Institute of Mathematical Statistics

LECTURE NOTES — MONOGRAPH SERIES

ESTIMATING FUNCTIONS FOR DISCRETELY OBSERVED DIFFUSIONS: A REVIEW

Michael Sørensen
University of Aarhus
Denmark

Abstract

Several estimating functions for discretely observed diffusion processes are reviewed. First we discuss simple explicit estimating functions based on Gaussian approximations to the transition density. The corresponding estimators often have considerable bias, a problem that can be avoided by using martingale estimating functions. These, on the other hand, are rarely explicit and therefore often require a considerable computational effort. We review results on how to choose an optimal martingale estimating function and on asymptotic properties of the estimators. Martingale estimating functions based on polynomials of the increments of the observed process or on eigenfunctions for the generator of the diffusion model are considered in more detail. The theory is illustrated by examples. In particular, the Cox-Ingersoll-Ross model is considered.

Key Words: Approximate likelihood function; asymptotic normality; bias; consistency; Cox-Ingersoll-Ross model; eigenfunctions; inference for diffusion processes; martingale estimating functions; optimal inference; polynomial estimating functions; quasi likelihood.

1 Introduction

Diffusion processes often provide a useful alternative to the discrete time stochastic processes traditionally used in time series analysis as models for observations at discrete time points of a phenomenon that develops dynamically in time. In many fields of application it is natural to model the dynamics in continuous time, whereas dynamic modelling in discrete time contains an element of arbitrariness. This is particularly so when the time between observations is not equidistant.

Statistical inference for diffusion processes based on discrete time observations can only rarely be based on the likelihood function as this is usually not explicitly available. The likelihood function is a product of transition densities, as follows easily from the fact that diffusions are Markov processes, but explicit expressions for the transition densities are only known in some special cases. One way around this problem is to find good approximations to the likelihood function by means of simulation methods for diffusions. This computer-intensive approach has been pursued in Pedersen (1995a, 1995b). Another solution is to base the inference on estimating functions. In this paper we review a number of recent contributions to this approach.

The likelihood theory for continuously observed diffusions is well studied. In practice, however, diffusions are not observed continuously, but only at discrete time points or for instance through an electronic filter. There is therefore a need of methods which are applicable in statistical practice, and in recent years this has inspired quite a lot of work on estimation for discretely observed diffusions. The need has been particularly acute in finance where diffusion models must be fitted to time series of stock prices, interest rates or currency exchange rates in order to price derivative assets such as options.

In Section 2 we discuss simple explicit estimating functions based on Gaussian approximations to the transition density. The corresponding estimators often have considerable bias, a problem which we discuss in some detail. When the distance between the observation times is sufficiently small, they are, however, useful in practice. Asymptotic results substantiating this claim are reviewed. The bias problems, to a large extend, can be avoided by using martingale estimating functions instead, which are treated in Section 3. Martingale estimating functions are, on the other hand, rarely explicit, and therefore often requires a considerable computational effort. We review results on how to choose an optimal martingale estimating function and on asymptotic properties of the estimators. Martingale estimating functions based on polynomials of the increments of the observed process or on eigenfunctions for the generator of the diffusion model are considered in more detail. A different kind of estimating functions, by which the bias problems discussed in Section 2 can also be avoided, and which have the advantage of being explicit, were recently proposed by Kessler (1996). Unfortunately, these can not be discussed in this relatively short review paper.

2 Simple explicit estimating functions

We consider one-dimensional diffusion processes defined as solutions of the following class of stochastic differential equations

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

where W is a standard Wiener process. We assume that the drift b and the diffusion coefficient σ are known apart from the parameter θ which varies in a subset Θ of $\mathbb{R}^d$. They are assumed to be smooth enough to ensure the existence of a unique weak solution for all θ in Θ. The assumption that the drift and the diffusion coefficient do not depend on time is not essential for several of the estimating functions discussed in this paper which can be modified in a straightforward way to diffusions that are not time-homogeneous. Also the assumption that X is one-dimensional is in several cases not needed, but is made to simplify the exposition. The statistical problem considered in this paper is to draw inference about the parameter θ on the basis of observations of the diffusion X at discrete time points: $X_{t_0}, X_{t_1}, \cdots, X_{t_n}$, $t_0 = 0 < t_1 < \cdots < t_n$. The likelihood function for θ based on $X_{t_0}, X_{t_1}, \cdots, X_{t_n}$ is

$$L_n(\theta) = \prod_{i=1}^{n} p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta), \tag{2.2}$$

where $\Delta_i = t_i - t_{i-1}$ and where $y \mapsto p(\Delta, x, y; \theta)$ is the density of X_Δ given $X_0 = x$ when θ is the true parameter value.

The transition density p is only rarely explicitly known, and when Δ is not small, it can be far from Gaussian. We can, however, obtain a number of useful estimating functions by replacing p by approximations. When Δ is small, we can approximate p by a normal density function. Expressions for the conditional moments of X_Δ given X_0 can usually not be found, so in order to get an explicit estimating function, the mean value is approximated by $x + b(x;\theta)\Delta$ and the variance by $\sigma^2(x;\theta)\Delta$. By using this approximate Gaussian transition density, we obtain an approximate likelihood function, which equals the likelihood function for the Euler-Maruyama approximation (see Kloeden and Platen, 1992) to the solution of (2.1). The corresponding score function is

$$H_n(\theta) = \sum_{i=1}^{n} \left\{ \frac{\partial_\theta b(X_{t_{i-1}};\theta)}{v(X_{t_{i-1}};\theta)} [X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}};\theta)\Delta_i] \right. \tag{2.3}$$

$$\left. + \frac{\partial_\theta v(X_{t_{i-1}};\theta)}{2v^2(X_{t_{i-1}};\theta)\Delta_i} [(X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}};\theta)\Delta_i)^2 - v(X_{t_{i-1}};\theta)\Delta_i] \right\},$$

where $v(x;\theta) = \sigma^2(x;\theta)$, and where $\partial_\theta b$ denotes the vector of partial derivatives with respect to θ. Vectors are column vectors. It is, of course, assumed that the partial derivatives in (2.3) exist. Throughout this paper, whenever a derivative appears its existence is implicitly assumed in order to avoid statements of obvious conditions. The d-dimensional estimating function (2.3) is biased because we have used rather crude approximations for the mean value and the variance of the transition distribution. Therefore it can only be expected to yield reasonable estimators when the Δ_i's are small, and we can only expect these estimators to be consistent and asymptotically normal if the asymptotics is not only that the length of the observation interval, t_n, goes to infinity, but also that the Δ_i's go to zero.

First consider the estimating function obtained by deleting the quadratic terms from (2.3):

$$\tilde{H}_n(\theta) = \sum_{i=1}^{n} \frac{\partial_\theta b(X_{t_{i-1}};\theta)}{v(X_{t_{i-1}};\theta)}[X_{t_i} - X_{t_{i-1}} - b(X_{t_{i-1}};\theta)\Delta]. \tag{2.4}$$

To simplify the exposition, we have here assumed that the observation times are equidistant, i.e. that $\Delta_i = \Delta$ for all i. This is the form (2.3) takes in cases where the diffusion coefficient is completely known, i.e. when it does not depend on θ, but (2.4) can obviously also be used when the diffusion coefficient depends on θ. Another way of obtaining this estimating function is by discretizing the score function based on continuous observation of the diffusion process X in the time interval $[0, t_n]$ (see Liptser and Shiryayev, 1977). The discretization is done by replacing Ito-integrals and Riemann-integrals by Ito-Riemann sums. The estimator $\tilde{\theta}_n$ obtained from (2.4), which can also be thought of as a weighted least squares estimator, was studied by Dorogovicev (1976), Prakasa Rao (1983, 1988) and Florens-Zmirou (1989) in the case where the diffusion coefficient is constant and the parameter θ is one-dimensional. Under various regularity conditions these authors showed that $\tilde{\theta}_n$ is consistent provided $\Delta_n \to 0$ and $n\Delta_n \to \infty$, where it is assumed that the time between observations, Δ_n, depends on the sample size n. Note that $n\Delta_n = t_n$ is the lenght of the observation interval. To prove asymptotic normality a stronger condition is needed. Prakasa Rao (1983, 1988) assumed that Δ_n tends to zero sufficiently fast that $n\Delta_n^2 \to 0$, and referred to this condition as a rapidly increasing experimental design assumption. Florens-Zmirou made the slightly weaker assumption that $n\Delta_n^3 \to 0$ in her result on asymptotic normality. We shall not state the results of these authors in details as a more general result will be given below.

A different type of asymptotics, which has turned out to be relevant in several applications, was studied by Genon-Catalot (1990). She considered the situation where the length of the observation interval $n\Delta_n$ is fixed and the diffusion coefficient is a constant σ^2 tending to zero as the number of observations n tends to infinity. Under reasonable regularity conditions she showed that the estimator $\tilde{\theta}_n$ based on (2.4) is consistent provided $\sigma \sim n^{-\beta}$ where $\beta \geq 0.5$ and asymptotically normal (and asymptotically efficient) under the additional condition $\beta < 1$.

These various asymptotic results indicate that estimators based on (2.3) or (2.4) behave reasonably well in practice when the time between observations Δ is sufficiently small. This has been confirmed in simulation studies, see e.g. Kloeden et al. (1996). However, when Δ is not small, the estimators can be severely biased, as demonstrated in simulation studies by Pedersen (1995a) and Bibby and Sørensen (1995). In practice it can be difficult to determine whether in concrete models Δ is sufficiently small for the estimators to work well.

Estimation based on (2.3) or (2.4) has been popular in the econometric literature under the name the generalized method of moments, a somewhat odd name, as the method is obviously not a method of moments, except approximately.

The problem with the simple estimating functions (2.4) and (2.3) is that they can be strongly biased. An idea about the magnitude of the bias can be obtained from the expansions (Florens-Zmirou, 1989 and Kessler 1997)

$$\begin{aligned} E_\theta(X_\Delta|X_0 = x) = x &+ \Delta b(x;\theta) + \tfrac{1}{2}\Delta^2\{b(x;\theta)\partial_x b(x;\theta) \\ &+ \tfrac{1}{2}v(x;\theta)\partial_x^2 b(x;\theta)\} + O(\Delta^3) \end{aligned} \tag{2.5}$$

and

$$\begin{aligned} \mathrm{Var}_\theta(X_\Delta|X_0 = x) &= \Delta v(x;\theta) + \Delta^2[\tfrac{1}{2}b(x;\theta)\partial_x v(x;\theta) \\ &+ v(x;\theta)\{\partial_x b(x;\theta) + \tfrac{1}{4}\partial_x^2 v(x;\theta)\}] + O(\Delta^3), \end{aligned} \tag{2.6}$$

where E_θ and Var_θ denote expectation and variance, respectively, when θ is the true parameter value, and where ∂_x^2 denotes the second partial derivative with respect to x. Suppose X is an ergodic diffusion with invariant probability measure μ_θ when θ is the true parameter value. If $X_0 \sim \mu_\theta$, we find, by (2.5), the following expression for the bias of the estimating function (2.4)

$$E_\theta(\tilde{H}_n(\theta)) = \tfrac{1}{2}\Delta^2 n E_{\mu_\theta}\{\partial_\theta b(\theta)[b(\theta)\partial_x b(\theta)/v(\theta) + \tfrac{1}{2}\partial_x^2 b(\theta)]\} + O(n\Delta^3). \tag{2.7}$$

For a function $(x,\theta) \mapsto g(x;\theta)$ we use the notation $E_{\mu_\theta}(g(\theta)) = \int g(x;\theta)d\mu_\theta(x)$. When the initial distribution is different from μ_θ, (2.7) is, by the ergodic theorem (see e.g. Billingsley, 1961 or Florens-Zmirou, 1989), still a good estimate of the bias provided the number of observations is sufficiently large. Under weak standard regularity conditions (e.g. conditions similar to Condition 3.3 below) it follows that the asymptotic bias ($n \to \infty$, Δ fixed) of the estimator $\tilde{\theta}_n$ derived from (2.4) is

$$\Delta \frac{E_{\mu_\theta}\{\partial_\theta b(\theta)[b(\theta)\partial_x b(\theta)/v(\theta) + \tfrac{1}{2}\partial_x^2 b(\theta)]\}}{2E_{\mu_\theta}\{(\partial_\theta b(\theta))^2/v(\theta)\}} + O(\Delta^2) \tag{2.8}$$

in the case of a one-dimensional parameter. The expression analogous to (2.7) for the estimating function (2.3) is

$$E_\theta(H_n(\theta)) = \tfrac{1}{2}\Delta n E_{\mu_\theta}\{\partial_\theta \log v(\theta) \quad [\ \tfrac{1}{2}b(\theta)\partial_x \log v(\theta) + \partial_x b(\theta) \\ + \tfrac{1}{4}\partial_x^2 v(\theta)]\} + O(n\Delta^2), \tag{2.9}$$

as is easily seen from (2.6). The fact that the bias of the estimating function (2.3) is of order $n\Delta$ when the diffusion coefficient depends on θ indicates that the corresponding estimator has a considerable bias even for small values of Δ. The reason is that in deriving (2.3) we used an approximation of the variance of the transition distribution that was too crude, see the discussion below.

Example 2.1 Consider the Cox-Ingersoll-Ross model, which is widely used in mathematical finance to model interst rates (Cox, Ingersoll and Ross, 1985). The model is given by the stochastic differential equation

$$dX_t = (\alpha + \theta X_t)dt + \sigma\sqrt{X_t}dW_t, \quad X_0 = x_0 > 0,$$

where $\theta < 0$ and $\sigma > 0$. The model has also been used in other applications, e.g. mathematical biology, for a long time. The state space is $(0, \infty)$.

It is not difficult to derive an estimator for the parameter vector $(\alpha, \theta, \sigma^2)$ from (2.3). To simplify things we assume equidistant sampling times.

$$\begin{aligned}
\tilde{\alpha}_n &= \frac{(X_{t_n} - x_0)(\frac{1}{n}\sum_{i=1}^n X_{t_{i-1}})^{-1} - \sum_{i=1}^n X_{t_{i-1}}^{-1}(X_{t_i} - X_{t_{i-1}})}{\Delta[n^2(\sum_{i=1}^n X_{t_{i-1}})^{-1} - \sum_{i=1}^n X_{t_{i-1}}^{-1}]} \\
\tilde{\theta}_n &= \frac{\sum_{i=1}^n X_{t_{i-1}}^{-1}(X_{t_i} - X_{t_{i-1}}) - \frac{1}{n}(X_{t_n} - x_0)\sum_{i=1}^n X_{t_{i-1}}^{-1}}{\Delta[n - (\sum_{i=1}^n X_{t_{i-1}})(\sum_{i=1}^n X_{t_{i-1}}^{-1})/n]} \\
\tilde{\sigma}_n^2 &= \frac{1}{n\Delta}\sum_{i=1}^n X_{t_{i-1}}^{-1}[X_{t_i} - X_{t_{i-1}} - (\tilde{\alpha}_n + \tilde{\theta}_n X_{t_{i-1}}\Delta)]^2.
\end{aligned}$$

For parameter values where X is ergodic, an expression for the bias of these estimators when n is large can easily be found using the ergodic theorem and the fact that the invariant probability measure for the Cox-Ingersoll-Ross model is a gamma distribution. The bias can for some parameter values be dramatic even for rather small values of Δ, see Bibby and Sørensen (1995).

□

The bias considerations above raise the question whether better estimators can be obtained by improving the approximations of the mean and variance in the Gaussian approximation to the transition distribution. Useful approximations were derived by Kessler (1997) under the following condition.

Condition 2.2 *For every θ the functions $b(x;\theta)$ and $\sigma(x;\theta)$ are K times continuously differentiable with respect to x and the derivatives are of polynomial growth in x uniformly in θ.*

In order to formulate Kessler's expansions we need the *generator* of the diffusion process given by (2.1), i.e. the differential operator

$$L_\theta = b(x;\theta)\frac{d}{dx} + \tfrac{1}{2}v(x;\theta)\frac{d^2}{dx^2}. \tag{2.10}$$

With the definition

$$r_k(\Delta, x;\theta) = \sum_{i=0}^{k} \frac{\Delta^i}{i!} L_\theta^i f(x), \tag{2.11}$$

where $f(x) = x$, and where L_θ^i denotes i-fold application of the differential operator L_θ, Kessler (1997) proved that

$$E_\theta(X_\Delta | X_0 = x) = r_k(\Delta, x;\theta) + O(\Delta^{k+1}), \tag{2.12}$$

provided $k \leq K/2 + 1$. Note that (2.5) is a particular case of (2.12). The dependence of the O-term on x and θ has been suppressed here. Kessler (1997) gave an upper bound for the term $O(\Delta^{k+1})$ which is uniform in θ.

For fixed x, y and θ the function $(y - r_k(\Delta, x;\theta))^2$ is a polynomial of order $2k$ in Δ. Define $g_{x,\theta}^j(y)$, $j = 0, 1, \cdots, k$ by

$$(y - r_k(\Delta, x;\theta))^2 = \sum_{j=0}^{k} \Delta^j g_{x,\theta}^j(y) + O(\Delta^{k+1}),$$

and $\Gamma_k(\Delta, x;\theta)$ by

$$\Gamma_k(\Delta, x;\theta) = \sum_{j=0}^{k} \Delta^j \sum_{r=0}^{k-j} \frac{\Delta^r}{r!} L_\theta^r g_{x,\theta}^j(x). \tag{2.13}$$

Kessler (1997) showed that

$$E_\theta([X_\Delta - r_k(\Delta, x;\theta)]^2 | X_0 = x) = \Gamma_k(\Delta, x;\theta) + O(\Delta^{k+1}) \tag{2.14}$$

for $k \leq K/2 + 1$. Also in this case he gave an upper bound for the term $O(\Delta^{k+1})$.

We can now obtain an approximation to the likelihood function (2.2), which is considerably better than the approximation we used above, by replacing the transition density $y \mapsto p(\Delta, x, y;\theta)$ by a normal density with mean value $r_k(\Delta, x; \theta)$ and variance $\Gamma_{k+1}(\Delta, x; \theta)$ with $k \leq K/2$. The corresponding estimating function (approximate score function) is

$$\begin{aligned} H_n^{(k)}(\theta) &= \sum_{i=1}^{n} \frac{\partial_\theta r_k(\Delta_i, X_{t_{i-1}}; \theta)}{\Gamma_{k+1}(\Delta_i, X_{t_{i-1}}; \theta)} [X_{t_i} - r_k(\Delta_i, X_{t_{i-1}}; \theta)] \qquad (2.15) \\ &+ \sum_{i=1}^{n} \frac{\partial_\theta \Gamma_{k+1}(\Delta_i, X_{t_{i-1}}; \theta)}{2\Gamma_{k+1}(\Delta_i, X_{t_{i-1}}; \theta)^2} [(X_{t_i} - r_k(\Delta_i, X_{t_{i-1}}; \theta))^2 - \Gamma_{k+1}(\Delta_i, X_{t_{i-1}}; \theta)]. \end{aligned}$$

We have again allowed the time between observations to vary.

Example 2.3 To avoid complicated expressions we consider as an example the Ornstein-Uhlenbeck process

$$dX_t = \theta X_t dt + \sigma dW_t, \ \ X_0 = x_0,$$

where $\theta \in \mathbb{R}$ and $\sigma > 0$. Long, but easy, calculations show that the estimators for θ and σ^2 based on $H_n^{(2)}$ are

$$\begin{aligned} \tilde{\theta}_{2,n} &= \Delta^{-1}(\sqrt{2Q_n - 1} - 1) \\ \tilde{\sigma}^2_{2,n} &= \frac{\frac{1}{n}\sum_{i=1}^{n}(X_{t_i} - X_{t_{i-1}}Q_n)}{\Delta + \tilde{\theta}_{2,n}\Delta^2 + \frac{2}{3}\tilde{\theta}^2_{2,n}\Delta^3}, \end{aligned}$$

provided

$$Q_n = \frac{\sum_{i=1}^{n} X_{t_i} X_{t_{i-1}}}{\sum_{i=1}^{n} X^2_{t_{i-1}}} \geq \frac{1}{2}.$$

To simplify matters we have assumed that the observation times are equidistant. There are, in fact, two solution for θ, but a moments reflection reveals that the other solution is not a good estimator.

□

Suppose X is ergodic with invariant probability measure μ_θ, all moments of which are finite. Then we find, using (2.12) and (2.14), that the bias of the estimating function $H_n^{(k)}$ is of order $O(n\Delta^{k+1})$. This indicates that for k sufficiently large the estimator obtained from $H_n^{(k)}$ is only slightly biased when Δ is not too large. This is indeed the case.

In order to avoid technical problems Kessler (1997) modified the approximate Gaussian likehood function we used to derive the estimating function (2.15) by replacing the functions $\log\{\Gamma_{k+1}(\Delta_i, x; \theta)/(\Delta_i v(x; \theta))\}$ and $\Delta_i v(x; \theta)/\Gamma_{k+1}(\Delta_i, x; \theta)$ by Taylor expansions to order k. The estimating function derived from Kessler's approximate likelihood function of order k

differs only from $H_n^{(k)}$ by terms of order $O(\Delta^{k+1})$. Therefore the estimator based on $H_n^{(k)}$ behaves in the same way as the estimator based on Kessler's (1997) approximate likelihood function, for which he gave results under (essentially) the following conditions.

Condition 2.4
1) For every θ there exists a constant C_θ such that

$$|b(x;\theta) - b(y;\theta)| + |\sigma(x;\theta) - \sigma(y;\theta)| \leq C_\theta |x - y|$$

for all x and y in the state space.

2) $\inf_{x,\theta} v(x,\theta) > 0$.

3) The functions $b(x;\theta)$ and $\sigma(x;\theta)$ and all their partial x-derivatives up to order K are three times differentiable with respect to θ for all x in the state space. All these derivatives with respect to θ are of polynomial growth in x uniformly in θ.

4) The process X is ergodic for every θ with invariant probability measure μ_θ. All polynomial moments of μ_θ are finite.

5) For all $p \geq 0$ and for all θ $\sup_t E_\theta(|X_t|^p) < \infty$.

Kessler further assumed that $\theta = (\alpha, \beta)$ belongs to a compact subset Θ of $\mathbb{R}^2$, that the drift depends only on α and that the diffusion coefficient depends only on β. Moreover, he imposed an obvious identifiability condition. The assumption that θ belongs to a compact set is only made to avoid technical problems concerning the existence of a maximum of the approximate likelihood function. Kessler (1997) proved the following result about the asymptotic properties of the estimator $\hat{\theta}_{k,n}$ which maximizes his approximate likelihood function. The observation times are assumed to be equidistant with spacing Δ_n, which depends on the sample size.

Theorem 2.5 *Assume that $k \leq K/2$ and that Condition 2.2 and Condition 2.4 hold. Then for all $\theta \in \Theta$*

$$\hat{\theta}_{k,n} \to \theta \tag{2.16}$$

in P_θ-probability as $n \to \infty$, provided $\Delta_n \to 0$ and $n\Delta_n \to \infty$.
If, in addition, $n\Delta^{2k+1} \to 0$ and $\theta \in \operatorname{int} \Theta$, then as $n \to \infty$

$$(\sqrt{n\Delta}(\hat{\alpha}_{k,n} - \alpha), \sqrt{n}(\hat{\beta}_{k,n} - \beta)) \to N(0, I(\theta)^{-1}), \tag{2.17}$$

in distribution under P_θ, where

$$I(\theta) = \begin{pmatrix} E_{\mu_\theta}[(\partial_\alpha b(\alpha))^2/v(\beta)] & 0 \\ 0 & \frac{1}{2}E_{\mu_\theta}[(\partial_\beta \log v(\beta))^2] \end{pmatrix}.$$

The estimating functions considered in this section were all derived from an approximate (or pseudo) likelihood function. This has the advantage that if there are more than one solution to the estimating equation, we can choose the one that is the global maximum point for the pseudo likelihood function. The estimating functions considered in the next section do not generally have this property.

3 Martingale estimating functions

The problems caused by the bias of the estimating functions considered in Section 2 can most conveniently be avoided by using martingale estimating functions. We shall therefore in this section, for the same kind of data as those considered in Section 2, study estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \tag{3.1}$$

where the function $g(\Delta, x, y; \theta)$ satisfies

$$\int g(\Delta, x, y; \theta) p(\Delta, x, y; \theta) dy = 0 \tag{3.2}$$

for all x, Δ and θ. Here, as in the previous section, $y \mapsto p(\Delta, x, y; \theta)$ denotes the transition density, i.e. the density of X_Δ given $X_0 = x$. In most cases it is not easy to find g's that satisfy (3.2) since p is usually not known, but such g's can always be found numerically, as we shall see later. Under (3.2) $G_n(\theta)$ is a martingale when θ is the true parameter value. In particular, $G_n(\theta)$ is an unbiased estimating function. If θ is d-dimensional, we usually take g to be d-dimensional too.

With the bias problem out of the way, the question of how to choose the estimating function in an optimal way becomes more interesting. Godambe and Heyde (1987) gave criteria for choosing within a class of martingale estimating functions the one which is closest to the true (but for diffusion models usually not explicitly known) score function (fixed sample criterion) or the one which has the smallest asymptotic variance as the number of observations tends to infinity (asymptotic criterion).

Suppose we have N real valued functions $h_j(\Delta, x, y; \theta)$, $j = 1, \cdots, N$, each of which satisfies (3.2) and which are all natural choices for defining a martingale estimating function. Then every function of the form

$$g(\Delta, x, y; \theta) = \sum_{j=1}^{N} \alpha_j(\Delta, x; \theta) h_j(\Delta, x, y; \theta), \tag{3.3}$$

where $\alpha_j(\Delta, x; \theta)$, $j = 1, \cdots, N$, are arbitrary functions, can be used to define a martingale estimating function by (3.1). If θ is d-dimensional, we will usually try to find d-dimensional α's. Let $\mathcal{G}$ denote the class of d-dimensional martingale estimating functions of the form (3.1) with g given by (3.3). The following result by Kessler (1995) tells how to find the optimal estimating function in the sense of Godambe and Heyde (1987) within the class $\mathcal{G}$. We need the further assumption that for fixed Δ, x and θ the functions $h_j(\Delta, x, y; \theta)$, $j = 1, \cdots, N$ are square integrable with respect to the transition distribution. Then the set of all real-valued functions of the form (3.3) is a (finite dimensional and hence closed) linear sub-space of $L^2(p(\Delta, x, y; \theta)dy)$. We denote this subspace by $\mathcal{H}(\Delta, x; \theta)$.

Theorem 3.1 *Suppose the transition density p is differentiable with respect to θ and that for all fixed Δ, x and θ the functions $\partial_{\theta_i} \log p$, $i = 1, \cdots, d$, belong to $L^2(p(\Delta, x, y; \theta)dy)$. Denote by*

$$g_i^*(\Delta, x, y; \theta) = \sum_{j=1}^{N} \alpha_{ji}^*(\Delta, x; \theta) h_j(\Delta, x, y; \theta) \tag{3.4}$$

the projection in $L^2(p(\Delta, x, y; \theta)dy)$ of $\partial_{\theta_i} \log p$ onto $\mathcal{H}(\Delta, x; \theta)$, and define

$$G_n^*(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{N} \alpha_j^*(\Delta_i, X_{t_{i-1}}; \theta) h_j(\Delta, X_{t_{i-1}}, X_{t_i}; \theta), \tag{3.5}$$

where α_j^* is the d-dimensional vector $(\alpha_{j1}^*, \cdots, \alpha_{jd}^*)^T$ (T denotes transposition). If $g_i^*(\Delta, x, y; \theta)$ is continuously differentiable with respect to θ for all fixed Δ, x and y, then $G_n^*(\theta)$ is the optimal estimating function within $\mathcal{G}$ with respect to the asymptotic criterion as well as to the fixed sample criterion of Godambe and Heyde (1987).

The α_{ji}^*'s are determined by the following linear equations

$$C(\Delta, x; \theta) \begin{pmatrix} \alpha_{1i}^*(\Delta, x; \theta) \\ \vdots \\ \alpha_{Ni}^*(\Delta, x; \theta) \end{pmatrix} = B_i(\Delta, x; \theta), \tag{3.6}$$

for $i = 1, \cdots, d$, where $C = \{c_{kl}\}$ and $B_i = (b_1^{(i)}, \cdots, b_N^{(i)})^T$ are given by

$$c_{kl}(\Delta, x; \theta) = \int h_k(\Delta, x, y; \theta) h_l(\Delta, x, y; \theta) p(\Delta, x, y; \theta) dy \tag{3.7}$$

and

$$b_j^{(i)}(\Delta, x; \theta) = \int h_j(\Delta, x, y; \theta)\partial_{\theta_i} p(\Delta, x, y; \theta)dy. \tag{3.8}$$

When the functions $h_j(\Delta, x, y; \theta)$, $j = 1, \cdots, N$ are linearly independent in $L^2(p(\Delta, x, y; \theta)dy)$, the matrix C is obviously invertible. The condition that the estimating function is differentiable with respect to θ is really only a technical matter in the Godambe-Heyde theory, and the estimating function given by (3.5) is no doubt also the most efficient in the class $\mathcal{G}$ under a weaker condition. From (3.6), (3.7) and (3.8) we see that it is not difficult to impose conditions on the functions h_j, $j = 1, \cdots, N$ which ensure that g^* is continuously differentible with respect to θ. Note that under weak conditions ensuring that differentiation and integration can be interchanged (e.g. Condition 3.3 below), the b_j's given by (3.8) can also be expressed as

$$b_j^{(i)}(\Delta, x; \theta) = -\int \partial_{\theta_i} h_j(\Delta, x, y; \theta) p(\Delta, x, y; \theta)dy. \tag{3.9}$$

Results similar to Theorem 3.1 hold for general Markov processes and for more general classes of martingale estimating functions than those given by (3.3), see Kessler (1995).

We next give a result about the asymptotic behaviour of the estimator obtained from a general martingale estimating function $G_n(\theta)$ of the form (3.1) with g given by (3.3), where the α_j's are d-dimensional and the h_j's satisfy (3.2). We do this under the assumption that the diffusion is ergodic, which is ensured by the following condition. Here $s(x;\theta)$ denotes the density of the scale measure of X:

$$s(x;\theta) = \exp\left(-2\int_{x^\#}^{x} \frac{b(y;\theta)}{v(y;\theta)}dy\right), \tag{3.10}$$

where $x^\#$ is an arbitrary point in the interior of the state space of X.

Condition 3.2 *The following holds for all $\theta \in \Theta$:*

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

and

$$\int_{-\infty}^{\infty} [s(x;\theta)v(x;\theta)]^{-1}dx = A(\theta) < \infty.$$

If the state space of X is not the whole real line, the integration limits $-\infty$ and ∞ should be changed accordingly. Under Condition 3.2 the process X is ergodic with an invariant probability measure μ_θ which has density

$[A(\theta)s(x;\theta)v(x;\theta)]^{-1}$ with respect to the Lebesgue measure. Define a probability measure Q_θ^Δ on $\mathbb{R}^2$ by

$$Q_\theta^\Delta(x,y) = \mu_\theta(x) \times p(\Delta, x, y; \theta). \tag{3.11}$$

For a function $g : \mathbb{R}^2 \mapsto \mathbb{R}$ we use the notation $Q_\theta^\Delta(g) = \int g dQ_\theta^\Delta$.

The predictable quadratic variation of the martingale $G_n(\theta)$, when θ is the true parameter value, is

$$\langle G(\theta)\rangle_n = \sum_{i=1}^n A(\Delta, X_{t_{i-1}}; \theta)^T C(\Delta, X_{t_{i-1}}; \theta) A(\Delta, X_{t_{i-1}}; \theta), \tag{3.12}$$

where $A(\Delta, x; \theta)_{ij} = \alpha_{ij}(\Delta, x; \theta)$. As above α_{ij} denotes the j'th coordinate of the d-dimensional vector α_i. We impose the following condition on the estimating functions. From now on θ_0 will denote the true value of θ.

Condition 3.3 *The following holds for all $\theta \in \Theta$:*
1) The function g is continuously differentiable with respect to θ for all Δ, x and y. The functions $(x,y) \mapsto \partial_{\theta_i} g_j(\Delta, x, y; \theta)$, $i, j = 1, \cdots, d$, where g_j denotes the j'th coordinate of g, are locally dominated square integrable with respect to $Q_{\theta_0}^\Delta$, and the matrix $D(\theta_0)$ given by

$$D(\theta_0)_{i,j} = Q_{\theta_0}^\Delta(\partial_{\theta_j} g_i(\Delta; \theta_0)) = \sum_{k=1}^N Q_{\theta_0}^\Delta[\alpha_{ki}(\Delta; \theta_0)\partial_{\theta_j} h_k(\Delta; \theta_0)]$$

is invertible.

2) Each coordinate of the function $(x,y) \mapsto g(\Delta, x, y; \theta)$ is in $L^2(Q_{\theta_0}^\Delta)$.

Theorem 3.4 *Suppose $\theta_0 \in \operatorname{int}\Theta$ and that the Conditions 3.2 and 3.3 hold. Then an estimator $\hat\theta_n$ that solves the estimating equation*

$$G_n(\theta) = 0 \tag{3.13}$$

exists with a probability tending to one as $n \to \infty$ under P_{θ_0}. Moreover, as $n \to \infty$,

$$\hat\theta_n \to \theta_0$$

in probability under P_{θ_0}, and

$$\sqrt{n}(\hat\theta_n - \theta_0) \to N(0, D(\theta_0)^{-1} V(\theta_0)(D(\theta_0)^{-1})^T)$$

in distribution under P_{θ_0}, where

$$V(\theta_0) = E_{\mu_{\theta_0}}(A(\Delta;\theta)^T C(\Delta;\theta) A(\Delta;\theta)). \tag{3.14}$$

Theorem 3.4 can be proved along the same lines as Theorem 3.3 in Bibby and Sørensen (1995), see also Kessler (1995) and Kessler and Sørensen (1995). Similar proofs of similar results can be found in several papers. Here Condition 3.1 (c) in Bibby and Sørensen (1995) has been omitted because Lemma 3.1 in Bibby and Sørensen (1995) remains valid without this condition as follows from Theorem 1.1 in Billingsley (1961a) and the central limit theorem for martingales in Billingsley (1961b). In fact, a multivariate version of the central limit theorem is needed here, but in the relatively simple ergodic case considered here this easily follows from the one-dimensional result by applying the Cramér-Wold device.

Under Condition 3.3 the b_j's given by (3.8) can also be expressed by (3.9), so $D(\theta_0) = -V(\theta_0)$ for the optimal estimating function $G_n^*(\theta)$ since here the α's are given by (3.6). Hence the asymptotic covariance matrix of the estimator based on $G_n^*(\theta)$ is given by $V(\theta_0)^{-1}$.

3.1 Polynomial estimating functions

Let us first consider linear martingale estimating functions, i.e. estimating functions of the type

$$K_{1,n}(\theta) = \sum_{i=1}^{n} \alpha(\Delta_i, X_{t_{i-1}}; \theta)[X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)], \tag{3.15}$$

where α is d-dimensional and where

$$F(\Delta, x; \theta) = E_\theta(X_\Delta | X_0 = x). \tag{3.16}$$

In most cases the mean value of the transition distribution is not explicitly known so that it must be determined numerically. This is, however, relatively easy to do using suitable methods from Kloeden and Platen (1992). It is certainly much easier than to determine the entire transition density numerically. Estimating functions of the type (3.15) were studied in Bibby and Sørensen (1995).

The optimal linear estimating function is (Bibby and Sørensen, 1995)

$$K_{1,n}^*(\theta) = \sum_{i=1}^{n} \partial_\theta F(\Delta_i, X_{t_{i-1}}; \theta) \Phi(\Delta_i, X_{t_{i-1}}; \theta)^{-1} [X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)], \tag{3.17}$$

where

$$\Phi(\Delta, x; \theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x). \tag{3.18}$$

Calculation of a derivative of a function that has to be determined numerically is a considerably more demanding numerical problem than determination of the function itself. Pedersen (1994) proposed a numerical procedure for determining $\partial_\theta F(\Delta_i, x; \theta)$ by simulation, which works in practice, but it is easier to use the following approximation to the optimal estimating function:

$$\tilde{K}_{1,n}(\theta) = \sum_{i=1}^{n} \partial_\theta b(X_{t_{i-1}}; \theta) v(X_{t_{i-1}}; \theta)^{-1} [X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)], \qquad (3.19)$$

which is obtained from $K^*_{i,n}$ by inserting in the weight function $\partial_\theta F/\Phi$ the first order approximations to F and Φ given by (2.5) and (2.6). The estimating function $\tilde{K}_{1,n}$ can also be obtained from the estimating function (2.4) by subtracting its compensator in order to turn it into a martingale and thus remove its bias, see Bibby and Sørensen (1995).

It is very important that we have only made approximations in the weight function and not in the term $X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)$, since such an approximation would destroy the martingale property, and hence the unbiasedness, and would thus reintroduce the problems encountered in Section 2. An approximation of the weights $\partial_\theta F/\Phi$ only implies a certain loss of efficiency. Bibby and Sørensen (1995) showed that expansions in powers of Δ of the asymptotic variances of the estimators based on $K^*_{1,n}$ and $\tilde{K}_{1,n}$ agree up to and including terms of order $O(\Delta^2)$, so for small values of Δ there is not much loss of efficiency in using the approximation. Calculations and simulations for a number of examples indicate that the loss of efficiency is often rather small, see Bibby and Sørensen (1995).

The linear estimating functions are useful when mainly the drift depends on the parameter θ. If only the diffusion coefficient depends on θ, while the drift is completely known, the linear estimating equations do not work. If the diffusion coefficient depends considerably on θ, it is an advantage to use second order polynomial estimating functions of the type

$$\begin{aligned} K_{2,n}(\theta) &= \sum_{i=1}^{n} \{\alpha(\Delta_i, X_{t_{i-1}}; \theta)[X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)] \qquad (3.20) \\ &+ \beta(\Delta_i, X_{t_{i-1}}; \theta)[(X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta))^2 - \Phi(\Delta_i, X_{t_{i-1}}; \theta)]\}. \end{aligned}$$

The optimal estimating function, $K^*_{2,n}$, of this type is given by

$$\alpha^*(x;\theta) = \frac{\partial_\theta \Phi(x;\theta)\eta(x;\theta) - \partial_\theta F(x;\theta)\Psi(x;\theta)}{\Phi(x;\theta)\Psi(x;\theta) - \eta(x;\theta)^2}, \qquad (3.21)$$

and

$$\beta^*(x;\theta) = \frac{\partial_\theta F(x;\theta)\eta(x;\theta) - \partial_\theta \Phi(x;\theta)\Phi(x;\theta)}{\Phi(x;\theta)\Psi(x;\theta) - \eta(x;\theta)^2}, \qquad (3.22)$$

where the Δ's have been omitted,

$$\eta(x;\theta) = E_\theta([X_\Delta - F(x;\theta)]^3 | X_0 = x) \tag{3.23}$$

and

$$\Psi(x;\theta) = E_\theta([X_\Delta - F(x;\theta)]^4 | X_0 = x) - \Phi(x;\theta)^2. \tag{3.24}$$

An approximation to the optimal quadratic estimating function is

$$\begin{aligned} \tilde{K}_{2,n}(\theta) &= \sum_{i=1}^{n} \left\{ \frac{\partial_\theta b(X_{t_{i-1}};\theta)}{v(X_{t_{i-1}};\theta)} [X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta)] \right. \\ &+ \left. \frac{\partial_\theta v(X_{t_{i-1}};\theta)}{2v^2(X_{t_{i-1}};\theta)\Delta_i} [(X_{t_i} - F(\Delta_i, X_{t_{i-1}};\theta))^2 - \Phi(\Delta_i, X_{t_{i-1}};\theta)] \right\}. \end{aligned} \tag{3.25}$$

This estimating function is similar to (2.3), but it is unbiased and therefore generally gives a far better estimator. It is obtained from the optimal quadratic estimating function, $K^*_{2,n}$, by using Gaussian approximations to (3.23) and (3.24), i.e. $\eta(x;\theta) \doteq 0$ and $\Psi(x;\theta) \doteq 2\Phi(x;\theta)^2$, and then using the first order approximations given by (2.5) and (2.6). Again it is important that we only make approximations in the weights α and β, so that the unbiasedness is preserved.

Quadratic estimating functions were treated in Bibby (1994) and Bibby and Sørensen (1996). Higher order polynomial estimating functions were investigated by Pedersen (1994) and Kessler (1995). Some times there can be good reasons to omit lower order terms in a polynomial estimating function, for an example of this see Bibby and Sørensen (1997).

Example 3.5 Let us return to the Cox-Ingersoll-Ross model considered in Example 2.1. For this model the optimal estimating function given by (3.21) and (3.22) can be explicitly found (Bibby and Sørensen, 1996), but the corresponding estimating equation must be solved numerically. In the case of equidistant sampling times the approximately optimal estimating function (3.25) yields the following explicit estimators (Bibby and Sørensen, 1995, 1996):

$$\begin{aligned} e^{\tilde{\theta}_n \Delta} &= \frac{n \sum_{i=1}^n X_{t_i}/X_{t_{i-1}} - (\sum_{i=1}^n X_{t_i})(\sum_{i=1}^n X_{t_{i-1}}^{-1})}{n^2 - (\sum_{i=1}^n X_{t_{i-1}})(\sum_{i=1}^n X_{t_{i-1}}^{-1})} \\ \tilde{\alpha}_n &= \frac{\tilde{\theta}_n (n e^{\tilde{\theta}_n \Delta} - \sum_{i=1}^n X_{t_i}/X_{t_{i-1}})}{(1 - e^{\tilde{\theta}_n \Delta}) \sum_{i=1}^n X_{t_{i-1}}^{-1}} \\ \tilde{\sigma}_n^2 &= \frac{\sum_{i=1}^n X_{t_{i-1}}^{-1} (X_{t_i} - F(\Delta, X_{t_{i-1}}; \tilde{\alpha}_n, \tilde{\theta}_n))^2}{\sum_{i=1}^n X_{t_{i-1}}^{-1} \phi^{\#}(\Delta, X_{t_{i-1}}; \tilde{\alpha}_n, \tilde{\theta}_n)}, \end{aligned}$$

where $F(\Delta, x; \alpha, \theta) = [(\alpha + \theta x)e^{\theta\Delta} - \alpha]/\theta$ and $\phi^{\#}(\Delta, x; \alpha, \theta) = \frac{1}{2}[(\alpha + 2\theta x)e^{2\theta\Delta} - 2(\alpha + \theta x)e^{\theta\Delta} + \alpha]\theta^{-2}$. The estimators exist provided the expression for $e^{\tilde{\theta}_n\Delta}$ is positive. A simulation study in Bibby and Sørensen (1995) indicates that these estimators are quite good.

□

3.2 Estimating equations based on eigenfunctions

The polynomial estimating functions are a generalization of the method of moments to Markov processes. They can also be thought of as approximations to the true score function, which are likely to be good when the time between observations is small enough that the transition density is not too far from being Gaussian. There is therefore no reason to believe that polynomial estimating functions are in general the best possible choise when the time between observations is large and the transition distribution is far from Gaussian. We shall therefore conclude this paper by discussing a type of martingale estimating functions that can be more closely tailored to the type of diffusion model under consideration. These estimating functions were proposed and studied by Kessler and Sørensen (1995).

A twice differentiable function $\phi(x; \theta)$ is called an *eigenfunction* for the generator L_θ (given by (2.10)) of the diffusion process (2.1) if

$$L_\theta \phi(x; \theta) = -\lambda(\theta)\phi(x; \theta), \tag{3.26}$$

where the real number $\lambda(\theta)$ is called the *eigenvalue* corresponding to $\phi(x; \theta)$. Under weak regularity conditions, see Kessler and Sørensen (1995),

$$E_\theta(\phi(X_\Delta; \theta)|X_0 = x) = e^{-\lambda(\theta)\Delta}\phi(x; \theta). \tag{3.27}$$

We can therefore define a martingale estimating function by (3.1) with

$$g(\Delta, x, y; \theta) = \sum_{j=1}^{N} \alpha_j(\Delta, x; \theta)[\phi_j(y; \theta) - e^{-\lambda(\theta)\Delta}\phi_j(x; \theta)], \tag{3.28}$$

where $\phi_1(\cdot; \theta), \cdots, \phi_N(\cdot; \theta)$ are eigenfunctions for L_θ with eigenvalues $\lambda_1(\theta), \cdots, \lambda_N(\theta)$.

The optimal estimating function of this type is given by (3.6) with

$$\begin{aligned} c_{kl}(\Delta, x; \theta) &= \int \phi_k(y; \theta)\phi_l(y; \theta)p(\Delta, x, y; \theta)dy \\ &- e^{-[\lambda_k(\theta)+\lambda_l(\theta)]\Delta}\phi_k(x, \theta)\phi_l(x, \theta) \end{aligned} \tag{3.29}$$

and

$$b_j^{(i)}(\Delta, x; \theta) = -\int \partial_{\theta_i}\phi_j(y;\theta)p(\Delta, x, y;\theta)dy + \partial_{\theta_i}[e^{-\lambda_j(\theta)\Delta}\phi_j(x,\theta)]. \quad (3.30)$$

Statistical inference based on this optimal estimating function is invariant under twice continuously differentiable transformations of data, see Kessler and Sørensen (1995). After such a transformation the data are, by Ito's formula, still observations from a certain diffusion process, and the eigenfunctions transform in exactly the way needed to keep the optimal estimating function invariant. Inference based on polynomial estimating functions is obviously not invariant under transformations of the data.

Apart from this theoretical advantage, the optimal estimating functions discussed here have clear numerical advantages over the optimal polynomial estimating functions. As discussed earlier, determination of quantities like $\partial_\theta F$ in (3.17) is a difficult numerical problem. In (3.30) the derivative is under the integral sign, which makes determination of the optimal weights in estimating functions of the type (3.28) a much simpler numerical problem than the similar problem for polynomial estimating functions. Moreover, $E_\theta(\phi(X_\Delta;\theta)|X_0 = x)$ is explicitly known, so numerical inaccuracies cannot destroy the martingale property and the unbiasedness of these estimating functions. It might in some applications be reasonable to obtain a quick estimator by reducing the numerical accuracy when determining the weights, α_j, $j = 1, \cdots, N$. For the estimating equations based on eigenfunctions this only implies a certain loss of efficiency, whereas the consistency of the estimators is preserved. It is also worth noting that for models where all eigenfunctions are polynomials or polynomials of the same function, the optimal weights given by (3.29) and (3.30) can be explicitly calculated, see Kessler and Sørensen (1995). The disadvantage of these estimating functions, on the other hand, is that it is not always possible to find eigenfunction for the generator of a given diffusion model. In such cases the polynomial estimating functions, in particular the quadratic, provide a very useful alternative.

Example 3.6 For the Cox-Ingersoll-Ross model the eigenfunctions are the Laguerre polynomials, and we obtain the polynomial estimating functions discussed in the previous subsection, see Example 3.5.

□

Example 3.7 A more interesting example is the class of diffusions which solve

$$dX_t = -\theta \tan(X_t)dt + dW_t, \; X_0 = x_0.$$

For $\theta \geq \frac{1}{2}$ the process X is an ergodic diffusion on the interval $(-\pi/2, \pi/2)$, which can be thought of as an Ornstein-Uhlenbeck process on a finite interval. The eigenfunctions are $\phi_i(x;\theta) = C_i^\theta(\sin(x))$, $i = 0, 1, \cdots$, with eigenvalues $i(\theta + i/2)$, $i = 0, 1, \cdots$, where C_i^θ is the Gegenbauer polynomial of order i. The optimal estimating function based on any set of eigenfunctions can be found explicitly, see Kessler and Sørensen (1995). The optimal estimating function based on the first non-trivial eigenfunction, $\sin(x)$, is

$$G_n^*(\theta) = \sum_{i=1}^n \frac{\sin(X_{t_{i-1}})[\sin(X_{t_i}) - e^{-(\theta+\frac{1}{2})\Delta}\sin(X_{t_{i-1}})]}{\frac{1}{2}(e^{2(\theta+1)\Delta} - 1)/(\theta+1) - (e^\Delta - 1)\sin^2(X_{t_{i-1}})}.$$

When Δ is small the optimal estimating function can be approximated by

$$\tilde{G}_n(\theta) = \sum_{i=1}^n \sin(X_{t_{i-1}})[\sin(X_{t_i}) - e^{-(\theta+\frac{1}{2})\Delta}\sin(X_{t_{i-1}})],$$

which yields the explicit estimator

$$\tilde{\theta}_n = -\Delta^{-1}\log\left(\frac{\sum_{i=1}^n \sin(X_{t_{i-1}})\sin(X_{t_i})}{\sum_{i=1}^n \sin^2(X_{t_{i-1}})}\right) - 1/2,$$

provided the numerator is positive. Simulations indicate that this estimator is often almost as efficient as the optimal estimator based on G^*, see Kessler and Sørensen (1995). □

Acknowledgement: Thanks are due to a referee for a careful reading of the manuscript.

References

Bibby, B.M. (1994): Optimal combinations of martingale estimating functions for discretely observed diffusion processes. Research Report No. 298, Department of Theoretical Statistics, University of Aarhus.

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

Bibby, B.M. and Sørensen, M. (1996): On estimation for discretely observed diffusions: A review. To appear in *Theory of Stochastic Processes* **2 (18)**, 49-56.

Bibby, B.M. and Sørensen, M. (1997): A hyperbolic diffusion model for stock prices. *Finance and Stochastics* **1**, 25-41.

Billingsley, P. (1961a): *Statistical Inference for Markov Processes.* The University of Chicago Press, Chicago.

Billingsley, P. (1961b): The Lindeberg-Lévy theorem for martingales. *Proc. Amer. Math. Soc.* **12**, 788-792.

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

Dorogovcev, A.Ja. (1976): The consistency of an estimate of a parameter of a stochastic differential equation. *Theor. Probability and Math. Statist.* **10**, 73-82.

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

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

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

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

Kessler, M. (1995): Martingale estimating functions for a Markov chain. Preprint, Laboratoire de Probabilités, Université Paris VI.

Kessler, M. (1996): Simple and explicit estimating functions for a discretely observed diffusion process. Research Report No. 336, Department of Theoretical Statistics, University of Aarhus.

Kessler, M. (1997): Estimation of an ergodic diffusion from discrete observations. *Scand. J. Statist.* **24**, 211-229.

Kessler, M. and Sørensen, M. (1995): Estimating equations based on eigenfunctions for a discretely observed diffusion process. Research Report No. 332, Department of Theoretical Statistics, University of Aarhus.

Kloeden P.E. and Platen, E. (1996): *Numerical Solution of Stochastic Differential Equations.* Springer-Verlag, New York.

Kloeden, P.E., Platen, E., Schurz, H. and Sørensen, M. (1996): On effects of discretization on estimators of drift parameters for diffusion processes. *J.*

Appl. Prob. **33**, 1061-1076.

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

Pedersen, A.R. (1994): Quasi-likelihood inference for discretely observed diffusion processes. Research Report No. 295, Department of Theoretical Statistics, University of Aarhus.

Pedersen, A.R. (1995a): A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *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. *Bernoulli* **1**, 257-279.

Prakasa Rao, B.L.S. (1983): Asymptotic theory for non-linear least squares estimator for diffusion processes. *Math. Operationsforsch. u. Statist., Ser. Statist.* **14**, 195-209.

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