Small $\Delta-$optimal martingale estimating functions for discretely observed diffusions: a simulation study

Martin Jacobsen*
Department of Theoretical Statistics and Operations Research
University of Copenhagen

March 20, 2001

Abstract

The problem of estimating the parameters of a discretely observed diffusion is discussed through two one-dimensional examples. Based on simulations, the parameters are estimated using small $\Delta-$optimal and other unbiased estimating functions. Small $\Delta-$optimality implies that the estimation is nearly efficient when the discrete observations are close together in time, and this effect is clearly visible from the simulations. It is also seen that the small $\Delta-$optimal estimating functions perform well, even when the observations are not close together, and that they are quite robust when the true observations are replaced by rounded ones.

KEY WORDS AND PHRASES: unbiased estimating functions based on conditional expectations, asymptotic efficiency, a generalised Cox-Ingersoll-Ross process, the Pedersen-Bibby-Sørensen-Kessler example, estimation using rounded observations.

*MaPhySto – Centre for Mathematical Physics and Stochastics, funded by a grant from the Danish National Research Foundation.

1. Introduction and background

The concept of small $\Delta-$optimal estimating functions was introduced by Jacobsen [2] and discussed further, Jacobsen [3], for the case of martingale estimating functions based on conditional expectations, see (1.2) below.

In this paper we study the behaviour of the martingale estimating functions through a simulation study based on two examples of one-dimensional diffusion models,

$$dX_t = b_\theta\left(X_t\right)\, dt + \sigma_\theta\left(X_t\right)\, dB_t \tag{1.1}$$

(with B denoting standard one-dimensional Brownian motion) where both the drift b_θ and the squared diffusion coefficient σ_θ^2 depend on an unknown parameter $\theta \in \Theta \subset \mathbb{R}^p$. It is assumed that X moves everywhere on a given open interval $I \subset \mathbb{R}$ (not depending on θ) with b_θ and $\sigma_\theta^2 > 0$ both continuous on I, and it is also assumed that there is for every $\theta \in \Theta$ an invariant density μ_θ, i.e. if X_0 is independent of B and has density μ_θ, then the solution X to (1.1) is strictly stationary. As is well known, μ_θ is then proportional to

$$\mu_\theta(x) \propto \frac{1}{\sigma_\theta^2(x) S_\theta'(x)}$$

with S_θ' the derivative of a scale function,

$$S_\theta'(x) = \exp\left(-2\int_{x_0}^{x} \frac{b_\theta(y)}{\sigma_\theta^2(y)}\, dy\right).$$

We consider the situation where X is observed at discrete points in time, equidistant, $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$ and then wish to estimate θ. For a given Δ the estimation is based on an *unbiased* estimating function of the form

$$(\theta, x, y) \mapsto g_{\Delta,\theta}\left(x, y\right)$$

from $\Theta \times I \times I$ to $\mathbb{R}^{p\times 1}$, i.e. one finds the estimator $\widehat{\theta}_n$ by solving the estimating equation

$$\sum_{m=1}^{n} g_{\Delta,\theta}\left(X_{(m-1)\Delta}, X_{m\Delta}\right) = 0$$

for θ, where $g_{\Delta,\theta}$ satisfies the unbiasedness condition

$$E_\theta^\mu g_{\Delta,\theta}\left(X_0, X_\Delta\right) = 0$$

for all θ, and for identification purposes

$$E_{\theta}^{\mu} g_{\Delta,\theta'}\left(X_0, X_\Delta\right) \neq 0$$

if $\theta \neq \theta'$. Here E_θ^μ signifies that X solves (1.1) for the prescribed value of θ with X_0 having the invariant density $\mu = \mu_\theta$.

The *martingale estimating functions* based on conditional expectations were introduced by Bibby and Sørensen [1] and in general have the form, for the k'th coordinate of $g_{\Delta,\theta}$,

$$g_{\Delta,\theta}^{k}\left(x, y\right) = \sum_{q=1}^{r} h_{\Delta,\theta}^{qk}(x)\left(f_\theta^q(y) - \pi_{\Delta,\theta} f^q(x)\right) \tag{1.2}$$

with

$$\pi_{\Delta,\theta} f^q(x) = E_\theta^\mu\left(f_\theta^q\left(X_\Delta\right)|X_0 = x\right). \tag{1.3}$$

In (1.2), r is the *order* of the martingale estimating function, $(f_\theta)_{1\leq q\leq r}$ denotes the *base* and $\left(h_{\Delta,\theta}^{qk}\right)_{1\leq q\leq r, 1\leq k\leq p}$ the *weights* determining $g_{\Delta,\theta}$. As stressed in Jacobsen [3], Assumption A, the members of the base must be affinely independent, i.e. if

$$\sum_{q=1}^{r} a_\theta^q f_\theta^q(x) + \alpha_\theta = 0$$

for all $x \in I$ and some constants a_θ^q and α_θ, then $a_\theta^1 = \cdots = a_\theta^r = \alpha_\theta = 0$, and also the weights must satisfy that the p $r-$variate functions $x \mapsto \left(h_{\Delta,\theta}^{1k}(x), \ldots, h_{\Delta,\theta}^{rk}(x)\right)$ forming the columns of the $r \times p-$matrix valued function $h_{\Delta,\theta}$, are linearly independent on I.

In (1.2) it is tacitly assumed that the base does not depend on Δ, while the weights may. There is nothing wrong in allowing f_θ^q to depend on Δ, but for practical purposes it does not seem relevant: in practice (1.2) is used for a given base for which explicit expressions for the conditional expectations (1.3) are available for all Δ.

With the base given, the problem arises of how to choose the weights $h_{\Delta,\theta}$. This may be done by minimizing the asymptotic covariance for the estimator $\widehat{\theta}_n$: under fairly mild regularity conditions it holds that $\sqrt{n}\left(\widehat{\theta}_n - \theta\right) \overset{\mathcal{D}_\theta}{\to} N\left(0, \Sigma_{\Delta,\theta}\right)$ as $n \to \infty$, θ denoting the true parameter value, and here it is actually possible to minimize the asymptotic covariance $\Sigma_{\Delta,\theta}$ when the weights vary (using the standard partial ordering on the space of covariance matrices: $A \prec B$ if $B - A$ is a covariance matrix) by choosing $h_{\Delta,\theta}$ as

$$h_{\Delta,\theta}^{\mathrm{opt}} = \left(\Pi_{\Delta,\theta} f_\theta\right)^{-1}\left(\partial_\theta\left(\pi_{\Delta,\theta} f_\theta\right) - \pi_{\Delta,\theta}\left(\dot{f}_\theta\right)\right), \tag{1.4}$$

see Bibby and Sørensen [1] for special cases and Jacobsen [3], Proposition 1, for the general result. The notation used here is

$$\begin{aligned}
\Pi_{\Delta,\theta} f_\theta &= \pi_{\Delta,\theta}\left(f_\theta^{\otimes 2}\right) - \left(\pi_{\Delta,\theta} f_\theta\right)^{\otimes 2} \in \mathbb{R}^{r\times r}, \\
\partial_\theta\left(\pi_{\Delta,\theta} f_\theta\right) &= \left(\partial_{\theta_\ell}\pi_{\Delta,\theta} f_\theta^q\right)_{1\le q\le r, 1\le \ell\le p} \in \mathbb{R}^{r\times p}, \\
\pi_{\Delta,\theta}\left(\dot{f}_\theta\right) &= \left(\pi_{\Delta,\theta}\left(\partial_{\theta_\ell} f_\theta^q\right)\right)_{1\le q\le r, 1\le \ell\le p} \in \mathbb{R}^{r\times p}.
\end{aligned}$$

Thus, to find the explicit expressions for the optimal weights – that of course depend on Δ – one must know the conditional expectations

$$\pi_{\Delta,\theta}\left(f_\theta^q f_\theta^{q'}\right) \quad \text{and} \quad \pi_{\Delta,\theta}\left(\partial_{\theta_\ell} f_\theta^q\right)$$

for all q, q', ℓ and must then invert the $r\times r-$matrix $\Pi_{\Delta,\theta} f_\theta$ analytically.

For the examples here we shall not use $h_{\Delta,\theta}^{\text{opt}}$, but instead find weights that are *small* $\Delta-$*optimal* as described in Jacobsen [3]. For this it is sufficient to look at weights that do not depend on Δ and consider for arbitrary $t>0$,

$$g_{t,\theta}^k(x,y) = \sum_{q=1}^r h_\theta^{qk}(x)\left(f_\theta^q(y) - \pi_{t,\theta} f^q(x)\right)$$

which defines a *flow* of unbiased martingale estimating functions, with $g_{t,\theta}$ to be used for data $(X_{mt})_{0\le m\le n}$. The small $\Delta-$optimality property is expressed exclusively through the limit, for $1\le k\le p$,

$$\begin{aligned}
g_{0,\theta}^k(x,y) &= \lim_{t\to 0} g_{t,\theta}(x,y) \\
&= \sum_{q=1}^r h_\theta^{qk}(x)\left(f_\theta^q(y) - f^q(x)\right),
\end{aligned} \tag{1.5}$$

viz. one should choose (for the one-dimensional diffusions considered here) a minimal value 1 or 2 for the order r and then find the weights $h_\theta\in\mathbb{R}^{r\times p}$ as specified in the following three cases (Jacobsen [3], Theorem 2 and Remark 2):

(i) if $\sigma_\theta^2 = \sigma^2$ does not depend on θ, use $r=1$ and

$$h_\theta^T(x) = \frac{\dot{b}_\theta^T(x)}{\sigma^2(x)}\frac{1}{\partial_x f_\theta(x)}.$$

(ii) if σ_θ^2 depends on all the parameters $\theta_1,\ldots,\theta_p$, use $r=2$ and

$$h_\theta^T(x) = \begin{pmatrix} 0_{p\times 1} & \dfrac{\left(\dot{\sigma}_\theta^2(x)\right)^T}{\sigma_\theta^4(x)}\end{pmatrix}\begin{pmatrix}\partial_x f_\theta(x) & \partial_{xx}^2 f_\theta(x)\end{pmatrix}^{-1}. \tag{1.6}$$

(iii) if (possibly after a reparametrization) σ_θ^2 depends on the parameters $\theta_1, \ldots, \theta_{p'}$ but not on $\theta_{p'+1}, \ldots, \theta_p$ (where $1 \leq p' < p$), use $r = 2$ and $h_\theta = (h_{1,\theta} \quad h_{2,\theta})$ with $h_{1,\theta} \in \mathbb{R}^{2\times p'}$ and $h_{2,\theta} \in \mathbb{R}^{2\times(p-p')}$ given by

$$h_{1,\theta}^T(x) = \begin{pmatrix} 0_{p'\times 1} & \frac{\left(\dot\sigma_{1,\theta}^2(x)\right)^T}{\sigma_\theta^4(x)} \end{pmatrix} \begin{pmatrix} \partial_x f_\theta(x) & \partial_{xx}^2 f_\theta(x) \end{pmatrix}^{-1}, \tag{1.7}$$

$$h_{2,\theta}^T(x) = \begin{pmatrix} \frac{\dot b_{2,\theta}^T(x)}{\sigma_\theta^2(x)} \frac{1}{\partial_x f_\theta^1(x)} & 0_{(p-p')\times 1} \end{pmatrix}. \tag{1.8}$$

Notation. $\dot b_\theta(x)$, $\dot\sigma_\theta^2(x)$ are the $p-$dimensional row vectors $(\partial_{\theta_\ell} b_\theta(x))$ and $(\partial_{\theta_\ell}(\sigma_\theta^2(x)))$ for $1 \leq \ell \leq p$. In case (iii), $\dot\sigma_{1,\theta}^2(x)$ is the $p'-$dimensional row vector obtained by differentiating $\sigma_\theta^2(x)$ after θ_ℓ for $1 \leq \ell \leq p'$, and similarly $\dot b_{2,\theta}(x)$ is the $(p-p')-$dimensional row vector derived by differentiating $b_\theta(x)$ with respect to $\theta_{p'+1}, \ldots, \theta_p$. f_θ^1 is of course just the first element of the base – one could as well have used f_θ^2 and obtained a different small $\Delta-$optimal estimating function.

Remark 1. *It is a main result from Jacobsen [3] that the flow of optimal martingale estimating functions determined by a given base and with weights as in (1.4) (with Δ replaced by t) is small $\Delta-$optimal essentially only if the order of the base is as in the relevant of the three cases (i), (ii) and (iii). In subsection 2.2 below we shall study the behaviour of an optimal flow of order 1, that is not small $\Delta-$optimal.*

The reader is reminded that (subject to a host of integrability conditions, blithely ignored in the present paper) what is achieved by small $\Delta-$optimality is the following: for any $t > 0$, the estimator $\widehat\theta_n(t)$ obtained from observations $(X_0, X_t, \ldots, X_{nt})$ using the martingale estimating function $g_{t,\theta}$ satisfies that, as $n \to \infty$,

$$\sqrt{n}\left(\widehat\theta_n(t) - \theta\right) \xrightarrow{\mathcal{D}_\theta} N\left(0, \Sigma_{t,\theta}\right)$$

where the asymptotic covariance matrix is given by

$$\Sigma_{t,\theta} = \Lambda_{t,\theta}^{-1}(g) E_\theta^\mu g_{t,\theta} g_{t,\theta}^T\left(X_0, X_t\right)\left(\Lambda_{t,\theta}^{-1}(g)\right)^T$$

where, writing $\left(\dot g_{t,\theta}\right)_{k\ell} = \partial_{\theta_\ell} g_{t,\theta}^k$

$$\Lambda_{t,\theta}(g) = E_\theta^\mu \dot g_{t,\theta}\left(X_0, X_t\right),$$

cf. Jacobsen [2] or [3]. Furthermore, as $t \to 0$, $\Sigma_{t,\theta}$ admits a series expansion

$$\Sigma_{t,\theta} = t^{-1} v_{-1,\theta}(g) + v_{0,\theta}(g) + o(1) \tag{1.9}$$

with coefficient matrices $v_{-1,\theta}(g)$ and $v_{0,\theta}(g)$ that can be expressed in terms of the limiting estimating function $g_{0,\theta}$, see (1.5). If the order of the base and the weights are chosen according to (i)-(iii), then

(i) $v_{-1,\theta}(g)$ is minimised globally;

(ii) $v_{-1,\theta}(g) = 0$ and $v_{0,\theta}(g)$ is minimised globally;

(iii) in block matrix notation

$$v_{-1,\theta}(g) = \begin{pmatrix} 0_{p' \times p'} & 0_{p' \times (p-p')} \\ 0_{(p-p') \times p'} & v_{22,-1,\theta}(g) \end{pmatrix}$$

with $v_{22,-1,\theta}(g)$ minimised globally, and furthermore the upper $p' \times p'-$left block of $v_{0,\theta}(g)$ is minimised globally.

The phrase 'minimised globally' used above means that the coefficient in question is the smallest possible (in the partial ordering of covariance matrices mentioned above) when *all* unbiased estimating functions are considered. Thus, for small Δ, the estimator obtained from a small $\Delta-$optimal estimating function is (almost) as good as the maximum-likelihood estimator (which comes from the martingale estimating function flow given by $g^k_{t,\theta}(x, y) = \partial_{\theta k} \log p_{t,\theta}(x, y)$ with $p_{t,\theta}$ the transition density $P_\theta (X_t \in dy \,| X_0 = x) = p_{t,\theta}(x, y) \, dy$).

The reader is reminded that the possibility of obtaining $v_{-1,\theta}(g) = 0$ (or parts of $v_{-1,\theta}(g) = 0$) in cases (ii) or (iii), stems from the singularity between the distributions of $(X_s)_{0\le s\le t}$ when observed completely in continuous time, whenever σ^2_θ depends on all or some of the parameters: for $\Delta > 0$ small we are close to observation in continuous time and the parameters appearing in σ^2_θ can (almost) be read off from the observations, e.g. through the quadratic variation for X. For general unbiased estimating functions, the term $v_{-1,\theta}(g)$ may well be present unless one is careful and hence lead to estimators of efficiency close to 0 if Δ is small!

Since small $\Delta-$optimality refers to small values of Δ there is of course no guarantee that a small $\Delta-$optimal flow of estimating functions will behave well for moderate or large values of Δ – also there are lots of small $\Delta-$optimal flows, and they may behave quite differently for large values of Δ. If however, as is the case with the martingale estimating functions we consider here, the flow $(g_{t,\theta})_{t>0}$ depends in a natural fashion on t, we nevertheless claim that if the flow is small $\Delta-$optimal, then the resulting estimators perform rather well at least for moderate values of Δ, a claim we shall substantiate by the simulation studies in the next section.

2. Simulations

2.1. The generalised Cox-Ingersoll-Ross process

The first example we consider is the four parameter model

$$dX_t = \left(aX_t^{2\gamma-1} + bX_t\right)\, dt + \sigma X_t^{\gamma}\, dB_t \tag{2.1}$$

of one-dimensional, strictly positive diffusions (so $I =]0, \infty[$ with some constraints on the parameters to ensure positivity) proposed by Jacobsen [3]. The model arises by considering powers of the ordinary Cox-Ingersoll-Ross (CIR) process,

$$d\widetilde{X}_t = \left(\widetilde{a} + \widetilde{b}\widetilde{X}_t\right)\, dt + \widetilde{\sigma}\widetilde{X}_t^{1/2}\, dB_t, \tag{2.2}$$

i.e., if X solves (2.1), then $\widetilde{X} = X^{2-2\gamma}$ solves (2.2) with

$$\widetilde{b} = (2-2\gamma)\, b, \quad \widetilde{\sigma}^2 = (2-2\gamma)^2\, \sigma^2, \quad \widetilde{a} - \tfrac{1}{2}\widetilde{\sigma}^2 = (2-2\gamma)\left(a - \tfrac{1}{2}\sigma^2\right). \tag{2.3}$$

The parameter set Θ consists of those $\theta = (a, b, \gamma, \sigma^2)$ for which X has an invariant probability, i.e.

$$\Theta = \left\{ \left(a, b, \gamma, \sigma^2\right) : \sigma^2 > 0 \text{ and either } \begin{array}{l} \gamma < 1,\ b < 0,\ 2a \geq \sigma^2 \\ \text{or} \quad \gamma > 1,\ b > 0,\ 2a \leq \sigma^2 \end{array} \right\}.$$

Note that $\gamma = 1$ is not included: for $\gamma = 1$, (2.1) describes a geometric Brownian motion which is never ergodic, and also, (2.1) for $\gamma = 1$ can never be obtained by considering $\widetilde{X}^p$ for some p, where $\widetilde{X}$ is a CIR-process as in (2.2).

In the simulations we either (i) estimate all four parameters, or (ii) estimate γ and σ^2 only, assuming a, b to be known. ((ii) is used to illustrate the effect of small $\Delta-$optimality, as well as to see how the estimates are affected when using rounded observations, for the latter see subsection 2.3 below). We use martingale estimating functions with the base of order 2 given by

$$f_\theta^1(x) = x^{2-2\gamma}, \quad f_\theta^2(x) = x^{4-4\gamma},$$

i.e. the base corresponding to considering the first and second order conditional moments for the CIR-process $\widetilde{X}$, which are of course well known. The explicit values of $\pi_{t,\theta} f_\theta(x)$ are, writing $\pi_{t,\theta} x^q$ for $\pi_{t,\theta}\phi(x)$ when $\phi(y) = y^q$,

$$\pi_{t,\theta} x^{2-2\gamma} = \widetilde{\xi}_1 + e^{\widetilde{b}t}\left(x^{2-2\gamma} - \widetilde{\xi}_1\right), \tag{2.4}$$

$$\pi_{t,\theta} x^{4-4\gamma} = \widetilde{\xi}_2 + \frac{2\widetilde{\xi}_2}{\widetilde{\xi}_1}\left(e^{\widetilde{b}t} - e^{2\widetilde{b}t}\right)\left(x^{2-2\gamma} - \widetilde{\xi}_1\right) + e^{2\widetilde{b}t}\left(x^{4-4\gamma} - \widetilde{\xi}_2\right) \tag{2.5}$$

where for $m=1,2$, $\widetilde{\xi}_m = E^{\mu}_{\theta} X_0^{(2-2\gamma)m}$ is the m'th moment for the CIR-process $\widetilde{X}$ when started with its invariant distribution, the $\Gamma-$distribution with shape parameter $2\widetilde{a}/\widetilde{\sigma}^2$ and scale parameter $-\widetilde{\sigma}^2/2\widetilde{b}$, so that

$$\widetilde{\xi}_1 = -\frac{\widetilde{a}}{\widetilde{b}}, \quad \widetilde{\xi}_2 = \frac{\widetilde{a}\left(2\widetilde{a}+\widetilde{\sigma}^2\right)}{2\widetilde{b}^2}. \tag{2.6}$$

For estimating all four parameters we use the small $\Delta-$optimal martingale estimating function with base $\left(f^q_\theta\right)_{q=1,2}$ as described by case (iii) above, see (1.7), (1.8). We have $p=4$, $p'=2$ and writing the parameters in the order (γ,σ^2,a,b) find

$$\begin{aligned} \dot{\sigma}^2_{1,\theta}(x) &= \begin{pmatrix} 2\sigma^2 x^{2\gamma}\log x & x^{2\gamma} \end{pmatrix}, \\ \dot{b}_{2,\theta}(x) &= \begin{pmatrix} x^{2\gamma-1} & x \end{pmatrix} \end{aligned}$$

and

$$\begin{pmatrix} \partial_x f_\theta(x) & \partial^2_{xx} f_\theta(x) \end{pmatrix} = (2-2\gamma)\begin{pmatrix} x^{1-2\gamma} & (1-2\gamma)\,x^{-2\gamma} \\ 2x^{3-4\gamma} & 2\,(3-4\gamma)\,x^{2-4\gamma} \end{pmatrix}.$$

Listing the parameters in the order γ,σ^2,a,b and finding the weights $h_\theta(x)$ from (1.7) and (1.8) and ignoring unimportant proportionalities, this leads to the small $\Delta-$optimal flow of estimating functions $g_{t,\theta}(x,y) = h^T_\theta(x)\left(f_\theta(y) - \pi_{t,\theta} f_\theta(x)\right)$ given by

$$g_{t,\theta}(x,y) = \begin{pmatrix} -2\log x & x^{2\gamma-2}\log x \\ -2 & x^{2\gamma-2} \\ x^{2\gamma-2} & 0 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} y^{2-2\gamma} - \pi_{t,\theta} x^{2-2\gamma} \\ y^{4-4\gamma} - \pi_{t,\theta} x^{4-4\gamma} \end{pmatrix} \tag{2.7}$$

with $\pi_{t,\theta} x^{(2-2\gamma)m}$ for $m=1,2$ given by (2.4), (2.5) and where explicit expressions in terms of θ of all the quantities appearing are provided by (2.6) and (2.3).

Because one-dimensional diffusions are reversible, an improved flow of unbiased estimating functions is obtained by considering the symmetrized flow

$$\overline{g}_{t,\theta}(x,y) = \tfrac{1}{2}\left(g_{t,\theta}(x,y) + g_{t,\theta}(y,x)\right),$$

see Jacobsen [3], Proposition 6.1. For the case at hand though, the gain achieved by symmetrisation appears negligeable.

Remark 2. *For the model (2.1), in Jacobsen [3] a different small $\Delta-$optimal flow was proposed as given by Theorem 2 there. The flow (2.7) above corresponds to the flow in [3], Remark 2 with the two bases f_θ and $\widetilde{f}_\theta$ there ($\widetilde{f}_\theta$ of order 1) given by the f_θ used here and $\widetilde{f}_\theta = f^1_\theta$ respectively.*

Remark 3. *An improvement on the estimates resulting from the estimating function (2.7) would of course have been obtained by using the optimal flow (1.4) with base f_θ. However, finding $h_{t,\theta}^{opt}$ requires explicit expressions for the conditional moments $\pi_{t,\theta}x^{(2-2\gamma)m}$ for $m = 3, 4$ in addition to the formulas for $m = 1, 2$ already used, in order to determine the matrix $\Pi_{t,\theta}$* and, *because f_θ depends on θ through γ, expressions for $\pi_{t,\theta}x^{(2-2\gamma)m}\log x$ for $m = 1, 2$ are also needed. Especially the latter may cause problems – finding the matrix $\Pi_{t,\theta}$ and its inverse is just a matter of grinding it out. What is certain is that the analytic expression for the optimal martingale estimating function will be complicated, yielding estimating equations that are a great deal more complicated than (2.7).*

For the simulation study, the true parameter values were set at

$$a = 1, \quad b = -1, \quad \gamma = \tfrac{1}{2}, \quad \sigma^2 = 1, \tag{2.8}$$

so X itself is a CIR-process. The initial value, X_0, was simulated from the invariant distribution, the $\Gamma-$distribution with shape parameter 2 and scale parameter $\frac{1}{2}$. The simulations were done using an Euler scheme to simulate $\sqrt{X}$, which is more accurate than the Euler scheme for X itself since the diffusion coefficient for $\sqrt{X}$ is a constant.

Simulations were done for different values of Δ. For each Δ, the number of observations was $n+1 = 501$ and 50 samples were made of each observation series. The summary statistics for the 50 sets of estimates obtained for each Δ, using the symmetrised version of (2.7), are given in Table 1.

Δ	success		mean	std.dev.	small	large	correlations	
0.01	50/50	a	1.77	0.864	0.737	4.51	-0.675^{ab}	$0.276^{\gamma\sigma^2}$
		b	-1.88	0.872	-4.84	-0.612	$0.298^{a\gamma}$	$-0.137^{b\gamma}$
		γ	0.493	0.054	0.396	0.641	$0.046^{a\sigma^2}$	$0.035^{b\sigma^2}$
		σ^2	1.00	0.073	0.806	1.174		
0.1	50/50	a	1.04	0.207	0.685	1.62	-0.812^{ab}	$0.209^{\gamma\sigma^2}$
		b	-1.08	0.262	-2.01	-0.662	$0.440^{a\gamma}$	$-0.208^{b\gamma}$
		γ	0.494	0.050	0.393	0.571	$0.494^{a\sigma^2}$	$-0.310^{b\sigma^2}$
		σ^2	1.00	0.086	0.786	1.18		
0.5	45/50	a	1.22	0.335	0.597	1.92	-0.975^{ab}	$0.286^{\gamma\sigma^2}$
		b	-1.22	0.308	-1.93	-0.674	$0.890^{a\gamma}$	$-0.818^{b\gamma}$
		γ	0.545	0.081	0.361	0.68	$0.380^{a\sigma^2}$	$-0.421^{b\sigma^2}$
		σ^2	0.995	0.087	0.730	1.24		

Table 1. Summary statistics for estimates in the four parameter generalised Cox-Ingersoll-Ross process.

According to the theory, the estimates for a and b should deteriorate when $\Delta \to 0$, as is certainly confirmed by Table 1, while the estimates for γ and σ^2 should remain stable, as also appears to be the case. For all three values of Δ, the estimates for γ and σ^2 are much more precise than those for a and b. But for $\Delta = 0.5$ there is some numerical instability and that would get worse for larger values of Δ. In particular, for $\Delta = 0.5$ in 5 samples out of 50 no solution to the estimating equations were found, or to be more specific, no solution with $\gamma \neq 1$ was found: a general numerical problem when solving the estimating equations is that a pseudo solution may be found corresponding to $\gamma = 1$, a value of γ that creates a singularity in the equations as is clear from (2.7).

To inspect more closely the effect of small $\Delta-$optimality, a second simulation study was conducted with a, b known and only γ and σ^2 assumed unknown. The small $\Delta-$optimal estimating function from case (ii), see (1.6), using the base $x^{2-2\gamma}$, $x^{4-4\gamma}$ as before, is given by (apart from irrelevant proportionality factors) the top two rows of (2.7), i.e.

$$g^I_{t,\theta}(x,y) = \begin{pmatrix} -2\log x & x^{2\gamma-2}\log x \\ -2 & x^{2\gamma-2} \end{pmatrix} \begin{pmatrix} y^{2-2\gamma} - \pi_{t,\theta} x^{2\gamma-2} \\ y^{4-4\gamma} - \pi_{t,\theta} x^{4\gamma-4} \end{pmatrix}. \tag{2.9}$$

The estimates obtained using $g^I_{t,\theta}$ (or its symmetrised version) were compared to those found by combining a small $\Delta-$optimal flow for estimating γ alone (the first row of $g^I_{t,\theta}$) with a flow for estimating σ^2 that is not small $\Delta-$optimal, viz. a simple unbiased estimating function $\widetilde{g}_{t,\theta}$ of the type studied by Kessler [4], $\widetilde{g}_{t,\theta}\,(x,y) = A_\theta \phi(x)$ depending neither on t nor on y. Here A_θ is the infinitesimal generator for X,

$$A_\theta \phi(x) = \left(a x^{2\gamma-1} + bx\right)\phi'(x) + \tfrac{1}{2}\sigma^2 x^{2\gamma}\phi''(x).$$

We chose $\phi(x) = x^2$ and thus used the flow

$$g^{II}_{t,\theta}\,(x,y) = \begin{pmatrix} -2\log x\,(y^{2-2\gamma} - \pi_{t,\theta}x^{2\gamma-2}) + x^{2\gamma-2}\log x\,(y^{4-4\gamma} - \pi_{t,\theta}x^{4\gamma-4}) \\ (2a+\sigma^2)\,x^{2\gamma} + 2bx^2 \end{pmatrix}$$

or its symmetrised version.

Again, for different values of Δ, observation series with $n+1 = 501$ were simulated corresponding to the true parameter values θ_0 given by (2.8), in some cases with X_0 constant and $\equiv E^{\mu}_{\theta_0} X_0 = 1$ (fixed start), in some cases with X_0 simulated from the invariant $\Gamma-$distribution (stationary start). The number of samples were 50 or 100. Summary statistics for the estimates are presented in Table 2, where e.g. g^I signifies that $g^I_{\Delta,\theta}$ was used precisely as given by (2.9), while 'g^I symm' means that the symmetrised version was used. For each value of Δ, the same data were used for all estimating functions that were applied for that value of Δ. In the table 'stat.' means stationary start and 'fixed' means fixed start.

Δ	g	success		mean	std.dev.	small	large
0.02	g^I	50/50	γ	0.495	0.047	0.353	0.585
stat.			σ^2	0.993	0.079	0.826	1.36
	g^{II}	34/50	γ	0.420	0.104	0.044	0.591
			σ^2	0.841	0.551	0.041	2.08
	g^I(symm)	50/50	γ	0.489	0.050	0.345	0.603
			σ^2	0.999	0.064	0.859	1.12
	g^{II}(symm)	32/50	γ	0.410	0.109	0.163	0.645
			σ^2	0.975	0.850	0.071	3.81
0.05	g^I	100/100	γ	0.486	0.058	0.280	0.599
fixed			σ^2	1.01	0.076	0.827	1.21
	g^{II}	93/100	γ	0.429	0.103	0.068	0.624
			σ^2	0.907	0.679	-0.259	3.10
0.1	g^I	100/100	γ	0.497	0.048	0.367	0.609
fixed			σ^2	1.01	0.072	0.869	1.22
	g^{II}	99/100	γ	0.474	0.061	0.267	0.603
			σ^2	0.940	0.456	0.007	2.60
0.2	g^I	98/100	γ	0.512	0.044	0.385	0.617
fixed			σ^2	1.00	0.071	0.802	1.16
	g^{II}	99/100	γ	0.496	0.056	0.245	0.599
			σ^2	0.939	0.305	-0.136	1.99
0.5	g^I	48/50	γ	0.513	0.044	0.425	0.585
fixed			σ^2	0.977	0.118	0.714	1.30
	g^{II}	48/50	γ	0.505	0.052	0.343	0.579
			σ^2	0.930	0.188	0.447	1.34
1	g^I	38/50	γ	0.536	0.056	0.434	0.675
			σ^2	1.02	0.138	0.646	1.34
	g^{II}	37/50	γ	0.530	0.057	0.419	0.675
			σ^2	0.943	0.136	0.750	1.29

Table 2. Summary statistics for the estimates of γ, σ^2 in the generalised Cox-Ingersoll-Ross process.

The effect of small $\Delta-$optimality is evident: for small values of Δ, g^I is much better than g^{II}, especially with regards to estimating σ^2. The g^{II}-estimates of γ are quite reasonable despite the wild fluctuations in the estimates of σ^2. Starting with $\Delta = 0.1$ and certainly for $\Delta = 0.2, 0.5$ and 1, there is not much difference between g^I and g^{II} when estimating γ, but for estimating σ^2 only for $\Delta = 0.5$ and 1 are the g^I and g^{II} estimates comparable in quality – for $\Delta = 0.05$ and 0.2 it is somewhat disturbing to see a (at least one) negative σ^2-estimate returned for g^{II}. For $\Delta = 1$ the success rate for both g^I and g^{II} is quite low, suggesting perhaps that some adjustment to the estimating functions should be made to accomodate larger values of Δ – how to do this in practice is however not clear. Only for $\Delta = 0.02$ were the symmetrised versions of g^I and g^{II} used, but no improvement was visible.

2.2. The Pedersen Bibby-Sørensen-Kessler example

The following model, originally proposed A.R. Pedersen, was used by Bibby and Sørensen [1] and by Kessler [4],

$$dX_t = -\theta X_t \, dt + \sqrt{\theta + X_t^2} \, dB_t \tag{2.10}$$

which is well defined for $\theta > 0$. We have $I = \mathbb{R}$ and the diffusion is ergodic for all $\theta > 0$ with invariant density

$$\mu_\theta(x) = \frac{\theta^{\theta+\frac{1}{2}} \Gamma(\theta+1)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\theta + \frac{1}{2}\right)} \frac{1}{\left(\theta + x^2\right)^{\theta+1}} \tag{2.11}$$

This distribution has heavy tails with only the absolute moments of order $< 2\theta+1$ being finite. Note that large values of θ signifies less heavy tails.

It is easy to find conditional moments for the model (2.10) (provided they exist). Thus

$$\pi_{t,\theta} x = e^{-\theta t} x \quad (\theta > 0), \qquad \pi_{t,\theta} x^2 = \frac{\theta}{2\theta - 1} \left(1 - e^{(1-2\theta)t}\right) + e^{(1-2\theta)t} x^2 \quad \left(\theta > \tfrac{1}{2}\right). \tag{2.12}$$

Both Bibby and Sørensen [1] and Kessler [4] did simulations for $\theta = 10$. In [1] the optimal martingale estimating function of order 1 with base $f^1(x) = x$ was used, i.e. (cf. (1.4))

$$g_{t,\theta}^{BS}(x, y) = \frac{x}{\pi_{t,\theta} x^2 - \left(\pi_{t,\theta} x\right)^2} \left(y - \pi_{t,\theta} x\right), \tag{2.13}$$

while Kessler [4] used the simple estimating function $A_\theta x^2$ with A_θ the generator $A_\theta \phi(x) = -\theta x \phi' + \frac{1}{2}\left(\theta + x^2\right) \phi''(x)$, i.e.

$$g_{t,\theta}^{KES}(x, y) = \theta\left(1 - 2x^2\right) + x^2. \tag{2.14}$$

The simulations in [1] and [4] were done for $\Delta = 0.05, 0.1, 0.2$ ([1] only), $0.3, 0.5$ ([4]) only. The results for Kessler's simple estimating function are quite awful – Kessler used this example to illustrate that the simple estimating functions proposed by him, that works fine for many models, can prove useless in some cases. By contrast, the estimating function used by Bibby and Sørensen performs well.

Here, apart from g^{BS} and g^{KES}, we shall also consider the small $\Delta-$optimal martingale estimating function of order 2 derived from the base $f^1(x) = x$, $f^2(x) = x^2$ derived from case (ii), (1.6), i.e.

$$g_{t,\theta}^{MJ}(x,y) = -\frac{2x}{\left(\theta + x^2\right)^2}\left(y - \pi_{t,\theta}x\right) + \frac{1}{\left(\theta + x^2\right)^2}\left(y^2 - \pi_{t,\theta}x^2\right). \tag{2.15}$$

For the simulations presented in Table 3 below, for each data set three estimates $\theta_{MJ,}$ θ_{BS} and θ_{KES} were found corresponding to the symmetrised versions of (2.15), (2.13) and (2.14) respectively (θ_{KES} only because it is so easy to compute). Throughout $\theta = 10$ is the true parameter value and each sample has $n+1 = 501$ observations with the initial value X_0 fixed at 0.

It should be noted that with $\theta = 10$ the process is racing towards equilibrium, cf. (2.12), which means that for even moderate values of Δ ($= 1$ say), the observations form almost an iid sample from the invariant distribution μ_θ. By elementary but tedious computations, the Fisher information about θ in μ_θ given by (2.11) can be found explicitly and evaluating the result for $\theta = 10$, the value $3.47 \cdot 10^{-5}$ is obtained. Thus there is hardly any information about θ in μ_θ which explains why the estimates presented below should deteriorate for Δ large enough – $\Delta \sim 0.3$ or thereabouts judging by Table 3.

The effect of small $\Delta-$optimality is once again clealy visible: for the smaller values of Δ, θ_{MJ} is better than θ_{BS}. For $\Delta = 0.1$ they behave very much the same, for $\Delta = 0.3$ the quality of both starts to deteriorate and there is evidence of some bias. For $\Delta = 0.5$ both θ_{MJ} and θ_{BS} are unreliable with θ_{MJ} in particular being too small (when it is found at all). In fact, for $\Delta = 0.5$, in 9 out of the 28 successful cases, the equation producing θ_{MJ} had 2 solutions and the smallest one was then used – in the 22 cases where no solution to the estimating equation was found, $\theta_{MJ} = \infty$ appeared as the best value (bring the value of the estimating function as close to 0 as possible). As should be the case, θ_{KES} is hopeless for all values of Δ, cf. the values for the asymptotic variance given by Kessler [4], p. 80.

Δ	success	θ	mean	std.dev.	small	large
0.002	50/50	θ_{MJ}	9.876	0.639	8.710	11.59
	50/50	θ_{BS}	10.96	4.38	3.059	24.84
	50/50	θ_{KES}	-2.611	9.46	-48.18	14.33
0.01	50/50	θ_{MJ}	10.02	0.586	8.731	11.32
	50/50	θ_{BS}	11.20	2.41	5.571	16.56
	50/50	θ_{KES}	2.457	24.2	-51.13	119.3
0.05	50/50	θ_{MJ}	9.877	0.671	8.023	10.95
	50/50	θ_{BS}	10.07	1.18	7.012	13.26
	50/50	θ_{KES}	4.076	23.3	-63.19	98.49
0.1	50/50	θ_{MJ}	10.07	1.13	8.247	12.33
	50/50	θ_{BS}	10.05	1.08	7.999	12.67
	50/50	θ_{KES}	52.33	242	-101.2	1454
0.3	39/50	θ_{MJ}	10.30	3.70	5.578	21.21
	41/50	θ_{BS}	10.76	3.42	7.155	22.26
	50/50	θ_{KES}	1.616	43.8	-264.5	102.0
0.5	28/50	θ_{MJ}	5.223	1.64	2.726	8.205
	29/50	θ_{BS}	7.257	2.57	3.801	14.12
	50/50	θ_{KES}	2.021	43.4	-217.8	93.35

Table 3. Summary statistics for estimates of the parameter in the Pedersen-Bibby-Sørensen-Kessler model.

2.3. Rounding

It is a consequence of small $\Delta-$optimality that for small values of Δ, parameters appearing in the (squared) diffusion coefficient can be estimated much more precisely than those appearing only in the drift. As discussed above this is related to the singularity between the distributions of X when the process is observed completely on a finite time interval, which means that in theory the parameters entering in the diffusion coefficient can be read off from e.g. the quadratic variation of the process. In practice however, computing anything like $\lim_{N\to\infty} \sum_{k=1}^{2^N} \left(X_{k2^{-N}} - X_{(k-1)2^{-N}}\right)^2$ is impossible: not all the X values are available and even if they were, since the observations are recorded with finite accuracy the sum of squared differences becomes numerically unstable for N large.

In order to see better if small $\Delta-$optimal estimating functions are reliable in practice, we shall therefore investigate the effect on the estimates arising from using rounded observations. In the tables below a *rounding factor* ρ appears, meaning that an exact observation z is replaced by the value

$$\frac{1}{\rho}\left[\rho z + \frac{1}{2}\right],$$

$[a]$ denoting the largest integer $\leq a$. No rounding corresponds to $\rho = \infty$, while e.g. $\rho = 10$ means that z is recorded to the first digit after the decimal point.

2.3.1. The generalised Cox-Ingersoll-Ross process

Simulations from the model (2.1) were done for $\Delta = 0.02$ and 0.05 when estimating γ, σ^2 only, using the symmetrised version of the small $\Delta-$optimal estimating function g^I from (2.9), assuming a and b to be known. As before the true parameter values were $a = 1$, $b = -1$, $\gamma = \frac{1}{2}$ and $\sigma^2 = 1$. Also as before $n + 1 = 501$ observations were used while 25 samples were made. Of course, for a given Δ, the same data were used for all factors.

Δ	factor	success		mean	std.dev.	small	large
0.02	∞	25/25	γ	0.493	0.050	0.391	0.605
			σ^2	0.999	0.086	0.816	1.173
	1000	25/25	γ	0.493	0.050	0.392	0.605
			σ^2	0.999	0.086	0.817	1.174
	100	25/25	γ	0.492	0.051	0.385	0.598
			σ^2	0.991	0.086	0.820	1.176
	20	19/25	γ	0.464	0.054	0.370	0.545
			σ^2	1.022	0.092	0.808	1.213
	10	16/25	γ	0.437	0.055	0.337	0.545
			σ^2	1.081	0.090	0.914	1.243
0.05	∞	25/25	γ	0.479	0.047	0.386	0.563
			σ^2	0.996	0.093	0.834	1.181
	1000	25/25	γ	0.479	0.047	0.386	0.563
			σ^2	0.996	0.093	0.834	1.181
	100	25/25	γ	0.478	0.047	0.384	0.560
			σ^2	0.996	0.093	0.837	1.181
	20	16/25	γ	0.467	0.041	0.387	0.537
			σ^2	1.014	0.085	0.899	1.198

Table 4. The effect of rounding when estimating γ and σ^2 in the generalise Cox-Ingersoll-Ross model.

The effect of the rounding is not visible for $\rho = 1000$ and barely for $\rho = 100$. For $\rho = 20$ the estimates still behave quite well, but a special problem due to the structure of the model appears: X is of course strictly positive but with a coarse rounding mechanism it may still happen that a rounded observation value of 0 is returned in which case the estimating function g^I from (2.9) breaks down – both $\log x$ and $x^{2\gamma-2}$ (for $\gamma = \frac{1}{2}$) are infinite. The problem persists of course for $\rho = 10$

and proved too severe for $\Delta = 0.05$. For $\Delta = 0.02$ there are perhaps some signs of bias, but the estimates are still useable.

That rounded values of 0 arise for $\rho = 10$ or 20 is clear when one considers the length of the observation series together with the form of the invariant distribution, which here is the $\Gamma-$distribution with shape parameter 2 and scale parameter $\frac{1}{2}$. If F is the distribution function, the F-values corresponding to the observation values rounded down to 0 for $\rho = 10, 20, 100$ respectively are $F(0.05) = 0.0047$, $F(0.025) = 0.0012$ and $F(0.005) = 0.0005$. Thus, by the ergodic theorem, one would expect approximately one observation in 200 to be rounded down to 0 if the rounding is to the first digit after the decimal point.

Arguably, the rounding mechanism is somewhat unfair for the particular model considered here but in conclusion it appears, that apart from the problem with rounded values of 0, the estimates perform quite well, even when the rounding is to the first decimal. For a process moving typically in the range $]0, 5]$ this is, in the view of the author, surprisingly good.

2.3.2. The Pedersen-Bibby-Sørensen-Kessler model

For this model the problem with invalid or useless rounded observation values does not arise, so here we investigate the effect of rounding more thoroughly. As before different values of Δ are used with observation series of length $n+1 = 501$. For each Δ the same data sets are used corresponding to different values of the rounding factor ρ. Throughout the true model corresponds to $\theta = 10$, exactly as in subsection 2.2, and for estimating θ the symmetrised version of the small $\Delta-$optimal flow (2.15) was used. The results of the simulations are as shown in Table 5.

The complete range for the $50 \cdot 501 = 25050$ observations simulated for each value of Δ were as follows: for $\Delta = 0.002$ from -2.969 to 2.292; for $\Delta = 0.01$ from -3.794 to 3.889; for $\Delta = 0.1$ from -3.358 to 3.551. Keeping this in mind, it is of course a joke to consider rounding factors 1 and 2.

For $\Delta = 0.002$, the estimate is fine until the factor 20. With factor 10 bias begins to show and for the smaller factors this bias renders the estimates useless. For $\Delta = 0.01$ and $\Delta = 0.1$ the same tendency is seen, except of course that with larger Δ, a coarser rounding can be allowed before the estimates collapse: for $\Delta = 0.01$ the critical factor is 10, bias appearing if the factor is 5 or smaller, while for $\Delta = 0.1$, even for factor 5 the estimate behaves sensibly.

In conclusion it would seem that the small $\Delta-$optimal estimating flow is quite robust against rounding – and certainly much more resilient than the author thought in advance. But the data also suggests that any kind of rounding may well introduce a systematic bias, that is negligeable if the rounding factor is large enough, but increases when the factor decreases.

Δ	factor	success	mean	std.dev.	small	large
0.002	∞	50/50	9.903	0.694	8.320	11.22
	100		9.912	0.692	8.327	11.22
	50		9.933	0.696	8.358	11.28
	20		10.107	0.673	8.507	11.27
	10		10.686	0.787	8.853	12.16
	5		13.415	0.928	11.33	15.51
	2		29.416	2.42	22.84	33.74
	1		61.149	12.46	34.14	86.42
0.01	∞	50/50	9.938	0.711	8.490	11.92
	100		9.941	0.713	8.485	11.94
	50		9.946	0.712	8.495	11.95
	20		9.988	0.724	8.507	12.01
	10		10.124	0.719	8.728	12.16
	5		10.680	0.747	9.125	12.82
	2		14.606	1.124	12.62	17.74
	1		27.859	2.898	21.95	35.42
0.1	∞	50/50	10.171	1.144	8.153	12.65
	100		10.171	1.144	8.152	12.63
	50		10.171	1.147	8.163	12.66
	20		10.187	1.151	8.160	12.73
	10		10.190	1.156	8.091	12.70
	5		10.335	1.190	8.112	13.02
	2		11.218	1.332	8.826	14.03
	1		14.612	1.980	9.888	19.00

Table 5. The effect of rounding when estimating θ in the Pedersen-Bibby-Sørensen-Kessler model.

References

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

[2] Jacobsen, M. (1998). Discretely observed diffusions: classes of estimating functions and small $\Delta-$optimality. Preprint 11, Department of Theoretical Statistics, University of Copenhagen. (To appear in Scand. J. Statist.)

[3] Jacobsen, M. (2000). Optimality and small $\Delta-$optimality of martingale estimating functions. Preprint 5, Department of Theoretical Statistics, University of Copenhagen.

[4] Kessler, M. (2000). Simple and explicit estimating functions for a discretely observed diffusion process. *Scand. J. Statist.* **27**, 65-82.