

# I

---

## Discretely Observed Diffusions: Approximation of the Continuous-time Score Function

---

### *Abstract*

We discuss parameter estimation for discretely observed, ergodic diffusion processes where the diffusion coefficient does not depend on the parameter. We propose using an approximation of the continuous-time score function as an estimating function. The estimating function can be expressed in simple terms through the drift and the diffusion coefficient and is thus easy to calculate. Simulation studies show that the method performs well.

### *Key words*

Continuous-time score function; diffusion process; discrete observations; estimating function.

### *Publication details*

This paper has been accepted for publication in *Scandinavian Journal of Statistics*. An earlier version was printed as Preprint no. 8, 1998 at Department of Theoretical Statistics, University of Copenhagen (Sørensen 1998a). In this version a few notational changes have been made.

## I.1 Introduction

This paper is about parameter estimation for discretely observed diffusion models with known diffusion function. The idea is to use an approximation of the continuous-time score function as estimating function.

This idea is very much in the spirit of the early work by Le Breton (1976) and Florens-Zmirou (1989). They both studied the usual Riemann-Itô discretization of the continuous-time log-likelihood function, and Florens-Zmirou (1989) showed that the corresponding estimator is inconsistent when the length of the time interval between observations is constant.

More recently, various methods providing consistent estimators have been developed, *e.g.* methods based on approximations of the true, discrete-time likelihood function (Pedersen 1995*b*, Aït-Sahalia 1998); methods based on auxiliary models (Gallant & Tauchen 1996, Gourieroux *et al.* 1993); and methods based on estimating functions (Bibby & Sørensen 1995, Hansen & Scheinkman 1995, Kessler 2000, Jacobsen 1998).

The estimating function discussed in this paper is of the simple, explicit type discussed by Hansen & Scheinkman (1995) and Kessler (2000), that is, on the form  $\sum_{i=1}^n \mathcal{A}_\theta h(X_{t_{i-1}}, \theta)$  where  $\mathcal{A}_\theta$  is the diffusion generator. Hansen & Scheinkman (1995) focus on identifiability and asymptotic behaviour of the estimating function whereas Kessler (2000) focuses on asymptotic behaviour and efficiency of the estimator.

The main contribution of this paper is to recognize that, with a special choice of  $h$ , the corresponding estimating function can be interpreted as an approximation to the continuous-time score function. The approximating estimating function is unbiased, it is invariant to data transformations, it provides consistent and asymptotically normal estimators, and it can be explicitly expressed in terms of the drift and diffusion coefficient. The estimating function is also — at least in some cases — available for multi-dimensional processes.

The main objection against the method is the need for a completely known diffusion function. In case of a parameter dependent diffusion function the suggested estimating function is still unbiased and can thus in principle be used, but there is no longer justification for using it since the continuous-time likelihood function does not exist.

We present the model and the basic assumptions in Section I.2, and the estimating function is derived in Section I.3. We give the asymptotic results in Section I.4, and examples and simulation studies in Section I.5. Sections I.2–I.5 discuss one-dimensional diffusion processes exclusively; we study the multi-dimensional case in Section I.6.

## I.2 Model and notation

In this section we present the diffusion model, state the assumptions and introduce some notation.

(I.2)

We consider the one-dimensional, time-homogeneous stochastic differential equation

$$dX_t = b(X_t, \theta) dt + \sigma(X_t) dW_t, \quad X_0 = x_0 \quad (\text{I.1})$$

where  $\theta$  is an unknown  $p$ -dimensional parameter from the parameter space  $\Theta \subseteq \mathbb{R}^p$  and  $W$  is a one-dimensional Brownian motion. The functions  $b: \mathbb{R} \times \Theta \rightarrow \mathbb{R}$  and  $\sigma: \mathbb{R} \rightarrow (0, \infty)$  are known, and the derivatives  $\partial\sigma/\partial x$  and  $\partial^2 b/\partial\theta_j\partial x$  are assumed to exist for all  $j = 1, \dots, p$ . Note that  $\sigma$  does not depend on  $\theta$ .

We assume that for any  $\theta$ , (I.1) has a unique, strong solution  $X$  and that the range of  $X$  does not depend on  $\theta$ . Assume furthermore that there exists a unique invariant distribution  $\mu_\theta = \mu(x, \theta) dx$  such that a solution to (I.1) with  $X_0 \sim \mu_\theta$  (instead of  $X_0 = x_0$ ) is strictly stationary. Sufficient conditions for these assumptions to hold can be found in Karatzas & Shreve (1991).

The invariant density is given by

$$\mu(x, \theta) = (C(\theta) s(x, \theta) \sigma^2(x))^{-1} \quad (\text{I.2})$$

where  $C(\theta)$  is a normalizing constant and  $s(\cdot, \theta)$  is the density of the scale measure, *i.e.*  $\log s(x, \theta) = -2 \int^x b(y, \theta) / \sigma^2(y) dy$ .

For all  $\theta \in \Theta$ , the distribution of  $X$  is denoted  $P_\theta$  if  $X_0 = x_0$  (as in (I.1)) and  $P_\theta^\mu$  if  $X_0 \sim \mu_\theta$ .<sup>1</sup> Under  $P_\theta^\mu$  all  $X_t \sim \mu_\theta$ .  $E_\theta^\mu$  is the expectation wrt.  $P_\theta^\mu$ .

The objective is to estimate  $\theta$  from observations of  $X$  at discrete time-points  $t_1 < \dots < t_n$ . Define  $t_0 = 0$  and  $\Delta_i = t_i - t_{i-1}$  and let  $\theta_0$  be the true parameter.

Finally, we need some matrix notation: Vectors in  $\mathbb{R}^p$  are considered as  $p \times 1$  matrices, and  $A^T$  is the transpose of  $A$ . For a function  $f = (f_1, \dots, f_q)^T: \mathbb{R} \times \Theta \rightarrow \mathbb{R}^q$  we let  $f'(x, \theta)$  be the  $q \times 1$  matrix of partial derivatives with respect to  $x$  and  $\dot{f}(x, \theta) = D_\theta f(x, \theta)$  be the  $q \times p$  matrix of partial derivatives with respect to  $\theta$ , *i.e.*  $\dot{f}_{jk} = \partial f_j / \partial \theta_k$ .

### I.3 The estimating function

In this section we derive a simple, unbiased estimating function as an approximation of the continuous-time score function.

First a comment on the model: It is important that  $\sigma$  does not depend on  $\theta$ . Otherwise the distributions of  $(X_s)_{0 \leq s \leq t}$  corresponding to two different parameter values are typically singular for all  $t \geq 0$ . If  $Y$  is the solution to  $dY_t = b(Y_t, \theta) dt + \tilde{\sigma}(Y_t, \theta) dW_t$ , then the process  $(\int^{Y_t} 1/\tilde{\sigma}(y, \theta) dy)_{t \geq 0}$  is the solution to (I.1) with  $\sigma \equiv 1$ , but this is of no help for estimation purposes since the transformation depends on the (unknown) parameter.

<sup>1</sup>Note the difference in notation from Chapter 2 and Paper II where  $P_\theta$  is the distribution of  $X$  when  $X_0$  is started according to the stationary distribution (and where we have no notation for the distribution of  $X$  given a particular value of  $X_0$ ).

When  $\sigma$  is completely known as we have assumed, it follows from Lipster & Shirayev (1977) that the likelihood function for a continuous observation  $(X_s)_{0 \leq s \leq t}$  exists and that the corresponding score process  $S^c$  is given by

$$S_t^c(\theta) = \int_0^t \frac{\dot{b}(X_s, \theta)}{\sigma^2(X_s)} dX_s - \int_0^t \frac{b(X_s, \theta) \dot{b}(X_s, \theta)}{\sigma^2(X_s)} ds.$$

Using (I.1) we find that

$$dS_t^c(\theta) = \frac{\dot{b}(X_t, \theta)}{\sigma(X_t)} dW_t. \quad (\text{I.3})$$

This shows that  $S^c(\theta)$  is a local martingale and that it is a genuine martingale if  $E_\theta^\mu \int_0^t (\dot{b}_j(X_s, \theta)/\sigma(X_s))^2 ds < \infty$  for all  $t \geq 0$  and all  $j = 1, \dots, p$ , i.e. if

$$E_\theta^\mu \left( \frac{\dot{b}_j(X_0, \theta)}{\sigma(X_0)} \right)^2 = \int \left( \frac{\dot{b}_j(x, \theta)}{\sigma(x)} \right)^2 \mu(x, \theta) dx < \infty \quad (\text{I.4})$$

for all  $j = 1, \dots, p$ . In particular,  $E_\theta^\mu S_t^c = 0$  for all  $t \geq 0$  if (I.4) holds.

If  $X$  was observed continuously on the interval  $[0, t_n]$  we would estimate  $\theta$  by solving the equation  $S_{t_n}^c(\theta) = 0$ . For discrete observations at time-points  $t_1, \dots, t_n$ , the idea is to use an approximation of  $S_{t_n}^c$  as estimating function.

The most obvious approximation is obtained by simply replacing the integrals in (I.3) with the corresponding Riemann and Itô sums,

$$R_n(\theta) = \sum_{i=1}^n \frac{\dot{b}(X_{t_{i-1}}, \theta)}{\sigma^2(X_{t_{i-1}})} (X_{t_i} - X_{t_{i-1}}) - \sum_{i=1}^n \Delta_i \frac{b(X_{t_{i-1}}, \theta) \dot{b}(X_{t_{i-1}}, \theta)}{\sigma^2(X_{t_{i-1}})}. \quad (\text{I.5})$$

Note that this would be the score function if the conditional distributions of the increments  $X_{t_i} - X_{t_{i-1}}$ , given the past, were Gaussian with expectation  $\Delta_i b(X_{t_{i-1}}, \theta)$  and variance  $\Delta_i \sigma^2(X_{t_{i-1}})$ . However, usually  $E_\theta^\mu R_n(\theta) \neq 0$ , and  $R_n$  provides inconsistent estimators unless  $\sup_{i=1, \dots, n} \Delta_i \rightarrow 0$  (Florens-Zmirou 1989).

We now propose an unbiased approximation of  $S_{t_n}^c$ . Let  $\mathcal{A}_\theta$  denote the differential operator associated with the infinitesimal generator for  $X$ , that is,

$$\mathcal{A}_\theta f(x, \theta) = b(x, \theta) f'(x, \theta) + \frac{1}{2} \sigma^2(x) f''(x, \theta)$$

for functions  $f : \mathbb{R} \times \Theta \rightarrow \mathbb{R}^p$  that are twice continuously differentiable wrt.  $x$ .

Recall that  $\mu$  is the invariant density and assume that the derivatives

$$h^* = D_\theta \log \mu : \mathbb{R} \times \Theta \rightarrow \mathbb{R}^p$$

wrt. the coordinates of  $\theta$  exist and are twice continuously differentiable wrt.  $x$ , such that  $\mathcal{A}_\theta h^*$  is well-defined. The connection between  $h^*$  and  $S^c$  is given in the following proposition:

(I.4)

**Proposition I.1** *With respect to  $P_\theta$  and  $P_\theta^\mu$ , it holds for all  $t \geq 0$  that*

$$2S_t^c(\theta) = h^*(X_t, \theta) - h^*(X_0, \theta) - \int_0^t \mathcal{A}_\theta h^*(X_s, \theta) ds. \quad (\text{I.6})$$

*Proof* We show that

$$dh^*(X_t, \theta) = \mathcal{A}_\theta h^*(X_t, \theta) dt + 2dS_t^c(\theta). \quad (\text{I.7})$$

Then (I.6) follows immediately since  $S_0^c = 0$ . Using (I.2) we easily find the first derivative of  $h^* = D_\theta \log \mu$  in terms of  $b$  and  $\sigma$ ;

$$h^{*'}(x, \theta) = D_x D_\theta \log \mu(x, \theta) = -D_\theta D_x \log s(x, \theta) = 2 \frac{\dot{b}(x, \theta)}{\sigma^2(x)}. \quad (\text{I.8})$$

Now simply apply Itô's formula on  $h^*$ . □

The proposition suggests that we use

$$F_n(\theta) = \frac{1}{2} \sum_{i=1}^n \Delta_i \mathcal{A}_\theta h^*(X_{t_{i-1}}, \theta)$$

as an approximation to  $-S_{t_n}^c$  (since the term  $h^*(X_{t_n}, \theta) - h^*(X_0, \theta)$  is negligible when  $n$  is large) and hence solve the equation  $F_n(\theta) = 0$  in order to find an estimator for  $\theta$ .

The right hand side of (I.6), with an arbitrary function  $h \in \mathcal{C}^2(I)$  substituted for  $h^*$ , is a martingale if  $E_\theta^\mu (h' \sigma)^2 < \infty$ . Hence,

$$E_\theta^\mu \mathcal{A}_\theta h(X_0, \theta) = 0 \quad (\text{I.9})$$

if furthermore  $h$  and  $\mathcal{A}_\theta h$  are in  $L^1(\mu_\theta)$ . In particular  $F_n$  is unbiased, *i.e.*  $E_\theta^\mu F_n(\theta) = 0$ , if (I.4) holds and if  $h^*$  and  $\mathcal{A}_\theta h^*$  are in  $L^1(\mu_\theta)$ .

The moment condition (I.9) was used by Hansen & Scheinkman (1995) to construct general method of moments estimators (their condition C1) and by Kessler (2000) and Jacobsen (1998) to construct unbiased estimating functions. Kessler particularly suggests choosing polynomials  $h$  of low degree — regardless of the model. Instead, we suggest the *model-dependent* choice  $h = h^*$ . Intuitively, this should be good for small  $\Delta_i$ 's since  $F_n \approx -S_{t_n}^c$ . Indeed, for  $\Delta_i \equiv \Delta$ ,  $F_n$  is small  $\Delta$ -optimal in the sense of Jacobsen (1998).

It should be clear, though, that moment conditions like (I.9) cannot achieve asymptotically efficient estimators for a given  $\Delta > 0$  since each term in the discrete-time score function involves *pairs* of observations. Note that if the observations were independent and identically  $\mu_\theta$ -distributed, then the score function would equal  $\sum_{i=1}^n h^*(X_{t_i}, \theta)$  which is thus optimal for  $\Delta \rightarrow \infty$ . Kessler (2000) discussed this estimating function.

When  $\Delta_i \equiv \Delta$ ,  $F_n$  is a *simple* estimating function, *i.e.* a function of the form  $\sum_{i=1}^n f(X_{t_{i-1}}, \theta)$  where  $E_\theta^\mu f(X_0, \theta) = 0$  (Kessler 2000). In general, the  $\Delta_i$ 's can be

interpreted as weights compensating for the dependence between observations that are close in time: an observation is given much weight if it is far in time from the previous one, and little weight if it is close in time to the previous one.

Note that (I.9) holds so that  $F_n$  is unbiased even if  $\sigma$  depends on  $\theta$ . However, the interpretation of  $F_n$  as an approximation of minus the continuous-time score function is of course no longer valid, and the method will be non-optimal even for small  $\Delta_i$ 's (Jacobsen 1998).

A nice property of  $F_n$  is that it is invariant to transformations of data; the estimator does not change if we observe  $\varphi(X_{t_1}), \dots, \varphi(X_{t_n})$  instead of  $X_{t_1}, \dots, X_{t_n}$ . This is not the case for the polynomial martingale estimating functions discussed by Bibby & Sørensen (1995).

To prove the invariance, we need some further notation: For a diffusion process  $Y$  satisfying a stochastic differential equation similar to (I.1), we write  $\mu_Y$  and  $\mathcal{A}_Y$  for the corresponding invariant density and the differential operator, and define  $h_Y^* = D_\theta \log \mu_Y$ .

**Proposition I.2** *Let  $\varphi : I \rightarrow J \subseteq \mathbb{R}$  be a bijection from  $\mathcal{C}^2(I)$  with inverse  $\varphi^{-1}$ , and let  $Y = \varphi(X)$ . Then*

$$\mathcal{A}_Y h_Y^*(y, \theta) = \mathcal{A}_X h_X^*(\varphi^{-1}(y), \theta). \quad (\text{I.10})$$

*Proof* By Itô's formula  $Y$  is the solution to

$$dY_t = b_Y(Y_t, \theta) dt + \sigma_Y(Y_t) dW_t$$

where, with obvious notation,

$$\begin{aligned} b_Y(y, \theta) &= b_X(\varphi^{-1}(y), \theta) \varphi'(\varphi^{-1}(y)) + \frac{1}{2} (\sigma_X^2 \varphi'')(\varphi^{-1}(y)), \\ \sigma_Y(y) &= (\sigma_X \varphi')(\varphi^{-1}(y)). \end{aligned}$$

One can now either check directly from (I.11) below that (I.10) holds or argue as follows. The density for the invariant distribution of  $Y = \varphi(X)$  is given by  $\mu_Y(y, \theta) = \mu_X(\varphi^{-1}(y), \theta) |(\varphi^{-1})'(y)|$  and thus

$$h_Y^*(y, \theta) = D_\theta \log \mu_Y(y, \theta) = D_\theta \log \mu_X(\varphi^{-1}(y), \theta) = h_X^*(\varphi^{-1}(y), \theta).$$

Finally, note that  $\mathcal{A}_Y(f \circ \varphi^{-1})(y) = \mathcal{A}_X f(\varphi^{-1}(y))$  for all  $f \in \mathcal{C}^2(I)$  which concludes the proof.  $\square$

In the following we write  $f^*$  for  $(\mathcal{A}_\theta h^*)/2 = (\mathcal{A}_\theta D_\theta \log \mu)/2$ . It is important to note that we can express  $f^*$  — and thus  $F_n = \sum_{i=1}^n \Delta_i f^*(X_{t_{i-1}}, \cdot)$  — explicitly in terms of  $b$  and  $\sigma$ , even if we have no explicit expression for the normalizing constant  $C(\theta)$ : from (I.8) we get

$$f^* = \mathcal{A}_\theta h^*/2 = \left( \frac{bb'}{\sigma^2} + \frac{1}{2} b'' - \frac{b\sigma'}{\sigma} \right). \quad (\text{I.11})$$

(I.6)

## I.4 Asymptotic properties

In this section we state the asymptotic results for  $F_n$ . We consider equidistant observations,  $t_i = i\Delta$  where  $\Delta$  does not depend on  $n$ , and let  $n \rightarrow \infty$ .

Under suitable regularity conditions, a solution  $\hat{\theta}_n$  to  $F_n(\theta) = 0$  exists with a  $P_{\theta_0}$ -probability tending to 1, and  $\hat{\theta}_n$  is a consistent, asymptotically normal estimator for  $\theta$ . The asymptotic distribution of  $\hat{\theta}_n$  is given by

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0, A(\theta_0)^{-1}V(\theta_0)(A(\theta_0)^{-1})^T\right)$$

wrt.  $P_{\theta_0}$  as  $n \rightarrow \infty$ , where  $A(\theta_0) = E_{\theta_0}^{\mu} \dot{f}^*(X_0, \theta_0)$  and

$$V(\theta_0) = E_{\theta_0}^{\mu} f^*(X_0, \theta_0) f^*(X_0, \theta_0)^T + 2 \sum_{k=1}^{\infty} E_{\theta_0}^{\mu} f^*(X_0, \theta_0) f^*(X_{k\Delta}, \theta_0)^T.$$

Conditions that ensure convergence of the sum in are given by Kessler (2000). If (I.9) holds for each  $\partial h_j^* / \partial \theta_k$ , then

$$A(\theta_0) = 2 E_{\theta_0}^{\mu} \left( \frac{\dot{b}(X_0, \theta_0)}{\sigma(X_0)} \right)^T \left( \frac{\dot{b}(X_0, \theta_0)}{\sigma(X_0)} \right),$$

and  $A(\theta_0)$  is symmetric and positive semidefinite. It must be positive definite. We will not go through the additional regularity conditions here but refer to Kessler (2000) and particularly to Sørensen (1998b).

## I.5 Examples

As already mentioned  $F_n$  can always be expressed explicitly in terms of  $b$  and  $\sigma$ . When  $b$  is linear wrt. the parameter we even get explicit estimators. Assume that  $b(x, \theta) = b_0(x) + \sum_{j=1}^p b_j(x)\theta_j$  for known functions  $b_0, b_1, \dots, b_p : \mathbb{R} \rightarrow \mathbb{R}$  such that the assumptions of Sections I.2 and I.3 hold. From (I.11) we easily deduce that the  $k$ 'th coordinate of  $f^*$  is given by

$$f_k^*(x, \theta) = \sum_{j=1}^p \frac{b_k(x)b_j(x)}{\sigma^2(x)} \theta_j + \frac{b_0(x)b_k(x)}{\sigma^2(x)} + \frac{1}{2} b_k'(x) - \frac{b_k(x)\sigma'(x)}{\sigma(x)}.$$

It follows that  $F_n$  is linear in  $\theta$  and it is easy to show that the estimating equation has a unique, explicit solution if and only if  $b_1, \dots, b_p$  are linearly independent.

The Ornstein-Uhlenbeck model and the Cox-Ingersoll-Ross model are special cases of this setup. Several authors have studied inference for these models, see Bibby & Sørensen (1995), Gouriéroux *et al.* (1993), Kessler (2000), Jacobsen (1998), Overbeck & Rydén (1997), and Pedersen (1995b), for example. From now on, we consider equidistant observations,  $\Delta_i \equiv \Delta$ .

**Example (The Ornstein-Uhlenbeck process)** Let  $X$  be the solution to

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

where  $\theta < 0$ . The estimator is given by  $\hat{\theta}_n = -n/(2\sum_{i=1}^n X_{(i-1)\Delta}^2)$ . Since  $h^*$  is an eigenfunction for  $\mathcal{A}_\theta$ , the simple estimating functions corresponding to  $f = h^*$  and  $f = f^*$  are proportional (and hence provide the same estimator). Kessler (2000) showed that, for all  $\Delta$ ,  $\hat{\theta}_n$  has the least asymptotic variance among estimators obtained from simple estimating equations. This also follows from results in Jacobsen (1998).  $\square$

**Example (The Cox-Ingersoll-Ross process)** Consider the solution  $X$  to

$$dX_t = (\alpha + \beta X_t) dt + \sqrt{X_t} dW_t, \quad X_0 = x_0$$

where  $\beta < 0$  and  $\alpha \geq 1/2$ . The estimating function  $F_n$  is given by

$$F_n(\alpha, \beta) = \begin{pmatrix} (\alpha - \frac{1}{2}) \sum_{i=1}^n 1/X_{(i-1)\Delta} + n\beta \\ \beta \sum_{i=1}^n X_{(i-1)\Delta} + n\alpha \end{pmatrix}.$$

To see how the estimator performs we have compared it to three other estimators in a simulation study. We have simulated 500 processes on the interval  $[0, 500]$  by the Euler scheme with time-step  $1/1000$ . The number of observations is  $n = 500$  and  $\Delta = 1$ . For each simulation we have calculated four estimators: those obtained from  $F_n$ ,  $R_n$  given by (I.5),  $H_n = \sum_{i=1}^n h^*(X_{i-1}, \cdot)$ , and the martingale estimating function suggested by Bibby & Sørensen (1995).

The estimating function  $H_n$  is given by

$$H_n(\theta) = 2 \begin{pmatrix} \sum_{i=1}^n \log X_{(i-1)\Delta} - n\Psi(2\alpha) + n \log(-2\beta) \\ \sum_{i=1}^n X_{(i-1)\Delta} + n\alpha/\beta \end{pmatrix}$$

where  $\Psi$  is the Digamma function,  $\Psi = \partial \log \Gamma / \partial x$ . Note that the second coordinates of  $F_n$  and  $H_n$  are equivalent and that  $H_n(\theta) = 0$  cannot be solved explicitly.

The empirical means and standard errors of the four estimators are listed in Table I.1. The true parameter values are  $\alpha_0 = 10$  and  $\beta_0 = -1$ .

The estimator from  $R_n$  is biased (as we knew).  $F_n$  and  $H_n$  seem to be almost equally good and are both better than the martingale estimating function.  $\square$

Finally, we consider an example where the parameter of interest enters as an exponent in the drift function.

**Example (A generalized Cox-Ingersoll-Ross model)** Let  $X$  be the solution to

$$dX_t = (\alpha + \beta X_t^\theta) dt + \sqrt{X_t} dW_t$$

(I.8)



Estimating function	$\hat{\alpha}_n$		$\hat{\beta}_n$	
	mean	s.e.	mean	s.e.
$F_n$	10.1271	0.7218	-1.0126	0.0737
$H_n$	10.1543	0.7151	-1.0154	0.0729
$R_n$	6.3691	0.4279	-0.6368	0.0430
Martingale	10.2000	1.1900	-1.0200	0.1200

**Table I.1:** Empirical means and standard errors for 500 realizations of various estimators for  $(\alpha, \beta)$  in the Cox-Ingersoll-Ross model. The number of observations is  $n = 500$  and  $\Delta = 1$ . The true value is  $(\alpha_0, \beta_0) = (10, -1)$ .

where  $\alpha \geq \frac{1}{2}$  and  $\beta < 0$  are known and  $\theta > 0$  is the unknown parameter. Note that  $X$  is a Cox-Ingersoll-Ross process if  $\theta = 1$ ; for  $\theta \neq 1$  the mean reverting force is stronger or weaker.

From (I.11) it follows that  $f^* = (\mathcal{A}_\theta h^*)/2$  is given by

$$f^*(x, \theta) = \beta x^{\theta-1} \log x \left( \alpha + \beta x^\theta + \frac{\theta}{2} - \frac{1}{2} \right) + \frac{1}{2} \beta x^{\theta-1}.$$

The estimating equation must be solved numerically. For comparison we have also considered the simple estimating function corresponding to

$$\tilde{f}(x, \theta) = x - E_\theta^\mu X_0 = x - \frac{\Gamma((2\alpha + 1)/\theta)}{\Gamma(2\alpha/\theta)} \left( -\frac{\theta}{2\beta} \right)^{1/\theta} \quad (\text{I.12})$$

As above, we have simulated 500 processes by means of the Euler scheme;  $n = 500$  and  $\Delta = 1$ . The true value of  $\theta$  is  $\theta_0 = 1.5$  and  $\alpha = 2$ ,  $\beta = -1$ . The means (standard errors) of the estimators are 1.5028 (0.0508) when using  $F_n$  and 1.5001 (0.0590) when using  $\sum \tilde{f}(X_{(i-1)\Delta}, \cdot)$ .

Both estimators are very precise. The estimator obtained from (I.12) is closer to the true value but has larger standard error than the estimator obtained from  $F_n$ .  $\square$

## I.6 Multi-dimensional processes

So far, we have only studied one-dimensional diffusion processes. In this section we discuss to what extent the ideas carry over to the multi-dimensional case.

We consider a  $d$ -dimensional stochastic differential equation

$$dX_t = b(X_t, \theta) dt + \sigma(X_t) dW_t, \quad X_0 = x_0. \quad (\text{I.13})$$

The parameter  $\theta$  is still  $p$ -dimensional,  $\theta \in \Theta \subseteq \mathbb{R}^p$ , but  $X$  and  $W$  are now  $d$ -dimensional. The functions  $b: \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d$  and  $\sigma: \mathbb{R}^d \rightarrow \mathbb{R}^{d \times d}$  are known,  $\sigma(x)$  is regular for all  $x \in \mathbb{R}^d$ , and  $x_0 \in \mathbb{R}^d$ .

Let  $\dot{b}$  be the  $d \times p$  matrix of derivatives;  $\dot{b}_{ij} = \partial b_i / \partial \theta_j$ , and let  $D_i g = \partial g / \partial x_i$  and  $D_{ij}^2 g = \partial^2 g / \partial x_i \partial x_j$  for functions  $g : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$  with  $g(\cdot, \theta)$  in  $\mathcal{C}^2(\mathbb{R}^d)$ . With this notation, the analogue of  $\mathcal{A}_\theta$  is given by

$$\mathcal{A}_\theta g(x, \theta) = \sum_{i=1}^d b_i(x, \theta) D_i g(x, \theta) + \frac{1}{2} \sum_{i,j=1}^d (\sigma(x) \sigma^T(x))_{ij} D_{ij}^2 g(x, \theta)$$

and the score process is given by

$$S_t^c(\theta) = \int_0^t \dot{b}^T(X_s, \theta) \Sigma(X_s) dX_s - \int_0^t \dot{b}^T(X_s, \theta) \Sigma(X_s) b(X_s, \theta) ds,$$

where  $\Sigma(x) = (\sigma(x) \sigma^T(x))^{-1}$ , see Lipster & Shiriyayev (1977). Using (I.13) we find  $dS_t^c(\theta) = \dot{b}^T(X_t, \theta) (\sigma^{-1})^T(X_t) dW_t$ .

Now, similarly to (I.7) we look for functions  $h_1^*, \dots, h_p^* : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$  such that for each  $k$ ,  $dh_k^*(X_t, \theta) = \mathcal{A}_\theta h_k^*(X_t, \theta) dt + 2dS_{k,t}^c(\theta)$ . Arguing as above, this leads to the equations

$$D_i h_k^*(x, \theta) = 2 \left[ \dot{b}^T(x, \theta) \Sigma(x) \right]_{ki} = 2 \sum_{r=1}^d \dot{b}_{rk}(x, \theta) \Sigma_{ir}(x), \quad i = 1, \dots, d \quad (\text{I.14})$$

and thus

$$\mathcal{A}_\theta h_k^* = 2 \sum_{i,r=1}^d \dot{b}_{rk} \Sigma_{ir} b_i + \sum_{j=1}^d \frac{\partial \dot{b}_{jk}}{\partial x_j} + \sum_{i,j,r=1}^d (\sigma \sigma^T)_{ij} \dot{b}_{rk} \frac{\partial \Sigma_{ir}}{\partial x_j} \quad (\text{I.15})$$

The equations (I.14) may, however, not any have solutions; differentiation wrt.  $x_j$  yields

$$D_{ij}^2 h_k^* = 2 \sum_{r=1}^d \left( \frac{\partial^2 b_r}{\partial \theta_k \partial x_j} \Sigma_{ir} + \frac{\partial b_r}{\partial \theta_k} \frac{\partial \Sigma_{ir}}{\partial x_j} \right),$$

but the right hand side is not necessarily symmetric wrt.  $i$  and  $j$ , see the example below.

If there are solutions, then (I.15) has expectation zero and the simple estimating function with  $f = (\mathcal{A}_\theta h_1^*, \dots, \mathcal{A}_\theta h_p^*)^T$  may be used. Otherwise, the right hand side of (I.15) is typically biased.

**Example (Homogeneous Gaussian diffusions)** Let  $B$  be a  $2 \times 2$  matrix with eigenvalues with strictly negative parts and let  $A$  be an arbitrary  $2 \times 1$  matrix. Consider the stochastic differential equation

$$dX_t = (A + BX_t) dt + \sigma dW_t, \quad X_0 = x_0$$

where  $\sigma > 0$  is known,  $W$  is a two-dimensional Brownian motion, and  $x_0 \in \mathbb{R}^2$ .

Solutions to all the equations (I.14) exist if and only if  $B$  is symmetric. Let  $\alpha_1, \alpha_2, \beta_{11}, \beta_{22}, \beta_{12}$  denote the entries of  $A$  and  $B$  and let  $S_1 = \sum X_{1,(i-1)\Delta}$ ,  $S_2 = \sum X_{2,(i-1)\Delta}$ ,  $S_{11} = \sum X_{1,(i-1)\Delta}^2$ ,  $S_{22} = \sum X_{2,(i-1)\Delta}^2$ , and  $S_{12} = \sum X_{1,(i-1)\Delta} X_{2,(i-1)\Delta}$ ; all sums are from 1 to  $n$ . Then the estimating equation is given by

$$\frac{1}{\sigma^2} \begin{pmatrix} n & 0 & S_1 & 0 & S_2 \\ 0 & n & 0 & S_2 & S_1 \\ S_1 & 0 & S_{11} & 0 & S_{12} \\ 0 & S_2 & 0 & S_{22} & S_{12} \\ S_2 & S_1 & S_{12} & S_{12} & S_{11} + S_{22} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \beta_{11} \\ \beta_{22} \\ \beta_{12} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -n/2 \\ -n/2 \\ 0 \end{pmatrix}.$$

We have simulated 500 processes (by exact simulation), each of a length of 500 with  $\Delta = 1$  and  $\sigma = \sqrt{2}$ . The true matrices are

$$A_0 = \begin{pmatrix} 4 \\ 1 \end{pmatrix} \quad \text{and} \quad B_0 = \begin{pmatrix} -2 & 1 \\ 1 & -3 \end{pmatrix}.$$

The means and the standard errors (to the right) are

$$\hat{A}_n = \begin{pmatrix} 4.0349 \\ 1.0035 \end{pmatrix} \quad \begin{matrix} 0.2904 \\ 0.2891 \end{matrix}$$

and

$$\hat{B}_n = \begin{pmatrix} -2.0155 & 1.0078 \\ 1.0078 & -3.0247 \end{pmatrix} \quad \begin{matrix} 0.1248 & 0.1177 \\ 0.1177 & 0.1978 \end{matrix}$$

The estimators are satisfactory. □

**Acknowledgements** I am grateful to my supervisor Martin Jacobsen for valuable discussions and suggestions, and to Jens Lund, Martin Richter and the referees for comments on earlier versions of the manuscript.

---

# Bibliography

---

- Aït-Sahalia, Y. (1996), 'Nonparametric pricing of interest rate derivative securities', *Econometrica* **64**, 527–560.
- Aït-Sahalia, Y. (1998), Maximum likelihood estimation of discretely sampled diffusions: a closed-form approach, Revised version of Working Paper 467, Graduate School of Business, University of Chicago.
- Andersen, T. G. & Lund, J. (1997), 'Estimating continuous-time stochastic volatility models of the short-term interest rate', *Journal of Econometrics* **77**, 343–377.
- Andersen, T. G. & Sørensen, B. E. (1996), 'GMM estimation of a stochastic volatility model: A Monte Carlo study', *Journal of Business and Economic Statistics* **14**, 328–352.
- Arcones, M. A. & Yu, B. (1994), 'Central limit theorems for empirical and  $U$ -processes of stationary mixing sequences', *Journal of Theoretical Probability* **7**, 47–71.
- Banon, G. (1978), 'Nonparametric identification for diffusion processes', *Siam J. Control and Optimization* **16**, 380–395.
- Barndorff-Nielsen, O. E., Jensen, J. L. & Sørensen, M. (1998), 'Some stationary processes in discrete and continuous time', *Advances in Applied Probability* **30**, 989–1007.
- Barndorff-Nielsen, O. E. & Shephard, N. (1999), Non-Gaussian OU based models and some of their uses in financial economics, Working Paper 37, Centre for Analytical Finance, University of Aarhus.
- Bera, A. K. & Higgins, M. L. (1993), 'ARCH models: properties, estimation and testing', *Journal of Economic Surveys* **7**, 305–366.
- 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.

- Bibby, B. M. & Sørensen, M. (1998), Simplified estimating functions for diffusion models with a high-dimensional parameter, Preprint 1998-10, Department of Theoretical Statistics, University of Copenhagen. To appear in *Scandinavian Journal of Statistics*.
- Bickel, P. J. & Ritov, Y. (1996), 'Inference in hidden Markov models I: Local asymptotic normality in the stationary case', *Bernoulli* **2**, 199–228.
- Bickel, P. J., Ritov, Y. & Rydén, T. (1998), 'Asymptotic normality of the maximum-likelihood estimator for general hidden Markov models', *Annals of Statistics* **26**, 1614–1635.
- Billingsley, P. (1961), 'The Lindeberg-Levy Theorem for martingales', *Proceedings of the American Mathematical Society* **12**, 788–792.
- Billingsley, P. (1968), *Convergence of Probability Measures*, Wiley.
- Black, F. & Scholes, M. (1973), 'The pricing of options and corporate liabilities', *Journal of Political Economy* **81**, 637–654.
- Bradley, R. C. (1986), Basic properties of strong mixing conditions, in E. Eberlein & M. S. Taqqu, eds, 'Dependence in Probability and Statistics: A Survey of Recent Results', Birkhäuser, Boston, pp. 165–192.
- Le Breton, A. (1974), Parameter estimation in a linear stochastic differential equation, in 'Transactions of the Seventh Prague Conference and of the European Meeting of Statisticians', pp. 353–366.
- Le Breton, A. (1976), 'On continuous and discrete sampling for parameter estimation in diffusion type processes', *Mathematical Programming Study* **5**, 124–144.
- 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.
- Chesney, M. & Scott, L. (1989), 'Pricing European currency options: a comparison of the modified Black-Scholes model and a random variance model', *Journal of Financial and Quantitative Analysis* **24**, 267–284.
- Dacunha-Castelle, D. & Duflo, M. (1986), *Probability and Statistics*, Vol. 2, Springer-Verlag, New York.
- Dacunha-Castelle, D. & Florens-Zmirou, D. (1986), 'Estimation of the coefficients of a diffusion from discrete observations', *Stochastics* **19**, 263–284.
- Danielsson, J. (1994), 'Stochastic volatility in asset prices: estimation with simulated maximum likelihood', *Journal of Econometrics* **64**, 375–400.

- Danielsson, J. & Richard, J.-F. (1993), 'Accelerated Gaussian importance sampler with application to dynamic latent variable models', *Journal of Applied Econometrics* **8**, S153–S173.
- Doukhan, P. (1994), *Mixing: Properties and Examples*, Lecture Notes in Statistics **85**, Springer-Verlag, New York.
- Elerian, O., Chib, S. & Shephard, N. (2000), Likelihood inference for discretely observed non-linear diffusions, Economics discussion paper 146, Nuffield College, Oxford. To appear in *Econometrica*.
- Eraker, B. (1998), MCMC analysis of diffusion models with application to finance, Discussion paper 1998-5, Department of Finance and Management Science, Norwegian School of Economics and Business Administration.
- Florens-Zmirou, D. (1989), 'Approximate discrete-time schemes for statistics of diffusion processes', *Statistics* **20**, 547–557.
- Florens-Zmirou, D. (1993), 'On estimating the diffusion coefficient from discrete observations', *Journal of Applied Probability* **30**, 790–804.
- Gallant, A. R., Hsieh, D. & Tauchen, G. (1997), 'Estimation of stochastic volatility models with diagnostics', *Journal of Econometrics* **81**, 159–192.
- Gallant, A. R. & Long, J. R. (1997), 'Estimating stochastic differential equations efficiently by minimum chi-squared', *Biometrika* **84**, 125–141.
- Gallant, A. R. & Tauchen, G. (1996), 'Which moments to match?', *Econometric Theory* **12**, 657–681.
- Genon-Catalot, V., Jeantheau, T. & Laredo, C. (1998a), 'Limit theorems for discretely observed stochastic volatility models', *Bernoulli* **4**, 283–303.
- Genon-Catalot, V., Jeantheau, T. & Laredo, C. (1998b), Stochastic volatility models as hidden Markov models and statistical applications, Preprint 1998-22, Equipe d'Analyse et de Mathématiques Appliquées, Université de Marne-la-Vallée.
- Genon-Catalot, V., Jeantheau, T. & Laredo, C. (1999), 'Parameter estimation for discretely observed stochastic volatility models', *Bernoulli* **5**, 855–872.
- Ghysels, E., Harvey, A. C. & Renault, E. (1996), Stochastic volatility, in G. S. Madala & C. R. Rao, eds, 'Statistical Methods in Finance', Vol. 14 of *Handbook of Statistics*, North-Holland, Amsterdam, pp. 119–191.
- Gourieroux, C., Monfort, A. & Renault, E. (1993), 'Indirect inference', *Journal of Applied Econometrics* **8**, S85–S118.
- Hall, P. & Heyde, C. C. (1980), *Martingale Limit Theory and its Application*, 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.
- Harvey, A., Ruiz, E. & Shephard, N. (1994), 'Multivariate stochastic variance models', *Review of Economic Studies* **61**, 247–264.
- Heston, S. L. (1993), 'A closed-form solution for options with stochastic volatility with applications to bond and currency options', *The Review of Financial Studies* **6**, 327–343.
- Heyde, C. C. (1997), *Quasi-Likelihood and its Application: A General Approach to Optimal Parameter Estimation*, Springer-Verlag, New York.
- Honoré, P. (1997), Maximum likelihood estimation of non-linear continuous-time term-structure models, Working paper 1997-7, Department of Finance, Aarhus School of Business.
- Hull, J. & White, A. (1987), 'The pricing of options on assets with stochastic volatilities', *The Journal of Finance* **42**, 281–300.
- Hull, J. & White, A. (1988), 'An analysis of the bias in option pricing caused by a stochastic volatility', *Advances in Futures and Options Research* **3**, 29–61.
- Jacobsen, M. (1998), Discretely observed diffusions: classes of estimating functions and small  $\Delta$ -optimality, Preprint 1998-11, Department of Theoretical Statistics, University of Copenhagen. To appear in *Scandinavian Journal of Statistics*.
- Jacquier, E., Polson, N. G. & Rossi, P. E. (1994), 'Bayesian analysis of stochastic volatility models (with discussion)', *Journal of Business and Economic Statistics* **12**, 371–417.
- Jensen, J. L. & Petersen, N. V. (1999), 'Asymptotic normality of the maximum likelihood estimator in state space models', *Annals of Statistics* **27**, 514–535.
- Jiang, G. J. & Knight, J. L. (1997), 'A nonparametric approach to the estimation of diffusion processes, with an application to a short-term interest rate model', *Econometric Theory* **13**, 615–645.
- Karatzas, I. & Shreve, S. E. (1991), *Brownian Motion and Stochastic Calculus*, 2nd edn, Springer-Verlag, New York.
- Karlin, S. & Taylor, H. M. (1981), *A Second Course in Stochastic Processes*, Academic Press, New York.

- Kessler, M. (2000), 'Simple and explicit estimating functions for a discretely observed diffusion process', *Scandinavian Journal of Statistics* **27**, 65–82.
- Kessler, M. & Paredes, S. (1999), Computational aspects related to martingale estimating functions for a discretely observed diffusion, Manuscript, Departamento de Matemática Aplicada y Estadística, Universidad Politécnica de Cartagena.
- Kessler, M. & Sørensen, M. (1999), 'Estimating equations based on eigenfunctions for a discretely observed diffusion process', *Bernoulli* **5**, 299–314.
- Kim, S., Shephard, N. & Chib, S. (1998), 'Stochastic volatility: likelihood inference and comparison with ARCH models', *Review of Economic Studies* **65**, 361–393.
- Lipster, R. S. & Shiriyayev, A. N. (1977), *Statistics of Random Processes*, Vol. 1, Springer-Verlag, New York.
- Lo, A. W. (1988), 'Maximum likelihood estimation of generalized Itô processes with discretely sampled data', *Econometric Theory* **4**, 231–247.
- Nelson, D. B. (1990), 'ARCH models as diffusion approximations', *Journal of Econometrics* **45**, 7–38.
- Nelson, D. B. (1992), 'Filtering and forecasting with misspecified ARCH models I', *Journal of Econometrics* **52**, 61–90.
- Nielsen, J. N., Vestergaard, M. & Madsen, H. (2000), Estimation in continuous-time stochastic volatility models using nonlinear filters. To appear in *International Journal of Theoretical and Applied Finance*.
- Overbeck, L. & Rydén, T. (1997), 'Estimation in the Cox-Ingersoll-Ross model', *Econometric Theory* **13**, 430–461.
- Pedersen, A. R. (1995a), 'Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes', *Bernoulli* **1**(3), 257–279.
- Pedersen, A. R. (1995b), 'A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations', *Scandinavian Journal of Statistics* **22**, 55–71.
- Pollard, D. (1984), *Convergence of Stochastic Processes*, Springer-Verlag, New York.
- Poulsen, R. (1999), Approximate maximum likelihood estimation of discretely observed diffusion processes, Working paper 29, Centre for Analytical Finance, Aarhus.
- Rogers, L. C. G. & Williams, D. (1987), *Diffusions, Markov Processes, and Martingales*, Vol. 2: Itô Calculus, Wiley.



- Rydén, T. (1994), 'Consistent and asymptotically normal parameter estimates for hidden Markov models', *Annals of Statistics* **22**, 1884–1895.
- Scott, L. O. (1987), 'Option pricing when the variance changes randomly: theory, estimation and an application', *Journal of Financial and Quantitative analysis* **22**, 419–438.
- Shephard, N. (1996), Statistical aspects of ARCH and stochastic volatility, in D. R. Cox, D. V. Hinkley & O. E. Barndorff-Nielsen, eds, 'Time Series Models in Econometrics, Finance and Other Fields', Chapman & Hall, London, pp. 1–67.
- Sørensen, H. (1998a), Approximation of the score functions for diffusion processes, Preprint 1998-8, Department of Theoretical Statistics, University of Copenhagen.
- Sørensen, H. (2000), Estimation of diffusion parameters for discretely observed diffusion processes, Preprint 2000-1, Department of Theoretical Statistics, University of Copenhagen.
- Sørensen, M. (1997), Estimating functions for discretely observed diffusions: A review, in I. V. Basawa, V. P. Godambe & R. L. Taylor, eds, 'Selected Proceedings of the Symposium on Estimating Functions', Vol. 32, IMS Lecture notes, pp. 305–325.
- Sørensen, M. (1998b), On asymptotics of estimating functions, Preprint 1998-6, Department of Theoretical Statistics, University of Copenhagen. To appear in *Brazilian Journal of Probability and Statistics*.
- Sørensen, M. (1999), Prediction-based estimating functions, Preprint 1999-5, Department of Theoretical Statistics, University of Copenhagen.
- Stein, E. M. & Stein, J. C. (1991), 'Stock price distributions with stochastic volatility: an analytic approach', *The Review of Financial Studies* **4**, 727–752.
- van der Vaart, A. W. & Wellner, J. A. (1996), *Weak Convergence and Empirical Processes*, Springer-Verlag, New York.
- Wiggins, J. B. (1987), 'Options values under stochastic volatility', *Journal of Financial Economics* **19**, 351–372.