta}^2}, \tag{2.16}$$

provided that $\sum_{i=1}^{n} X_{(i-1)\Delta}X_{i\Delta} > 0$.

Both $G^\dagger$ and G^* are proportional to $\tilde{G}$, so that all three estimating functions give $\tilde{\theta}_n$ as estimator for θ.

It turns out that the estimator, $\tilde{\theta}_n$, found by solving either of the three estimating equations is equal to the maximum likelihood estimator of θ in the model where σ^2 is unknown.

Example 2.2 *The mean-reverting process*
Consider the class of diffusion processes emerging as solutions to the following stochastic differential equation:

$$\mathrm{d}X_t = (\alpha + \theta X_t)\,\mathrm{d}t + \psi(X_t)\,\mathrm{d}W_t, \qquad X_0 = x_0, \tag{2.17}$$

where ψ is a positive real-valued function. Processes satisfying (2.17) are often referred to as *mean-reverting processes* in the financial literature.

We consider both α and θ as unknown parameters. Clearly $f(t) = \mathrm{E}_{\alpha,\theta}(X_t \mid X_0)$ solves the ordinary differential equation $f'(t) = \alpha + \theta f(t)$, so

$$F(x;\alpha,\theta) = x\,\mathrm{e}^{\theta\Delta} + \frac{\alpha}{\theta}(\mathrm{e}^{\theta\Delta} - 1).$$

Using the explicit form of the conditional mean we get the following expression for $\tilde{G}$ and G^* (see (2.14) and (2.15)):

$$\tilde{G}_n(\alpha,\theta) = \Bigg\{\sum_{i=1}^{n}\frac{1}{\psi^2(X_{(i-1)\Delta})}\Big(X_{i\Delta} - X_{(i-1)\Delta}\mathrm{e}^{\theta\Delta} + \frac{\alpha}{\theta}(1-\mathrm{e}^{\theta\Delta})\Big),$$
$$\sum_{i=1}^{n}\frac{X_{(i-1)\Delta}}{\psi^2(X_{(i-1)\Delta})}\Big(X_{i\Delta} - X_{(i-1)\Delta}\mathrm{e}^{\theta\Delta} + \frac{\alpha}{\theta}(1-\mathrm{e}^{\theta\Delta})\Big)\Bigg\}^{\mathrm{T}},$$

$$G_n^*(\alpha,\theta) = \Bigg\{\sum_{i=1}^{n}\frac{\mathrm{e}^{\theta\Delta}-1}{\theta\phi(X_{(i-1)\Delta};\alpha,\theta)}\Big(X_{i\Delta} - X_{(i-1)\Delta}\mathrm{e}^{\theta\Delta} + \frac{\alpha}{\theta}(1-\mathrm{e}^{\theta\Delta})\Big),$$
$$\sum_{i=1}^{n}\frac{\Delta\mathrm{e}^{\theta\Delta}\Big(X_{(i-1)\Delta} + \frac{\alpha}{\theta}\Big) + \frac{\alpha}{\theta^2}(1-\mathrm{e}^{\theta\Delta})}{\phi(X_{(i-1)\Delta};\alpha,\theta)}$$
$$\Big(X_{i\Delta} - X_{(i-1)\Delta}\mathrm{e}^{\theta\Delta} + \frac{\alpha}{\theta}(1-\mathrm{e}^{\theta\Delta})\Big)\Bigg\}^{\mathrm{T}}.$$

The third estimating function, $G^\dagger$, can be obtained from G^* by expanding $\mathrm{e}^{\theta\Delta}$ in $\dot{F}$, but it is of no interest to consider $G^\dagger$ since F is known.

In the case of $\tilde{G}$ we can solve the estimation equations directly and get

$$e^{\tilde{\theta}_n \Delta} = \frac{\left(\sum_{i=1}^{n} \frac{X_{(i-1)\Delta}}{\psi^2(X_{(i-1)\Delta})}\right)\left(\sum_{i=1}^{n} \frac{X_{i\Delta}}{\psi^2(X_{(i-1)\Delta})}\right) - \left(\sum_{i=1}^{n} \frac{X_{(i-1)\Delta}X_{i\Delta}}{\psi^2(X_{(i-1)\Delta})}\right)\left(\sum_{i=1}^{n} \frac{1}{\psi^2(X_{(i-1)\Delta})}\right)}{\left(\sum_{i=1}^{n} \frac{X_{(i-1)\Delta}}{\psi^2(X_{(i-1)\Delta})}\right)^2 - \left(\sum_{i=1}^{n} \frac{X^2_{(i-1)\Delta}}{\psi^2(X_{(i-1)\Delta})}\right)\left(\sum_{i=1}^{n} \frac{1}{\psi^2(X_{(i-1)\Delta})}\right)}, \quad (2.18)$$

and

$$\tilde{\alpha}_n = \frac{\tilde{\theta}_n}{1 - e^{\tilde{\theta}_n \Delta}} \frac{\left(\sum_{i=1}^{n} \frac{X_{(i-1)\Delta}}{\psi^2(X_{(i-1)\Delta})}\right) e^{\tilde{\theta}_n \Delta} - \left(\sum_{i=1}^{n} \frac{X_{i\Delta}}{\psi^2(X_{(i-1)\Delta})}\right)}{\left(\sum_{i=1}^{n} \frac{1}{\psi^2(X_{(i-1)\Delta})}\right)}.$$

In the case where $\psi(x) = \sigma\sqrt{x}$ we have that $E_{\alpha,\theta}(X_t^2 \mid X_0)$, the conditional second moment, solves the ordinary differential equation $f'(t) = (2\alpha + \sigma^2)E_{\alpha,\theta}(X_t \mid X_0) + 2\theta f(t)$ so that

$$\phi(x; \alpha, \theta) = \frac{\sigma^2}{2\theta^2}\{(\alpha + 2\theta x)\, e^{2\theta\Delta} - 2(\alpha + \theta x)\, e^{\theta\Delta} + \alpha\}.$$

This specification of ψ has often been used in the financial literature (see, for example, Cox *et al.* 1985).

An example of a diffusion process where the parameter of interest enters the diffusion coefficient is the mean-reverting process with $\alpha = 0$, θ replaced by $-\theta$ in the drift, and $\psi(x) = \sqrt{\theta + x^2}$ where $\theta > 0$. Again the conditional second moment solves an ordinary differential equation, and the conditional variance takes the form

$$\phi(x; \alpha, \theta) = x^2\, e^{-2\theta\Delta}(e^{\Delta} - 1) + \frac{\theta}{2\theta - 1}(1 - e^{(1-2\theta)\Delta}).$$

If the diffusion coefficient is just slightly more complicated than these two examples, it is not possible to find a closed expression for the conditional variance.

Example 2.3 *The hyperbolic diffusion process*
The solution to the following stochastic differential equation

$$dX_t = \theta \frac{X_t}{\sqrt{1 + X_t^2}}\, dt + \sigma\, dW_t, \qquad X_0 = x_0,$$

is called the *hyperbolic diffusion process* because it has a hyperbolic stationary distribution when $\theta < 0$. In fact, the stationary density is proportional to $\exp(\theta\sqrt{1 + x^2}/\sigma^2)$.

It is not possible to determine the conditional expectation and conditional variance for the hyperbolic diffusion process, so the estimation functions look like this:

$$\tilde{G}_n(\theta) = \sum_{i=1}^{n} \frac{X_{(i-1)\Delta}}{\sigma^2\sqrt{1 + X^2_{(i-1)\Delta}}}\{X_{i\Delta} - F(X_{(i-1)\Delta}; \theta)\},$$

and

$$G_n^{\dagger}(\theta) = \sum_{i=1}^{n} \frac{X_{i\Delta} - F(X_{(i-1)\Delta};\theta)}{\phi(X_{(i-1)\Delta};\theta)}$$

$$\left\{ \Delta \frac{X_{(i-1)\Delta}}{\sqrt{1+X_{(i-1)\Delta}^2}} + \Delta^2 \left(\frac{\theta X_{(i-1)\Delta}}{(1+X_{(i-1)\Delta}^2)^2} - \frac{3}{4} \frac{\sigma^2 X_{(i-1)\Delta}}{(1+X_{(i-1)\Delta}^2)^{5/2}} \right) \right\},$$

whereas $G_n^*(\theta)$ can only be written in the general form (2.10).

3. Consistency and asymptotic normality

In this section we show that the estimators obtained from the estimating equations derived in the previous section are consistent and asymptotically normal. In fact, we give a result for a general class of martingale estimating equations. In addition, questions of asymptotic efficiency will be discussed. We mainly confine the discussion to the case of ergodic diffusion processes, but at the end of the section we will briefly comment on non-ergodic models.

Our framework is similar to that of Florens-Zmirou (1989). We consider discrete observations of a process X that solves (2.1) for some $\theta \in \Theta \subseteq \mathbb{R}$, where Θ is open. The observations are $\{X_{i\Delta}: i = 1, \ldots, n\}$. Let $s(x;\theta)$ denote the density of the scale measure

$$s(x;\theta) = \exp\left(-2 \int_0^x \frac{b(y;\theta)}{\sigma^2(y;\theta)} \, dy \right).$$

From now on we shall work under the assumptions below.

Condition 3.1
The following hold for all $\theta \in \Theta$.
(a)

$$\int_0^{\infty} s(x;\theta)\, dx = \int_{-\infty}^{0} s(x;\theta)\, dx = \infty$$

(b)

$$\int_{-\infty}^{\infty} \{s(x;\theta)\sigma^2(x;\theta)\}^{-1} \, dx = A(\theta) < \infty$$

(c) There exist constants $M > 0$ and $C > 0$ such that

$$\{b(x;\theta)/\sigma(x;\theta) - \tfrac{1}{2}\sigma'(x;\theta)\}\, \mathrm{sign}(x) \le -C \qquad \text{for } |x| > M.$$

Under assumptions (a) and (b) X is ergodic, and with respect to the Lebesgue measure its invariant measure μ_θ has density $\{A(\theta)\sigma^2(x;\theta)s(x;\theta)\}^{-1}$. Condition 3.1 is formulated for diffusions with state space equal to the real line. For diffusions confined to a smaller interval the conditions should be reformulated by replacing the boundaries $\pm\infty$ by the relevant boundaries, and possibly replacing zero by another interior point in the state space in (a) of Condition 3.1.

Let P_θ^μ (respectively P_θ) denote the distribution of $\{X_t\}$ when $X_0 \sim \mu_\theta$ (respectively $X_0 = x_0$). Furthermore, denote by π_t^θ the transition kernel of X under P_θ, i.e. $\pi_t^\theta(\mathrm{d}y, x) = P_\theta(X_t \in \mathrm{d}y \mid X_0 = x)$, and define a measure Q_t^θ on $\mathbb{R}^2$ by

$$Q_t^\theta = \pi_t^\theta \times \mu_\theta.$$

For a function $g: \mathbb{R}^2 \to \mathbb{R}$, we use the notation

$$Q_t^\theta(g) = \int g dQ_t^\theta,$$

and

$$\pi_t^\theta(g)(x) = \int g(x, y)\pi_t^\theta(\mathrm{d}y, x).$$

From Florens-Zmirou (1989) we have the following lemma.

Lemma 3.1 Suppose Condition 3.1 holds, and let the function $f: \mathbb{R}^2 \to \mathbb{R}$ satisfy $Q_\Delta^\theta(f^2) < \infty$. Then

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}) \to Q_\Delta^\theta(f) \tag{3.1}$$

in $L^2(P_\theta)$ as $n \to \infty$. Suppose, moreover, that $\pi_\Delta^\theta(f)$ is identically equal to zero. Then

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} f(X_{(i-1)\Delta}, X_{i\Delta}) \to N(0, Q_\Delta^\theta(f^2)) \tag{3.2}$$

in distribution under P_θ as $n \to \infty$.

Result (3.1) follows from Lemma 2 in Florens-Zmirou (1989), while (3.2) follows from Theorem 1 of the same paper. The results in that paper are only given for processes with a constant diffusion coefficient. They can, however, easily be generalized to our case by the well-known transformation

$$t(x) = \int_0^x \sigma(y; \theta)^{-1}\, \mathrm{d}y$$

(see also Florens-Zmirou, 1984).

We consider a general estimating function of the form

$$G_n(\theta) = \sum_{i=1}^{n} g(X_{(i-1)\Delta}; \theta)\{X_{i\Delta} - F(X_{(i-1)\Delta}; \theta)\}, \tag{3.3}$$

where F is given by (2.4). Recall also in the following that ϕ is given by (2.7).

We make the following assumptions about $G_n(\theta)$. Let $\theta_0 \in \Theta$ denote the true value of θ.

Condition 3.2
For all $\theta \in \Theta$ we have:

(a) The functions $g(x;\theta)$ and $F(x;\theta)$ are continuously differentiable with respect to θ for all x.
(b) The function

$$h(\theta; x, y) = \dot{g}(x;\theta)\{y - F(x;\theta)\} - g(x;\theta)\dot{F}(x;\theta) \tag{3.4}$$

is locally dominated square integrable with respect to $Q_\Delta^{\theta_0}$, and

$$f(\theta_0) = Q_\Delta^{\theta_0}\{h(\theta_0)\} = -\mathrm{E}_{\mu_{\theta_0}}\{g(\theta_0)\dot{F}(\theta_0)\} \neq 0.$$

(c) The function $g(x;\theta_0)\{y - F(x;\theta_0)\}$ is in $L^2(Q_\Delta^{\theta_0})$, and

$$v(\theta_0) = \mathrm{E}_{\mu_{\theta_0}}\{g(\theta_0)^2\phi(\theta_0)\} > 0.$$

Under these assumptions we can prove the following theorem.

Theorem 3.2 Under Conditions 3.1 and 3.2 an estimator $\hat{\theta}_n$ exists for every n, which on a set C_n solves the equation

$$G_n(\hat{\theta}_n) = 0, \tag{3.5}$$

where $P_{\theta_0}(C_n) \to 1$ as $n \to \infty$. Moreover, as $n \to \infty$,

$$\hat{\theta}_n \to \theta_0$$

in probability under P_{θ_0} and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \to N(0, v(\theta_0)/f(\theta_0)^2) \tag{3.6}$$

in distribution under P_{θ_0}.

Proof
The existence and consistency of $\hat{\theta}_n$ follow from Theorem A.1 in Barndorff-Nielsen and Sørensen (1994) if we can verify that under P_{θ_0}

$$\sup_{\theta \in M_{c,n}} |n^{-1}\dot{G}_n(\theta) - Q_\Delta^{\theta_0}\{h(\theta_0)\}| \to 0 \tag{3.7}$$

in probability as $n \to \infty$, where

$$M_{c,n} = \{\theta : |\theta - \theta_0| \leq c/\sqrt{n}\},$$

for $c > 0$ and small enough that $M_{c,1} \subseteq \Theta$, and

$$\frac{1}{\sqrt{n}} G_n(\theta_0) \to N(0, v(\theta_0)) \tag{3.8}$$

in distribution as $n \to \infty$. The result in Barndorff-Nielsen and Sørensen (1994) is formulated in terms of the maximum-likelihood estimator, but the proof does not use the fact that the estimating function is the score function.

Result (3.8) follows immediately from (3.2) in Lemma 3.1. To prove (3.7), note that

$$\dot{G}_n(\theta) = \sum_{i=1}^{n} h(\theta; X_{(i-1)\Delta}, X_{i\Delta}),$$

where h is given by (3.4), and that

$$\sup_{\theta \in M_{c,n}} |n^{-1}\dot{G}_n(\theta) - f(\theta_0)| \leq \sup_{\theta \in M_{c,1}} \left| n^{-1}\sum_{i=1}^{n} h(\theta; X_{(i-1)\Delta}, X_{i\Delta}) - f(\theta) \right| + \sup_{\theta \in M_{c,n}} |f(\theta) - f(\theta_0)|,$$

where

$$f(\theta) = Q_{\Delta}^{\theta_0}\{h(\theta)\}. \tag{3.9}$$

Now apply Lemma 3.3 below.

The asymptotic normality (3.6) follows by the standard expansion

$$G_n(\theta) = -\dot{G}_n(\tilde{\theta}_n)(\hat{\theta}_n - \theta),$$

where $\tilde{\theta}_n$ is between θ_0 and $\hat{\theta}_n$, and where we have used (3.5). By (A.5) in Barndorff-Nielsen and Sørensen (1994)

$$n^{-1}\dot{G}_n(\tilde{\theta}_n) \to f(\theta_0)$$

in probability under P_{θ_0} as $n \to \infty$. This, together with (3.8), proves (3.6). □

Lemma 3.3

(a) The function $f: \Theta \to \mathbb{R}$ given by (3.9) is continuous.

(b) Let K be a compact subset of Θ, and define

$$f_n(\theta) = n^{-1}\sum_{i=1}^{n} h(\theta; X_{(i-1)\Delta}, X_{i\Delta}). \tag{3.10}$$

Then

$$\sup_{\theta \in K} |f_n(\theta) - f(\theta)| \to 0$$

in probability under P_{θ_0} as $n \to \infty$.

Proof

From the local dominated integrability of h it follows that

$$Q_{\Delta}^{\theta_0}\{k(\theta, \delta)\} \to 0 \qquad \text{for } \delta \to 0,$$

where

$$k(\theta, \delta; x, y) = \sup_{|\tilde{\theta} - \theta| \leq \delta} |h(\tilde{\theta}; x, y) - h(\theta; x, y)|.$$

In particular, we see that f is continuous. From the local dominated square integrability of h it follows that for every $\theta \in \Theta$ there exists a $\delta_\theta > 0$ such that $k(\theta, \delta; x, y)$ is in $L^2(Q_\Delta^{\theta_0})$ for $0 < \delta < \delta_\theta$.

Fix $\epsilon > 0$. For every $\theta \in K$, we can choose $\lambda_\theta \in (0, \delta_\theta]$ such that

$$|f(\tilde{\theta}) - f(\theta)| < \epsilon/4 \qquad \text{for } |\tilde{\theta} - \theta| < \lambda_\theta$$

and

$$Q_\Delta^{\theta_0}\{k(\theta, \lambda_\theta)\} < \epsilon/4.$$

Of course, $K \subseteq \bigcup_{\theta \in K} B(\theta, \lambda_\theta)$, where $B(\theta, \lambda) = \{\tilde{\theta}: |\theta - \tilde{\theta}| < \lambda\}$. Since K is compact, we can find $\theta_1, \ldots, \theta_r$ such that

$$K \subseteq \bigcup_{j=1}^{r} B(\theta_j, \lambda_{\theta_j}).$$

For $\theta \in K$, we choose θ_j such that $|\theta - \theta_j| < \lambda_{\theta_j}$. Then

$$\begin{aligned}|f_n(\theta) - f(\theta)| &\le |f_n(\theta) - f_n(\theta_j)| + |f_n(\theta_j) - f(\theta_j)| + |f(\theta_j) - f(\theta)| \\ &\le \left| n^{-1}\sum_{i=1}^{n} k(\theta_j, \lambda_{\theta_j}; X_{(i-1)\Delta}, X_{i\Delta}) - Q_\Delta^{\theta_0}\{k(\theta_j, \lambda_{\theta_j})\}\right| \\ &\quad + |f_n(\theta_j) - f(\theta_j)| + \epsilon/2,\end{aligned}$$

so

$$\begin{aligned}\sup_{\theta \in K} |f_n(\theta) - f(\theta)| \le{}& \max_{1 \le j \le r} \left| \frac{1}{n}\sum_{i=1}^{n} k(\theta_j, \lambda_{\theta_j}; X_{(i-1)\Delta}, X_{i\Delta}) - Q_\Delta^{\theta_0}\{k(\theta_j, \lambda_{\theta_j})\}\right| \\ &+ \max_{1 \le j \le r} |f_n(\theta_j) - f(\theta_j)| + \epsilon/2.\end{aligned}$$

Now (b) follows by applying (3.1) in Lemma 3.1 twice. □

Not surprisingly, we find that the asymptotic expected information on the inverse asymptotic variance of the estimator corresponding to the optimal estimating function G^* is

$$f(\theta_0) = \mathrm{E}_{\mu_{\theta_0}}\{\dot{F}(\theta_0)^2/\phi(\theta_0)\}. \tag{3.11}$$

By inserting the expansions (2.11) and (2.12) for F and ϕ, respectively, we find that

$$\begin{aligned}f(\theta_0) ={}& \Delta \mathrm{E}_{\mu_{\theta_0}}\left(\frac{\dot{b}(\theta_0)^2}{\sigma^2(\theta_0)}\right) \\ &+ \tfrac{1}{2}\Delta^2 \mathrm{E}_{\mu_{\theta_0}}\left[\dot{b}(\theta_0)\dot{b}''(\theta_0) + 2\frac{\dot{b}(\theta_0)}{\sigma^2(\theta_0)}\{b(\theta_0)\dot{b}'(\theta_0) + \dot{\sigma}(\theta_0)\sigma(\theta_0)b''(\theta_0)\}\right. \\ &\left. - \frac{\dot{b}(\theta_0)^2}{\sigma^2(\theta_0)}\{2b(\theta_0)\sigma'(\theta_0)/\sigma(\theta_0) + \sigma'(\theta_0)^2 + \sigma(\theta_0)\sigma''(\theta_0)\}\right] + O(\Delta^3).\end{aligned} \tag{3.12}$$

For the estimating function $\tilde{G}$ we find that

$$v(\theta_0) = \mathrm{E}_{\mu_{\theta_0}}\{\dot{b}(\theta_0)^2\phi(\theta_0)/\sigma^4(\theta_0)\},$$

and

$$f(\theta_0) = -\mathrm{E}_{\mu_{\theta_0}}\{\dot{b}(\theta_0)\dot{F}(\theta_0)/\sigma^2(\theta_0)\}.$$

Hence by (2.11) and (2.12) it follows that

$$\begin{aligned} f(\theta_0)^2/v(\theta_0) = {} & \Delta\mathrm{E}_{\mu_{\theta_0}}\left(\frac{\dot{b}(\theta_0)^2}{\sigma^2(\theta_0)}\right) \\ & + \tfrac{1}{2}\Delta^2\mathrm{E}_{\mu_{\theta_0}}\left[\dot{b}(\theta_0)\dot{b}''(\theta_0) + 2\frac{\dot{b}(\theta_0)}{\sigma^2(\theta_0)}\{b(\theta_0)\dot{b}'(\theta_0) + \dot{\sigma}(\theta_0)\sigma(\theta_0)b''(\theta_0)\}\right. \\ & \left. - \frac{\dot{b}(\theta_0)^2}{\sigma^2(\theta_0)}\{2b(\theta_0)\sigma'(\theta_0)/\sigma(\theta_0) + \sigma'(\theta_0)^2 + \sigma(\theta_0)\sigma''(\theta_0)\}\right] + O(\Delta^3). \end{aligned} \tag{3.13}$$

Thus terms up to order $O(\Delta^2)$ in the expressions (3.12) and (3.13) are identical. For σ constant, Ducunha-Castelle and Florens-Zmirou (1986) found a similar expansion for the Fisher information. The term of order $O(\Delta)$ in their expression equals the corresponding term in (3.12) and (3.13), but they found that the term of order $O(\Delta^2)$ is zero. This is rarely the case for (3.12) and (3.13). For a discussion of conditions ensuring integrability of the remainder term in (2.11) and (2.12), see Florens-Zmirou (1989).

Example 3.1 *The mean-reverting process, continued*
Consider again the mean-reverting process (2.17) with $\psi(x) = \sigma\sqrt{x}$. For simplicity we put $\sigma = 1$. In this case the invariant distribution is a $\Gamma(2\alpha, -2\theta)$ distribution, provided $\alpha > 0$ and $\theta < 0$.

If $\alpha > 1$, then for $\tilde{G}$ the asymptotic expected information for α and θ is given by

$$\frac{f(\alpha)^2}{v(\alpha)} = \frac{4(\alpha-1)}{2\alpha-1}\,\frac{1-\mathrm{e}^{\theta\Delta}}{\alpha+(\alpha-2)\mathrm{e}^{\theta\Delta}},$$

$$\frac{f(\theta)^2}{v(\theta)} = \frac{2\alpha}{\theta^2}\,\frac{(1-\mathrm{e}^{\theta\Delta})}{(1+\mathrm{e}^{\theta\Delta})}.$$

For G^* the asymptotic expected information is given by the expectation under the invariant measure of the ratio of a second-order polynomial and a first-order polynomial. In this case the asymptotic expected information can easily be calculated numerically.

Figures 3.1 and 3.2 show the asymptotic expected information as a function of the size of the time-step in the case of the mean-reverting process with $\psi(x) = \sqrt{x}$, $\alpha = 10$ and $\theta = -1$. The figures give no reason to prefer G^* over $\tilde{G}$ from a practical point of view.

A result for the existence and consistency of $\hat{\theta}_n$ and for the asymptotic normality of

$$\dot{G}_n(\theta)^{1/2}(\hat{\theta}_n - \theta_0) \tag{3.14}$$

also holds for many non-ergodic models. It can be proved via Theorem A.1 in Barndorff-Nielsen

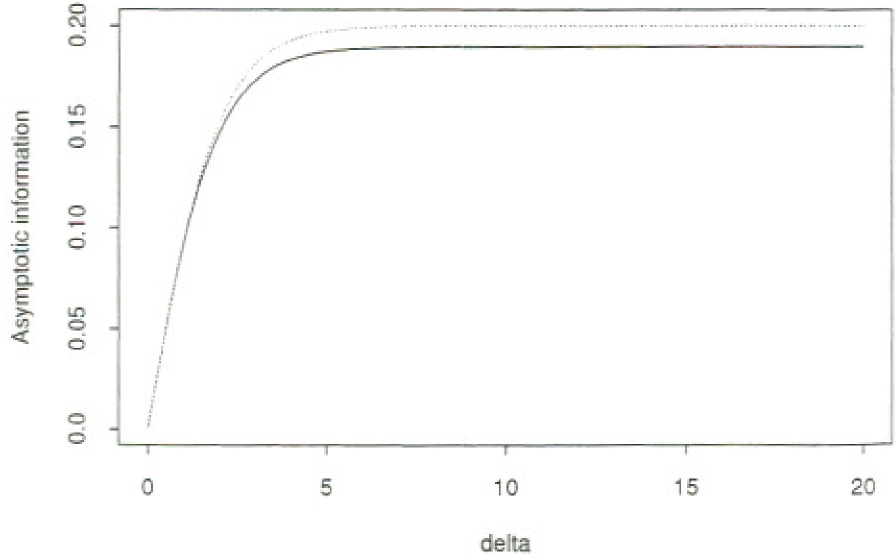

Figure 3.1. The asymptotic expected information for α in the mean-reverting process with $\psi(x) = \sqrt{x}$, $\alpha = 10$ and $\theta = -1$. The dotted line corresponds to G^*, and the solid line corresponds to $\tilde{G}$

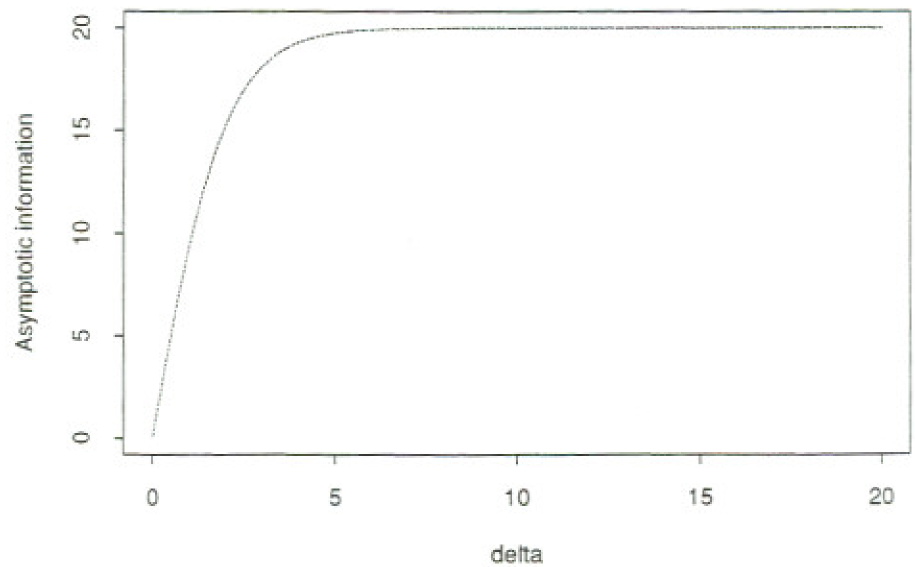

Figure 3.2. The asymptotic expected information for θ in the mean-reverting process with $\psi(x) = \sqrt{x}$, $\alpha = 10$ and $\theta = -1$. The dotted line corresponds to G^* and the solid line corresponds to $\tilde{G}$. Note that the two curves are virtually indistinguishable

and Sørensen (1994), but of course we cannot use ergodic theory to prove that sums such as $\dot{G}_n(\theta)$, properly normalized, converge to a non-negative random variable. This must in general be assumed. Similarly, the central limit theorem for ergodic diffusions used in this paper must be replaced by a central limit theorem for martingales. This is rather straightforward to do, but since the necessary assumptions are not easy to check in practice, we will not state the result. The Ornstein–Uhlenbeck process (Example 2.1) with $\theta > 0$ is an example of a non-ergodic model where $\hat{\theta}_n$ is consistent and where (3.14) is asymptotically normal.

4. Practical considerations and simulations

When wanting to use one or more of the estimation functions $\tilde{G}$, $G^{\dagger}$, and G^* in practice, one immediate problem is the unknown functions F and ϕ, the conditional mean and the conditional variance (see Section 2). Only for very simple models, such as when the drift depends linearly on the state of the process, is F explicitly known (see Examples 2.1 and 2.2). An explicit expression for the conditional variance is even more rarely known.

Our approach is to substitute the conditional mean and the conditional variance by the sample mean and sample variance of a large number of simulated realizations of the process in question at the relevant time-point. More precisely, we use the approximation

$$F(x;\theta) \doteq \frac{1}{m}\sum_{i=1}^{m} Y_\Delta^{(i)}, \tag{4.1}$$

and

$$\phi(x;\theta) \doteq \frac{1}{m}\sum_{i=1}^{m}\left(Y_\Delta^{(i)} - \frac{1}{m}\sum_{i=1}^{m} Y_\Delta^{(i)}\right)^2, \tag{4.2}$$

where $Y_\Delta^{(i)}$ is the ith replication of the simulated value of the diffusion process at the time-point Δ given that the process starts out in x at time 0.

Most often it is not possible to simulate the diffusion process in question exactly. An approximate simulation scheme is then necessary (see Kloeden and Platen 1992). In order to obtain a reliable simulated value of the process at time Δ given the starting point of the process, the interval $[0, \Delta]$ should be split up into a large number, N, of subintervals and the process simulated at the end-point of every subinterval.

Example 4.1 *The Ornstein–Uhlenbeck process, continued*
The transition probability is known for the Ornstein–Uhlenbeck process so it can be simulated exactly. Figure 4.1 shows a typical trajectory of $\tilde{\theta}$ (see (2.16)). After some initial fluctuations it settles around the true value $\theta = -1$.

To study the behaviour of $\tilde{\theta}$ further, 500 observations of the Ornstein–Uhlenbeck process were simulated in the time interval $[0, 200]$. This was repeated 500 times, yielding Table 4.1. In the table θ_n^{d} is the estimator based on the discretized continuous-time likelihood function. See (3.6) for an expression for the asymptotic variance of $\tilde{\theta}_n$. The table shows that $\tilde{\theta}_n$ has only a slight bias and is considerably closer to the true value than θ_n^{d}.

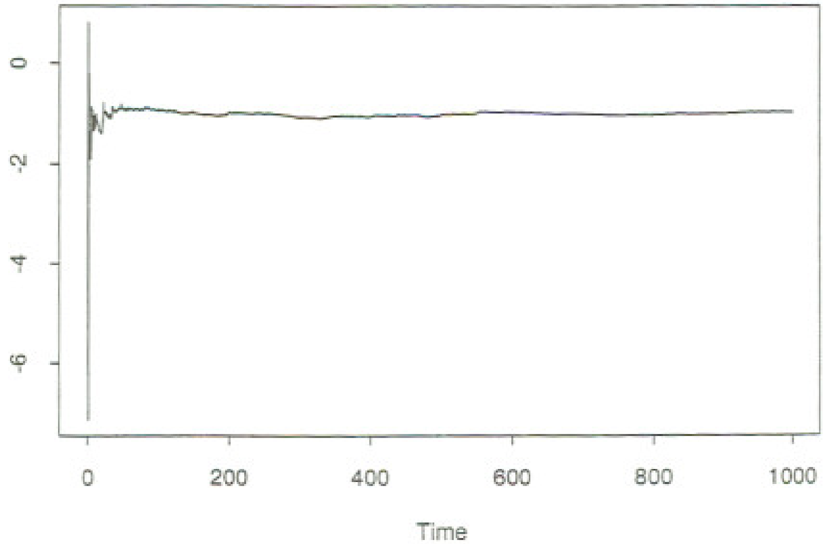

Figure 4.1. A typical simulated trajectory for $\tilde{\theta}$ in the case of the Ornstein–Uhlenbeck process. Here $\theta = -1$, $\sigma = 1$, $x_0 = 0$, and $\Delta = 0.2$

Figure 4.2 shows a normal quantile plot of the 500 simulated values of $\tilde{\theta}_n$. Since the points show a straight-line behaviour, we conclude that $\tilde{\theta}_n$ is, to a very good approximation, normally distributed for $n = 500$ and $\Delta = 0.4$.

Example 4.2 *The mean-reverting process, continued*
For simulation purposes we restrict ourselves to the special cases where $\psi(x) = \sigma\sqrt{x}$ or $\psi(x) = \sqrt{\theta + x^2}$, $\alpha = 0$, and θ is replaced by $-\theta$ in the drift (see (2.17)). For these models F and ϕ can be found explicitly (Example 2.2). Realizations of the mean-reverting process are obtained using the order 1.5 strong Taylor scheme (see Kloeden and Platen 1992, p. 351). In the case where $\psi(x) = \sigma\sqrt{x}$, the scheme is given by the following iterative formula:

$$\begin{aligned}Y_{n+1} = {} & Y_n + \Delta(\alpha + \theta Y_n) + \sigma\sqrt{Y_n}\Delta W + \frac{\sigma^2}{4}\{(\Delta W)^2 - \Delta\} \\ & + \theta\sigma\sqrt{Y_n}\Delta Z + \tfrac{1}{2}\Delta^2\theta(\alpha + \theta Y_n) + \frac{\sigma}{2\sqrt{Y_n}}\left(\alpha + \theta Y_n - \frac{\sigma^2}{4}\right)(\Delta W \Delta - \Delta Z),\end{aligned}$$

where $\Delta W = \sqrt{\Delta}U_1$, $\Delta Z = \Delta^{3/2}(U_1 + U_2/\sqrt{3})/2$, and U_1 and U_2 are independent standard normal random variables.

Table 4.1. Mean and standard error of the estimator based on the discretized continuous-time likelihood function and $\tilde{\theta}$ for the Ornstein–Uhlenbeck process. Here the true parameter value is $\theta = -1$, and $\sigma = 1$, $x_0 = 0$, and $\Delta = 0.4$

tmax	num. obs.	num. sim.	mean θ_n^d	SE θ_n^d	mean $\tilde{\theta}_n$	SE $\tilde{\theta}_n$	asymp. SE
200	500	500	−0.8362	0.0836	−1.021	0.1259	0.1238

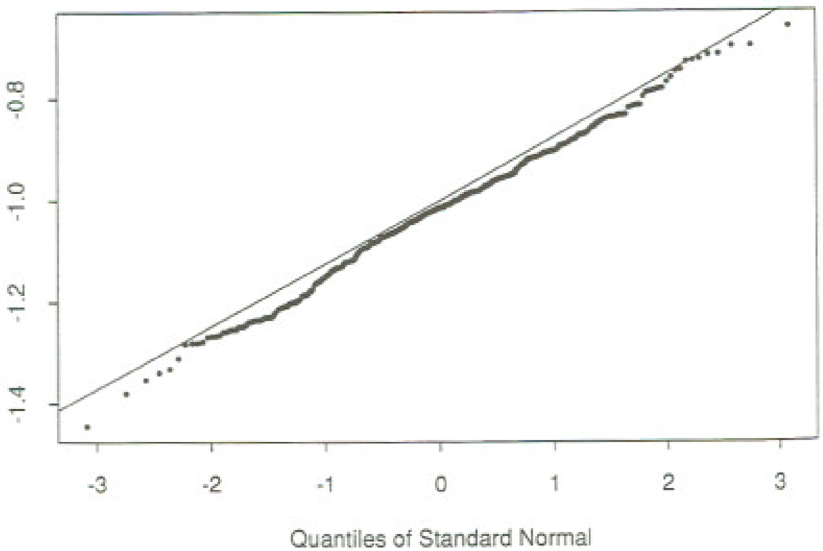

Figure 4.2. A normal quantile plot of the 500 values of $\tilde{\theta}_n$ ($n = 500$, $\Delta = 0.4$) in the case of the Ornstein–Uhlenbeck process with $\theta = -1$, $\sigma = 1$, and $x_0 = 0$. The straight line corresponds to the asymptotic normal distribution. The discrepancy is mainly due to a slight bias

Various numbers of observations of the mean-reverting process in different time intervals were produced according to the order 1.5 strong Taylor scheme. For each set of simulated observations the parameters α and θ were estimated using the discretized continuous-time likelihood function, $\tilde{G}$, and G^*. The estimating equations corresponding to G^* were solved using a generalization of Newton's method called linearization (see Reverchon and Ducamp 1993, p. 313).

From Tables 4.2–4.4 we see that both $\tilde{G}$ and G^* give much better estimates than the discretized continuous-time likelihood function. We note, however, that as Δ is decreased, α^{d} and θ^{d} come closer to their true values. The relatively small bias of the estimators based on $\tilde{G}$ and G^* is almost the same. It decreases with sample size in most cases. The standard error increases with increasing values of Δ and with decreasing number of observations. Note that the estimators based on $\tilde{G}$ have a standard error that is only slightly larger than that of the estimators based on G^*. The reason why the number of simulations are different is that runs resulting in negative values on the right-hand side of (2.18) were discarded. In such cases the estimators do not exist.

The normal quantile plot in Fig. 4.3 shows that the distribution of $\tilde{\theta}$ is close to normal in the case where the number of observations is 500 and $\Delta = 0.5$. The estimator $\tilde{\alpha}$ behaves similarly.

In the case where $\psi(x) = \sqrt{\theta + x^2}$, $\alpha = 0$, and θ is replaced by $-\theta$ in the drift, the order 1.5 strong Taylor scheme is as follows:

$$Y_{n+1} = (1 - \Delta\theta + \tfrac{1}{2}\Delta^2\theta^2)Y_n + \tfrac{1}{2}\sqrt{\theta + Y_n^2}\{\tfrac{1}{3}(\Delta W)^2 - \Delta + 2\}\Delta W$$
$$+ \tfrac{1}{2}Y_n\{(\Delta W)^2 - \Delta\} - \theta\sqrt{\theta + Y_n^2}\Delta Z + \frac{\theta}{\sqrt{\theta + Y_n^2}}(\tfrac{1}{2} - Y_n^2)(\Delta W\Delta - \Delta Z).$$

Table 4.2. Mean and standard error of the estimator based on the discretized continuous-time likelihood function for the mean-reverting process with $\psi(x) = \sigma\sqrt{x}$. Here $x_0 = 10$ and $\sigma = 1$, and the true parameter values are $\alpha = 10$ and $\theta = -1$

Δ	num. obs.	num. sim.	mean α_n^d	SE α_n^d	mean θ_n^d	SE θ_n^d
0.5	200	500	8.106	1.186	−0.8104	0.1189
0.5	500	500	8.021	0.7067	−0.8029	0.0714
0.5	1000	500	7.927	0.4984	−0.7923	0.0498
0.5	1500	500	7.912	0.4141	−0.7919	0.0412
1.0	200	500	6.458	0.6828	−0.6457	0.0689
1.0	500	500	6.404	0.4206	−0.6409	0.0420
1.0	1000	500	6.328	0.2958	−0.6328	0.0297
1.0	1500	500	6.352	0.2516	−0.6352	0.0252
1.5	200	496	5.209	0.4559	−0.5209	0.0451
1.5	500	500	5.204	0.2999	−0.5201	0.0299
1.5	1000	500	5.176	0.2141	−0.5177	0.0213
1.5	1500	500	5.180	0.1786	−0.5179	0.0179
2.0	200	466	4.297	0.2990	−0.4297	0.0298
2.0	500	495	4.337	0.2228	−0.4338	0.0223
2.0	1000	500	4.330	0.1606	−0.4331	0.0160
2.0	1500	500	4.325	0.1381	−0.4325	0.0138

Table 4.3. Mean and standard error of the estimator corresponding to $\tilde{G}$ for the mean-reverting process with $\psi(x) = \sigma\sqrt{x}$. Here $x_0 = 10$ and $\sigma = 1$, and the true parameter values are $\alpha = 10$ and $\theta = -1$

Δ	num. obs.	num. sim.	mean $\tilde{\alpha}_n$	SE $\tilde{\alpha}_n$	mean $\tilde{\theta}_n$	SE $\tilde{\theta}_n$
0.5	200	500	10.50	2.064	−1.049	0.2072
0.5	500	500	10.29	1.192	−1.030	0.1204
0.5	1000	500	10.11	0.8293	−1.011	0.0830
0.5	1500	500	10.08	0.6872	−1.009	0.0685
1.0	200	500	10.59	2.111	−1.059	0.2127
1.0	500	500	10.30	1.199	−1.031	0.1199
1.0	1000	500	10.05	0.8158	−1.005	0.0818
1.0	1500	500	10.11	0.6969	−1.011	0.0698
1.5	200	496	10.50	2.344	−1.050	0.2339
1.5	500	500	10.26	1.474	−1.025	0.1476
1.5	1000	500	10.06	0.9841	−1.006	0.0984
1.5	1500	500	10.05	0.8273	−1.005	0.0830
2.0	200	466	10.35	2.493	−1.035	0.2500
2.0	500	495	10.46	2.029	−1.046	0.2036
2.0	1000	500	10.22	1.390	−1.022	0.1390
2.0	1500	500	10.13	1.097	−1.013	0.1097

Table 4.4. Mean and standard error of the estimator corresponding to G^* for the mean-reverting process with $\psi(x) = \sigma\sqrt{x}$. Here $x_0 = 10$ and $\sigma = 1$, and the true parameter values are $\alpha = 10$ and $\theta = -1$

Δ	num. obs.	num. sim.	mean α_n^*	SE α_n^*	mean θ_n^*	SE θ_n^*
0.5	200	500	10.49	2.058	−1.048	0.2066
0.5	500	500	10.29	1.190	−1.030	0.1202
0.5	1000	500	10.11	0.8272	−1.010	0.0828
0.5	1500	500	10.09	0.6865	−1.009	0.0684
1.0	200	500	10.59	2.109	−1.059	0.2127
1.0	500	500	10.30	1.193	−1.031	0.1194
1.0	1000	500	10.05	0.8165	−1.005	0.0819
1.0	1500	500	10.11	0.6968	−1.011	0.0698
1.5	200	496	10.48	2.327	−1.048	0.2322
1.5	500	500	10.26	1.444	−1.025	0.1446
1.5	1000	500	10.06	0.9751	−1.006	0.0975
1.5	1500	500	10.05	0.8171	−1.005	0.0820
2.0	200	466	10.30	2.389	−1.030	0.2395
2.0	500	495	10.44	1.993	−1.045	0.2000
2.0	1000	500	10.22	1.336	−1.022	0.1335
2.0	1500	500	10.12	1.085	−1.012	0.1085

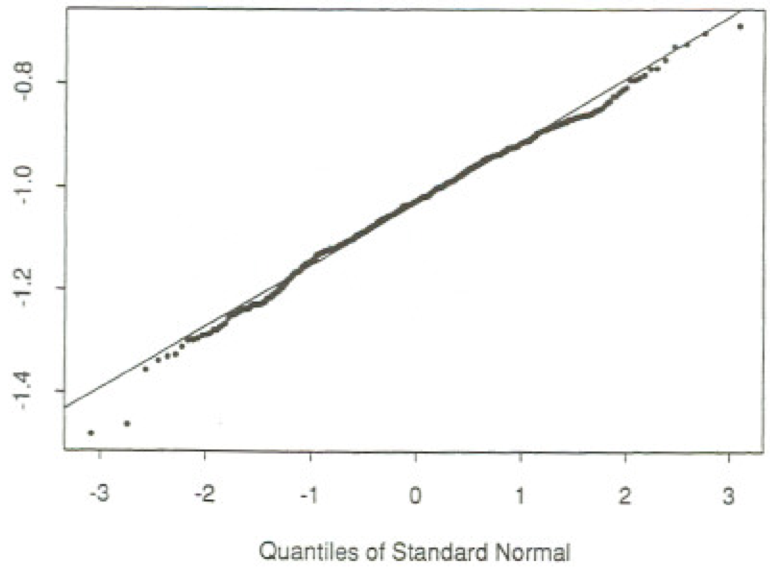

Figure 4.3. A normal quantile plot of 500 values of $\tilde{\theta}_n$ for the mean-reverting process with $\psi(x) = \sigma\sqrt{x}$, $\sigma = 1$, $\alpha = 10$, $\theta = -1$, and $x_0 = 10$. The number of observations is 500 and $\Delta = 0.5$. The straight line is based on the sample mean and the sample variance

Table 4.5. Mean and standard error of the estimator based on the discretized continuous-time likelihood function and the estimators corresponding to $\tilde{G}$ and G^* for the mean-reverting process with $\psi(x) = \sqrt{\theta + x^2}$, $\alpha = 0$, and θ replaced by $-\theta$ in the drift. Here $x_0 = 0$ and $\sigma = 1$, and the true parameter value is $\theta = 10$

Δ	num. obs.	num. sim.	mean θ_n^d	SE θ_n^d	mean $\tilde{\theta}_n$	SE $\tilde{\theta}_n$	mean θ_n^*	SE θ_n^*
0.05	500	500	7.952	0.7136	10.18	1.197	10.18	1.197
0.05	1000	500	7.877	0.5117	10.03	0.8507	10.03	0.8524
0.1	500	500	6.337	0.4228	10.12	1.175	10.13	1.174
0.1	1000	500	6.341	0.3056	10.09	0.8366	10.09	0.8370
0.2	500	493	4.316	0.2179	10.26	2.043	10.26	2.105
0.2	1000	500	4.318	0.1625	10.12	1.309	10.13	1.317
0.3	1000	463	3.160	0.0892	10.38	2.122	10.39	2.182

Table 4.5 shows that the estimator based on the discretized continuous-time likelihood function is strongly biased and that it becomes more and more inaccurate as Δ is increased. The same tendency is found for $\tilde{\theta}_n$ and θ_n^*, but not nearly to the same extent. The estimators $\tilde{\theta}_n$ and θ_n^* behave almost identically, and their slight bias decreases with increasing sample size. The standard error decreases with increasing number of observations.

Normal quantile plots show that for $\Delta = 0.1$ normality for $\tilde{\theta}_n$ is approximately reached when the number of observations is 1000. When the number of observations is 500, there are deviations from a normal distribution, particularly in the tails.

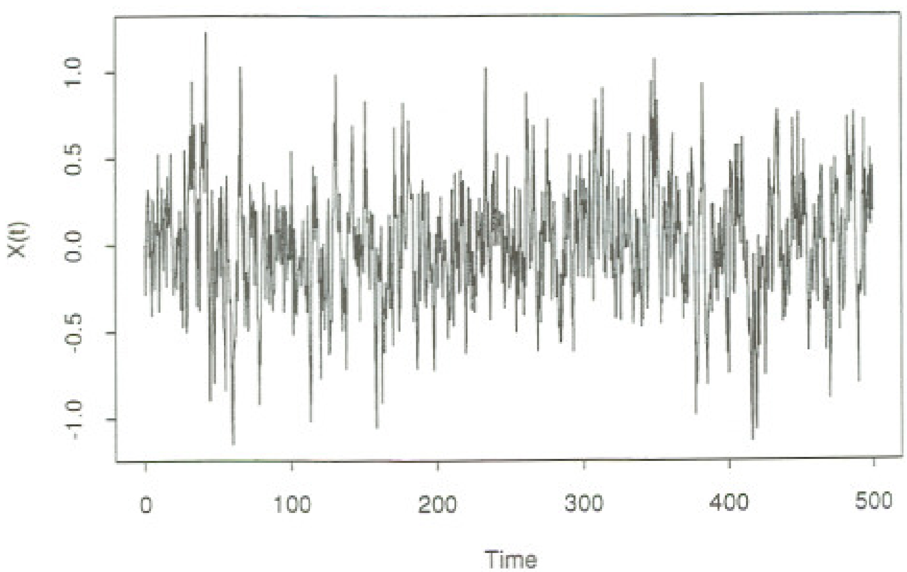

Figure 4.4. A typical simulated trajectory for the hyperbolic diffusion process. The number of simulated points is 1000 and $\Delta = 0.5$. The parameters are $\theta = -1$, $\sigma = 0.5$, and $x_0 = 0$

Table 4.6. Mean and standard error of the estimator based on the discretized continuous-time likelihood function and the estimators corresponding to $\tilde{G}$ and $G^{\dagger}$ in the case of the hyperbolic diffusion process. Here $x_0 = 0$ and $\sigma = 0.5$, and the true parameter value is $\theta = -1$

Δ	numb. obs.	m	N	mean θ_n^{d}	SE θ_n^{d}	mean $\tilde{\theta}_n$	SE $\tilde{\theta}_n$	mean $\theta_n^{\dagger}$	SE $\theta_n^{\dagger}$
0.25	200	25	25	−0.9457	0.1969	−1.070	0.2572	−1.070	0.2595
0.25	500	25	25	−0.9222	0.1249	−1.034	0.1581	−1.035	0.1563
0.25	1000	25	25	−0.9114	0.0823	−1.019	0.1037	−1.023	0.1161
0.5	200	25	25	−0.7958	0.1120	−0.9837	0.1744	−1.001	0.1934
0.5	500	25	25	−0.8093	0.0740	−1.003	0.1165	−1.005	0.1296
0.5	1000	25	25	−0.8082	0.0552	−0.9992	0.0870	−1.001	0.0985
1.0	200	25	25	−0.6776	0.0688	−1.036	0.1724	−0.7508	0.1300
1.0	500	25	25	−0.6774	0.0483	−1.027	0.1187	−0.7883	0.1128
1.0	1000	25	25	−0.6719	0.0321	−1.012	0.0755	−0.8044	0.0897

Example 4.3 *The hyperbolic diffusion process, continued*
As for the mean-reverting process we use the order 1.5 strong Taylor scheme to simulate realizations of the hyperbolic diffusion process. Here it is of the form

$$Y_{n+1} = Y_n + \theta\Delta\frac{Y_n}{\sqrt{1+Y_n^2}} + \sigma\Delta W + \frac{\theta\Delta^2 Y_n}{2(1+Y_n^2)^2}\left(\theta - \frac{3\sigma^2}{2\sqrt{1+Y_n^2}}\right) + \frac{\sigma\theta}{(1+Y_n^2)^{3/2}}\Delta Z,$$

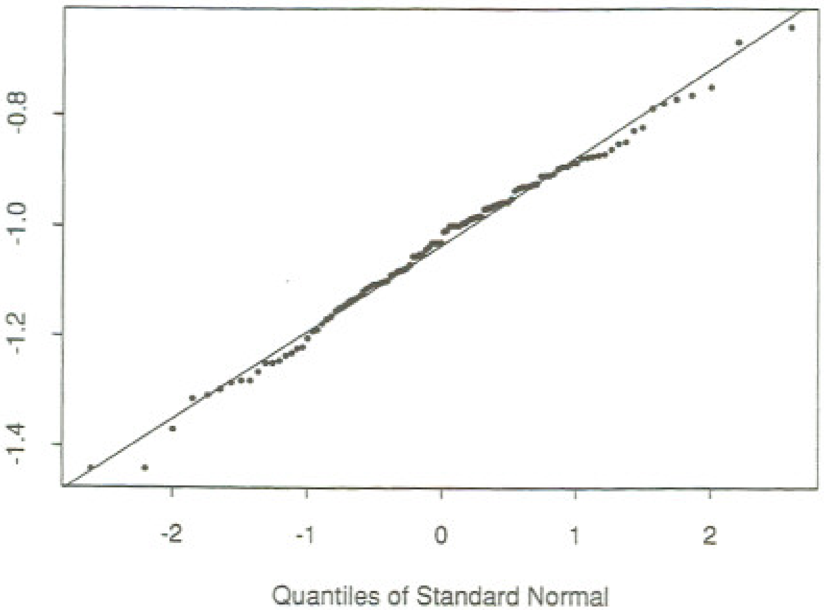

Figure 4.5. A normal quantile plot of 100 values of $\tilde{\theta}_n$ for the hyperbolic diffusion process with $\theta = -1$, $\sigma = 0.5$, and $x_0 = 0$. The number of observations is 500 and $\Delta = 0.25$. The straight line is based on the sample mean and the sample variance

where $\Delta W = \sqrt{\Delta} U_1$, $\Delta Z = \Delta^{3/2}(U_1 + U_2/\sqrt{3})/2$, and U_1 and U_2 are independent standard normal random variables. Figure 4.4 shows a typical realization of the hyperbolic diffusion process.

Because of numerical instability when calculating $\dot{F}$, we focus our attention on $\tilde{G}$ and $G^{\dagger}$. The parameter θ was estimated based on the discretized continuous-time likelihood function, $\tilde{G}$, and $G^{\dagger}$. The mean and the standard error in Table 4.6 are based on 100 simulations. The method used for solving the estimating equations is bisection.

Looking at Table 4.6, we note that, except for large values of Δ, the estimators for θ have much the same bias for $\tilde{G}$ and $G^{\dagger}$ and are very close to the true value. When Δ is la·ge $\tilde{\theta}$ behaves better than $\theta^{\dagger}$. The estimator for θ based on the discretized continuous-time likelihood function, θ^{d}, does well for small values of Δ, but poorly for large values of Δ. The standard error is decreasing with increasing number of observations and with increasing values of Δ.

The size of the number of repeats, m, and the number of points in between, N (see (4.1), (4.2) and the proceeding paragraph), has almost no effect on the estimates when N is sufficiently large.

Figure 4.5 shows that the distribution of $\tilde{\theta}_n$ is close to normal in the case where the number of observations is 500 and $\Delta = 0.25$. A similar plot shows that when the number of observations is 200 normality has not yet been reached.

Acknowledgement

The first named author gratefully acknowledges the support of the Research Centre Foulum.

References

Barndorff-Nielsen, O.E. and Sørensen, M. (1994) A review of some aspects of asymptotic likelihood theory for stochastic processes. *Internat. Statist. Rev.*, **62**(1).

Brown, B.M. and Hewitt, J.T. (1975) Asymptotic likelihood theory for diffusion processes. *J. Appl. Prob.*, **12**, 228–238.

Cox, J.C., Ingersoll, Jr, J.E. and Ross, S.A. (1985) A theory of the term structure of interest rates. *Econometrica*, **53**(2), 385–407.

Dacunha-Castelle, D. and Florens-Zmirou, D. (1986) Estimation of the coefficients of a diffusion from discrete observations. *Stochastics*, **19**, 263–284.

Florens-Zmirou, D. (1984) Théorème de limite centrale pour une diffusion et pour sa discrétisée. *C.R. Acad. Sci. Paris, Ser. I*, **299**(19), 995–998.

Florens-Zmirou, D. (1989) Approximate discrete-time schemes for statistics of diffusion processes. *Statistics*, **20**(4), 547–557.

Genon-Catalot, V. (1990) Maximum contrast estimation for diffusion processes from discrete observations. *Statistics*, **21**(1), 99–116.

Godambe, V.P. and Heyde, C.C. (1987) Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.*, **55**, 231–244.

Heyde, C.C. (1988) Fixed sample and asymptotic optimality for classes of estimating functions. *Contemporary Mathematics*, **80**, 241–247.

Ibragimov, I.A. and Has'minskii, R.Z. (1981) *Statistical Estimation – Asymptotic Theory*. New York: Springer-Verlag.

Kloeden, P.E. and Platen, E. (1992) *Numerical Solution of Stochastic Differential Equations*. Springer-Verlag.

Kloeden, P.E., Platen, E., Schurz, H. and Sørensen, M. (1992) On effects of discretization on estimators of drift parameters for diffusion processes. Research Report No. 249, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus.

Kutoyants, A.Y. (1984) *Parameter Estimation for Stochastic Processes*. Berlin: Heldermann Verlag.

Le Breton, A. (1976) On continuous and discrete sampling for parameter estimation in diffusion type processes. *Math. Programming Studies*, **5**, 124–144.

Liptser, R.S. and Shiryayev, A.N. (1977) *Statistics of Random Processes I, II*. New York: Springer-Verlag.

Pedersen, A.R. (1995a) A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Research Report No. 264, Department of Theoretical Statistics, Institute of Mathematics, University of Aarhus. *Scand. J. Statist.* **22**, 55–71.

Pedersen, A.R. (1995b) Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. To appear in *Bernoulli*.

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

Reverchon, A. and Ducamp, M. (1993) *Mathematical Software Tools in C++*. Wiley.

Sørensen, M. (1991) Likelihood methods for diffusions with jumps. In N.U. Prabhu and I.V. Basawa (eds), *Statistical Inference in Stochastic Processes*, pp. 67–105. New York: Marcel Dekker.

Taraskin, A.F. (1974) On the asymptotic normality of vector valued stochastic integrals and estimates of multi-dimensional diffusion processes. *Theory Probab. Math. Statist.*, **2**, 209–224.

Yoshida, N. (1990) Estimation for diffusion processes from discrete observations. *J. Multivariate. Anal.* **41**, 220–242.

Received March 1994 and revised August 1994