CREATES
Center for Research in Econometric Analysis of Time Series

CREATES Research Paper 2008-18

Parametric inference for discretely sampled stochastic differential equations

Michael Sørensen

School of Economics and Management
University of Aarhus
Building 1322, DK-8000 Aarhus C
Denmark

Aarhus School of Business
University of Aarhus
Handelshøjskolen
Aarhus Universitet

UNIVERSITY OF COPENHAGEN

Parametric inference for discretely sampled stochastic differential equations

Michael Sørensen*
Dept. of Mathematical Sciences, Univ. of Copenhagen and CREATES

Abstract

A review is given of parametric estimation methods for discretely sampled multivariate diffusion processes. The main focus is on estimating functions and asymptotic results. Maximum likelihood estimation is briefly considered, but the emphasis is on computationally less demanding martingale estimating functions. Particular attention is given to explicit estimating functions. Results on both fixed frequency and high frequency asymptotics are given. When choosing among the many estimators available, guidance is provided by simple criteria for high frequency efficiency and rate optimality that are presented in the framework of approximate martingale estimating functions.

Key words: Asymptotic results, discrete time observation of a diffusion, efficiency, eigenfunctions, explicit inference, generalized method of moments, likelihood inference, martingale estimating functions, high frequency asymptotics, Pearson diffusions.

JEL codes: C22, C32.

1 Introduction

In this chapter we consider parametric inference based on observations $X_0, X_\Delta, \ldots, X_{n\Delta}$ from a d-dimensional diffusion process given by

$$dX_t = b(X_t; \theta)dt + \sigma(X_t; \theta)dW_t, \tag{1.1}$$

where σ is a $d \times d$-matrix and W a d-dimensional standard Wiener process. The drift b and the diffusion matrix σ depend on a parameter θ which varies in a subset Θ of $\mathbb{R}^p$. The main focus is on estimating functions and asymptotic results.

The true (data-generating) model is supposed to be the stochastic differential equation (1.1) with the parameter value θ_0, and the coefficients b and σ are assumed to be sufficiently smooth functions of the state to ensure the existence of a unique weak solution

*Michael Sørensen acknowledges support from Center for Research in Econometric Analysis of Time Series, CREATES, www.creates.au.dk, funded by the Danish National Research Foundation and from the Danish Center for Accounting and Finance funded by the Danish Social Science Research Council.

for all θ in Θ. The state space of X is denoted by D. When $d = 1$, the state space is an interval (ℓ, r), where ℓ could possibly be $-\infty$, and r might be ∞. We suppose that the transition distribution has a density $y \mapsto p(\Delta, x, y; \theta)$ with respect to the Lebesgue measure on D, and that $p(\Delta, x, y; \theta) > 0$ for all $y \in D$. The transition density is the conditional density of $X_{t+\Delta}$ given that $X_t = x$. Since the data are equidistant, we will often suppress the argument Δ in the transition density and write $p(x, y; \theta)$.

It is assumed that the diffusion is ergodic, and that its invariant probability measure has density function μ_θ for all $\theta \in \Theta$. The initial value of the diffusion is assumed to be either known, $X_0 = x_0$, or $X_0 \sim \mu_\theta$. In the latter case the diffusion is stationary.

2 Asymptotics: fixed frequency

We consider the asymptotic properties of an estimator $\hat{\theta}_n$ obtained by solving the estimating equation

$$G_n(\hat{\theta}_n) = 0, \tag{2.1}$$

where G_n is an estimating function of the form

$$G_n(\theta) = \sum_{i=1}^{n} g(\Delta, X_{\Delta i}, X_{\Delta(i-1)}; \theta) \tag{2.2}$$

for some suitable function $g(\Delta, y, x; \theta)$ with values in $\mathbb{R}^p$. All estimators discussed below can be represented in this way. An estimator, $\hat{\theta}_n$, which solves (2.1) with probability approaching one as $n \to \infty$, is called a G_n–estimator. A priori there is no guarantee that a unique solution to (2.1) exists. In this section, we consider the standard asymptotic scenario, where the time between observations Δ is fixed and the number of observations goes to infinity. In most cases we suppress Δ in the notation and write for example $g(y, x; \theta)$.

We have assumed that the diffusion is ergodic and denote the density function of the invariant probability measure by μ_θ. Let Q_θ denote the probability measure on D^2 with density function $\mu_\theta(x)p(\Delta, x, y; \theta)$. This is the density function of two consecutive observations $(X_{\Delta(i-1)}, X_{\Delta i})$ when the diffusion is stationary, i.e. when $X_0 \sim \mu_\theta$. We impose the following condition on the function g

$$Q_\theta(g_j(\theta)^2) = \int_{D^2} g_j(y, x; \theta)^2 \mu_\theta(x) p(x, y; \theta) dy dx < \infty, \quad j = 1, \ldots, p, \tag{2.3}$$

for all $\theta \in \Theta$, where g_j denotes the jth coordinate of g. The quantity $Q_\theta(g_j(\theta))$ is defined similarly. Under the assumption of ergodicity and (2.3), it follows that

$$\frac{1}{n} \sum_{i=1}^{n} g(X_{\Delta i}, X_{\Delta(i-1)}; \theta) \xrightarrow{P_\theta} Q_\theta(g(\theta))^1. \tag{2.4}$$

When the diffusion, X, is one-dimensional, the following simple conditions ensure *ergodicity*, and an explicit expression exists for the density of the invariant probability

[1] $Q_\theta(g(\theta))$ denotes the vector $(Q_\theta(g_j(\theta)))_{j=1,\ldots,p}$.

measure. The *scale measure* of X has Lebesgue density

$$s(x;\theta) = \exp\left(-2\int_{x^\#}^{x} \frac{b(y;\theta)}{\sigma^2(y;\theta)}dy\right), \quad x \in (\ell, r), \tag{2.5}$$

where $x^\# \in (\ell, r)$ is arbitrary.

Condition 2.1 *The following holds for all $\theta \in \Theta$:*

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

and

$$\int_{\ell}^{r} [s(x;\theta)\sigma^2(x;\theta)]^{-1}dx = A(\theta) < \infty.$$

Under Condition 2.1 the process X is ergodic with an invariant probability measure with Lebesgue density

$$\mu_\theta(x) = [A(\theta)s(x;\theta)\sigma^2(x;\theta)]^{-1}, \quad x \in (\ell, r). \tag{2.6}$$

For details see e.g. Skorokhod (1989).

For the following asymptotic results to hold, we also need to assume that under P_θ the estimating function (2.2) satisfies a central limit theorem

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n} g(X_{\Delta i}, X_{\Delta(i-1)};\theta) \xrightarrow{\mathcal{D}} N(0, V(\theta)) \tag{2.7}$$

for some $p \times p$-matrix $V(\theta)$. For (2.7) to hold, it is obviously necessary that $Q_\theta(g(\theta)) = 0$.

Theorem 2.2 *Assume that $\theta_0 \in \operatorname{int}\Theta$ and that a neighbourhood N of θ_0 in Θ exists, such that:*
(1) The function $g(\theta) : (x,y) \mapsto g(x,y;\theta)$ is integrable with respect to the probability measure Q_{θ_0} for all $\theta \in N$, and

$$Q_{\theta_0}(g(\theta_0)) = 0. \tag{2.8}$$

(2) The function $\theta \mapsto g(x,y;\theta)$ is continuously differentiable on N for all $(x,y) \in D^2$.
(3) The functions[2] $(x,y) \mapsto \partial_{\theta_j} g_i(x,y;\theta)$, $i,j = 1,\ldots,p$, are dominated for all $\theta \in N$ by a function which is integrable with respect to Q_{θ_0}.
(4) The $p \times p$ matrix[3]

$$W = Q_{\theta_0}\left(\partial_{\theta^T} g(\theta_0)\right) \tag{2.9}$$

is invertible.

Then a consistent G_n–estimator $\hat{\theta}_n$ exists, and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N_p\left(0, W^{-1}VW^{T^{-1}}\right) \tag{2.10}$$

[2] $\partial_{\theta_j} g_i$ denotes the partial derivative $\frac{\partial g_i}{\partial \theta_j}$.

[3] In this chapter T denotes transposition, vectors are column vectors, and $Q_{\theta_0}\left(\partial_{\theta^T} g(\theta_0)\right)$ denotes the matrix $\{Q_{\theta_0}(\partial_{\theta_j} g_i(\theta_0))\}$, where i is the row number and j the column number.

under P_{θ_0}, where $V = V(\theta_0)$. If, moreover, the function $g(x,y;\theta)$ is locally dominated integrable[4] *with respect to Q_{θ_0} and*

$$Q_{\theta_0}(g(\theta)) \neq 0 \quad \textit{for all} \quad \theta \neq \theta_0,$$

then the estimator $\hat{\theta}_n$ is unique on any bounded subset of Θ containing θ_0 with probability approaching one as $n \to \infty$.

A proof of this theorem can be found in Jacod & Sørensen (2008). Related asymptotic results formulated in the language of the generalized method of moments were given by Hansen (1982).

If an estimating function does not satisfy (2.8), the obtained estimator is not consistent, but will converge to the solution $\bar{\theta}$ to the equation

$$Q_{\theta_0}(g(\bar{\theta})) = 0. \tag{2.11}$$

If the estimating function $G_n(\theta)$ is a martingale under P_θ, the asymptotic normality in (2.7) follows without further conditions from the central limit theorem for martingales, see Hall & Heyde (1980). This result goes back to Billingsley (1961). In the martingale case the matrix $V(\theta)$ is given by

$$V(\theta) = Q_{\theta_0}\left(g(\theta)g(\theta)^T\right), \tag{2.12}$$

and the asymptotic covariance matrix of the estimator $\hat{\theta}_n$ can be consistently estimated by means of the matrices W_n and V_n given in the following theorem; see Jacod & Sørensen (2008).

Theorem 2.3 *Under the conditions (2) – (4) of Theorem 2.2,*

$$W_n = \frac{1}{n}\sum_{i=1}^{n} \partial_\theta g(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_n) \xrightarrow{P_{\theta_0}} W, \tag{2.13}$$

and the probability that W_n is invertible approaches one as $n \to \infty$. If, moreover, the functions $(x,y) \mapsto g_i(x,y;\theta)$, $i = 1, \ldots, p$, are dominated for all $\theta \in N$ by a function which is square integrable with respect to Q_{θ_0}, then in the martingale case

$$V_n = \frac{1}{n}\sum_{i=1}^{n} g(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_n) g(X_{(i-1)\Delta}, X_{i\Delta}; \hat{\theta}_n)^T \xrightarrow{P_{\theta_0}} V. \tag{2.14}$$

When the estimating function $G_n(\theta)$ is not a martingale under P_θ, further conditions on the diffusion process must be imposed to ensure the asymptotic normality in (2.7). If the diffusion is stationary and geometrically α-mixing[5], (2.7) holds with

$$\begin{aligned} V(\theta) \;=\; & Q_{\theta_0}\left(g(\theta)g(\theta)^T\right) + \sum_{k=1}^{\infty}\Big[E_{\theta_0}\left(g(X_\Delta, X_0) g(X_{(k+1)\Delta}, X_{k\Delta})^T\right) \\ & \qquad + E_{\theta_0}\left(g(X_{(k+1)\Delta}, X_{k\Delta}) g(X_\Delta, X_0)^T\right)\Big], \end{aligned} \tag{2.15}$$

[4] A function $g: D^2 \times \Theta \mapsto \mathbb{R}$ is called locally dominated integrable with respect to Q_{θ_0} if for each $\theta' \in \Theta$ there exists a neighbourhood $U_{\theta'}$ of θ' and a non-negative Q_{θ_0}-integrable function $h_{\theta'}: D^2 \mapsto \mathbb{R}$ such that $|\, g(x,y;\theta)\,| \leq h_{\theta'}(x,y)$ for all $(x,y,\theta) \in D^2 \times U_{\theta'}$.

[5] α-mixing with mixing coefficients that tend to zero geometrically fast.

provided that $V(\theta)$ is strictly positive definite, and that $Q_\theta(g_i(\theta)^{2+\epsilon}) < \infty$ for some $\epsilon > 0$, see e.g. Doukhan (1994). Genon-Catalot, Jeantheau & Larédo (2000) gave the following simple sufficient condition for a one-dimensional diffusion to be geometrically α-mixing.

Condition 2.4
(i) The function b is continuously differentiable with respect to x and σ is twice continuously differentiable with respect to x, $\sigma(x;\theta) > 0$ for all $x \in (\ell, r)$, and there exists a constant $K_\theta > 0$ such that $|b(x;\theta)| \leq K_\theta(1 + |x|)$ and $\sigma^2(x;\theta) \leq K_\theta(1 + x^2)$ for all $x \in (\ell, r)$.

(ii) $\sigma(x;\theta)\mu_\theta(x) \to 0$ as $x \downarrow \ell$ and $x \uparrow r$.

(iii) $1/\gamma(x;\theta)$ has a finite limit as $x \downarrow \ell$ and $x \uparrow r$, where $\gamma(x;\theta) = \partial_x\sigma(x;\theta) - 2b(x;\theta)/\sigma(x;\theta)$.

Other conditions for geometric α-mixing were given by Veretennikov (1987), Hansen & Scheinkman (1995), and Kusuoka & Yoshida (2000).

3 Likelihood inference

The diffusion process X is a Markov process, so the likelihood function based on the observations $X_{t_0}, X_{t_1}, \cdots, X_{t_n}$ $(t_0 = 0)$, conditional on X_0, is

$$L_n(\theta) = \prod_{i=1}^{n} p(t_i - t_{i-1}, X_{t_{i-1}}, X_{t_i}; \theta), \tag{3.1}$$

where $y \mapsto p(s, x, y; \theta)$ is the transition density. Under weak regularity conditions the maximum likelihood estimator is efficient, i.e. it has the smallest asymptotic variance among all estimators. The transition density is only rarely explicitly known, but several numerical approaches make likelihood inference feasible for diffusion models. Pedersen (1995) proposed a method for obtaining an approximation to the likelihood function by rather extensive simulation. Pedersen's method was very considerably improved by Durham & Gallant (2002), whose method is computationally much more efficient. Poulsen (1999) obtained an approximation to the transition density by numerically solving a partial differential equation, whereas Aït-Sahalia (2002), Aït-Sahalia (2003) proposed to approximate the transition density by means of expansions. A Gaussian approximation to the likelihood function obtained by local linearization of (1.1) was proposed by Ozaki (1985), while Forman & Sørensen (2008) proposed to use an approximation in terms of eigenfunctions of the generator of the diffusion. Bayesian estimators with the same asymptotic properties as the maximum likelihood estimator can be obtained by Markov chain Monte Carlo methods, see Elerian, Chib & Shephard (2001), Eraker (2001), and Roberts & Stramer (2001). Finally, exact and computationally efficient likelihood-based estimation methods were presented by Beskos et al. (2006). These various approaches to maximum likelihood estimation will not be considered further in this chapter. Some of them are treated in Phillips & Yu (2008). Asymptotic results for the maximum likelihood estimator were established by Dacunha-Castelle & Florens-Zmirou (1986), while asymptotic results when the observations are made at random time points were obtained by Aït-Sahalia & Mykland (2003).

The vector of partial derivatives of the log-likelihood function with respect to the coordinates of θ,

$$U_n(\theta) = \partial_\theta \log L_n(\theta) = \sum_{i=1}^{n} \partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta), \tag{3.2}$$

where $\Delta_i = t_i - t_{i-1}$, is called the *score function* (or score vector). The maximum likelihood estimator usually solves the estimating equation $U_n(\theta) = 0$. The score function is a martingale under P_θ, which is easily seen provided that the following interchange of differentiation and integration is allowed:

$$\begin{aligned} &E_\theta\left(\partial_\theta \log p(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta) \middle| X_{t_1}, \ldots, X_{t_{i-1}} \right) \\ &= \int_D \frac{\partial_\theta p(\Delta_i, X_{t_{i-1}}, y; \theta)}{p(\Delta_i, X_{t_{i-1}}, y; \theta)} p(\Delta_i, X_{t_{i-1}}, y, \theta) dy = \partial_\theta \int_D p(\Delta_i, X_{t_{i-1}}, y; \theta) dy = 0. \end{aligned}$$

A simple approximation to the likelihood function is obtained by approximating the transition density by a Gaussian density with the correct first and second conditional moments. For a one-dimensional diffusion we get

$$p(\Delta, x, y; \theta) \approx q(\Delta, x, y; \theta) = \frac{1}{\sqrt{2\pi\phi(\Delta, x; \theta)}} \exp\left[-\frac{(y - F(\Delta, x; \theta))^2}{2\phi(\Delta, x; \theta)} \right]$$

where

$$F(\Delta, x; \theta) = \mathrm{E}_\theta(X_\Delta | X_0 = x) = \int_\ell^r y p(\Delta, x, y; \theta) dy. \tag{3.3}$$

and

$$\phi(\Delta, x; \theta) = \mathrm{Var}_\theta(X_\Delta | X_0 = x) = \int_\ell^r [y - F(\Delta, x; \theta)]^2 p(\Delta, x, y; \theta) dy. \tag{3.4}$$

In this way we obtain the *quasi-likelihood*

$$L_n(\theta) \approx QL_n(\theta) = \prod_{i=1}^{n} q(\Delta_i, X_{t_{i-1}}, X_{t_i}; \theta),$$

and by differentiation with respect to the parameter vector, we obtain the quasi-score function

$$\begin{aligned} \partial_\theta \log QL_n(\theta) = \sum_{i=1}^{n} \Bigg\{ & \frac{\partial_\theta F(\Delta_i, X_{t_{i-1}}; \theta)}{\phi(\Delta_i, X_{t_{i-1}}; \theta)} [X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta)] \\ & + \frac{\partial_\theta \phi(\Delta_i, X_{t_{i-1}}; \theta)}{2\phi(\Delta_i, X_{t_{i-1}}; \theta)^2} \left[(X_{t_i} - F(\Delta_i, X_{t_{i-1}}; \theta))^2 - \phi(\Delta_i, X_{t_{i-1}}; \theta) \right] \Bigg\}, \end{aligned} \tag{3.5}$$

which is clearly a martingale under P_θ. It is a particular case of the quadratic martingale estimating functions considered by Bibby & Sørensen (1995) and Bibby & Sørensen (1996). Maximum quasi-likelihood estimation was considered by Bollerslev & Wooldridge (1992).

4 Martingale estimating functions

In this section we present a rather general way of obtaining approximations to the score function by means of martingales of a similar form. Suppose we have a collection of real valued functions $h_j(x, y, ; \theta)$, $j = 1, \ldots, N$ satisfying

$$\int_D h_j(x, y; \theta) p(x, y; \theta) dy = 0 \tag{4.1}$$

for all $x \in D$ and $\theta \in \Theta$. Each of the functions h_j could be used separately to define an estimating function of the form (2.2), but a better approximation to the score function, and hence a more efficient estimator, is obtained by combining them in an optimal way. Therefore we consider estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^{n} a(X_{(i-1)\Delta}, \theta) h(X_{(i-1)\Delta}, X_{i\Delta}; \theta), \tag{4.2}$$

where $h = (h_1, \ldots, h_N)^T$, and the $p \times N$ weight matrix $a(x, \theta)$ is a function of x such that (4.2) is P_θ-integrable. It follows from (4.1) that $G_n(\theta)$ is a martingale under P_θ for all $\theta \in \Theta$. An estimating function with this property is called a *martingale estimating function.*

The matrix a determines how much weight is given to each of the h_js in the estimation procedure. This weight matrix can be chosen in an optimal way rather straightforwardly using the theory of optimal estimating functions, see Godambe (1960), Durbin (1960), Godambe & Heyde (1987) and Heyde (1997). The optimal weight matrix a^* gives the estimating function of the form (4.2) that provides the best possible approximation to the score function (3.2) in a mean square sense. Moreover, the optimal $g^*(x, y; \theta) = a^*(x; \theta) h(x, y; \theta)$ is obtained from $\partial_\theta \log p(x, y; \theta)$ by projection in a certain space of square integrable functions, see Kessler (1996) and Sørensen (1997).

The choice of the functions h_j, on the other hand, is an art rather than a science. The ability to tailor these functions to a given model or to particular parameters of interest is a considerable strength of the estimating functions methodology. It is, however, also a source of weakness, since it is not always clear how best to choose the h_js. In this and the next section, we shall present ways of choosing these functions that usually work well in practice.

Example 4.1 The martingale estimating function (3.5) is of the type (4.2) with $N = 2$ and

$$\begin{aligned} h_1(x, y; \theta) &= y - F(\Delta, x; \theta), \\ h_2(x, y; \theta) &= (y - F(\Delta, x; \theta))^2 - \phi(\Delta, x, \theta), \end{aligned}$$

where F and ϕ are given by (3.3) and (3.4). The weight matrix is

$$\left(\frac{\partial_\theta F(\Delta, x; \theta)}{\phi(\Delta, x; \theta)}, \ \frac{\partial_\theta \phi(\Delta, x; \theta)}{2\phi^2(\Delta, x; \theta)\Delta} \right), \tag{4.3}$$

which we shall see is approximately optimal. □

In the econometrics literature, a popular way of using functions like $h_j(x, y, ; \theta)$, $j = 1, \ldots, N$, to estimate the parameter θ is the *generalized method of moments* (GMM) of Hansen (1982). The method is usually implemented as follows, see e.g. Campbell, Lo & MacKinlay (1997). Consider

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

Under weak conditions, cf. Theorem 2.3, a consistent estimator of the asymptotic covariance matrix M of $\sqrt{n} F_n(\theta)$ is

$$M_n = \frac{1}{n} \sum_{i=1}^{n} h(X_{(i-1)\Delta}, X_{i\Delta}; \tilde{\theta}_n) h(X_{(i-1)\Delta}, X_{i\Delta}; \tilde{\theta}_n)^T,$$

where $\tilde{\theta}_n$ is a consistent estimator of θ (for instance obtained by minimizing $F_n(\theta)^T F_n(\theta)$). The GMM-estimator is obtained by minimizing the function

$$H_n(\theta) = F_n(\theta)^T M_n^{-1} F_n(\theta).$$

The corresponding estimating function is obtained by differentiation with respect to θ

$$\partial_\theta H_n(\theta) = D_n(\theta) M_n^{-1} F_n(\theta),$$

where by (2.4)

$$D_n(\theta) = \frac{1}{n} \sum_{i=1}^{n} \partial_\theta h(X_{(i-1)\Delta}, X_{i\Delta}; \theta)^T \xrightarrow{P_{\theta_0}} Q_{\theta_0} \left(\partial_\theta h(\theta)^T \right).$$

Hence the estimating function $\partial_\theta H_n(\theta)$ is asymptotically equivalent to an estimating function of the form (4.2) with a constant weight matrix

$$a(x, \theta) = Q_{\theta_0} \left(\partial_\theta h(\theta)^T \right) M^{-1},$$

and we see that GMM-estimators are covered by the theory for martingale estimating functions presented in this chapter.

We now return to the problem of finding the optimal estimating function $G_n^*(\theta)$, i.e. of the form (4.2) with the *optimal weight matrix*. To do so we assume that the functions h_j satisfy the following condition.

Condition 4.2

(1) The functions h_j, $j = 1, \ldots N$, are linearly independent.
(2) The functions $y \mapsto h_j(x, y; \theta)$, $j = 1, \ldots N$, are square integrable with respect to $p(x, y; \theta)$ for all $x \in D$ and $\theta \in \Theta$.
(3) $h_j(x, y; \theta)$, $j = 1, \ldots N$, are differentiable with respect to θ.
(4) The functions $y \mapsto \partial_\theta h_j(x, y; \theta)$ are integrable with respect to $p(x, y; \theta)$ for all $x \in D$ and $\theta \in \Theta$.

According to the theory of optimal estimating functions, the optimal choice of the weight matrix a is given by

$$a^*(x;\theta) = B_h(x;\theta)\, V_h(x;\theta)^{-1}, \tag{4.4}$$

where

$$B_h(x;\theta) = \int_D \partial_\theta h(x,y;\theta)^T p(x,y;\theta)dy \tag{4.5}$$

and

$$V_h(x;\theta) = \int_D h(x,y;\theta)h(x,y;\theta)^T p(x,y;\theta)dy. \tag{4.6}$$

The asymptotic variance of an optimal estimator, i.e. a G_n^*–estimator, is simpler than the general expression in (2.10) because in this case the matrices W and V given by (2.9) and (2.12) are equal and given by (4.8), as can easily be verified. Thus we have the following corollary to Theorem 2.2:

Corollary 4.3 *Assume that $g^*(x,y,\theta) = a^*(x;\theta)h(x,y;\theta)$ satisfies the conditions of Theorem 2.2. Then a sequence $\hat{\theta}_n$ of G_n^*–estimators has the asymptotic distribution*

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N_p\left(0, V^{-1}\right), \tag{4.7}$$

where

$$V = \mu_{\theta_0}\left(B_h(\theta_0)V_h(\theta_0)^{-1}B_h(\theta_0)^T\right) \tag{4.8}$$

with B_h and V_h given by (4.5) *and* (4.6).

Example 4.4 Consider the martingale estimating function of form (4.2) with $N = 2$ and with h_1 and h_2 as in Example 4.1, where the diffusion is one-dimensional. The optimal weight matrix has columns given by

$$\begin{aligned} a_1^*(x;\theta) &= \frac{\partial_\theta\phi(x;\theta)\eta(x;\theta) - \partial_\theta F(x;\theta)\psi(x;\theta)}{\phi(x;\theta)\psi(x;\theta) - \eta(x;\theta)^2} \\ a_2^*(x;\theta) &= \frac{\partial_\theta F(x;\theta)\eta(x;\theta) - \partial_\theta\phi(x;\theta)\phi(x;\theta)}{\phi(x;\theta)\psi(x;\theta) - \eta(x;\theta)^2}, \end{aligned}$$

where

$$\eta(x;\theta) = \mathrm{E}_\theta([X_\Delta - F(x;\theta)]^3|X_0 = x)$$

and

$$\psi(x;\theta) = \mathrm{E}_\theta([X_\Delta - F(x;\theta)]^4|X_0 = x) - \phi(x;\theta)^2.$$

We can simplify these expressions by making the Gaussian approximations

$$\eta(t,x;\theta) \approx 0 \quad \text{and} \quad \psi(t,x;\theta) \approx 2\phi(t,x;\theta)^2. \tag{4.9}$$

If we insert these approximations into the expressions for a_1^* and a_2^*, we obtain the weight functions in (3.5). When Δ is not large this can be justified, because the transition distribution is not far from Gaussian. □

In the next subsection we shall present a class of martingale estimating functions for which the matrices $B_h(x;\theta)$ and $V_h(x;\theta)$ can be found explicitly, but for most models these matrices must be found by simulation. If a^* is determined by a relatively time consuming numerical method, it might be preferable to use the estimating function

$$G_n^\bullet(\theta) = \sum_{i=1}^{n} a^*(X_{(i-1)\Delta};\tilde{\theta}_n)h(X_{(i-1)\Delta}, X_{i\Delta};\theta), \tag{4.10}$$

where $\tilde{\theta}_n$ is a weakly $\sqrt{n}$-consistent estimator of θ, for instance obtained by some simple choice of the weight matrix a. In this way a^* needs to be calculated only once per observation point. Under weak regularity conditions, the estimator obtained from $G_n^\bullet(\theta)$ has the same efficiency as the optimal estimator; see e.g. Jacod & Sørensen (2008).

Most martingale estimating functions proposed in the literature are of the form

$$G_n(\theta) = \sum_{i=1}^{n} a(X_{(i-1)\Delta},\theta)\left[f(X_{i\Delta};\theta) - \pi_\Delta^\theta(f(\theta))(X_{(i-1)\Delta})\right], \tag{4.11}$$

where $f = (f_1,\dots,f_N)^T$, and π_Δ^θ denotes the *transition operator*

$$\pi_s^\theta(f)(x) = \int_D f(y)p(s,x,y;\theta)dy = \mathrm{E}_\theta(f(X_s)\,|\,X_0 = x). \tag{4.12}$$

The polynomial estimating functions given by $f_j(y) = y^j$, $j = 1,\dots,N$, are an example. For martingale estimating functions of the special form (4.11), the expression for the optimal weight matrix simplifies to some extent to

$$B_h(x;\theta)_{ij} = \pi_\Delta^\theta(\partial_{\theta_i} f_j(\theta))(x) - \partial_{\theta_i}\pi_\Delta^\theta(f_j(\theta))(x), \tag{4.13}$$

$i = 1,\dots p$, $j = 1,\dots,N$, and

$$V_h(x;\theta)_{ij} = \pi_\Delta^\theta(f_i(\theta)f_j(\theta))(x) - \pi_\Delta^\theta(f_i(\theta))(x)\pi_\Delta^\theta(f_j(\theta))(x), \tag{4.14}$$

$i,j = 1,\dots,N$. Often the functions f_j can be chosen such that they do not depend on θ, in which case

$$B_h(x;\theta)_{ij} = -\partial_{\theta_i}\pi_\Delta^\theta(f_j)(x). \tag{4.15}$$

A useful approximations to the optimal weight matrix can be obtained by applying the formula

$$\pi_s^\theta(f)(x) = \sum_{i=0}^{k} \frac{s^i}{i!} A_\theta^i f(x) + O(s^{k+1}), \tag{4.16}$$

where A_θ denotes the generator of the diffusion

$$A_\theta f(x) = \sum_{k=1}^{d} b_k(x;\theta)\partial_{x_k} f(x) + \frac{1}{2}\sum_{k,\ell=1}^{d} C_{k\ell}(x;\theta)\partial^2_{x_k x_\ell} f(x), \tag{4.17}$$

where $C = \sigma\sigma^T$. The formula (4.16) holds for $2(k+1)$ times continuously differentiable functions under weak conditions which ensure that the remainder term has the correct order, see Kessler (1997). It is often enough to use the approximation $\pi^\theta_\Delta(f_j)(x) \approx f_j(x) + \Delta A_\theta f_j(x)$. When f does not depend on θ this implies that

$$B_h(x;\theta) \approx \Delta\left[\partial_\theta b(x;\theta)f'(x) + \tfrac{1}{2}\partial_\theta \sigma^2(x;\theta)f''(x)\right] \tag{4.18}$$

and (for $N = 1$)

$$V_h(x;\theta) \approx \Delta\left[A_\theta(f^2)(x) - 2f(x)A_\theta f(x)\right] = \Delta\,\sigma^2(x;\theta)f'(x)^2. \tag{4.19}$$

Example 4.5 If we simplify the optimal weight matrix found in Example 4.4 by (4.16) and the Gaussian approximation (4.9), we obtain the approximately optimal quadratic martingale estimating function

$$\begin{aligned} G_n^\circ(\theta) = \sum_{i=1}^n \Bigg\{ & \frac{\partial_\theta b(X_{(i-1)\Delta};\theta)}{\sigma^2(X_{(i-1)\Delta};\theta)}[X_{i\Delta} - F(X_{(i-1)\Delta};\theta)] \\ & + \frac{\partial_\theta \sigma^2(X_{(i-1)\Delta};\theta)}{2\sigma^4(X_{(i-1)\Delta};\theta)\Delta}\left[(X_{i\Delta} - F(X_{(i-1)\Delta};\theta))^2 - \phi(X_{(i-1)\Delta};\theta)\right]\Bigg\}. \end{aligned} \tag{4.20}$$

For the CIR-model

$$dX_t = -\beta(X_t - \alpha)dt + \tau\sqrt{X_t}dW_t, \tag{4.21}$$

where $\beta, \tau > 0$, the approximately optimal quadratic martingale estimating function is

$$\begin{pmatrix} \displaystyle\sum_{i=1}^n \frac{1}{X_{(i-1)\Delta}}\left[X_{i\Delta} - X_{(i-1)\Delta}e^{-\beta\Delta} - \alpha(1 - e^{-\beta\Delta})\right] \\ \displaystyle\sum_{i=1}^n \left[X_{i\Delta} - X_{(i-1)\Delta}e^{-\beta\Delta} - \alpha(1 - e^{-\beta\Delta})\right] \\ \displaystyle\sum_{i=1}^n \frac{1}{X_{(i-1)\Delta}}\Big[\left(X_{i\Delta} - X_{(i-1)\Delta}e^{-\beta\Delta} - \alpha(1 - e^{-\beta\Delta})\right)^2 \\ \displaystyle\quad -\frac{\tau^2}{\beta}\left\{\left(\alpha/2 - X_{(i-1)\Delta}\right)e^{-2\beta\Delta} - (\alpha - X_{(i-1)\Delta})e^{-\beta\Delta} + \alpha/2\right\}\Big] \end{pmatrix}. \tag{4.22}$$

This is obtained from (4.20) after multiplication by an invertible non-random matrix to obtain a simpler expression. This does not change the estimator. From this estimating function explicit estimators can easily be obtained. A simulation study and an investigation of the asymptotic variance of the estimators for α and β in Bibby & Sørensen (1995) show that they are not much less efficient than the estimators from the optimal estimating function; see also the simulation study in Overbeck & Rydén (1997). □

When the optimal weight matrix is approximated by means of (4.16), there is a certain loss of efficiency, which as in the previous example is often quite small; see Bibby & Sørensen (1995) and the section on high frequency asymptotics below. Therefore the relatively simple estimating function (4.20) is often a good choice in practice.

It is tempting to go on to approximate $\pi^\theta_\Delta(f_j(\theta))(x)$ in (4.11) by (4.16) in order to obtain an explicit estimating function, but as we shall see in the next section this is often a dangerous procedure. In general the conditional expectation in π^θ_Δ should therefore be approximated by simulations. Fortunately, Kessler & Paredes (2002) have established that, provided the simulation is done with sufficient accuracy, this does not cause any bias, only a minor loss of efficiency that can be made arbitrarily small. Moreover, as we shall also see in the next section, $\pi^\theta_\Delta(f_j(\theta))(x)$ can be found explicitly for a quite flexible class of diffusions.

5 Explicit inference

In this section we consider one-dimensional diffusion models for which estimation is particularly easy because an explicit martingale estimating function exists.

Kessler & Sørensen (1999) proposed estimating functions of the form (4.11) where the functions f_j, $i = 1, \ldots, N$ are *eigenfunctions* for the generator (4.17), i.e.

$$A_\theta f_j(x;\theta) = -\lambda_j(\theta) f_j(x;\theta),$$

where the real number $\lambda_j(\theta) \geq 0$ is called the *eigenvalue* corresponding to $f_j(x;\theta)$. Under weak regularity conditions, f_j is also an eigenfunction for the transition operator π^θ_t, i.e.

$$\pi^\theta_t(f_j(\theta))(x) = e^{-\lambda_j(\theta)t} f_j(x;\theta). \tag{5.1}$$

for all $t > 0$. Thus explicit martingales are obtained. Each of the following three conditions imply (5.1):

(i) $\sigma(x;\theta)$ and $\partial_x f_j(x;\theta)$ are bounded functions of $x \in (\ell, r)$.

(ii) $\int_\ell^r [\partial_x f_j(x;\theta)\sigma(x;\theta)]^2 \mu_\theta(dx) < \infty$.

(iii) b and σ are of linear growth, and $\partial_x f_j$ is of polynomial growth in $x \in (\ell, r)$.

Example 5.1 The model

$$dX_t = -\beta[X_t - (m + \gamma z)]dt + \sigma\sqrt{z^2 - (X_t - m)^2}dW_t \tag{5.2}$$

where $\beta > 0$ and $\gamma \in (-1, 1)$ has been proposed as a model for the random variation of the logarithm of an exchange rate in a target zone between realignments by De Jong, Drost & Werker (2001) ($\gamma = 0$) and Larsen & Sørensen (2007). This is a diffusion on the interval $(m - z, m + z)$ with mean reversion around $m + \gamma z$. The parameter γ quantifies the asymmetry of the model. When $\beta(1 - \gamma) \geq \sigma^2$ and $\beta(1 + \gamma) \geq \sigma^2$, X is an ergodic diffusion, for which the stationary distribution is a beta-distribution on $(m - z, m + z)$ with parameters $\kappa_1 = \beta(1 - \gamma)\sigma^{-2}$ and $\kappa_2 = \beta(1 + \gamma)\sigma^{-2}$.

The eigenfunctions for the generator of the diffusion (5.2) are

$$f_i(x; \beta, \gamma, \sigma, m, z) = P_i^{(\kappa_1 - 1, \kappa_2 - 1)}((x - m)/z), \quad i = 1, 2, \ldots$$

where $P_i^{(a,b)}(x)$ denotes the Jacobi polynomial of order i given by

$$P_i^{(a,b)}(x) = \sum_{j=0}^{i} 2^{-j} \binom{n+a}{n-j} \binom{a+b+n+j}{j} (x-1)^j, \quad -1 < x < 1,$$

as can easily be seem by direct calculation. For this reason, the process (5.2) is called a *Jacobi-diffusion*. The eigenvalue of f_i is $i\left(\beta + \frac{1}{2}\sigma^2(i-1)\right)$. Since condition (i) above is obviously satisfied because the state space is bounded, (5.1) holds. □

When the eigenfunctions are of the form

$$f_i(y;\theta) = \sum_{j=0}^{i} a_{i,j}(\theta)\,\kappa(y)^j \tag{5.3}$$

where κ is a real function defined on the state space and is independent of θ, the optimal weight matrix (4.4) can be found explicitly too, provided that $2N$ eigenfunctions are available. Specifically,

$$B_h(x,\theta)_{ij} = \sum_{k=0}^{j}\left(\partial_{\theta_i} a_{j,k}(\theta)\nu_k(x;\theta) - \partial_{\theta_i}[e^{-\lambda_j(\theta)\Delta}\phi_j(x;\theta)]\right)$$

and

$$V_h(x,\theta)_{i,j} = \\ \sum_{r=0}^{i}\sum_{s=0}^{j}\left(a_{i,r}(\theta)a_{j,s}(\theta)\nu_{r+s}(x;\theta) - e^{-[\lambda_i(\theta)+\lambda_j(\theta)]\Delta}\ \phi_i(x;\theta)\phi_j(x;\theta)\right),$$

where $\nu_i(x;\theta) = \pi_\Delta^\theta(\kappa^i)(x)$, $i = 1,\ldots,2N$, solve the following triangular system of linear equations

$$e^{-\lambda_i(\theta)\Delta} f_i(x;\theta) = \sum_{j=0}^{i} a_{i,j}(\theta)\nu_j(x;\theta) \quad i = 1,\ldots,2N, \tag{5.4}$$

with $\nu_0(x;\theta) = 1$. The expressions for B_h and V_h follow from (4.13) and (4.14), while (5.4) follows by applying π_Δ^θ to both sides of (5.3).

Example 5.2 A widely applicable class of diffusion models for which explicit polynomial eigenfunctions are available is the class of *Pearson diffusions,* see Wong (1964) and Forman & Sørensen (2008). A Pearson diffusion is a stationary solution to a stochastic differential equation of the form

$$dX_t = -\beta(X_t-\mu)dt + \sqrt{(aX_t^2 + bX_t + c)}dW_t, \tag{5.5}$$

where $\beta > 0$, and a, b and c are such that the square root is well defined when X_t is in the state space. The class of stationary distributions equals the full Pearson system of distributions, so a very wide spectrum of marginal distributions is available ranging from distributions with compact support to very heavy-tailed distributions. For instance Pearson's type IV distributions, a skew t-type distribution, which seems very useful in finance, see e.g. Nagahara (1996), is the stationary distribution of the diffusion

$$dZ_t = -\beta Z_t dt + \sqrt{2\beta(\nu-1)^{-1}\{Z_t^2 + 2\rho\nu^{\frac{1}{2}}Z_t + (1+\rho^2)\nu\}}dW_t,$$

with $\nu > 1$. The parameter ρ is a skewness parameter. For $\rho = 0$ a t-distribution with ν degrees of freedom is obtained. Well-known instances of (5.5) are the Ornstein-Uhlenbeck process, the square root (CIR) process, and the Jacobi diffusions.

For a diffusion $T(X)$ obtained from a solution X to (5.5) by a twice differentiable and invertible transformation T, the eigenfunctions of the generator are $p_n\{T^{-1}(x)\}$, where p_n is an eigenfunction of the generator of X. The eigenvalues are the same as for the original eigenfunctions. Since the original eigenfunctions are polynomials, the eigenfunctions of $T(X)$ are of the form (5.3) with $\kappa = T^{-1}$. Hence explicit optimal martingale estimating functions are available for transformed Pearson diffusions too.

As an example let X be the Jacobi-diffusion (5.2) with $m = 0$ and $z = 1$, and consider $Y_t = \sin^{-1}(X_t)$. Then

$$dY_t = -\rho \frac{\sin(Y_t) - \varphi}{\cos(Y_t)} dt + \sigma dW_t,$$

where $\rho = \beta - \frac{1}{2}\sigma^2$ and $\varphi = \beta\gamma/\rho$. The state space is $(-\pi/2, \pi/2)$. Note that Y has dynamics that are very different from those of (5.2): the drift is non-linear and the diffusion coefficient is constant. The process Y was proposed and studied in Kessler & Sørensen (1999) for $\varphi = 0$, where the drift is $-\rho \tan(x)$. The general asymmetric version was proposed in Larsen & Sørensen (2007) as a model for exchange rates in a target zone.
□

Explicit martingale estimating functions are only available for the relatively small, but versatile, class of diffusions for which explicit eigenfunctions for the generator are available. *Explicit non-martingale estimating functions* can be found for all diffusions, but cannot be expected to approximate the score functions as well as martingale estimating functions, and will therefore usually give less efficient estimators.

Hansen & Scheinkman (1995) proposed non-martingale estimating functions given by

$$g_j(\Delta, x, y; \theta) = h_j(y) A_\theta f_j(x) - f_j(x) \hat{A}_\theta h_j(y), \tag{5.6}$$

where A_θ is the generator (4.17), and the functions f_j and h_j satisfy weak regularity conditions ensuring that (2.8) holds. The differential operator

$$\hat{A}_\theta f(x) = \sum_{k=1}^{d} \hat{b}_k(x;\theta) \partial_{x_k} f(x) + \frac{1}{2} \sum_{k,\ell=1}^{d} C_{k\ell}(x;\theta) \partial^2_{x_k x_\ell} f(x),$$

where $C = \sigma\sigma^T$ and

$$\hat{b}_k(x;\theta) = -b_k(x;\theta) + \frac{1}{\mu_\theta(x)} \sum_{\ell=1}^{d} \partial_{x_\ell} \left(\mu_\theta C_{kl}\right)(x;\theta),$$

is the generator of the time reversal of the observed diffusion X. A simpler type of explicit non-martingale estimating functions is of the form $g(\Delta, x, y; \theta) = h(x;\theta)$. Hansen & Scheinkman (1995) and Kessler (2000) studied $h_j(x;\theta) = A_\theta f_j(x)$, which is a particular case of (5.6). Kessler (2000) also proposed $h(x;\theta) = \partial_\theta \log \mu_\theta(x)$, which corresponds to considering the observations as an i.i.d. sample from the stationary distribution. Finally, Sørensen (2001) derived the estimating function with $h(x;\theta) = A_\theta \partial_\theta \log \mu_\theta(X_{t_i})$ as an approximation to the continuous-time score function. In all cases weak regularity conditions are needed to ensure that (2.8) holds, i.e. that $\int h(x;\theta_0) \mu_{\theta_0}(x) dx = 0$.

Quite generally, an explicit *approximate martingale estimating function* can be obtained from a martingale estimating function of the form (4.11) by approximating $\pi_\Delta^\theta(f_j(\theta))(x)$ and the weight matrix by (4.16). The simplest version of this approach gives the same estimator as the Gaussian quasi-likelihood based on the Euler-approximation to (1.1). Estimators of this type have been considered by Dorogovcev (1976), Prakasa Rao (1988), Florens-Zmirou (1989), Yoshida (1992), Chan et al. (1992), Kessler (1997), and Kelly, Platen & Sørensen (2004). It is, however, important to note that there is a dangerous pitfall when using these simple approximate martingale estimating functions. They do not satisfy (2.8), and hence the estimators are inconsistent and converge to the solution to (2.11). The problem is illustrated by the following example.

Example 5.3 Consider again the CIR-model (4.21). If we insert the approximation $F(x;\alpha,\beta) = -\beta(x-\alpha)\Delta$ into (4.22) we obtain the following estimator for β

$$\hat{\beta}_n = \frac{\frac{1}{n}(X_{\Delta n} - X_0)\sum_{i=1}^n X_{\Delta(i-1)}^{-1} - \sum_{i=1}^n X_{\Delta(i-1)}^{-1}(X_{\Delta i} - X_{\Delta(i-1)})}{\Delta[n - (\sum_{i=1}^n X_{\Delta(i-1)})(\sum_{i=1}^n X_{\Delta(i-1)}^{-1})/n]}.$$

It follows from (2.4) that

$$\hat{\beta}_n \xrightarrow{P_\theta} (1 - e^{-\beta_0\Delta})/\Delta \le \Delta^{-1}.$$

Thus the estimator of the reversion parameter β is reasonable only when $\beta_0\Delta$ is considerably smaller than one. Note that the estimator will always converge to a limit smaller than the sampling frequency. When $\beta_0\Delta$ is large, the behaviour of the estimator is bizarre, see Bibby & Sørensen (1995). Without prior knowledge of the value of β_0 it is thus a dangerous estimator. □

The asymptotic bias given by (2.11) is small when Δ is sufficiently small, and the results in the following section on high frequency asymptotics show that in this case the approximate martingale estimating functions work well. However, how small Δ has to be depends on the parameter values, and without prior knowledge about the parameters, it is safer to use an exact martingale estimating function, which gives consistent estimators at all sampling frequencies.

6 High frequency asymptotics and efficient estimation

A large number of estimating functions have been proposed for diffusion models, and a large number of simulation studies have been performed to compare their relative merits, but the general picture has been rather confusing. By considering the high frequency scenario,

$$n \to \infty, \quad \Delta_n \to 0, \quad n\Delta_n \to \infty, \tag{6.1}$$

Sørensen (2007) obtained simple conditions for rate optimality and efficiency for ergodic diffusions, which allow identification of estimators that work well when the time between observations, Δ_n, is not too large. For financial data the speed of reversion is usually slow enough that this type of asymptotics works for daily, sometimes even weekly observations. A main result of this theory is that under weak conditions optimal martingale estimating functions give rate optimal and efficient estimators.

To simplify the exposition, we restrict attention to a one-dimensional diffusion given by

$$dX_t = b(X_t;\alpha)dt + \sigma(X_t;\beta)dW_t, \tag{6.2}$$

where $\theta = (\alpha,\beta) \in \Theta \subseteq \mathbb{R}^2$. The results below can be generalized to multivariate diffusions and parameters of higher dimension. We consider estimating functions of the general form (2.2), where the two-dimensional function $g = (g_1, g_2)$ for some $\kappa \geq 2$ and for all $\theta \in \Theta$ satisfies

$$E_\theta(g(\Delta_n, X_{\Delta_n i}, X_{\Delta_n(i-1)};\theta) \mid X_{\Delta_n(i-1)}) = \Delta_n^\kappa R(\Delta_n, X_{\Delta_n(i-1)};\theta). \tag{6.3}$$

Here and later $R(\Delta, y, x;\theta)$ denotes a function such that $|R(\Delta, y, x;\theta)| \leq F(y, x;\theta)$, where F is of polynomial growth in y and x uniformly for θ in a compact set[6]. We assume that the diffusion and the estimating functions satisfy the technical regularity Condition 6.3 given below.

Martingale estimating functions obviously satisfy (6.3) with $R = 0$, but for instance the approximate martingale estimating functions discussed at the end of the previous section satisfy (6.3) too.

Theorem 6.1 *Suppose that*

$$\begin{aligned} \partial_y g_2(0, x, x;\theta) &= 0, & (6.4)\\ \partial_y g_1(0, x, x;\theta) &= \partial_\alpha b(x;\alpha)/\sigma^2(x;\beta), & (6.5)\\ \partial_y^2 g_2(0, x, x;\theta) &= \partial_\beta \sigma^2(x;\beta)/\sigma^2(x;\beta)^2, & (6.6) \end{aligned}$$

for all $x \in (\ell, r)$ and $\theta \in \Theta$. Assume, moreover, that the following identifiability condition is satisfied

$$\int_\ell^r [b(x,\alpha_0) - b(x,\alpha)]\partial_y g_1(0, x, x;\theta)\mu_{\theta_0}(x)dx \neq 0 \quad \text{when } \alpha \neq \alpha_0,$$
$$\int_\ell^r [\sigma^2(x,\beta_0) - \sigma^2(x,\beta)]\partial_y^2 g_2(0, x, x;\theta)\mu_{\theta_0}(x)dx \neq 0 \quad \text{when } \beta \neq \beta_0,$$

and that

$$W_1 = \int_\ell^r \frac{(\partial_\alpha b(x;\alpha_0))^2}{\sigma^2(x;\beta_0)}\mu_{\theta_0}(x)dx \neq 0,$$
$$W_2 = \int_\ell^r \left[\frac{\partial_\beta \sigma^2(x;\beta_0)}{\sigma^2(x;\beta_0)}\right]^2 \mu_{\theta_0}(x)dx \neq 0.$$

Then a consistent G_n–estimator $\hat{\theta}_n = (\hat{\alpha}_n, \hat{\beta}_n)$ exists and is unique in any compact subset of Θ containing θ_0 with probability approaching one as $n \to \infty$. For a martingale estimating function, or more generally if $n\Delta^{2(\kappa-1)} \to 0$,

$$\begin{pmatrix} \sqrt{n\Delta_n}(\hat{\alpha}_n - \alpha_0) \\ \sqrt{n}(\hat{\beta}_n - \beta_0) \end{pmatrix} \xrightarrow{\mathcal{D}} N_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} W_1^{-1} & 0 \\ 0 & W_2^{-1} \end{pmatrix} \right). \tag{6.7}$$

[6]For any compact subset $K \subseteq \Theta$, there exist constants $C_1, C_2, C_3 > 0$ such that $\sup_{\theta\in K} |F(y, x;\theta)| \leq C_1(1 + |x|_2^C + |y|_3^C)$ for all x and y in the state space of the diffusion.

An estimator satisfying (6.7) is rate optimal and efficient, cf. Gobet (2002), who showed that the model considered here is locally asymptotically normal. Note that the estimator of the diffusion coefficient parameter, β, converges faster than the estimator of the drift parameter, α. Condition (6.4) implies rate optimality. If this condition is not satisfied, the estimator of the diffusion coefficient parameter converges at the slower rate $\sqrt{n\Delta_n}$. This condition is called *the Jacobsen condition*, because it appears in the theory of *small* Δ-*optimal estimation* developed in Jacobsen (2001) and Jacobsen (2002). In this theory the asymptotic covariance matrix in (2.10) is expanded in powers of Δ, the time between observations. The leading term is minimal when (6.5) and (6.6) are satisfied. The same expansion of (2.10) was used by Aït-Sahalia & Mykland (2004).

The assumption $n\Delta_n \to \infty$ in (6.1) is needed to ensure that the drift parameter, α, can be consistently estimated. If the drift is known and only the diffusion coefficient parameter, β, needs to be estimated, this condition can be omitted, see Genon-Catalot & Jacod (1993). Another situation where the infinite observation horizon, $n\Delta_n \to \infty$, is not needed for consistent estimation of α is when the high frequency asymptotic scenario is combined with the small diffusion scenario, where $\sigma(x;\beta) = \epsilon_n \zeta(x;\beta)$ and $\epsilon_n \to 0$, see Genon-Catalot (1990), Sørensen & Uchida (2003) and Gloter & Sørensen (2008).

The reader is reminded of the trivial fact that for any non-singular 2×2 matrix, M_n, the estimating functions $M_n G_n(\theta)$ and $G_n(\theta)$ give exactly the same estimator. We call them *versions* of the same estimating function. The matrix M_n may depend on Δ_n. Therefore a given version of an estimating function needs not satisfy (6.4) – (6.6). The point is that a version must exist which satisfies these conditions.

Example 6.2 Consider a quadratic martingale estimating function of the form

$$g(\Delta, y, x; \theta) = \begin{pmatrix} a_1(x,\Delta;\theta)[y - F(\Delta, x;\theta)] \\ a_2(x,\Delta;\theta)\left[(y - F(\Delta,x;\theta))^2 - \phi(\Delta,x;\theta)\right] \end{pmatrix}, \tag{6.8}$$

where F and ϕ are given by (3.3) and (3.4). By (4.16), $F(\Delta, x;\theta) = x + O(\Delta)$ and $\phi(\Delta, x;\theta) = O(\Delta)$, so

$$g(0, y, x;\theta) = \begin{pmatrix} a_1(x,0;\theta)(y-x) \\ a_2(x,0;\theta)(y-x)^2 \end{pmatrix}. \tag{6.9}$$

Since $\partial_y g_2(0,y,x;\theta) = 2a_2(x,\Delta;\theta)(y-x)$, the Jacobsen condition (6.4) is satisfied, so estimators obtained from (6.8) are rate optimal. Using again (4.16), it is not difficult to see that efficient estimators are obtained in three particular cases: the optimal estimating function given in Example 4.4 and the approximations (3.5) and (4.20). □

It follows from results in Jacobsen (2002) that to obtain a rate optimal and efficient estimator from an estimating function of the form (4.11), we need that $N \geq 2$ and that the matrix

$$D(x) = \begin{pmatrix} \partial_x f_1(x;\theta) & \partial_x^2 f_1(x;\theta) \\ \partial_x f_2(x;\theta) & \partial_x^2 f_2(x;\theta) \end{pmatrix}$$

is invertible for μ_θ-almost all x. Under these conditions, Sørensen (2007) showed that Godambe-Heyde optimal martingale estimating functions give rate optimal and efficient estimators. For a d-dimensional diffusion, Jacobsen (2002) gave the conditions $N \geq d(d+3)/2$, and that the $N \times (d+d^2)$-matrix $D(x) = (\partial_x f(x;\theta)\ \partial_x^2 f(x;\theta))$ has full rank $d(d+3)/2$.

We conclude this section by stating technical conditions under which the results in this section hold. The assumptions about polynomial growth are far too strong, but simplify the proofs. These conditions can most likely be weakened very considerably in a way similar to the proofs in Gloter & Sørensen (2008).

Condition 6.3 *The diffusion is ergodic and the following conditions hold for all $\theta \in \Theta$:*

(1) $\int_\ell^r x^k \mu_\theta(x)dx < \infty$ *for all* $k \in \mathbb{N}$.

(2) $\sup_t E_\theta(|X_t|^k) < \infty$ *for all* $k \in \mathbb{N}$.

(3) $b, \sigma \in C_{p,4,1}((\ell,r) \times \Theta)$.

(4) $g(\Delta, y, x; \theta) \in C_{p,2,6,2}(\mathbb{R}_+ \times (\ell,r)^2 \times \Theta)$ *and has an expansion in powers of* Δ*:*

$$g(\Delta, y, x; \theta) = \\ g(0,y,x;\theta) + \Delta g^{(1)}(y,x;\theta) + \tfrac{1}{2}\Delta^2 g^{(2)}(y,x;\theta) + \Delta^3 R(\Delta,y,x;\theta),$$

where

$$\begin{array}{rcl} g(0,y,x;\theta) & \in & C_{p,6,2}((\ell,r)^2 \times \Theta), \\ g^{(1)}(y,x;\theta) & \in & C_{p,4,2}((\ell,r)^2 \times \Theta), \\ g^{(2)}(y,x;\theta) & \in & C_{p,2,2}((\ell,r)^2 \times \Theta). \end{array}$$

We define $C_{p,k_1,k_2,k_3}(\mathbb{R}_+ \times (\ell,r)^2 \times \Theta)$ as the class of real functions $f(t,y,x;\theta)$ satisfying that

(i) $f(t,y,x;\theta)$ is k_1 times continuously differentiable with respect t, k_2 times continuously differentiable with respect y, and k_3 times continuously differentiable with respect α and with respect to β

(ii) f and all partial derivatives $\partial_t^{i_1}\partial_y^{i_2}\partial_\alpha^{i_3}\partial_\beta^{i_4} f$, $i_j = 1,\ldots k_j$, $j = 1,2$, $i_3 + i_4 \leq k_3$, are of polynomial growth in x and y uniformly for θ in a compact set (for fixed t).

The classes $C_{p,k_1,k_2}((\ell,r) \times \Theta)$ and $C_{p,k_1,k_2}((\ell,r)^2 \times \Theta)$ are defined similarly for functions $f(y;\theta)$ and $f(y,x;\theta)$, respectively.

References

Aït-Sahalia, Y. (2002). "Maximum likelihood estimation of discretely sampled diffusions: a closed-form approximation approach". *Econometrica*, 70:223–262.

Aït-Sahalia, Y. (2003). "Closed-form likelihood expansions for multivariate diffusions". Working paper, Princeton University. To appear in *Ann. Statist.*

Aït-Sahalia, Y. & Mykland, P. (2003). "The effects of random and discrete sampling when estimating continuous-time diffusions". *Econometrica*, 71:483–549.

Aït-Sahalia, Y. & Mykland, P. A. (2004). "Estimators of diffusions with randomly spaced discrete observations: a general theory". *Ann. Statist.*, 32:2186–2222.

Beskos, A.; Papaspiliopoulos, O.; Roberts, G. O. & Fearnhead, P. (2006). "Exact and computationally efficient likelihood-based estimation for discretely observed diffusion processes". *J. Roy. Statist. Soc. B*, 68:333–382.

Bibby, B. M. & Sørensen, M. (1995). "Martingale estimation functions for discretely observed diffusion processes". *Bernoulli*, 1:17–39.

Bibby, B. M. & Sørensen, M. (1996). "On estimation for discretely observed diffusions: a review". *Theory of Stochastic Processes*, 2:49–56.

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

Bollerslev, T. & Wooldridge, J. (1992). "Quasi-maximum likelihood estimators and inference in dynamic models with time-varying covariances". *Econometric Review*, 11:143–172.

Campbell, J. Y.; Lo, A. W. & MacKinlay, A. C. (1997). *The Econometrics of Financial Markets*. Princeton University Press, Princeton.

Chan, K. C.; Karolyi, G. A.; Longstaff, F. A. & Sanders, A. B. (1992). "An empirical comparison of alternative models of the short-term interest rate". *Journal of Finance*, 47:1209–1227.

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

De Jong, F.; Drost, F. C. & Werker, B. J. M. (2001). "A jump-diffusion model for exchange rates in a target zone". *Statistica Neerlandica*, 55:270–300.

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

Doukhan, P. (1994). *Mixing, Properties and Examples*. Springer, New York. Lecture Notes in Statistics 85.

Durbin, J. (1960). "Estimation of parameters in time-series regression models". *J. Roy. Statist. Soc. B*, 22:139–153.

Durham, G. B. & Gallant, A. R. (2002). "Numerical techniques for maximum likelihood estimation of continuous-time diffusion processes". *J. Business & Econom. Statist.*, 20:297–338.

Elerian, O.; Chib, S. & Shephard, N. (2001). "Likelihood inference for discretely observed non-linear diffusions". *Econometrica*, 69:959–993.

Eraker, B. (2001). “MCMC Analysis of Diffusion Models with Application to Finance”. *J. Business & Econom. Statist.*, 19:177–191.

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

Forman, J. L. & Sørensen, M. (2008). “The Pearson diffusions: A class of statistically tractable diffusion processes”. *Scand. J. Statist.* To appear.

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

Genon-Catalot, V. & Jacod, J. (1993). “On the estimation of the diffusion coefficient for multi-dimensional diffusion processes”. *Ann. Inst. Henri Poincaré, Probabilités et Statistiques*, 29:119–151.

Genon-Catalot, V.; Jeantheau, T. & Larédo, C. (2000). “Stochastic volatility models as hidden Markov models and statistical applications”. *Bernoulli*, 6:1051–1079.

Gloter, A. & Sørensen, M. (2008). “Estimation for stochastic differential equations with a small diffusion coefficient”. *Stoch. Proc. Appl.* To appear.

Gobet, E. (2002). “LAN property for ergodic diffusions with discrete observations”. *Ann. Inst. Henri Poincaré, Probabilités et Statistiques*, 38:711–737.

Godambe, V. P. (1960). “An optimum property of regular maximum likelihood estimation”. *Ann. Math. Stat.*, 31:1208–1212.

Godambe, V. P. & Heyde, C. C. (1987). “Quasi likelihood and optimal estimation”. *International Statistical Review*, 55:231–244.

Hall, P. & Heyde, C. C. (1980). *Martingale Limit Theory and Its Applications.* Academic Press, New York.

Hansen, L. P. (1982). “Large sample properties of generalized method of moments estimators”. *Econometrica*, 50:1029–1054.

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

Heyde, C. C. (1997). *Quasi-Likelihood and Its Application.* Springer-Verlag, New York.

Jacobsen, M. (2001). “Discretely observed diffusions; classes of estimating functions and small Δ-optimality”. *Scand. J. Statist.*, 28:123–150.

Jacobsen, M. (2002). “Optimality and small Δ-optimality of martingale estimating functions”. *Bernoulli*, 8:643–668.

Jacod, J. & Sørensen, M. (2008). “Asymptotic statistical theory for stochastic processes: a review.”. Preprint, Department of Mathematical Sciences, University of Copenhagen. In preparation.

Kelly, L.; Platen, E. & Sørensen, M. (2004). "Estimation for discretely observed diffusions using transform functions". *J. Appl. Prob.*, 41:99–118.

Kessler, M. (1996). *Estimation paramétrique des coefficients d'une diffusion ergodique à partir d'observations discrètes.* PhD thesis, Laboratoire de Probabilités, Université Paris VI.

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

Kessler, M. (2000). "Simple and explicit estimating functions for a discretely observed diffusion process". *Scand. J. Statist.*, 27:65–82.

Kessler, M. & Paredes, S. (2002). "Computational aspects related to martingale estimating functions for a discretely observed diffusion". *Scand. J. Statist.*, 29:425–440.

Kessler, M. & Sørensen, M. (1999). "Estimating equations based on eigenfunctions for a discretely observed diffusion process". *Bernoulli*, 5:299–314.

Kusuoka, S. & Yoshida, N. (2000). "Malliavin calculus, geometric mixing, and expansion of diffusion functionals". *Probability Theory and Related Fields*, 116:457–484.

Larsen, K. S. & Sørensen, M. (2007). "A diffusion model for exchange rates in a target zone". *Mathematical Finance*, 17:285–306.

Nagahara, Y. (1996). "Non-Gaussian distribution for stock returns and related stochastic differential equation". *Financial Engineering and the Japanese Markets*, 3:121–149.

Overbeck, L. & Rydén, T. (1997). "Estimation in the Cox-Ingersoll-Ross model". *Econometric Theory*, 13:430–461.

Ozaki, T. (1985). "Non-linear time series models and dynamical systems". In Hannan, E. J.; Krishnaiah, P. R. & Rao, M. M., editors, *Handbook of Statistics, Vol. 5*, pages 25–83. Elsevier Science Publishers.

Pedersen, A. R. (1995). "A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations". *Scand. J. Statist.*, 22:55–71.

Phillips, P. & Yu, J. (2008). "Gaussian/ML estimation of continuous time models in finance". In Andersen, T. G.; Davis, R. A.; Kreiss, J.-P. & Mikosch, T., editors, *Handbook of Financial Time Series.* Springer.

Poulsen, R. (1999). "Approximate maximum likelihood estimation of discretely observed diffusion processes". Working Paper 29, Centre for Analytical Finance, Aarhus.

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

Roberts, G. O. & Stramer, O. (2001). "On inference for partially observed nonlinear diffusion models using Metropolis-Hastings algorithms". *Biometrika*, 88:603–621.

Skorokhod, A. V. (1989). *Asymptotic Methods in the Theory of Stochastic Differential Equations*. American Mathematical Society, Providence, Rhode Island.

Sørensen, H. (2001). "Discretely observed diffusions: Approximation of the continuous-time score function". *Scand. J. Statist.*, 28:113–121.

Sørensen, M. (1997). "Estimating functions for discretely observed diffusions: a review". In Basawa, I. V.; Godambe, V. P. & Taylor, R. L., editors, *Selected Proceedings of the Symposium on Estimating Functions*, pages 305–325. Hayward: Institute of Mathematical Statistics. IMS Lecture Notes – Monograph Series, Vol. 32.

Sørensen, M. (2007). "Efficient estimation for ergodic diffusions sampled at high frequency". Preprint, Department of Mathematical Sciences, University of Copenhagen.

Sørensen, M. & Uchida, M. (2003). "Small-diffusion asymptotics for discretely sampled stochastic differential equations". *Bernoulli*, 9:1051–1069.

Veretennikov, A. Y. (1987). "Bounds for the mixing rate in the theory of stochastic equations". *Theory of Probability and its Applications*, 32:273–281.

Wong, E. (1964). "The construction of a class of stationary Markoff processes". In Bellman, R., editor, *Stochastic Processes in Mathematical Physics and Engineering*, pages 264–276. American Mathematical Society, Rhode Island.

Yoshida, N. (1992). "Estimation for diffusion processes from discrete observations". *Journal of Multivariate Analysis*, 41:220–242.

Research Papers 2008

2008-05 Annastiina Silvennoinen and Timo Teräsvirta: Modelling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model

2008-06 Annastiina Silvennoinen and Timo Teräsvirta: Multivariate GARCH models. To appear in T. G. Andersen, R. A. Davis, J.-P. Kreiss and T. Mikosch, eds. Handbook of Financial Time Series. New York: Springer.

2008-07 Changli He, Annastiina Silvennoinen and Timo Teräsvirta: Parameterizing unconditional skewness in models for financial time series

2008-08 Cristina Amado and Timo Teräsvirta: Modelling Conditional and Unconditional Heteroskedasticity with Smoothly Time-Varying Structure

2008-09 Søren Johansen and Bent Nielsen: An analysis of the indicator saturation estimator as a robust regression estimator

2008-10 Peter Christoffersen, Kris Jacobs, Christian Dorion and Yintian Wang: Volatility Components, Affine Restrictions and Non-Normal Innovations

2008-11 Peter Christoffersen, Kris Jacobs, Chayawat Ornthanalai and Yintian Wang: Option Valuation with Long-run and Short-run Volatility Components

2008-12 Tom Engsted and Stig V. Møller: An iterated GMM procedure for estimating the Campbell-Cochrane habit formation model, with an application to Danish stock and bond returns

2008-13 Lars Stentoft: Option Pricing using Realized Volatility

2008-14 Jie Zhu: Pricing Volatility of Stock Returns with Volatile and Persistent Components

2008-15 Jie Zhu: Testing for Expected Return and Market Price of Risk in Chinese A-B Share Market: A Geometric Brownian Motion and Multivariate GARCH Model Approach

2008-16: Jie Zhu: FIEGARCH-M and and International Crises: A Cross-Country Analysis

2008-17: Almut E. D. Veraart: Inference for the jump part of quadratic variation of Itô semimartingales

2008-18: Michael Sørensen: Parametric inference for discretely sampled stochastic differential equations