

# Estimating equations based on eigenfunctions for a discretely observed diffusion process

MATHIEU KESSLER<sup>1</sup> and MICHAEL SØRENSEN<sup>2</sup>

<sup>1</sup>*Departamento de Matemática Aplicada, Universidad de Murcia, Campus de Cartagena, Paseo Alfonso XIII, 30203 Cartagena, Spain. e-mail: kessler@plc.um.es*

<sup>2</sup>*Department of Theoretical Statistics, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark. e-mail: michael@math.ku.dk*

A new type of martingale estimating function is proposed for inference about classes of diffusion processes based on discrete-time observations. These estimating functions can be tailored to a particular class of diffusion processes by utilizing a martingale property of the eigenfunctions of the generators of the diffusions. Optimal estimating functions in the sense of Godambe and Heyde are found. Inference based on these is invariant under transformations of data. A result on consistency and asymptotic normality of the estimators is given for ergodic diffusions. The theory is illustrated by several examples and by a simulation study.

*Keywords:* generator; optimal estimating function; quasilikelihood; stochastic differential equation

## 1. Introduction

Martingale estimating functions have turned out to give good and relatively simple estimators for discretely observed diffusion models, for which the likelihood function is only explicitly known in special cases. By discrete observations we mean observations of the process at deterministic times  $t_0 < \dots < t_n$ . These estimators have the added virtue of being consistent and asymptotically normal when the number of observations tends to infinity without the (often unrealistic) assumption that the time between observations tends to zero which must be made for most other estimators proposed in the literature. Only the estimators proposed by Kessler (1995b) and Pedersen (1995a,b) are also consistent and asymptotically normal without this assumption.

Bibby (1994), Pedersen (1994), Bibby and Sørensen (1995) and Kessler (1995a) considered martingale estimating functions of the type

$$G_n(\theta) = \sum_{i=1}^n g(X_{t_i}, X_{t_{i-1}}; \theta), \quad (1.1)$$

where  $X_{t_0}, X_{t_1}, \dots, X_{t_n}$  are the data, and where  $g(y, x; \theta)$  is a polynomial in  $y$  such that

$$\mathbb{E}_\theta[g(X_{t_i}, X_{t_{i-1}}; \theta) | X_{t_{i-1}}] = 0 \quad \mathbb{P}_\theta - \text{a.s.} \quad (1.2)$$

for all values of the parameter  $\theta$ . Such polynomial estimating functions can be regarded as approximations of the true score function, which are very good when the time between observations is small. In fact, the transition distribution can be well approximated by a normal law when the time between observations is small. This implies a Gaussian approximate likelihood function and an approximate score function of the type (1.1) with a polynomial  $g$ . Some polynomial estimating functions indeed provide efficient estimators when the time between observations tends to zero with increasing number of observations (Bibby and Sørensen 1995). Polynomial estimating functions of the type (1.1) can be viewed as a natural generalization of the method of moments to Markov processes.

There is, however, no reason to believe that polynomial estimating functions are in general the best choice, or even natural, when the time between observations is large and thus the transition distribution is far from Gaussian. In this situation, we also cannot in general expect a generalized method of moments to behave particularly well. In the present paper we therefore study martingale estimating functions of the form (1.1), where the function  $g$  need not be a polynomial but is based on eigenfunctions for the generator of the diffusion model. This allows the estimating function to be more closely tailored to the diffusion model. A first justification for this choice of  $g$  is the importance of the eigenfunctions in the study of the transition density. Indeed, a classical result is the series expansion of the transition density for a diffusion process in terms of the eigenfunctions of its generator (Karlin and Taylor 1981). It should also be mentioned here that inference based on the optimal estimating functions studied in this paper is invariant under transformation of data, which is not the case for the polynomial estimating functions.

In some well-known simple diffusion models it turns out that the eigenfunctions are polynomials, such that we are led back to the polynomial estimating functions. This is, not surprisingly, the case for the Ornstein–Uhlenbeck process, where the eigenfunctions are the Hermite polynomials. Even in these cases it is, however, an advantage to follow the route via spectral theory since this gives more explicit estimating equations than have been found in earlier papers. Moreover, this puts the polynomial estimating functions into a broader perspective.

In Section 2 we review a few results from the spectral theory of diffusions and derive the martingale estimating functions based on eigenfunctions of the generator. In Section 3 we find optimal estimating functions of this type, while consistency and asymptotic normality of the estimators are proved for ergodic diffusions in Section 4. In Section 5 we give some results on eigenfunctions and martingales. In particular, we give sufficient conditions that the estimating functions introduced in Section 2 are martingales. The theory is illustrated by examples.

## **2. Martingale estimating functions based on eigenfunctions**

Consider a class of one-dimensional diffusion processes defined as solutions of the stochastic differential equations

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

where  $W$  is a Wiener process. The functions  $b$  and  $\sigma$  are known; the unknown parameter  $\theta$  varies in a subset  $\Theta$  of  $\mathbb{R}^d$ . They are assumed to be smooth enough to ensure for every  $\theta$  the uniqueness of the law of the solution, which we denote by  $\mathbb{P}_\theta$ . The statistical problem is to draw inference about  $\theta$  based on discrete observations  $X_{t_0}, X_{t_1}, \dots, X_{t_n}$  which, for convenience, we assume to be observed equidistantly, i.e.  $t_i - t_{i-1} = \Delta$ . The extension to arbitrary  $t_i$  is obvious.

The differential operator  $L_\theta$  defined by

$$L_\theta = \frac{1}{2}\sigma^2(x; \theta) \frac{d^2}{dx^2} + b(x; \theta) \frac{d}{dx} \tag{2.2}$$

for all twice differentiable functions is called the *generator* of the diffusion model (2.1). A twice continuously differentiable function  $\phi(x; \theta)$  is called an *eigenfunction* for  $L_\theta$  with *eigenvalue*  $\lambda(\theta)$  if

$$L_\theta \phi(x; \theta) = -\lambda(\theta)\phi(x; \theta) \tag{2.3}$$

for all  $x$  in the state space of  $X$  under  $\mathbb{P}_\theta$ .

It can be shown that under weak regularity conditions (see Section 5)

$$\mathbb{E}_\theta[\phi(X_{t_i}; \theta) | X_{t_{i-1}}] = e^{-\lambda(\theta)\Delta} \phi(X_{t_{i-1}}; \theta), \tag{2.4}$$

where  $\mathbb{E}_\theta$  denotes the expectation under  $\mathbb{P}_\theta$ . Therefore, we can define a martingale estimating function by (1.1) with

$$g(y, x; \theta) = \alpha(x; \theta)\{\phi(y; \theta) - e^{-\lambda(\theta)\Delta} \phi(x; \theta)\}, \tag{2.5}$$

where  $\alpha$  is an arbitrary function. More generally, given  $k$  eigenfunctions  $\phi_1(x; \theta), \dots, \phi_k(x; \theta)$  with distinct eigenvalues  $\lambda_1(\theta), \dots, \lambda_k(\theta)$ , we can define a martingale estimating function by

$$g(y, x; \theta) = \sum_{j=1}^k \alpha_j(x; \theta)\{\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta)\}. \tag{2.6}$$

Sufficient conditions for (2.4) to hold will be discussed in Section 5. In particular, (2.4) holds in the examples considered below. It is well known from the spectral theory of diffusion processes that the set  $\Lambda_\theta$  of all eigenvalues for  $L_\theta$ , the spectrum of  $L_\theta$ , is contained in  $[0, \infty)$ . For many diffusion models the spectrum is a discrete set  $\Lambda_\theta = \{\lambda_0(\theta), \lambda_1(\theta), \dots\}$ , where  $0 \leq \lambda_0(\theta) < \lambda_1(\theta) < \dots < \lambda_n(\theta) \uparrow \infty$ . Since the dependence on the past is mainly determined by the small eigenvalues, it seems natural to use the eigenfunctions associated with the first  $k$  eigenvalues to define an estimating function. A further justification for considering functions  $g$  of this form is that the eigenvalue problem (2.3) is a Sturm–Liouville problem. By a classical result of this theory, we have, for an ergodic diffusion with invariant probability  $\mu$ , a series expansion of any function in  $L^2(\mu)$  in terms of the eigenfunctions

$(\phi_i)_{i \geq 0}$  associated with the eigenvalues  $(\lambda_i)_{i \geq 0}$  (Coddington and Levinson 1955), i.e. a function  $f$  in  $L^2(\mu)$  can be written as

$$f(y) = \sum_{i=0}^{\infty} c_i \phi_i(y), \tag{2.7}$$

where  $(c_i)$  is a sequence of real numbers and the series converges in  $L^2(\mu)$ . There is a similar result for many non-ergodic diffusions. Thus, for a fixed  $x$ ,  $\sum_{j=0}^k \alpha_j(x; \theta) \phi_j(y; \theta)$  can be seen as a truncated series of the form (2.7). The estimating function (2.6) is obtained when one makes a martingale out of this sum. Expansions of transition densities of the form (2.7) mainly depend on the first eigenfunctions. In fact, the weights  $c_i$  decrease exponentially with the eigenvalues. In Section 3, we derive the optimal estimating function based on  $k$  given eigenfunctions. The optimal estimating function can be thought of as an approximation in terms of eigenfunctions to the unknown score function. We shall not discuss how best to choose  $k$  or the eigenfunctions for a given class of diffusions.

If the state space of the diffusion is a finite interval, the spectrum is discrete. Otherwise the nature of the spectrum depends on the behaviour at  $\pm\infty$  of the function

$$U_\theta(x) = \frac{b^2(x; \theta)}{2\sigma^2(x; \theta)} + \frac{1}{2}b'(x; \theta) - \frac{b(x; \theta)\sigma'(x; \theta)}{\sigma(x; \theta)} + \frac{1}{8}\sigma'(x; \theta)^2 - \frac{1}{4}\sigma(x; \theta)\sigma''(x; \theta),$$

where a prime denotes the derivative with respect to  $x$ . The function  $U_\theta(x)$  is closely related to the potential of a Schrödinger equation (Banon 1978). Define  $u_\theta = \min\{\lim_{x \rightarrow -\infty} U_\theta(x), \lim_{x \rightarrow \infty} U_\theta(x)\}$ . Then  $\Lambda_\theta \cap [0, u_\theta]$  is discrete (Goel and Richter-Dyn 1974; Banon 1978). In particular, if  $U_\theta(x)$  tends to  $\infty$  as  $x \rightarrow \pm\infty$ , the spectrum is discrete. When the spectrum is not discrete, it may still be a good idea to choose  $k$  eigenfunctions with associated eigenvalues  $0 < \lambda_1(\theta) < \dots < \lambda_k(\theta)$  and thus define a martingale estimating function with  $g$  of the type (2.6).

Further properties of the eigenfunctions are worth mentioning. Consider an ergodic diffusion with invariant measure  $\mu_\theta$  (although these properties hold in a wider generality). Then we have for any eigenfunction  $\phi(x; \theta)$  that  $\phi(\cdot; \theta)$  is in  $L^2(\mu_\theta)$  and

$$\int \phi(y; \theta) \mu_\theta(dy) = 0. \tag{2.8}$$

If  $\phi_i(y; \theta)$  and  $\phi_j(y; \theta)$  are two eigenfunctions with two distinct eigenvalues, we have

$$\int \phi_i(y; \theta) \phi_j(y; \theta) \mu_\theta(dy) = 0 \tag{2.9}$$

(Karlin and Taylor 1981).

**Example 2.1.** The class of diffusion processes which solve

$$dX_t = (\beta + \theta X_t) dt + \sigma X_t^{1/2} dW_t, \tag{2.10}$$

where  $\beta > 0$ ,  $\theta < 0$  and  $\sigma > 0$ , is known in the financial literature as the Cox–Ingersoll–Ross (1985) model. The model has been well known, for instance in mathematical biology, for

many years. For this model, the spectrum is  $\Lambda_\theta = \{-n\theta: n = 0, 1, \dots\}$  with corresponding eigenfunctions  $\phi_i(x) = L_i^{(\nu)}(-2\theta x\sigma^{-2})$ , where  $L_i^{(\nu)}$  is the  $i$ th order Laguerre polynomial with parameter  $\nu = 2\beta\sigma^{-2} - 1$  (Karlin and Taylor 1981). We are thus in the case of polynomial martingale estimating functions.

**Example 2.2.** Consider the class of diffusions given as solutions of

$$dX_t = -\theta \tan(X_t) dt + dW_t, \tag{2.11}$$

which for  $\theta \geq \frac{1}{2}$  is an ergodic diffusion on the interval  $(-\pi/2, \pi/2)$ . This diffusion can be thought of as an Ornstein–Uhlenbeck process on a finite interval. The invariant measure has a density proportional to  $\cos(x)^{2\theta}$ .

The spectrum of the model is  $\Lambda_\theta = \{n(\theta + n/2): n = 0, 1, \dots\}$  with associated eigenfunctions given by  $\phi_i(x; \theta) = C_i^\theta(\sin(x))$ , where  $C_i^\theta$  is a Gegenbauer polynomial of order  $i$ . This follows from the differential equation solved by the Gegenbauer polynomials.

By equation 8.934-2 of Gradshteyn and Ryzhik (1965),

$$\phi_i(x; \theta) = \sum_{m=0}^i \binom{\theta - 1 + m}{m} \binom{\theta - 1 + i - m}{i - m} \cos\left\{(2m - i)\left(\frac{\pi}{2} - x\right)\right\}.$$

The first non-trivial eigenfunction is  $\sin(x)$  (a constant is omitted) with eigenvalue  $\theta + \frac{1}{2}$ . From the martingale estimating function

$$\sum_{i=1}^n \sin(X_{t_{i-1}}) \{\sin(X_{t_i}) - e^{-(\theta+1/2)\Delta} \sin(X_{t_{i-1}})\}$$

we obtain the very simple estimator for  $\theta$ :

$$\tilde{\theta}_n = -\Delta^{-1} \ln \left\{ \left( \sum_{i=1}^n \sin(X_{t_{i-1}}) \sin(X_{t_i}) \right) / \sum_{i=1}^n \sin^2(X_{t_{i-1}}) \right\} - \frac{1}{2}, \tag{2.12}$$

which is, of course, only defined when the numerator is positive. We shall see later that  $\tilde{\theta}_n$  is defined with probability tending to 1 as  $n \rightarrow \infty$  and that it is  $n^{1/2}$  consistent and asymptotically normal, but not optimal.

**Example 2.3 (A generalized logistic diffusion).** In many areas of application, e.g. mathematical finance and turbulence, diffusion models with a marginal distribution with log-linear tails are needed. Such a diffusion model is obtained when the drift is

$$b(x; \theta_1, \theta_2) = \left\{ (\theta_1 - \theta_2) \cosh\left(\frac{x}{2}\right) - (\theta_1 + \theta_2) \sinh\left(\frac{x}{2}\right) \right\} \cosh\left(\frac{x}{2}\right),$$

where  $\theta_1 > 0$  and  $\theta_2 > 0$ , and the diffusion coefficient is given by  $\sigma(x) = 2 \cosh(x/2)$ . This diffusion is ergodic, and its invariant measure is the generalized logistic distribution with density

$$B(\theta_1 + 1, \theta_2 + 1) e^{(\theta_1+1)x} (1 + e^x)^{-(\theta_1+\theta_2+2)}.$$

Here  $B$  denotes the beta function. This distribution was studied by Barndorff-Nielsen *et al.*

(1982). The model has a discrete spectrum given by  $\Lambda_{\theta_1, \theta_2} = \{(n/2)(n + \theta_1 + \theta_2 + 1) : n = 0, 1, \dots\}$ . The associated eigenfunctions are  $\phi_i(x; \theta_1, \theta_2) = P_i^{\theta_2, \theta_1}(\tanh(x/2))$ , where  $P_i^{\theta_2, \theta_1}$  denotes the Jacobi polynomial of order  $i$ .

**Example 2.4. (A stopped Bessel process).** The Bessel processes solve the equation

$$dX_t = \theta X_t^{-1} dt + dW_t. \tag{2.13}$$

The state space is  $(0, \infty)$ . Suppose that the process is started at a point in the interval  $(0, 1)$  and stopped when it hits unity. Let us further assume that  $\theta \geq \frac{1}{2}$ . Then the eigenfunctions are  $\phi_i(x; \theta) = x^{-\theta+1/2} J_{\theta-1/2}(x \zeta_i(\theta))$ ,  $i = 1, 2, \dots$  where  $J_\nu$  is a Bessel function and  $\{\zeta_i(\theta) : i = 1, 2, \dots\}$  is the sequence of positive zeros of  $J_{\theta-1/2}$  (Karlin and Taylor 1981, p. 336). The spectrum is  $\{2\zeta_i(\theta)^2 : i = 1, 2, \dots\}$ .

### 3. Derivation of the optimal estimating functions

In this section we shall investigate how to construct from  $k$  given eigenfunctions,  $\phi_1(x; \theta), \dots, \phi_k(x; \theta)$ , an optimal martingale estimating function in the sense of Godambe and Heyde (1987). After first considering the general case, we shall study how the results simplify when the eigenfunctions are polynomials. Finally, we consider the case of data-independent weights, where completely explicit results can be given. To simplify the exposition we assume in this section that the parameter  $\theta$  is real.

#### 3.1. The general case

Let  $C$  be the space of functions  $\Theta \times \mathbb{R} \rightarrow \mathbb{R}$  which are continuously differentiable in  $\theta$ . For a non-negative integer  $k$ , define  $\Gamma_k$  to be the set of martingale estimating functions  $G^k$  obtained by choosing  $k$  functions  $\alpha_1, \dots, \alpha_k$  in  $C$  and defining for all  $n \geq 1$  and  $\theta$

$$G_n^k(\theta) = \sum_{i=1}^n \sum_{j=1}^k \alpha_j(X_{t_{i-1}}; \theta) \{ \phi_j(X_{t_i}; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(X_{t_{i-1}}; \theta) \}. \tag{3.1}$$

Let us stress that in (3.1) the sequences  $\phi_1, \dots, \phi_k$  and  $\lambda_1, \dots, \lambda_k$  of eigenfunctions and associated eigenvalues are fixed; hence  $\Gamma_k$  is obtained when  $\alpha_1, \dots, \alpha_k$  vary in (3.1).

We are now interested in finding within  $\Gamma_k$  the optimal element  $G_n^{*k}(\theta)$  in the sense of Godambe and Heyde (1987). This optimal element can be viewed as a projection in a suitable sense of the score function onto  $\Gamma_k$ . In our framework, its characterization is particularly simple. Let  $\Pi_\Delta(y, x; \theta)$  denote the transition density, i.e.  $\Pi_\Delta(\cdot, x; \theta)$  is the density under  $\mathbb{P}_\theta$  of  $X_\Delta$  given  $X_0 = x$  and define  $\gamma^k$  as the set of functions  $\Theta \times \mathbb{R}^2 \rightarrow \mathbb{R}$  such that  $\Gamma_k = \{ \sum_{i=1}^n g(X_{t_i}, X_{t_{i-1}}; \theta) : g \in \gamma^k \}$ . Finally, let  $\gamma^k(x; \theta)$  be the space of functions  $\mathbb{R} \rightarrow \mathbb{R}$  obtained by fixing  $x$  and  $\theta$  in the elements of  $\gamma^k$ . Then the optimal estimating function within the class  $\Gamma_k$  is given by  $G_n^{*k}(\theta) = \sum_{i=1}^n g^*(X_{t_i}, X_{t_{i-1}}; \theta)$  where  $g^*(\cdot, x; \theta)$  is the projection in  $L^2(\Pi_\Delta(y, x; \theta) dy)$  of  $(\partial_\theta \Pi_\Delta / \Pi_\Delta)(\cdot, x; \theta)$  onto  $\gamma^k(x; \theta)$ . By  $\partial_\theta$  we denote partial differentiation with respect to  $\theta$ . For details, see Kessler (1995a).

If we write  $g^*(\theta, x, y) = \sum_{j=1}^k \alpha_j^*(x; \theta) \{\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta)\}$ , the determination of  $\alpha^* = (\alpha_1^*, \dots, \alpha_k^*)^T$ , where the superscript T denotes transposition, is an easy consequence of this characterization. Indeed, the orthogonality property

$$\int \left( \frac{\partial_\theta \Pi_\Delta}{\Pi_\Delta} - g^* \right) g(y, x; \theta) \Pi_\Delta(y, x; \theta) dy = 0 \quad \text{for all } g(\cdot, x; \theta) \in \gamma^k(x; \theta) \quad (3.2)$$

is equivalent to the equation

$$A\alpha^* = B \quad (3.3)$$

with  $A = (a_{i,j})_{1 \leq i,j \leq k}$  where for  $1 \leq r, s \leq k$

$$a_{r,s}(x; \theta) := \int \{\phi_r(y; \theta) - e^{-\lambda_r(\theta)\Delta} \phi_r(x; \theta)\} \{\phi_s(y; \theta) - e^{-\lambda_s(\theta)\Delta} \phi_s(x; \theta)\} \times \Pi_\Delta(y, x; \theta) dy,$$

and  $B = (b_1, \dots, b_k)^T$  where for  $1 \leq j \leq k$

$$b_j(x; \theta) := - \int \partial_\theta (\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta)) \Pi_\Delta(y, x; \theta) dy.$$

Again details can be found in Kessler (1995a). Easy computations involving (2.3) simplify the expressions of  $a_{r,s}$  and  $b_j$  and allow us to state the following proposition.

**Proposition 3.1.** *The unique element which is optimal within  $\Gamma_k$  in both the fixed sample and the asymptotic sense is*

$$G_n^{*k}(\theta) = \sum_{i=1}^n \sum_{j=1}^k \alpha_j^*(X_{t_{i-1}}; \theta) (\phi_j(X_{t_i}; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(X_{t_{i-1}}; \theta)), \quad (3.4)$$

where the coefficients  $\alpha_1^*, \dots, \alpha_k^*$  satisfy (3.3) with

$$a_{r,s}(x; \theta) = \int \phi_r(y; \theta) \phi_s(y; \theta) \Pi_\Delta(y, x; \theta) dy - e^{-(\lambda_r(\theta) + \lambda_s(\theta))\Delta} \phi_r(x; \theta) \phi_s(x; \theta) \quad (3.5)$$

for  $1 \leq r, s \leq k$ , and

$$b_j(x; \theta) = - \int \partial_\theta \phi_j(y; \theta) \Pi_\Delta(y, x; \theta) dy + \partial_\theta (e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta)) \quad (3.6)$$

for  $1 \leq j \leq k$ .

In the case where  $\theta \in \Theta \subset \mathbb{R}^d$ , the optimal estimating function is a  $d$ -dimensional vector, the  $i$ th component of which is obtained as the projection in  $L^2(\Pi_\Delta(y, x; \theta) dy)$  of  $(\partial_{\theta_i} \Pi_\Delta / \Pi_\Delta)(\cdot, x; \theta)$  onto the subspace  $\gamma^k(x; \theta)$  defined previously. Thus it is determined by solving a system similar to (3.3) with  $\partial_\theta$  substituted by  $\partial_{\theta_i}$  in (3.6).

As mentioned in the introduction, inference based on the optimal estimating function (3.4) is *invariant under data transformations*. To see this note that, if  $u$  is a twice continuously differentiable bijection, then the diffusion  $Y_t = u(X_t)$  has eigenfunctions

$\phi_j(u^{-1}(y); \theta)$  with the same eigenvalues  $\lambda_j(\theta)$ ,  $j = 0, 1, 2, \dots$ , as the diffusion  $X$ . This can be proved by writing, using the Itô formula, the stochastic differential equation satisfied by  $Y$  and checking that, with obvious notation,  $L^Y(f \circ u^{-1})(y) = L^X f(u^{-1}(y))$  for any  $f$  in  $C^2(\mathbb{R})$ . Hence a straightforward change of variable in the system (3.3) proves that the optimal estimating function based on the transformed observations  $Y_{t_1} = u(X_{t_1}), \dots, Y_{t_n} = u(X_{t_n})$  equals that based on the original data  $X_{t_1}, \dots, X_{t_n}$ . This is an important statistical property which polynomial estimating functions do not possess.

**Example 3.2 (Example 2.2 continued).** For the class of diffusions given by (2.11) we can find the optimal estimating functions explicitly. We illustrate this by finding  $G^{*1}(\theta)$ . The first non-trivial eigenfunction is, up to a multiplicative function of  $\theta$ ,  $\sin(x)$  with eigenvalue  $\theta + \frac{1}{2}$ ; so  $b_1(x; \theta) = -\Delta e^{-(\theta+1/2)\Delta} \sin(x)$ ; see (3.6). Hence the only problem is the integral  $\int \sin^2(y)\Pi_\Delta(y, x; \theta) dy$ , but the second non-trivial eigenfunction is  $2(\theta + 1)\sin^2(x) - 1$  with eigenvalue  $2(\theta + 1)$ . Therefore, using (2.4), we find that

$$G^{*1}(\theta) = \sum_{i=1}^n \frac{\sin(X_{t_{i-1}})\{\sin(X_{t_i}) - e^{-(\theta+1/2)\Delta} \sin(X_{t_{i-1}})\}}{\frac{1}{2}(e^{2(\theta+1)\Delta} - 1)/(\theta + 1) - (e^\Delta - 1)\sin^2(X_{t_{i-1}})}, \tag{3.7}$$

where a constant has been omitted. The optimal estimating functions can, in a similar way, be explicitly found for the diffusion in Example 2.3.

Note that, when  $\Delta$  is small, we obtain the approximately optimal estimating function

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

which has the explicit solution

$$\tilde{\theta} = -\Delta^{-1} \ln \left( \frac{\sum_{i=1}^n \frac{\tan(X_{t_{i-1}})\sin(X_{t_i})}{\cos(X_{t_{i-1}})}}{\sum_{i=1}^n \tan^2(X_{t_{i-1}})} \right) - \frac{1}{2}. \tag{3.8}$$

The explicit estimator  $\tilde{\theta}$  can, for instance, be used as a starting value for finding the optimal estimator. It is also  $n^{1/2}$  consistent and asymptotically normal provided that  $\theta_0 > 1.5$ .

Let us finish this section by a remark of a practical nature. The optimal estimating function exhibited by Bibby and Sørensen (1995) is as follows:

$$G_n^*(\theta) = \sum_{i=1}^n \frac{\partial_\theta F}{\Phi}(X_{t_{i-1}}; \theta)(X_{t_i} - F(X_{t_{i-1}}; \theta)),$$

where  $F$  and  $\Phi$  are the first moment and second centred moment respectively of the transition distribution. Although there are, in general, no closed expressions for  $F$  and  $\Phi$ , Bibby and Sørensen demonstrated that a good approximation to these quantities can be obtained by simulating a number of replicates of  $X_\Delta$  and using the law of large numbers. However, the computation of  $\partial_\theta F$  is more difficult because of problems of numerical stability. In our situation, this problem disappears since, in the expressions for  $a_{r,s}$  and  $b_j$  (see (3.5) and (3.6)), the only terms that we have to approximate numerically are simple integrals with



respect to  $\Pi_\Delta$  and not derivatives of these integrals. Therefore our estimating functions can easily be found by simulation.

### 3.2. Polynomial estimating functions

For some classical diffusion processes (e.g. Example 2.1), the eigenfunctions  $\phi_j$  for  $j \geq 1$  are polynomials of degree  $j$ . In these cases, explicit calculation of  $A$  and  $B$  in (3.3) is possible.

It is straightforward that, if  $\phi_j(y; \theta) = \sum_{i=0}^j c_{j,i}(\theta)y^i$ , computing  $A$  and  $B$  is equivalent to computing the moments  $\int y^i \Pi_\Delta(y, x; \theta) dy$  for  $1 \leq i \leq 2k$ . Using (2.3), it turns out that these moments are the solutions of the linear system

$$e^{-\lambda_i(\theta)} \phi_i(x; \theta) = \sum_{j=0}^i c_{i,j}(\theta) \int y^j \Pi_\Delta(y, x; \theta) dy, \quad i = 1, \dots, k.$$

Hence, we have a closed expression for the moments of the transition density, and the expression for the optimal element  $G^{*k}$  is explicit.

**Example 3.3 (The radial Ornstein–Uhlenbeck process).** The class of solutions of

$$dX_t = (\theta X_t^{-1} - X_t) dt + dW_t,$$

where  $\theta > 0$ , is called the radial Ornstein–Uhlenbeck processes. The state space is the positive real line. The process is ergodic when  $\theta \geq \frac{1}{2}$ . The eigenfunctions are  $\phi_i(x; \theta) = L_i^{(\nu-1/2)}(x^2)$ , where  $L_i^{(\nu)}$  is the  $i$ th-order Laguerre polynomial with parameter  $\nu$ , and the associated eigenvalues are  $\{2i: i = 1, 2, \dots\}$  (Karlin and Taylor 1981, p. 333). Since the eigenfunctions are polynomials, the optimal estimating functions can be found explicitly. Using the results above, we find that the first optimal estimating function is

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

Note that as in Example 3.2 we can obtain an approximately optimal estimating function for  $\Delta$  small by omitting the last term in the numerator. This approximate estimating function is linear in  $\theta$  and hence has an explicit solution.

### 3.3. A simple estimating function

Let us finally consider the simple estimating function

$$H_n^k(\theta) = \sum_{i=1}^n \sum_{j=1}^k \beta_j(\theta) \{ \phi_j(X_{t_j}; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(X_{t_{j-1}}; \theta) \}, \quad (3.9)$$

where  $\beta_1, \dots, \beta_k$  are continuously differentiable functions of  $\theta$  only. We denote by  $\mathcal{H}_k$  the class of martingale estimating functions of this form. The choice of the “best” element of  $\mathcal{H}_k$  is more difficult. Indeed, the existence of an optimal element within  $\mathcal{H}_k$  in the sense of Godambe and Heyde is not ensured. Nevertheless, it is possible to minimize the asymptotic

variance amongst all elements of  $\mathcal{H}_k$ , provided that we restrict attention to elements of  $\mathcal{H}_k$  which yield a  $n^{1/2}$  consistent and asymptotically normal estimator (Kessler 1995a). We shall therefore only consider ergodic diffusions here. Conditions ensuring  $n^{1/2}$  consistency and asymptotic normality for ergodic diffusions are given in the next section. Let  $\mu_\theta$  denote the invariant probability and define  $Q_\Delta^\theta(x, y) = \mu_\theta(x)\Pi_\Delta(y, x; \theta)$ . The element  $H_n^{*k}(\theta) = \sum_{i=1}^n h^{*k}(X_{t_i}, X_{t_{i-1}}; \theta)$  which minimizes the asymptotic variance is found (Kessler 1995a) when  $h^{*k}(\cdot, \cdot; \theta)$  is the projection of  $\partial_\theta Q_\Delta^\theta / Q_\Delta^\theta$  in  $L^2(Q_\Delta^\theta(y, x) dy dx)$  onto the subspace  $\epsilon^k(\theta)$  defined by  $\mathcal{H}_k = \{H_n^k(\theta) = \sum_{i=1}^n h^k(X_{t_i}, X_{t_{i-1}}; \theta), h^k \in \epsilon^k(\theta)\}$ . From (2.9), we deduce that  $\phi_i(y; \theta) - e^{-\lambda_i(\theta)\Delta} \phi_i(x; \theta)$ ,  $i = 1, \dots, k$ , is an orthogonal sequence in  $L^2(Q_\Delta^\theta(y, x) dy dx)$ . Thus,  $h^{*k}(y, x; \theta) = \sum_{j=1}^k \beta_j^*(\theta)(\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta))$ , with

$$\begin{aligned} \beta_j^*(\theta) &= \frac{\int (\partial_\theta Q_\Delta^\theta / Q_\Delta^\theta)(y, x) \{ \phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta) \} Q_\Delta^\theta(y, x) dy dx}{\int (\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta))^2 Q_\Delta^\theta(y, x) dy dx} \\ &= - \frac{1 - e^{-\lambda_j(\theta)\Delta} \int \partial_\theta \phi_j(x; \theta) \mu_\theta(x) dx}{1 - e^{-2\lambda_j(\theta)\Delta} \int \phi_j^2(x; \theta) \mu_\theta(x) dx}. \end{aligned}$$

The coefficients  $\beta_j^*(\theta)$  are explicit, and we have a closed expression for the element in  $\mathcal{H}_k$  for which the asymptotic variance is minimum. Note the strong relationship of this criterion to the asymptotic optimality criterion considered by Heyde and Gay (1989).

### 4. Consistency and asymptotic normality

In this section we show that the estimators obtained from the estimating equations derived in the previous sections are consistent and asymptotically normal. We confine the discussion to the case of ergodic diffusion processes.

Our framework is similar to that of Florens-Zmirou (1989) and Bibby and Sørensen (1995). We consider discrete observations of a process  $X$  that solves (2.1) for some  $\theta \in \Theta \subseteq \mathbb{R}$ , where  $\Theta$  is open. The assumption that  $\theta$  is one dimensional is made only to simplify the exposition. Let  $s(x; \theta)$  denote the density of the scale measure:

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

We shall work under the assumptions that will now be given.

**Condition 4.1.** *The following holds for all  $\theta \in \Theta$ :*

$$\int_0^\infty s(x; \theta) dx = \int_{-\infty}^0 s(x; \theta) dx = \infty \quad \text{and} \quad \int_{-\infty}^\infty [s(x; \theta)\sigma^2(x; \theta)]^{-1} dx = A(\theta) < \infty.$$

Under Condition 4.1,  $X$  is ergodic, and with respect to the Lebesgue measure its invariant

measure  $\mu_\theta$  has the density  $[A(\theta)s(x; \theta)\sigma^2(x; \theta)]^{-1}$ . Condition 4.1 is formulated for diffusions whose state space is the real line. For diffusions confined to a smaller interval, the conditions should be reformulated appropriately. The probability measure  $Q_\Delta^\theta$  is defined as in Section 3.3. For a function  $g: \mathbb{R}^2 \rightarrow \mathbb{R}$ , we use the notation  $Q_\Delta^\theta(g) = \int g dQ_\Delta^\theta$ . We consider a general estimating function of the form (3.1) which satisfies the following condition where  $\theta_0$  denotes the true value of  $\theta$ , and where

$$g_j(x, y; \theta) = \alpha_j(x; \theta)\{\phi_j(y; \theta) - e^{-\lambda_j(\theta)\Delta} \phi_j(x; \theta)\}, \quad 1 \leq j \leq k. \tag{4.1}$$

**Condition 4.2.** *The following hold for all  $\theta \in \Theta$ .*

- (a) *The functions  $g_j(x, y; \theta)$ ,  $j = 1, \dots, k$ , are continuously differentiable with respect to  $\theta$  for all  $x$  and  $y$ . The functions  $\partial_\theta g_j(x, y; \theta)$  are locally dominated square integrable with respect to  $Q_\Delta^{\theta_0}$ , and  $f(\theta_0) = \sum_{j=1}^k Q_\Delta^{\theta_0}(\partial_\theta g_j(\theta_0)) \neq 0$ .*
- (b) *The functions  $g_j(x, y; \theta)$ ,  $j = 1, \dots, k$ , are in  $L^2(Q_\Delta^{\theta_0})$ .*

Under these assumptions the following theorem can be proved along the same lines as Theorem 3.3. of Bibby and Sørensen (1995). Here we have not included Condition 3.1 (c) of Bibby and Sørensen (1995) because Lemma 3.1. of Bibby and Sørensen (1995) holds without this assumption as follows from Theorem 1.1 of Billingsley (1961a) and the central limit theorem for martingales of Billingsley (1961b).

**Theorem 4.3.** *Under Conditions 4.1 and 4.2 an estimator  $\hat{\theta}_n$ , which solves the equation  $G_n^k(\hat{\theta}_n) = 0$ , exists with a probability tending to one as  $n \rightarrow \infty$  under  $P_{\theta_0}$ . Moreover, as  $n \rightarrow \infty$ ,  $\hat{\theta}_n \rightarrow \theta_0$  in probability under  $P_{\theta_0}$  and*

$$n^{1/2}(\hat{\theta}_n - \theta_0) \rightarrow \mathcal{N}\left(0, \frac{v(\theta_0)}{f^2(\theta_0)}\right)$$

*in distribution under  $P_{\theta_0}$ , where*

$$v(\theta_0) = \sum_{r=1}^k \sum_{s=1}^k \int \alpha_r(x; \theta_0)\alpha_s(x; \theta_0)a_{r,s}(x; \theta_0)\mu_{\theta_0}(x) dx.$$

From Theorem 4.3 we deduce that, if the optimal estimating function within  $\Gamma_k$  satisfies Condition 4.2, it provides an estimator  $\hat{\theta}_n^{*k}$  which is  $n^{1/2}$  consistent and asymptotically normal. Its asymptotic variance is the smallest within the class of all estimators derived from an estimating function in  $\Gamma_k$ . Let us briefly discuss the *efficiency* of the proposed estimators. Note that, for the optimal estimating function (3.4),  $v(\theta_0) = f(\theta_0)$  so that the minimal asymptotic variance within  $\Gamma_k$  equals  $v(\theta_0)^{-1}$ . From the definition of  $g^*$ , we have that  $v(\theta) = \int (g^*)^2 dQ_\Delta^\theta \leq \int (\partial_\theta \Pi / \Pi)^2 dQ_\Delta^\theta$ , where the right-hand side is the asymptotic Fisher information. Since  $v(\theta_0)$  is the upper bound for the asymptotic expected information within the class considered, the maximum-likelihood estimator will always be at least as efficient as any of our estimators. If the union  $\bigcup_{k=1}^\infty V_k$ , where  $V_k$  is the space spanned by  $\{\phi_1(\cdot; \theta), \dots, \phi_k(\cdot; \theta)\}$ , is dense in  $L^2(\Pi_\Delta(y, x; \theta) dy)$  for every  $x$ , then there exists a sequence  $k_n$  such that the asymptotic variance of the estimator  $\hat{\theta}_n^{*k_n}$  is the inverse of the

asymptotic Fisher information. For details, see Kessler (1995a). In particular in the case of a bounded state interval, where it is well known that the sequence  $\phi_1(\cdot; \theta), \phi_2(\cdot; \theta), \dots$  is complete in  $L^2(\mu_\theta)$ , the union  $\bigcup_{k=1}^\infty V_k$  is dense in  $L^2(\Pi_\Delta(y, x; \theta) dy)$ ; so in this case there exists quite generally a sequence  $k_n$  such that the estimator  $\hat{\theta}_n^{*k_n}$  is asymptotically efficient. In the case of an infinite state interval, the sequence  $\phi_1(\cdot; \theta), \phi_2(\cdot; \theta), \dots$  is also complete in  $L^2(\mu_\theta)$  when the spectrum is discrete but, in order to deduce denseness of  $\bigcup_{k=1}^\infty V_k$  in  $L^2(\Pi_\Delta(y, x; \theta) dy)$ , additional conditions are needed.

Comparison of the relative efficiency of optimal estimators obtained from two different classes of estimating equations usually cannot be done explicitly unless one of the classes is included in the other, in which case the estimator obtained from the larger class is the more efficient. Therefore, we cannot give general results above the relative efficiency of optimal estimators based on disjoint sets of eigenfunctions. However, the efficiency of  $\hat{\theta}_n^{*k}$  obviously increases with increasing  $k$  but so does the computational complexity. In most models the efficiency is presumably high and increases only slowly with  $k$ , provided that  $k$  is sufficiently large, and a compromise between efficiency and computational feasibility must be found.

Let us finish this section by testing the practical behaviour of some of our estimators for the process which is the solution of (2.11). We simulated, using the Milstein scheme, 200 trajectories of this process for two values of  $\theta$  and computed the mean value and the standard error of the simple estimator  $\check{\theta}_n$  given by (2.12), of  $\tilde{\theta}_n^*$  which solves the optimal estimating equation (3.7), and of  $\tilde{\theta}_n$  the estimator given by (3.8). Table 1 and Table 2 collect the results obtained for different values of  $\Delta$  if the true parameter values are 1.5 and 3 respectively. For each estimator, we have reported the number of trajectories for which the estimating equations had no solution. The mean and the standard error are computed for the remaining trajectories.

**Table 1.** Comparison of the mean and standard deviation for 200 realizations of  $\tilde{\theta}_n^*$ ,  $\check{\theta}_n$  and  $\tilde{\theta}_n$  for the process solving (2.11) with  $\theta = 1.5$  and  $x_0 = 0$ . In all cases there was a solution to the estimating equation

$\Delta$	$n$	$\tilde{\theta}_n^*$		$\check{\theta}_n$		$\tilde{\theta}_n$	
		Mean	Standard deviation	Mean	Standard deviation	Mean	Standard deviation
0.1	200	1.62	0.44	1.60	0.45	1.78	2.00
	500	1.55	0.27	1.55	0.29	1.59	0.49
	1000	1.54	0.19	1.53	0.21	1.57	0.31
0.3	200	1.55	0.34	1.55	0.35	1.59	0.45
	500	1.53	0.21	1.53	0.21	1.55	0.27
	1000	1.52	0.15	1.52	0.15	1.54	0.22
0.5	200	1.54	0.38	1.54	0.38	1.53	0.53
	500	1.50	0.22	1.50	0.22	1.51	0.32
	1000	1.49	0.15	1.49	0.15	1.50	0.25

**Table 2.** Comparison of the mean and standard deviation for 200 realizations of  $\tilde{\theta}_n^*$ ,  $\check{\theta}_n$  and  $\tilde{\theta}_n$  for the process solving (2.13) with  $\theta = 3$  and  $x_0 = 0$ , where  $N^n$  is the number of realizations for which there is no solution to the estimating equation

$\Delta$	$n$	$\tilde{\theta}_n^*$			$\check{\theta}_n$			$\tilde{\theta}_n$		
		$N^n$	Mean	Standard deviation	$N^n$	Mean	Standard deviation	$N^n$	Mean	Standard deviation
0.1	200	3	3.12	0.74	0	3.16	0.81	0	3.17	0.81
	500	0	3.04	0.42	0	3.04	0.43	0	3.04	0.43
	1000	0	3.04	0.30	0	3.04	0.31	0	3.05	0.30
0.3	200	3	3.16	0.62	0	3.20	0.70	0	3.18	0.74
	500	0	3.08	0.47	0	3.08	0.41	0	3.08	0.47
	1000	0	3.04	0.29	0	3.04	0.26	0	3.04	0.29
0.5	200	16	3.03	0.72	1	3.25	1.08	1	3.25	1.29
	500	2	3.03	0.50	0	3.05	0.55	0	3.06	0.61
	1000	0	3.00	0.38	0	3.00	0.38	0	3.01	0.43

Note that, by using the kind of arguments developed in Example 3.2, on the one hand, we can, for the simple estimator  $\check{\theta}_n$  compute explicitly the asymptotic variance of  $n^{1/2}(\check{\theta}_n - \theta_0)$ :

$$\frac{v(\theta_0)}{f^2(\theta_0)} = \frac{1 - e^{-2(\theta_0+1)\Delta}}{\Delta^2 e^{-2(\theta_0+1/2)\Delta}} + 3 \frac{\theta_0 + 1 (e^{-\Delta} - 1)}{\theta_0 + 2 \Delta^2},$$

and, on the other hand, we can compute numerically the optimal asymptotic variance. The approximately optimal estimating function  $\tilde{G}$  does not satisfy Condition 4.2 (b) if the true parameter value is less than or equal to 1.5. Nevertheless, as we see from Table 1, its behaviour for the considered sample sizes is reasonably satisfactory. For  $\theta_0 = 3$ , we can explicitly compute the asymptotic variance of  $n^{1/2}(\tilde{\theta}_n - \theta_0)$  to be

$$\frac{5}{\Delta^2} \left( \frac{1 - e^{-8\Delta}}{4 e^{-7\Delta}} + e^{-\Delta} - 1 \right).$$

Table 3 contains the asymptotic variances for the estimators  $\tilde{\theta}^*$  and  $\check{\theta}$  for the different values of  $\Delta$ , when  $\theta_0 = 1.5$ , whereas in Table 4 the true parameter value is taken to be 3 and the

**Table 3.** Asymptotic variances of  $n^{1/2}(\tilde{\theta}_n^* - \theta_0)$  and  $n^{1/2}(\check{\theta}_n - \theta_0)$  for the process solving (2.13) with  $\theta_0 = 1.5$

$\Delta$	$\tilde{\theta}^*$	$\check{\theta}$
0.1	34.39	38.31
0.3	21.95	22.49
0.5	23.61	23.76

**Table 4.** Asymptotic variances for  $n^{1/2}(\tilde{\theta}_n^* - \theta_0)$ ,  $n^{1/2}(\check{\theta}_n - \theta_0)$  and  $n^{1/2}(\tilde{\theta}_n - \theta_n)$  for the process solving (2.13) with  $\theta_0 = 3$

$\Delta$	$\tilde{\theta}^*$	$\check{\theta}$	$\tilde{\theta}$
0.1	85.37	88.05	91.03
0.3	75.34	75.59	88.73
0.5	126.21	126.26	154.66

asymptotic variances for the three estimators are given. Note how close  $\check{\theta}$  is to being optimal, particularly when  $\Delta$  is not too small. Note also that the standard deviations found for finite sample sizes by simulation are in good accordance with those given by the asymptotic variances.

## 5. Eigenfunctions and martingales

The generator  $L$  of a Markov process  $X$  is defined by

$$\Delta^{-1}\{\Pi_\Delta(f)(x) - f(x)\} \rightarrow Lf(x) \quad (5.1)$$

as  $\Delta \rightarrow 0$ , where  $\Pi_\Delta(f)(x) = \mathbb{E}[f(X_\Delta)|X_0 = x]$ . Often the convergence in (5.1) is taken to be uniformly bounded in  $x$  so that the limit is bounded. When the Markov process is a diffusion, the domain  $\mathcal{D}$  of the generator  $L$  is then the class of bounded, twice continuously differentiable functions with bounded derivatives. For an eigenfunction  $\phi \in \mathcal{D}$  with eigenvalue  $\lambda$  it is not difficult to prove, using the Markov property, that (5.1) implies that

$$\frac{\partial}{\partial \Delta} \Pi_\Delta(\phi)(x) = L\Pi_\Delta(\phi)(x) = -\lambda\Pi_\Delta(\phi)(x),$$

and that hence

$$\Pi_\Delta(\phi)(x) = e^{-\lambda\Delta} \phi(x),$$

which is (2.4). The domain  $\mathcal{D}$  is, however, too restricted for our purpose; so we shall instead consider the so-called extended domain  $\mathcal{D}^*$  defined as the class of twice continuously differentiable functions  $f$  for which the process

$$N_t = f(X_t) - f(X_0) - \int_0^t Lf(X_s) ds \quad (5.2)$$

is a martingale. If  $dX_t = b(X_t)dt + \sigma(X_t)dW_t$ , it follows by Itô's formula that

$$N_t = \int_0^t f'(X_s)\sigma(X_s)dW_s;$$

so a sufficient condition that  $f \in \mathcal{D}^*$  is

$$\int_0^t \mathbb{E}[f'(X_s)^2 \sigma^2(X_s)] ds < \infty \tag{5.3}$$

for all  $t > 0$ . Let  $\mathcal{D}^{**}$  denote the class of twice continuously differentiable functions satisfying (5.3).

Now suppose that  $\phi$  is an eigenfunction with eigenvalue  $\lambda$ , and consider the process  $Y_t = e^{\lambda t} \phi(X_t)$ . By Itô's formula,

$$\begin{aligned} Y_t &= Y_0 + \int_0^t e^{\lambda s} \{L\phi(X_s) + \lambda\phi(X_s)\} ds + \int_0^t e^{\lambda s} \phi'(X_s) \sigma(X_s) dW_s \\ &= Y_0 + \int_0^t e^{\lambda s} \phi'(X_s) \sigma(X_s) dW_s. \end{aligned}$$

Therefore, if  $\phi \in \mathcal{D}^{**}$ ,  $Y$  is a martingale, which implies (2.4). With a little more care, it can be shown that (2.4) holds for any eigenfunction in the extended domain  $\mathcal{D}^*$ .

We shall give two simple but useful sufficient conditions under which a function  $f$  satisfies (5.3). First suppose that  $X$  is ergodic with invariant measure  $\mu$ , and that  $X_0 \sim \mu$ . Then the condition

$$\int_{\mathbb{R}} f'(x)^2 \sigma^2(x) \mu(dx) < \infty, \tag{5.4}$$

implies (5.3) and hence that  $Y$  is a martingale. Thus (5.4) implies (2.4).

Now consider a not necessarily ergodic diffusion. If  $\sigma$  and  $f'$  are bounded on the state space of  $X$ , (5.3) is obviously satisfied. This is the case for any twice differentiable function  $f$  if the state space is bounded and if  $\sigma$  and  $f'$  are continuous on the closure of the state space. A more generally useful sufficient condition is that  $b$  and  $\sigma$  are of linear growth and that  $f'$  is of polynomial growth. Then an application of Gronwall's lemma implies (5.3). By means of these results, it is easy to verify that the estimating functions considered in the examples are indeed martingales.

## Acknowledgements

This work was partially supported by the EU Human Capital and Mobility programme. The authors are grateful to Jean Jacod for helpful comments. Thanks are also due to an associate editor for a detailed set of comments that has improved the presentation.

## References

- Banon, G. (1978) Non parametric identification for diffusion processes. *SIAM J. Control Optimization*, **16**, 380–395.
- Barndorff-Nielsen, O.E., Kent, J. and Sørensen, M. (1982) Normal variance-mean mixtures and  $z$  distributions. *Int. Statist. Rev.*, **50**, 145–159.
- Bibby, B.M. (1994) Optimal combination of martingale estimating functions for discretely observed

- diffusion processes. Research Report 264, Department of Theoretical Statistics, University of Aarhus.
- Bibby, B.M. and Sørensen, M. (1995) Martingale estimation functions for discretely observed diffusion processes. *Bernoulli*, **1**, 17–39.
- Billingsley, P. (1961a) *Statistical Inference for Markov Processes*. Chicago, IL: University of Chicago.
- Billingsley, P. (1961b) The Lindeberg–Lévy theorem for martingales. *Proc. Amer. Math. Soc.*, **12**, 788–792.
- Coddington, E.A. and Levinson, N. (1955) *Theory of Ordinary Differential Equations*. New York: McGraw-Hill.
- Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985) A theory of the term structure of interest rates. *Econometrica*, **53**, 385–407.
- Florens-Zmirou, D. (1989) Approximate discrete schemes for statistics of diffusion processes. *Statistics*, **20**, 547–557.
- Godambe, V.P. and Heyde, C.C. (1987) Quasi likelihood and optimal estimation. *Int. Statist. Rev.*, **55**, 231–244.
- Goel, N.S. and Richter-Dyn, N. (1974) *Stochastic Models in Biology*. New York: Academic Press.
- Gradshteyn, I.S. and Ryzhik, I.M. (1965) *Table of Integrals, Series, and Products*, 4th edn. New York: Academic Press.
- Heyde, C.C. and Gay, R. (1989) On asymptotic quasi-likelihood estimation. *Stochastic Processes Applic.*, **31**, 223–236.
- Karlin, S. and Taylor, H.M. (1981) *A Second Course in Stochastic Processes*. Orlando, FL: Academic Press.
- Kessler, M. (1995a) Quasi-likelihood inference for a discrete Markov chain. Preprint, Université de Paris VI.
- Kessler, M. (1995b) Simple estimating functions for a discretely observed diffusion process. Research Report 336, Department of Theoretical Statistics, University of Aarhus.
- Pedersen, A.R. (1994) Quasi-likelihood inference for discretely observed diffusion processes. Research Report 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.

Received December 1995 and revised November 1997