MPS-RR 2000-21

ISSN 1398-599X

Optimality and small $\Delta-$optimality of martingale estimating functions

Martin Jacobsen*

Abstract

Martingale estimating functions determined from a given collection (the base) of conditional expectations are considered for estimating the parameters of a discretely observed diffusion. It is discussed how to make the martingale estimating functions small $\Delta-$optimal, i.e. nearly efficient when the observations are close together, in particular it is shown that this is possible provided the base is large enough. It is also shown that the optimal martingale estimating function with a given base, is automatically small $\Delta-$optimal, provided only that the base is sufficiently large. In both cases the critical dimension of the base is the same and determined by the dimension of the diffusion, and on whether the squared diffusion matrix is parameter dependent or not; the critical number does not depend however on the dimension of the parameter.

KEYWORDS AND PHRASES: unbiased estimating functions, minimizing asymptotic covariances, conditional moments, a generalized Cox-Ingersoll-Ross process, Gaussian diffusions.

1 Introduction

Suppose a $d-$dimensional, ergodic, timehomogeneous diffusion X is observed at finitely many points $i\Delta$ in time, $i = 0, \ldots, n$. In order to estimate the unknown parameter θ that determines the distribution of X, rather than doing maximum-likelihood, which may well prove unfeasible, one often resorts to

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

the use of unbiased estimating functions, some of the most successful of which are based on conditional expectations, resulting in estimating equations of the form

$$\sum_{i=1}^{n}\sum_{q=1}^{r} h^q\left(X_{(i-1)\Delta}\right)\left(f^q\left(X_{i\Delta}\right) - E_\theta\left(f^q\left(X_{i\Delta}\right)\left|X_{(i-1)\Delta}\right.\right)\right) = 0 \qquad (1)$$

where, for the moment, we consider θ to be one-dimensional (see (6) and (2) below for the general setup). (1) is the prime example of an estimating equation obtained from an unbiased *martingale* estimating function.

The study of estimating equations of the form (1) was initiated by Bibby and Sørensen [1] who focused on the case $r = 1$ and $f(x) = x$ and for that case also showed that under mild conditions (1) has a consistent and $\sqrt{n}-$asymptotically Gaussian solution in θ (as $n \to \infty$ with $\Delta > 0$ fixed). These asymptotic results readily generalize to other types of unbiased estimating equations, see e.g. Sørensen [6].

In the present paper we consider estimating equations of the form (1) for $d-$dimensional diffusions X with a $p-$dimensional parameter θ. The main issue is the discussion of the choice of *base* $(f^q)_{1\leq q\leq r}$, in particular the choice of r, and the choice of *weights* h^q. For this we think of $\Delta > 0$ as arbitrary and then consider families of estimating equations of the form (1) where h^q, but not f^q, is allowed to depend on Δ, and such that for any Δ, (1) has a consistent and asymptoticallyGaussian solution as described above.

Given the base (f^q) there is an *optimal* choice of weights, see Proposition 1 below. The resulting estimator will typically not be efficient, while an estimator that is *small* $\Delta-$*optimal* will be nearly efficient for small values of Δ – and of course still consistent and asymptotically Gaussian for all Δ, although not optimal. As we shall see, while it may be difficult to find the optimal estimator (to obtain and use the weights one needs the explicit form of the inverse of a $r \times r-$matrix with all elements conditional variances), it is quite easy to determine weights that lead to small $\Delta-$optimal estimators.

The concept of small $\Delta-$optimality was introduced by Jacobsen [2], and the main purpose of the present paper is to discuss conditions for small $\Delta-$optimality for the type of martingale estimating functions underlying (1).

The first main result, Theorem 2, shows that given the base, provided only that the dimension r is large enough, there always exist weights such that small $\Delta-$optimality is achieved. Furthermore it is easy to find the

weights – it is just a matter of solving at any point x in the range for the diffusion, a set of linear equations, and in the statement of the theorem, a concrete solution is exhibited.

The second main result, Theorem 3, shows that for any base, again provided r is large enough, for the optimal choice of weights small $\Delta-$optimality is automatic.

In both theorems the same critical value r_0 for the dimension of the base appears: for $r \geq r_0$ small $\Delta-$optimality can be achieved for any base, while for $r < r_0$ this may only be possible, if at all, for a special choice of base. (The case $r < r_0$ is only mentioned, but not really explored below). The value of r_0 depends on the structure of the model but not on the dimension of the parameter with $r_0 = d$, the dimension of the diffusion, if the diffusion coefficient does not depend on the parameter, and $r_0 = d(d+3)/2$ otherwise. Thus one natural choice of base is $f^i(x_1, \ldots, x_d) = x_i$ if $r_0 = d$, and these f^i supplemented by $f^{ij}(x_1, \ldots, x_d) = x_i x_j$ for $1 \leq i \leq j \leq d$ if $r_0 = d(d+3)/2$.

The paper is concluded (Section 3) with two examples, one describing a generalized Cox-Ingersoll-Ross model, the second the finite-dimensional Ornstein-Uhlenbeck processes. For the latter it turns out, that using the base of first and second order moments (f^i, f^{ij}), the concrete small $\Delta-$optimal estimating function exhibited in Theorem 2 for any Δ yields the maximum-likelihood estimator.

Although both here and in Jacobsen [2], small $\Delta-$optimality is dicussed exclusively for diffusions, we wish to point out that the concept makes perfect sense for any model involving discrete observations from an ergodic, timehomogeneous Markov process in continuous time.

2 Optimality and small $\Delta-$optimality

Let $X = (X_t)_{t \geq 0}$ be a $d-$dimensional ergodic diffusion, solving the stochastic differential equation

$$dX_t = b_\theta(X_t)\, dt + \sigma_\theta(X_t)\, dB_t, \quad X_0 = U$$

where $b_\theta(x) \in \mathbb{R}^{d \times 1}$, $\sigma_\theta(x) \in \mathbb{R}^{d \times d}$, B is a standard $d-$dimensional Brownian motion and U is a $d-$dimensional random variable, independent of B. Both the drift b_θ and the diffusion coefficient σ_θ is allowed to depend on the $p-$dimensional parameter $\theta \in \Theta$. The invariant distribution for X is de-

noted μ_θ, i.e. if U has distribution μ_θ, then X is a strictly stationary Markov process.

We shall also assume that X takes its values within some open subset D of $\mathbb{R}^d$, of course not allowed to depend on θ. Also we assume that $C_\theta(x) := \sigma_\theta(x)\sigma_\theta^T(x) \succ 0$ for all θ and all $x \in D$, i.e. the symmetric positive semidefinite matrix $C_\theta(x)$ is assumed to be strictly positive definite always. (^{T} denotes matrix transposition).

We shall write μ_θ also for the density of μ_θ and assume that for all θ, $\mu_\theta > 0$ everywhere on D. The transition density is denoted $p_{t,\theta}(x,y)$,

$$P_\theta\left(X_{s+t} \in dy \,|X_s = x\right) = p_{t,\theta}\left(x,y\right)\, dy.$$

The underlying probability P_θ depends not only on θ but also on the distribution of X_0. It is denoted P_θ^μ if X_0 has distribution μ_θ and P_θ^x if $X_0 \equiv x$. The corresponding expectations are written E_θ^μ and E_θ^x.

We shall denote by $Q_{t,\theta}$ the joint distribution of (X_s, X_{s+t}) under P_θ^μ (for any s), and by $q_{t,\theta}$ the density of $Q_{t,\theta}$,

$$q_{t,\theta}\left(x,y\right) = \mu_\theta(x)p_{t,\theta}(x,y).$$

Finally, the transition operators for X are denoted $\pi_{t,\theta}$,

$$\pi_{t,\theta}f(x) = E_\theta^x f(X_t)$$

provided the integral makes sense (e.g. for f bounded or $f \in L^1\left(\mu_\theta\right)$), and the differential operator determining the infinitesimal generator is denoted A_θ,

$$A_\theta f_\theta(x) = \sum_{i=1}^{d} b_\theta^i(x)\partial_{x_i}f(x) + \tfrac{1}{2}\sum_{i,j=1}^{d} C_\theta^{ij}(x)\partial_{x_ix_j}^2 f(x)$$

for sufficiently smooth functions f.

Suppose now that X is observed at finitely many equidistant timepoints, $X_0, X_\Delta, \ldots, X_{n\Delta}$. We shall discuss optimality properties of estimators based on martingale estimating functions, i.e. the estimator $\hat{\theta}_n$ of θ is found by solving the estimating equation

$$G_{n,\Delta}\left(\theta\right) = \sum_{i=1}^{n} g_{\Delta,\theta}\left(X_{(i-1)\Delta}, X_{i\Delta}\right) = 0 \tag{2}$$

where $(t, \theta, x, y) \to g_{t,\theta}(x, y)$ is a $p-$variate function such that each coordinate $g^k_{t,\theta}$ satisfies the martingale condition

$$g^{k*}_{t,\theta}(x) = 0 \text{ for all } x, \quad g^{k*}_{t,\theta}(x) = E^x_\theta g^k_{t,\theta}(X_0, X_t) \tag{3}$$

ensuring that $(G_{n,\Delta}(\theta))_{n\geq 1}$ is a $p-$dimensional $P_\theta-$martingale (whatever the initial distribution of X).

Following the terminology in Jacobsen [2], we shall refer to $\mathcal{G} = (g_{t,\theta})_{t>0,\theta\in\Theta}$ as a *well behaved flow of martingale estimating functions*, $\mathcal{G} \subset \mathcal{M}$ (the space of flows of martingale estimating functions) if each $g^k_{t,\theta} \in L^2(Q_{t,\theta})$ with

$$E^\mu_\theta g_{t,\theta'}(X_0, X_t) = 0 \text{ if and only if } \theta = \theta'$$

(where the equality holds for $\theta = \theta'$ because of (3)) and if furthermore $E^\mu_\theta\left(g_{t,\theta} g^T_{t,\theta}\right)(X_0, X_t) \succ 0$ for all t, θ, $\partial_{\theta_\ell} g^k_{t,\theta} \in L^1(Q_{t,\theta})$ for all t, θ, all indices k, ℓ, and finally and most important if for every θ_0 and every $t = \Delta > 0$ there is with $P^\mu_{\theta_0}-$probability tending to 1 a consistent solution $\hat{\theta}_n$ to (2) such that $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ converges in distribution for $n \to \infty$ to the $p-$dimensional Gaussian distribution with mean vector 0 and covariance matrix $\mathrm{var}_{\Delta,\theta_0}\left(g, \hat{\theta}\right)$ given by

$$\mathrm{var}_{\Delta,\theta_0}\left(g, \hat{\theta}\right) = \Lambda^{-1}_{\Delta,\theta_0}(g) E^\mu_{\theta_0}\left(g_{\Delta,\theta_0} g^T_{\Delta,\theta_0}\right)(X_0, X_\Delta)\left(\Lambda^{-1}_{\Delta,\theta_0}(g)\right)^T. \tag{4}$$

Here

$$\Lambda_{t,\theta} := E^\mu_\theta \dot{g}_{t,\theta}(X_0, X_t), \tag{5}$$

the dot $\dot{}$ signifying differentiation with respect to θ so that $\dot{g}_{t,\theta}(x, y) \in \mathbb{R}^{p\times p}$ is given by

$$\left(\dot{g}_{t,\theta}(x, y)\right)_{k\ell} = \partial_{\theta_\ell} g^k_{t,\theta}(x, y).$$

The reader is reminded that asymptotic normality of $\hat{\theta}_n$, as specified above, holds under quite weak assumptions (see Sørensen [6]) and that certainly (4) is the natural expression for the asymptotic covariance. The most critical among the assumptions needed is that $\Lambda_{t,\theta} \in \mathbb{R}^{p\times p}$ must be non-singular for all t, θ.

Notation. Throughout the paper, derivatives are understood as matrices in analogy with (5): if ϕ is a $\rho-$variate function of a $\upsilon-$dimensional variable $z = (z_1, \ldots, z_\upsilon) \in \mathbb{R}^\upsilon$, $\partial_z \phi$ denotes the $\rho \times \upsilon-$matrix of partial derivatives

with κ'th row $(\partial_{z_1}\phi^\kappa, \ldots, \partial_{z_v}\phi^\kappa)$. The dot notation is used exclusively for differentiation with respect to θ, $\dot{\phi} = \partial_\theta \phi$.

In the remainder of the paper we shall focus on martingale estimating functions derived from conditional expectations of given functionals, i.e. we assume that

$$g_{t,\theta}^k(x,y) = \sum_{q=1}^{r} h_{t,\theta}^{qk}(x) \left(f_\theta^q(y) - (\pi_{t,\theta} f_\theta^q)(x)\right), \tag{6}$$

or, in matrix notation,

$$g_{t,\theta}(x,y) = h_{t,\theta}^T(x) \left(f_\theta(y) - (\pi_{t,\theta} f_\theta)(x)\right)$$

with $h_{t,\theta}(x) \in \mathbb{R}^{r\times p}$, $f_\theta(x) \in \mathbb{R}^{r\times 1}$. (The integrability assumptions imposed on general g above, makes it natural to assume here that f_θ^q and $h_{t,\theta}^{qk} \in L^4(\mu_\theta)$, while $\theta-$derivatives of f_θ^q and $h_{t,\theta}^{qk}$ must belong to $L^2(\mu_\theta)$. We shall not be too concerned about these conditions in the sequel – it is tacitly assumed everywhere that the flow $\mathcal{G}$ given by (6) is well behaved).

Estimating functions of the form (6) were first used by Bibby and Sørensen [1], see also Jacobsen [2], Section 3 for an overview.

We shall refer to the functions $f_\theta^1, \ldots, f_\theta^r$ as the *base* for the flow of estimating functions given by (6). The problem to be studied in this paper is that of finding good choices for the dimension of the base and for the *weights* $h_{t,\theta}$ given the base.

Assumption A. *The functions $f_\theta^q(x)$ are supposed to be differentiable in θ and twice differentiable in x. Also, the base $f_\theta^1, \ldots, f_\theta^r$ is supposed to have full affine rank r on the domain D for all θ, i.e. for an arbitrary θ the identity*

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

for some constants $a_\theta^q, \alpha_\theta$ implies $a_\theta^1 = \cdots = a_\theta^r = \alpha_\theta = 0$.

The functions $h_{t,\theta}^{qk}$ are supposed to satisfy that for any t, θ, the p $r-$variate functions $x \to \left(h_{t,\theta}^{1k}(x), \ldots, h_{t,\theta}^{rk}(x)\right)$ forming the columns of $h_{t,\theta}$ are linearly independent on D. ■

Note that if $f_\theta^1, \ldots, f_\theta^r$ does not have full affine rank, there is a representation (6) of the $g_{t,\theta}^k$ with r replaced by $r-1$. The condition that $f_\theta^1, \ldots, f_\theta^r$ has full rank is equivalent to assuming that the r $d-$variate functions $\partial_x f_\theta^q$

for $1 \leq q \leq r$ be linearly independent. In the main results, Theorems 2 and 3 below, Assumption A is supplemented by conditions on the pointwise behaviour of $\partial_x f_\theta$ and $\partial^2_{xx} f_\theta$.

If for some t it holds for all θ that the columns of $h_{t,\theta}$ are not linearly independent, i.e there is $\beta_{t,\theta} \in \mathbb{R}^{p\times 1}\backslash 0$ such that $h_{t,\theta}(x)\beta_{t,\theta} = 0$ for all x, then $\beta^T_{t,\theta} g_{t,\theta}(x,y) = 0$ for all x, y so that one of the p estimating equations in (2) can be obtained from the others and it is impossible to estimate all p parameters θ_ℓ – formally both matrices $\Lambda_{t,\theta}(g)$ and $E^\mu_\theta\left(g_{t,\theta} g^T_{t,\theta}\right)(X_0, X_t)$ become singular and (4) does not make sense.

Note that we allow the base $f^1_\theta, \ldots, f^r_\theta$ to depend on θ, but *not* on t.

For a given base, it is easy to determine the optimal choices for the $h^{qk}_{t,\theta}$, i.e. the choices minimizing $\mathrm{var}_{t,\theta}\left(g, \hat{\theta}\right)$. We use the notation $A \succeq B$ between symmetric, positive semidefinite matrices to signify that $A - B$ is also positive semidefinite.

Proposition 1 *Assume that for all x, t and θ the symmetric $r \times r-$matrix*

$$\Pi_{t,\theta} f_\theta(x) := \pi_{t,\theta}\left(f_\theta f^T_\theta\right)(x) - \left(\pi_{t,\theta} f_\theta\right)(x)\left(\pi_{t,\theta} f^T_\theta\right)(x)$$

is non-singular, and define

$$h^{\mathrm{opt}}_{t,\theta} = \left(\Pi_{t,\theta} f\right)^{-1}\left(\partial_\theta\left(\pi_{t,\theta} f_\theta\right) - \pi_{t,\theta}\left(\dot{f}_\theta\right)\right). \tag{7}$$

Then, provided differentiation and integration can be interchanged in

$$\partial_\theta \int p_{t,\theta}(x,y) f_\theta(y)\, dy = \int \partial_\theta\left(p_{t,\theta}(x,y) f_\theta(y)\right) dy$$

and the flow $\mathcal{G}^{\mathrm{opt}}$ given by

$$g^{\mathrm{opt}}_{t,\theta}(x,y) = \left(h^{\mathrm{opt}}_{t,\theta}\right)^T(x)\left(f_\theta(y) - \pi_{t,\theta} f(x)\right) \tag{8}$$

is well behaved, it holds that

$$\mathrm{var}_{t,\theta}\left(g^{\mathrm{opt}}, \hat{\theta}\right) \preceq \mathrm{var}_{t,\theta}\left(g, \hat{\theta}\right)$$

for any well behaved flow $\mathcal{G} = (g_{t,\theta})$ of the form (6) with base $f^1_\theta, \ldots, f^r_\theta$.

Proof. Since f is allowed to depend on θ, this extends (2.10) in Bibby and Sørensen [1] and Example 5.5 in Jacobsen [2], so we indicate the proof. By the projection theorem (Kessler [3], Proposition 1, Jacobsen [2], Proposition 5.3), $g_{t,\theta}^{k,\mathrm{opt}}$ is found by projecting the k'th coordinate of the score function, $\partial_{\theta_k} p_{t,\theta}(x,y)/p_{t,\theta}(x,y)$ onto the subspace of $L^2(Q_{t,\theta})$ spanned by functions of the form (6) with the f_θ^q fixed and arbitrary $h_{t,\theta}^{qk} \in L^2(\mu_\theta)$. Thus $h_{t,\theta}^{qk,\mathrm{opt}}$ satisfies for all $1 \leq q_0 \leq r$, $1 \leq k \leq p$ and all $h \in L^2(\mu_\theta)$ that

$$\begin{aligned} 0 &= E_\theta^\mu \left[\frac{\partial_{\theta k} p_{t,\theta}}{p_{t,\theta}}(X_0, X_t) - \sum_{q=1}^{r} h_{t,\theta}^{qk,\mathrm{opt}}(X_0) \left(f_\theta^q(X_t) - \pi_{t,\theta} f_\theta^q(X_0) \right) \right] \\ &\quad \times h(X_0) \left(f_\theta^{q_0}(X_t) - \pi_{t,\theta} f_\theta^{q_0}(X_0) \right) \end{aligned} \tag{9}$$

Using that $\dot{p}_{t,\theta}/p_{t,\theta}$ is a martingale estimating function and that

$$\begin{aligned} E_\theta^x \frac{\partial_{\theta k} p_{t,\theta}}{p_{t,\theta}}(X_0, X_t) f_\theta^{q_0}(X_t) &= \int \partial_{\theta_k} p_{t,\theta}(x,y) f_\theta^{q_0}(y)\, dy \\ &= \partial_{\theta_k} \int p_{t,\theta}(x,y) f_\theta^{q_0}(y)\, dy \\ &\quad - \int p_{t,\theta}(x,y) \partial_{\theta_k} f_\theta^{q_0}(y)\, dy, \end{aligned}$$

(9) may be written

$$\begin{aligned} &E_\theta^\mu h(X_0) [\partial_{\theta_k} (\pi_{t,\theta} f_\theta^{q_0})(X_0) - \pi_{t,\theta} (\partial_{\theta_k} f_\theta^{q_0})(X_0) \\ &- \sum_{q=1}^{r} h_{t,\theta}^{qk,\mathrm{opt}}(X_0) (\pi_{t,\theta} (f_\theta^q f_\theta^{q_0})(X_0) - \pi_{t,\theta} f_\theta^q(X_0) \pi_{t,\theta} f_\theta^{q_0}(X_0))] \end{aligned}$$

for all $h \in L^2(\mu_\theta)$, i.e. the expression in square brackets must vanish $P_\theta^\mu-$a.s. and the result follows. ■

Proposition 1 is a result on *local optimality*, i.e. it exhibits the best member from a given, restricted class of estimating functions, best from the point of view of minimizing the asymptotic covariance of the resulting estimator. But only in exceptional cases will this choice be *globally optimal*, i.e. the (locally) optimal estimator will be efficient against the maximum-likelihood estimator.

By contrast, the concept of small $\Delta-$optimality introduced by Jacobsen [2], Section 7, gives conditions for global optimality, not for any given $\Delta > 0$,

but only for $\Delta \to 0$. We shall briefly recapitulate the sufficient conditions for small $\Delta-$optimality of martingale estimating functions, Theorem 7.5 in Jacobsen [2].

With $\mathcal{G} \subset \mathcal{M}$ a well behaved flow of estimating functions, it is first of all essential to assume that there is a smooth extension of $g_{t,\theta}(x,y)$ (which is defined only for $t>0$) to allow $t=0$, i.e. after a possible renormalization of $g_{t,\theta}$ by a factor (non-zero scalar or non-singular $p \times p-$matrix) depending on t, θ but not on x, y (so the solution of (2) is not affected) the limit

$$g_{0,\theta}(x,y) = \lim_{t \to 0} g_{t,\theta}(x,y)$$

must exist with $(x,y) \to g_{0,\theta}(x,y)$ not identically 0, of full rank p in a suitable sense and sufficiently smooth as required by the conditions below.

With this smooth extension of $g_{t,\theta}$ available, it is shown in Jacobsen [2] that subject to important integrability conditons (see the paragraph below (13)), the asymptotic covariance for $\hat{\theta}_n$ has an expansion, as $\Delta \to 0$,

$$\mathrm{var}_{\Delta,\theta}\left(g, \hat{\theta}\right) = \frac{1}{\Delta} v_{-1,\theta}\left(g, \hat{\theta}\right) + v_{0,\theta}\left(g, \hat{\theta}\right) + o(1) \tag{10}$$

and three cases (i), (ii) and (iii) for the structure of the diffusion model are then considered for the discussion of small $\Delta-$optimality (to achieve the structure in (iii) it may be necessary first to reparametrize the model):

(i) $C_\theta = C$ does not depend on θ. Then the main term in (10) is always present and small $\Delta-$optimality is achieved by minimizing globally (over all g) $v_{-1,\theta}\left(g, \hat{\theta}\right)$. A sufficient condition for a given flow $(g_{t,\theta})$ to be small $\Delta-$optimal is that

$$\partial_y g_{0,\theta}(x,x) = K_\theta \dot{b}_\theta^T(x) C^{-1}(x) \tag{11}$$

for some non-singular $K_\theta \in \mathbb{R}^{p \times p}$. ($\partial_y g_{0,\theta}(x,x)$ evaluates $\partial_y g_{0,\theta}(x,y)$ along the diagonal $y=x$).

(ii) C_θ depends on all parameters $\theta_1, \ldots, \theta_p$. Then the main term in (10) vanishes provided $\partial_y g_{0,\theta}(x,x) \equiv 0$ and small $\Delta-$optimality is achieved by minimizing $v_{0,\theta}\left(g, \hat{\theta}\right)$. A sufficient condition for $(g_{t,\theta})$ to be small $\Delta-$optimal is that

$$\partial_y g_{0,\theta}(x,x) = 0, \quad \partial_{yy}^2 g_{0,\theta}(x,x) = K_\theta \dot{C}_\theta^T(x) \left(C_\theta^{\otimes 2}(x)\right)^{-1} \tag{12}$$

for some non-singular $K_\theta \in \mathbb{R}^{p\times p}$. (Here $\dot{C}_\theta(x) \in \mathbb{R}^{d^2\times p}$ with $\left(\dot{C}_\theta(x)\right)_{ij,k} = \partial_{\theta_k} C_\theta^{ij}(x)$).

(iii) C_θ depends on the parameters $\theta_1, \ldots, \theta_{p'}$ but not on $\theta_{p'+1}, \ldots, \theta_p$ for some p' with $1 \leq p' < p$. Then parts of the main term in (10) can be made to disappear so that

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

Furthermore the matrix $v_{22,-1,\theta}\left(g, \hat{\theta}\right) \in \mathbb{R}^{(p-p')\times(p-p')}$ can be minimized and small $\Delta-$optimality is achieved by in addition minimizing the upper left block $v_{11,0,\theta}\left(g, \hat{\theta}\right)$ of $v_{0,\theta}\left(g, \hat{\theta}\right)$. A sufficient condition for small $\Delta-$optimality is that

$$\begin{aligned} \partial_y g_{0,\theta}(x,x) &= c_\theta \begin{pmatrix} 0_{p'\times d} \\ \dot{b}_{2,\theta}^T(x) C_\theta^{-1}(x) \end{pmatrix} \\ \partial_{yy}^2 g_{1,0,\theta}(x,x) &= K'_\theta \dot{C}_{1,\theta}^T(x) \left(C_\theta^{\otimes 2}(x)\right)^{-1} \end{aligned} \tag{13}$$

for some constant $c_\theta \neq 0$ and some non-singular $K'_\theta \in \mathbb{R}^{p'\times p'}$. (Notation: $\dot{b}_{2,\theta} \in \mathbb{R}^{d\times(p-p')}$ comprises the last $p-p'$ columns of $\dot{b}_\theta$, $g_{1,0,\theta}$ the first p' coordinates of $g_{0,\theta}$ and $\dot{C}_{1,\theta} \in \mathbb{R}^{d^2\times p'}$ the first p' columns of $\dot{C}_\theta$).

As mentioned above, to check for small $\Delta-$optimality more is required than just checking (11), (12) or (13), viz. it must be verified that various matrices involving expectations of quantities related to $\dot{b}$, $\dot{C}$, $\partial_y g_{0,\theta}$ and $\partial_{yy}^2 g_{0,\theta}$ must be non-singular, see Theorem 7.5 in Jacobsen [2].

Remark 1 *The conditions on $\partial_y g_{0,\theta}$ in (12) and (13) are important: if not satisfied, the main term in (10) will not disappear and some of the components of the estimates resulting from g, will be totally inefficient – have efficiency close to 0 – if Δ is small.*

We shall now show that subject to these integrability conditions, small $\Delta-$optimality of martingale estimating functions is easy to achieve. The three cases refer to (i), (ii) and (iii) above.

Notation. Let $J := \{(i',j') : 1 \leq i' \leq j' \leq d\}$. Thus J has $|J| = d + \binom{d}{2}$ elements and can be used as an index set for characterizing the elements of

a symmetric $d \times d-$matrix. We write $R \in \mathbb{R}^{d^2 \times J}$ for the reduction matrix with elements

$$R_{ij,i'j'} = \delta_{ii'}\delta_{jj'} \quad (1 \leq i,j \leq d,\ (i',j') \in J)\,.$$

Thus, if $M \in \mathbb{R}^{A \times d^2}$, $MR \in \mathbb{R}^{A \times J}$ with $(MR)_{a,i'j'} = M_{a,i'j'}$ as is used frequently below.

As counterpart to R, the expansion matrix $\tilde{R} \in \mathbb{R}^{J \times d^2}$ is defined by

$$\tilde{R}_{i'j',ij} = \begin{cases} \delta_{i'i}\delta_{j'j} & \text{if } i \leq j \\ \delta_{i'j}\delta_{j'i} & \text{if } i > j. \end{cases}$$

Then

$$\tilde{R}R = I_{J \times J} \tag{14}$$

and for any matrix $N \in \mathbb{R}^{S \times d^2}$, symmetric in the sense that $N_{s,ij} = N_{s,ji}$ for all, s, i, j, it holds that

$$N\left(R\tilde{R}\right) = N. \tag{15}$$

Define

$$\dim(d) := d + |J| = d(d+3)/2,$$

a number that plays a critical role below.

Theorem 2 *Let $(f_\theta^1, \ldots, f_\theta^r)$ be a base for a martingale estimating function, of full affine rank r.*

(i) Suppose that $r \geq d$, that for $\mu_\theta-$a.a. x the matrix $\partial_x f_\theta(x) \in \mathbb{R}^{r \times d}$ is of full rank d, and that the p $d-$variate functions forming the columns of $\dot{b}_\theta$ are linearly independent. Then there exists $h_{t,\theta}(x) = h_\theta(x) \in \mathbb{R}^{r \times p}$ not depending on t such that $g_{t,\theta}(x,y) := h_\theta^T(x)\left(f_\theta(y) - \pi_{t,\theta}f(x)\right)$ satisfies the small $\Delta-$optimality condition (11). In particular, for $r = d$ one may choose

$$h_\theta^T(x) = \dot{b}_\theta^T(x)C^{-1}(x)\left(\partial_x f_\theta(x)\right)^{-1} \tag{16}$$

and this h_θ has linearly independent columns as required in Assumption A.

(ii) Suppose that $r \geq \dim(d)$, that for $\mu_\theta-$a.a. x, the matrix

$$\left(\ \partial_x f_\theta(x) \quad \partial_{xx}^2 f_\theta(x)\ \right) \in \mathbb{R}^{r \times \left(d + d^2\right)}$$

is of full rank $\dim(d)$ *and that the* p d^2-*variate functions forming the columns of* $\dot{C}_\theta$ *are linearly independent. Then there exists* $h_{t,\theta} = h_\theta \in \mathbb{R}^{r\times p}$ *not depending on* t *such that* $g_{t,\theta}(x,y) := h_\theta^T(x)\left(f_\theta(y) - \pi_{t,\theta}f(x)\right)$ *satisfies the small* $\Delta-$*optimality condition (12). In particular, for* $r = \dim(d)$ *one may choose*

$$h_\theta^T(x) = \left(\begin{array}{cc} 0_{p\times d} & \dot{C}_\theta^T(x)\left(C_\theta^{\otimes 2}(x)\right)^{-1} R \end{array} \right) \left(\begin{array}{cc} \partial_x f_\theta(x) & \partial^2_{xx} f_\theta(x) R \end{array} \right)^{-1}, \tag{17}$$

and this h_θ *has linearly independent columns as required in Assumption A.*

(iii) Suppose that $r \geq \dim(d)$*, that for* $\mu_\theta-$*a.a.* x*, the matrix*

$$\left(\begin{array}{cc} \partial_x f_\theta(x) & \partial^2_{xx} f_\theta(x) \end{array} \right) \in \mathbb{R}^{r\times\left(d+d^2\right)}$$

is of full rank $\dim(d)$*, that the* $p-p'$ $d-$*variate functions forming the columns of* $\dot{b}_{2,\theta}$ *are linearly independent, and that the* p' d^2-*variate functions forming the columns of* $\dot{C}_{1,\theta}$ *are linearly independent. Then there exists* $h_{t,\theta} = h_\theta \in \mathbb{R}^{r\times p}$ *not depending on* t *such that* $g_{t,\theta}(x,y) := h_\theta^T(x)\left(f_\theta(y) - \pi_{t,\theta}f(x)\right)$ *satisfies the small* $\Delta-$*optimality condition (13). In particular, for* $r = \dim(d)$ *one may choose*

$$h_\theta^T(x) = \left(\begin{array}{cc} 0_{p'\times d} & \dot{C}_{1,\theta}^T(x)\left(C_\theta^{\otimes 2}(x)\right)^{-1} R \\ \dot{b}_{2,\theta}^T(x) C_\theta^{-1}(x) & * \end{array} \right) \left(\begin{array}{cc} \partial_x f_\theta(x) & \partial^2_{xx} f_\theta(x) R \end{array} \right)^{-1}. \tag{18}$$

with $*$ *a* $(p-p')\times|J|$ *matrix, depending arbitrarily on* θ *and* x*. If* $*$ *is chosen* $= 0$*, then this* h_θ *has linearly independent columns as required in Assumption A.*

Proof. Since h_θ does not depend on t,

$$g_{0,\theta}(x,y) = h_\theta^T(x)\left(f_\theta(y) - f_\theta(x)\right)$$

whence

$$\partial_y g_{0,\theta}(x,x) = h_\theta^T(x)\partial_x f_\theta(x), \quad \partial^2_{yy} g_{0,\theta}(x,x) = h_\theta^T(x)\partial^2_{xx} f_\theta(x).$$

Thus, for each x, (11), (12) or (13) gives a system of linear equations for determining the elements of $h_\theta(x)$. The conditions of the theorem ensures

that these equations have at least one solution, and exactly one in case (i) if $r = d$ and in case (ii) if $r = \dim(d)$. (For case (ii), note that since $\partial^2_{x_i x_j} = \partial^2_{x_j x_i}$, the rank of $\partial^2_{xx} f_\theta$ is at most $|J|$. With h_θ given by (17), one now finds

$$\partial^2_{yy} g_{0,\theta}(x,x) R = \dot{C}^T_\theta(x) \left(C^{\otimes 2}_\theta(x)\right)^{-1} R$$

and using (15) this implies the second identity in (12)).

The assertions about h_θ having linearly independent columns follow readily from the assumptions made on the columns of $\dot{b}_\theta$ (case (i)), $\dot{C}_\theta$ (case (ii)) and $\dot{b}_{2,\theta}$ and $\dot{C}_{1,\theta}$ (case (iii)). ■

Theorem 2 only gives a solution for h_θ such that the relevant of (11), (12) or (13) is satisfied. To check small $\Delta-$optimality one further has to check the required integrability conditions (e.g. that all $h^{qk}_\theta \in L^2(\mu_\theta)$).

In Theorem 2 we have exhibited a concrete choice of small $\Delta-$optimal estimating functions from a given base (f^q_θ). But it is then easy to define a host of others that are also small $\Delta-$optimal, but may behave better for a given Δ, viz. flows $(g_{t,\theta})$ of the form

$$g^k_{t,\theta}(x,y) = \sum_{q=1}^{r} a^{qk}_\theta(t)\, h^{qk}_\theta(x) \left(f^q_\theta(y) - \pi_{t,\theta} f^q_\theta(x)\right) \tag{19}$$

with h^T_θ given by the relevant of (16), (17) or (18) and each $a^{qk}_\theta(t)$ a non-random function of t, continuous with $a^{qk}_\theta(0) = 1$: for this flow, $g_{0,\theta}$ is the same as for the original flow, so small $\Delta-$optimality still holds. However, there is no obvious optimal choice for the $a^{qk}_\theta(t)$: each $g^k_{t,\theta}$ given by (19) varies in a linear subspace when the $a^{qk}_\theta(t)$ are arbitrary, but the subspace depends on k so the projection technique from the proof of Proposition 1 does not apply. (To use the proposition and find optimal constants, one must consider a much larger space of estimating functions such as

$$g^k_{t,\theta}(x,y) = \sum_{k'=1}^{p} \sum_{q=1}^{r} a^{k,qk'}_\theta(t)\, h^{qk'}_\theta(x) \left(f^q_\theta(y) - \pi_{t,\theta} f^q_\theta(x)\right)$$

with arbitrary constants $a^{k,qk'}_\theta(t)$. The optimal constants can be found explicitly, but the expression involves the for practical purposes unwieldy inverse of the $pr \times pr-$matrix with $(kq, k'q')$'th element

$$E^\mu_\theta h^{qk}_\theta(X_0) \left(f^q_\theta(X_t) - \pi_{t,\theta} f^q_\theta(X_0)\right) h^{q'k'}_\theta(X_0) \left(f^{q'}_\theta(X_t) - \pi_{t,\theta} f^{q'}_\theta(X_0)\right)).$$

Remark 2 *In Theorem 2, case (iii), the expression (18) for $h_\theta^T(x)$ depends on the choice of $*$. A different small $\Delta-$optimal flow not involving this arbitrary matrix may be obtained as follows: take $r = \dim(d)$ and apart from the given base (f_θ^q) of dimension r, choose a second base $\left(\widetilde{f}_\theta^q\right)$ of dimension d. Assuming in addition to the assumptions from the theorem that $\partial_x \widetilde{f}_\theta(x)$ is non-singular, the flow $(\widetilde{g}_{t,\theta})$ given by the p', respectively $p-p'$ components*

$$\begin{aligned} \widetilde{g}_{1,t,\theta}(x,y) &= h_{1,\theta}^T(x)\left(f_\theta(y) - \pi_{t,\theta} f_\theta(x)\right) \\ \widetilde{g}_{2,t,\theta}(x,y) &= \widetilde{h}_\theta(x)\left(\widetilde{f}_\theta(y) - \pi_{t,\theta}\widetilde{f}_\theta(x)\right) \end{aligned}$$

with

$$\begin{aligned} h_{1,\theta}^T(x) &= \left(\begin{array}{cc} 0_{p'\times d} & \dot{C}_{1,\theta}^T(x)\left(C_\theta^{\otimes 2}(x)\right)^{-1} R \end{array}\right)\left(\begin{array}{cc} \partial_x f_\theta(x) & \partial_{xx}^2 f_\theta(x) R \end{array}\right)^{-1} \\ \widetilde{h}_\theta^T(x) &= \dot{b}_{2,\theta}^T(x) C_\theta^{-1}(x)\left(\partial_x \widetilde{f}_\theta(x)\right)^{-1} \end{aligned} \tag{21}$$

is small $\Delta-$optimal according to (13).

For general flows $(\widetilde{g}_{t,\theta})$ of the form (21), with the components of $\widetilde{g}_{1,t,\theta}$ and $\widetilde{g}_{2,t,\theta}$ varying in two different subspaces, optimal time dependent choices $h_{1,t,\theta}^T$ of $h_{1,\theta}^T$ and $\widetilde{h}_{t,\theta}^T$ of $\widetilde{h}_\theta^T$ cannot be found using e.g. Proposition 1. However partly optimal candidates may be determined using the proposition twice: first with base (f_θ^q) in the model with $\theta_{p'+1},\ldots,\theta_p$ assumed known, yielding $h_{1,t,\theta}^T$, second with base $\left(\widetilde{f}_\theta^{q'}\right)$ in the model with $\theta_1,\ldots,\theta_{p'}$ known, yielding $\widetilde{h}_{t,\theta}$.

We return now to the discussion of the optimal martingale estimating function (8) determined by the base $(f_\theta^1,\ldots,f_\theta^r)$. Since for any $t=\Delta>0$, $g_{\Delta,\theta}^{\mathrm{opt}}$ is better than a $g_{\Delta,\theta}$ where $h_{\Delta,\theta} = h_\theta$ is determined as in Theorem 2, the flow $\mathcal{G}^{\mathrm{opt}} = \left(g_{t,\theta}^{\mathrm{opt}}\right)$ should be small $\Delta-$optimal if $r \geq d$ in case (i) or $r \geq \dim(d)$ in cases (ii) or (iii). What however does not follow from Theorem 2 is that $\mathcal{G}^{\mathrm{opt}}$ satisfies the conditions (11), (12) or (13). We shall now verify that this is the case (for $r=d$ and $\dim(d)$ respectively) and also argue that the lower bounds d and $\dim(d)$ for r cannot be improved upon.

Theorem 3 *For the optimal flow $\mathcal{G}^{\mathrm{opt}} = \left(g_{t,\theta}^{\mathrm{opt}}\right)$ of martingale estimating functions with base $(f_\theta^1,\ldots,f_\theta^r)$ it holds that:*

(i) If $r = d$ and the matrix $\partial_x f_\theta(x)$ is non-singular for $\mu_\theta-$a.a. x, then $g^{\mathrm{opt}}_{0,\theta}(x,y) = \lim_{t\to 0} g^{\mathrm{opt}}_{t,\theta}(x,y)$ and (11) holds,

$$\partial_y g^{\mathrm{opt}}_{0,\theta}(x,x) = \dot{b}^T_\theta(x) C^{-1}(x).$$

(ii) If $r = \dim(d)$ and the matrix

$$\left(\begin{array}{cc} \partial_x f_\theta(x) & \partial^2_{xx} f_\theta(x) \end{array}\right) \in \mathbb{R}^{r\times\left(d+d^2\right)}$$

is of full rank $\dim(d)$ for $\mu_\theta-$a.a. x, then $g^{\mathrm{opt}}_{0,\theta}(x,y) = \lim_{t\to 0} t g^{\mathrm{opt}}_{t,\theta}(x,y)$ and (12) holds,

$$\partial_y g^{\mathrm{opt}}_{0,\theta}(x,x) = 0, \quad \partial^2_{yy} g^{\mathrm{opt}}_{0,\theta}(x,x) = \dot{C}^T_\theta(x)\left(C_\theta(x)^{\otimes 2}\right)^{-1}.$$

(iii) If $r = \dim(d)$ and the matrix

$$\left(\begin{array}{cc} \partial_x f_\theta(x) & \partial^2_{xx} f_\theta(x) \end{array}\right) \in \mathbb{R}^{r\times\left(d+d^2\right)}$$

is of full rank $\dim(d)$, then $g^{\mathrm{opt}}_{0,\theta}(x,y) = \lim_{t\to 0} \left(\begin{array}{c} t g^{\mathrm{opt}}_{1,t,\theta} \\ g^{\mathrm{opt}}_{2,t,\theta} \end{array}\right)$ and (13) holds,

$$\partial_y g^{\mathrm{opt}}_{0,\theta}(x,x) = \left(\begin{array}{c} 0_{p'\times d} \\ \dot{b}^T_{2,\theta}(x) C^{-1}_{\theta(x)} \end{array}\right), \quad \partial^2_{yy} g^{\mathrm{opt}}_{1,0,\theta}(x,x) = \dot{C}^T_{1,\theta}(x)\left(C^{\otimes 2}_\theta(x)\right)^{-1}. \tag{22}$$

If $r < d$ in (i) or $r < \dim(d)$ in (ii) or (iii), (11), resp. (12) or (13) will not be satisfied except possibly for a special choice of base $(f^1_\theta, \ldots, f^r_\theta)$.

Proof. The main difficulty consists in finding $g^{\mathrm{opt}}_{0,\theta}$ from (8) and (7). First note that for smooth functions ϕ, ψ, since

$$\begin{array}{rcl} \pi_{t,\theta}(\phi\psi) & = & \phi\psi + t A_\theta(\phi\psi) + o(t), \\ (\pi_{t,\theta}\phi)(\pi_{t,\theta}\psi) & = & (\phi + t A_\theta \phi)(\psi + t A_\theta \psi) + o(t) \end{array}$$

(with $o(t)/t = o(t,x)/t \to 0$ for each x) and

$$A_\theta(\phi\psi) = (A_\theta\phi)\,\psi + \phi\,(A_\theta\psi) + (\partial_x\phi)\, C_\theta\, (\partial_x\psi)^T + o(t) \tag{23}$$

it follows that

$$\pi_{t,\theta}(\phi\psi) - (\pi_{t,\theta}\phi)(\pi_{t,\theta}\psi) = t(\partial_x\phi) C_\theta (\partial_x\psi)^T + o(t). \tag{24}$$

Now use (24) with $\phi = f^q$, $\psi = f^{q'}$, $1 \le q, q' \le r$ (with r arbitrary at the moment) to obtain

$$\pi_{t,\theta}\left(f_\theta f_\theta^T\right) - (\pi_{t,\theta}f_\theta)(\pi_{t,\theta}f_\theta)^T = t(\partial_x f_\theta) C_\theta (\partial_x f_\theta)^T + o(t) \tag{25}$$

and it is seen that the main term on the right evaluated at x is a non-singular $r \times r-$matrix only if $r \le d$ and $\partial_x f(x) \in \mathbb{R}^{r\times d}$ is of full rank r.

To find $g_{0,\theta}^{\text{opt}}$ we also need to aproximate the factor $\partial_\theta \pi_{t,\theta} f_\theta - \pi_{t,\theta}\dot{f}$ from (7). Assuming that $\partial_\theta o(t) = o(t)$ and that $\dot{f}_\theta$ is smooth enough,

$$\begin{aligned}
\partial_\theta \pi_{t,\theta} f_\theta &= \partial_\theta (f_\theta + tA_\theta f_\theta + o(t)) \\
&= \dot{f}_\theta + t\left(A_\theta \dot{f}_\theta + (\partial_x f_\theta)\dot{b}_\theta + \tfrac{1}{2}\left(\partial_{xx}^2 f_\theta\right)\dot{C}_\theta\right) + o(t), \\
\pi_{t,\theta}\dot{f}_\theta &= \dot{f}_\theta + tA_\theta \dot{f}_\theta + o(t)
\end{aligned}$$

and thus

$$\partial_\theta \pi_{t,\theta} f_\theta - \pi_{t,\theta}\dot{f}_\theta = t\left((\partial_x f_\theta)\dot{b}_\theta + \tfrac{1}{2}\left(\partial_{xx}^2 f_\theta\right)\dot{C}_\theta\right) + o(t). \tag{26}$$

Case (i). Since $\dot{C}_\theta \equiv 0$, (26) reduces to

$$\partial_\theta \pi_{t,\theta} f_\theta - \pi_{t,\theta}\dot{f}_\theta = t(\partial_x f_\theta)\dot{b}_\theta + o(t),$$

and therefore, using (25) it follows that if $r \le d$,

$$\begin{aligned}
g_{0,\theta}^{\text{opt}}(x,y) &= \lim_{t\to 0} g_{t,\theta}^{\text{opt}}(x,y) \\
&= \dot{b}_\theta^T(x)(\partial_x f_\theta)^T(x)\left[\partial_x f_\theta(x) C(x)(\partial_x f_\theta)^T(x)\right]^{-1}(f_\theta(y) - f_\theta(x))
\end{aligned}$$

so that

$$\partial_y g_{0,\theta}^{\text{opt}}(x,x) = \dot{b}_\theta^T(x)(\partial_x f_\theta)^T(x)\left[\partial_x f_\theta(x) C(x)(\partial_x f_\theta)^T(x)\right]^{-1}\partial_x f_\theta(x). \tag{27}$$

For $r = d$ this reduces to $\dot{b}_\theta^T(x)C^{-1}(x)$ as wanted. For $r < d$ the $d\times d-$matrix appearing as a factor to the right of $\dot{b}_\theta^T(x)$, has rank r, hence can never equal the non-singular $d \times d-$matrix $C^{-1}(x)$. (However, it may still be possible to

obtain $\partial_y g_{0,\theta}^{\mathrm{opt}}(x,x) = \dot{b}_\theta^T(x)C^{-1}(x)$: multiplying from the right by $C(x)$, it is seen that this holds for $\partial_y g_{0,\theta}^{\mathrm{opt}}(x,x)$ given by (27) iff the $d \times d-$matrix

$$(\partial_x f_\theta)^T(x)\left[\partial_x f_\theta(x)C(x)\left(\partial_x f_\theta\right)^T(x)\right]^{-1}\partial_x f_\theta(x)C(x)$$

acts as the identity on the subspace $L_b = \mathrm{rowspan}\left(\dot{b}_\theta^T(x)\right)$ (multiplying by row vectors from the left). If $r < d$ this will be possible only if $\dim(L_b) < d$ (e.g. if $p < d$) and a special choice of f_θ determined by b_θ and C is required).

Case (ii). Assume that $r > d$. Then the main term on the right of (25) becomes singular and it is therefore necessary to expand further. But from the basic expansion

$$\pi_{t,\theta}\varphi = \varphi + tA_\theta\varphi + \tfrac{1}{2}t^2A_\theta^2\varphi + o(t^2),$$

using (23) repeatedly it eventually follows that

$$\pi_{t,\theta}\left(f_\theta f_\theta^T\right) - \left(\pi_{t,\theta}f_\theta\right)\left(\pi_{t,\theta}f_\theta\right)^T = t\left(\partial_x f_\theta\right)C_\theta\left(\partial_x f_\theta\right)^T + \tfrac{1}{2}t^2Q + o(t^2) \tag{28}$$

with Q of the form

$$Q = \left(\partial_{xx}^2 f_\theta\right)C_\theta^{\otimes 2}\left(\partial_{xx}^2 f_\theta\right)^T + \left(\partial_x f_\theta\right)S + S^T\left(\partial_x f_\theta\right)^T \tag{29}$$

for some $S(x) \in \mathbb{R}^{d\times r}$. By Lemma 4 from the appendix therefore (with $A = \left(\partial_x f_\theta\right)C_\theta\left(\partial_x f_\theta\right)^T$, $B = \frac{1}{2}Q$)

$$\lim_{t\to 0} t^2\left[\pi_{t,\theta}\left(f_\theta f_\theta^T\right) - \left(\pi_{t,\theta}f_\theta\right)\left(\pi_{t,\theta}f_\theta\right)^T\right]^{-1} = \mathcal{O}_2^T\left(\mathcal{O}_2\tfrac{1}{2}Q\mathcal{O}_2^T\right)^{-1}\mathcal{O}_2 \tag{30}$$

where $\mathcal{O}(x) = \binom{\mathcal{O}_1(x)}{\mathcal{O}_2(x)} \in \mathbb{R}^{r\times r}$ is orthogonal for each x, $\mathcal{O}_1(x)$ comprising the first d and $\mathcal{O}_2(x)$ the last $r-d$ rows of $\mathcal{O}(x)$, and satisfies

$$\mathcal{O}(x)\left(\partial_x f_\theta(x)\right)C_\theta(x)\left(\partial_x f_\theta\right)^T(x)\mathcal{O}^T(x) = \mathrm{diag}\left(\lambda_1(x),\ldots,\lambda_d(x),0,\ldots,0\right) \tag{31}$$

with $\lambda_1(x),\ldots,\lambda_d(x) > 0$ the non-zero eigenvalues for $\left(\partial_x f_\theta(x)\right)C_\theta(x)\left(\partial_x f_\theta\right)^T(x)$.

But from (31) follows that

$$\mathcal{O}_2(x)\left(\partial_x f_\theta(x)\right)C_\theta(x)\left(\partial_x f_\theta\right)^T(x)\mathcal{O}_2^T(x) = 0$$

or, since $C_\theta(x) \succ 0$, that

$$\mathcal{O}_2\partial_x f_\theta = 0. \tag{32}$$

Combining (30) with (26) and using (32) it follows that

$$\begin{aligned} g_{0,\theta}^{\mathrm{opt}}(x,y) &= \lim_{t\to 0} t g_{t,\theta}^{\mathrm{opt}}(x,y) \\ &= \dot{C}_\theta^T \left(\partial_{xx}^2 f_\theta\right)^T \mathcal{O}_2^T \left[\mathcal{O}_2 \left(\partial_{xx}^2 f_\theta\right) C_\theta^{\otimes 2} \left(\partial_{xx}^2 f_\theta\right)^T \mathcal{O}_2^T\right]^{-1} \\ &\quad \times \mathcal{O}_2 \left(f_\theta(y) - f_\theta(x)\right) \end{aligned}$$

with all factors to the left of $f_\theta(y)$ evaluated at x. Using (32) it is clear that $\partial_y g_{0,\theta}^{\mathrm{opt}}(x,x) = 0$ always and hence, to obtain (12), it remains to check whether (omitting the argument x with $\partial_{xx}^2 g_{0,\theta}^{\mathrm{opt}}$ short for $\partial_{yy}^2 g_{0,\theta}^{\mathrm{opt}}(x,x)$)

$$\partial_{xx}^2 g_{0,\theta}^{\mathrm{opt}} = \dot{C}_\theta^T \left(\partial_{xx}^2 f_\theta\right)^T \mathcal{O}_2^T \left[\mathcal{O}_2 \left(\partial_{xx}^2 f_\theta\right) C_\theta^{\otimes 2} \left(\partial_{xx}^2 f_\theta\right)^T \mathcal{O}_2^T\right]^{-1} \mathcal{O}_2 \partial_{xx}^2 f_\theta \quad (33)$$

equals $\dot{C}_\theta^T \left(C_\theta^{\otimes 2}\right)^{-1}$.

To achieve this we now assume that $r = \dim(d)$, so that $r - d = |J|$ and use the assumption from the theorem that $\partial_{xx}^2 f_\theta(x)$ has full rank $|J|$ for all x. Then $\Gamma := \left(\partial_{xx}^2 f_\theta\right) R$ also has rank $|J|$ and $\mathcal{O}_2\Gamma \in \mathbb{R}^{J\times J}$ is non-singular and using that $\partial_{xx}^2 f_\theta = \partial_{xx}^2 f_\theta \left(R\tilde{R}\right)$ (cf. (15)), (33) therefore gives

$$\begin{aligned} \left(\partial_{xx}^2 g_{0,\theta}^{\mathrm{opt}}\right) R &= \dot{C}_\theta^T \tilde{R}^T \Gamma^T \mathcal{O}_2^T \left[\mathcal{O}_2 \Gamma \tilde{R} C_\theta^{\otimes 2} \tilde{R}^T \Gamma^T \mathcal{O}_2^T\right]^{-1} \mathcal{O}_2 \Gamma \\ &= \dot{C}_\theta^T \tilde{R}^T \left(\tilde{R} C_\theta^{\otimes 2} \tilde{R}^T\right)^{-1}. \end{aligned} \quad (34)$$

That $\partial_{xx}^2 g_{0,\theta}^{\mathrm{opt}} = \dot{C}_\theta^T \left(C_\theta^{\otimes 2}\right)^{-1}$ will follow from (cf. (15))

$$\left(\partial_{xx}^2 g_{0,\theta}^{\mathrm{opt}}\right) R = \dot{C}_\theta^T \left(C_\theta^{\otimes 2}\right)^{-1} R,$$

and that the right hand side here indeed equals that of (34) is verified multiplying by $\tilde{R} C_\theta^{\otimes 2} \tilde{R}^T$ from the right again appealing to (15).

Case (iii). Here we initially proceed as in case (ii), arriving at (cf. (28))

$$\begin{aligned} g_{t,\theta}^{\mathrm{opt}}(x,y) &= \left(\dot{b}_\theta^T \left(\partial_x f_\theta\right)^T + \tfrac{1}{2} \dot{C}_\theta^T \left(\partial_{xx}^2 f_\theta\right)^T + o(1)\right) \quad (35) \\ &\quad \times \left(\left(\partial_x f_\theta\right) C_\theta \left(\partial_x f_\theta\right)^T + \tfrac{1}{2} t Q + o(t)\right)^{-1} \left(f_\theta(y) - f_\theta(x) + o(1)\right) \end{aligned}$$

with Q as in (29).

Considering first the last $p-p'$ components of $g_{t,\theta}^{\mathrm{opt}}$, since by assumption $\dot{C}_{2,\theta}\equiv 0$, it follows from Lemma 4 that

$$\begin{aligned} g_{2,t,\theta}^{\mathrm{opt}}(x,y) &= \dot{b}_{2,\theta}^{T}\left(\partial_x f_\theta\right)^{T}\left(\tfrac{1}{t}\mathcal{O}_2^{T}\left(\mathcal{O}_2\tfrac{1}{2}Q\mathcal{O}_2^{T}\right)^{-1}\mathcal{O}_2+N\right)\\ &\quad\times\left(f_\theta\left(y\right)-f_\theta\left(x\right)\right)+o(1), \end{aligned}$$

which because of (32) in the limit reduces to

$$g_{2,0,\theta}^{\mathrm{opt}}(x,y)=\lim_{t\to 0} g_{2,t,\theta}^{\mathrm{opt}}(x,y)=\dot{b}_{2,\theta}^{T}\left(\partial_x f_\theta\right)^{T}N\left(f_\theta\left(y\right)-f_\theta\left(x\right)\right)$$

with N of the form

$$N=\mathcal{O}_1^{T}\left(\mathcal{O}_1\left(\partial_x f_\theta\right)C_\theta\left(\partial_x f_\theta\right)^{T}\mathcal{O}_1^{T}\right)^{-1}\mathcal{O}_1+\mathcal{O}_2^{T}\widetilde{S}+\widetilde{S}^{T}\mathcal{O}_2.$$

But then, again using (32) and since $\mathcal{O}_1\left(\partial_x f_\theta\right)\in\mathbb{R}^{d\times d}$ is non-singular

$$\begin{aligned} \partial_y g_{2,0,\theta}^{\mathrm{opt}}(x,x) &= \dot{b}_{2,\theta}^{T}\left(\partial_x f_\theta\right)^{T}\mathcal{O}_1^{T}\left(\mathcal{O}_1\left(\partial_x f_\theta\right)C_\theta\left(\partial_x f_\theta\right)^{T}\mathcal{O}_1^{T}\right)^{-1}\mathcal{O}_1\left(\partial_x f_\theta\right)\\ &= \dot{b}_{2,\theta}^{T}C_\theta^{-1} \end{aligned}$$

as wanted in the first part of (22).

As for the first p' components of $g_{t,\theta}^{\mathrm{opt}}$, obtain from (35) that

$$\begin{aligned} tg_{1,t,\theta}^{\mathrm{opt}}(x,y) &= \left(\dot{b}_{1,\theta}^{T}\left(\partial_x f_\theta\right)^{T}+\tfrac{1}{2}\dot{C}_{1,\theta}^{T}\left(\partial_{xx}^{2}f_\theta\right)^{T}\right)\mathcal{O}_2^{T}\left(\mathcal{O}_2\tfrac{1}{2}Q\mathcal{O}_2^{T}\right)^{-1}\mathcal{O}_2\\ &\quad\times\left(f_\theta\left(y\right)-f_\theta\left(x\right)\right)+o(1), \end{aligned}$$

whence

$$\begin{aligned} g_{1,0,\theta}^{\mathrm{opt}}(x,y) &= \lim_{t\to 0} tg_{1,t,\theta}^{\mathrm{opt}}(x,y)\\ &= \tfrac{1}{2}\dot{C}_{1,\theta}^{T}\left(\partial_{xx}^{2}f_\theta\right)^{T}\mathcal{O}_2^{T}\left(\mathcal{O}_2\tfrac{1}{2}\left(\partial_{xx}^{2}f_\theta\right)C^{\otimes 2}\left(\partial_{xx}^{2}f_\theta\right)^{T}\mathcal{O}_2^{T}\right)^{-1}\\ &\quad\times\mathcal{O}_2\left(f_\theta\left(y\right)-f_\theta\left(x\right)\right) \end{aligned}$$

once again using (32). But then (32) also gives

$$\partial_y g_{1,0,\theta}^{\mathrm{opt}}(x,x)=0.$$

and arguing exactly as in the last part of case (ii), one finally finds that

$$\partial_{yy}^{2}g_{1,0,\theta}^{\mathrm{opt}}(x,x)=\dot{C}_{1,\theta}^{T}\left(C_\theta^{\otimes 2}\right)^{-1},$$

and we have completed the proof of (22). ■

We have not shown that (11), respectively (12), (13) are satisfied for the optimal martingale estimating function when $r > d$, resp. $r > \dim(d)$. For case (i) with $r > d$ one may copy the argument involving $g^{\text{opt}}_{2,\theta}$ given in case (iii) above. For cases (ii) and (iii), if $r > \dim(d)$ a further expansion of (28) together with a refinement of the lemma is required, since e.g. the columns of $A = (\partial_x f_\theta)\, C_\theta\, (\partial_x f_\theta)^T$ and $B = \frac{1}{2}Q$ cannot span a subspace of dimension r. We believe however that the relevant of (12) or (13) is still valid for the optimal martingale estimating function, even if $r > \dim(d)$.

Remark 3 *For $d = 1$, Bibby and Sørensen [1] studied martingale estimating functions with the one-dimensional ($r = 1$) base $f(x) = x$, and apart from deriving the optimal estimating function G^*_n ((2.15) in [1]), also suggested the use of the approximately optimal $\widetilde{G}_n$ ((2.14) in [1]). As they point out, the weights for $\widetilde{G}_n$ are arrived at by replacing the true transition probabilities as they appear in (7) by the Gaussian approximations corresponding to the Euler scheme, i.e. the conditional distribution of X_t given $X_0 = x$ is approximated by the normal distribution $n_{t,\theta}(x, \cdot)$ with mean $x + tb_\theta(x)$ and variance $tC_\theta(x)$. We shall now discuss how the use of this approximation may still lead to estimating functions that are small $\Delta-$optimal. We shall only consider one-dimensional diffusions but allow p and r to be arbitrary.*

In terms of $n_{t,\theta}(x, \cdot)$, the approximation to $\pi_{t,\theta}\phi$ becomes

$$\begin{aligned}
\pi_{t,\theta}\phi(x) &\approx \int n_{t,\theta}(x, dy)\ \phi(y) \\
&= \phi(x) + t\left(b_\theta(x)\phi'(x) + \tfrac{1}{2}C_\theta(x)\phi''(x)\right) \\
&\quad + \tfrac{1}{2}t^2\left(b^2_\theta(x)\phi''(x) + b_\theta(x)C_\theta(x)\phi'''(x) + \tfrac{1}{4}C^2_\theta(x)\phi''''(x)\right) + o(t^2)
\end{aligned}$$

resulting in the approximations (omitting the argument x),

$$\pi_{t,\theta}(\phi\psi) - (\pi_{t,\theta}\phi)\,(\pi_{t,\theta}\psi) \approx n_{t,\theta}(\phi, \psi)$$

where

$$\begin{aligned}
n_{t,\theta}(\phi, \psi) &= tC_\theta\phi'\psi' \\
&\quad + \tfrac{1}{2}t^2[C^2_\theta\phi''\psi'' + 2b_\theta C_\theta(\phi''\psi' + \phi'\psi'') \\
&\quad + C^2_\theta(\phi'''\psi' + \phi'\psi''')] + o(t^2),
\end{aligned} \tag{36}$$

and

$$\partial_\theta \left(\pi_{t,\theta}\phi\right) \approx t\left(\dot{b}_\theta \phi' + \tfrac{1}{2}\dot{C}_\theta \phi''\right) + o(t).$$

Given a base $(f^q)_{1\le q\le r}$ *(for convenience assumed not to depend on* θ*), now consider the martingale estimating function (cf. (7) and (8))*

$$g^{\text{norm}}_{t,\theta}(x,y) = \left(h^{\text{norm}}_{t,\theta}\right)^T(x)\left(f(y) - \pi_{t,\theta}f(x)\right) \tag{37}$$

where

$$h^{\text{norm}}_{t,\theta} = \left(N_{t,\theta}f\right)^{-1} t\left(\partial_x f\, \dot{b}_\theta + \tfrac{1}{2}\partial^2_{xx} f\, \dot{C}_\theta\right)$$

writing $N_{t,\theta}f$ *for the* $r\times r-$*matrix with* (q,q')*'th element* $n_{t,\theta}\left(f^q, f^{q'}\right)$*. From the approximations above it follows that*

$$N_{t,\theta}f = t\left(\partial_x f\right) C_\theta \left(\partial_x f\right)^T + \tfrac{1}{2}t^2\widetilde{Q} + o(t^2)$$

with $\widetilde{Q}$ *of the same form as* Q *in (29), and this is enough for the proof of Theorem 3 to carry over and yield the following result:* the estimating function (37) is small $\Delta-$optimal (in the sense of (11), (12) or (13)) in case (i) if $r = 1$ and in cases (ii) or (iii) if $r = \dim(1) = 2$. In case (i) with $r = 1$, instead of using $N_{t,\theta}f = n_{t,\theta}(f,f)$ one may use

$$N_{t,\theta}f = tC_\theta\left(\partial_x f\right)^2$$

corresponding to the main term in (36). In cases (ii) and (iii) it is essential to use (36) as it stands and that $r \ge 2$.

3 Examples

We shall ilustrate the results of the previous sections through two examples.

3.1 A generalized Cox-Ingersoll-Ross process

Consider the one-dimensional ($d = 1$) SDE

$$dX_t = \left(aX_t^{2\gamma-1} + bX_t\right)\, dt + \sigma X_t^\gamma\, dB_t \tag{38}$$

where $a, b \in \mathbb{R}$, $\gamma \ne 1$ and $\sigma > 0$. For $\gamma = \frac{1}{2}$ this is the SDE for the Cox-Ingersoll-Ross process (CIR–process, see (39) below). The generalization

(38) is arrived at by considering all powers $\widetilde{X}^\rho$ of a CIR with $\rho \neq 0$, more precisely if X solves (38), then the associated CIR–process is $\widetilde{X} = X^{2-2\gamma}$ solving

$$d\widetilde{X}_t = \left(\widetilde{a} + \widetilde{b}X_t\right)\, dt + \widetilde{\sigma}\sqrt{\widetilde{X}_t}\, dB_t \tag{39}$$

where

$$\widetilde{b} = (2-2\gamma)\, b, \quad \widetilde{\sigma}^2 = (2-2\gamma)^2\, \sigma^2, \quad \widetilde{a} - \tfrac{1}{2}\widetilde{\sigma}^2 = (2-2\gamma)\left(a - \tfrac{1}{2}\sigma^2\right), \tag{40}$$

(which also explains why $\gamma = 1$ is not allowed in (38)).

Because of the connection to the CIR–process, the model described by (38) is much simpler to handle than the more standard CKLS–model,

$$dX_t = (a + bX_t)\, dt + \sigma X_t^\gamma\, dB_t,$$

in particular, for (38) it is easy to find martingale estimating functions of the type considered in the preceding sections.

In (38) the parameter space has dimension $p = 4$. We shall want X to be strictly positive and ergodic, which happens iff the associated CIR–process $\widetilde{X}$ is strictly positive and ergodic, i.e. $\widetilde{b} < 0$ and $2\widetilde{a} \geq \widetilde{\sigma}^2$, or equivalently, either $\gamma < 1$, $b < 0$, $2a \geq \sigma^2$ or $\gamma > 1$, $b > 0$, $2a \leq \sigma^2$. As *open* parameter set we shall therefore use

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

Note that if $\theta = (a, b, \gamma, \sigma^2) \in \Theta$ and $\rho \neq 0$, then X^ρ solves (38) with parameters $\theta^* = (a^*, b^*, \gamma^*, \sigma^{*2})$ given by

$$b^* = \rho b, \quad \sigma^{*2} = \rho^2\sigma^2, \quad 2 - 2\gamma^* = \frac{1}{\rho}(2-2\gamma), \quad a^* - \tfrac{1}{2}\sigma^{*2} = \rho\left(a - \tfrac{1}{2}\sigma^2\right).$$

In particular, taking $\rho < 0$ corresponds to a switch from $\gamma < 1$ to $\gamma^* > 1$ (or from $\gamma > 1$ to $\gamma^* < 1$).

Since the invariant distribution for $\widetilde{X}$ is a $\Gamma-$distribution, the invariant distribution for X is that of a $\Gamma-$distributed random variable raised to the power $(2-2\gamma)^{-1}$. The density is

$$\mu_\theta(x) = \frac{|2-2\gamma|}{\Gamma\left(\frac{2\widetilde{a}}{\widetilde{\sigma}^2}\right)\left(\frac{(2\gamma-2)\sigma^2}{2b}\right)^{2\widetilde{a}/\widetilde{\sigma}^2}}\, x^{\frac{2a}{\sigma^2}-2\gamma} \exp\left(-\tfrac{2b}{(2\gamma-2)\sigma^2}\, x^{2-2\gamma}\right) \tag{41}$$

for $x > 0$, where (cf (40))

$$\frac{2\widetilde{a}}{\widetilde{\sigma}^2} = \frac{2a}{(2-2\gamma)\,\sigma^2} + \frac{1-2\gamma}{2-2\gamma}.$$

(For $\gamma = \frac{1}{2}$ the familiar invariant $\Gamma-$density for the CIR–process is obtained).

Because a $\Gamma-$distribution has finite moments of all orders $m \in \mathbb{N}$ we have $E_\theta^\mu X_0^{(2\gamma-2)m} < \infty$ for all $m \in \mathbb{N}$, and since

$$E_\theta^\mu X_t^{(2\gamma-2)m} = \int_0^\infty dx\, \mu_\theta(x) \pi_{t,\theta} x^{(2\gamma-2)m}$$

(where $\pi_{t,\theta} x^\beta$ is short for $\pi_{t,\theta} f(x)$ for $f(y) = y^\beta$), also

$$\pi_{t,\theta} x^{(2\gamma-2)m} < \infty$$

for all $t > 0$, $m \in \mathbb{N}$ and (Lebesgue almost all) $x > 0$.

The conditional moments for a CIR–process are known and in any case easy to find using polynomial martingales: for $m \in \mathbb{N}$, let $\widetilde{\xi}_m$ be the m'th moment in the invariant distribution for $\widetilde{X}$,

$$\widetilde{\xi}_m = \left(-\frac{\widetilde{\sigma}^2}{2\widetilde{b}}\right)^m \frac{\Gamma\left(\frac{2\widetilde{a}}{\widetilde{\sigma}^2} + m\right)}{\Gamma\left(\frac{2\widetilde{a}}{\widetilde{\sigma}^2}\right)}$$

and verify, for instance using induction on m and Itô's formula, that $M^{(m)}$ is a mean-zero martingale (see the note below) under each P_θ^x, where

$$M_t^{(m)} = e^{-\widetilde{b}mt} \sum_{i=1}^m \beta_i^{(m)} \left(X_t^{(2-2\gamma)i} - \widetilde{\xi}_i\right) \tag{42}$$

with

$$\beta_i^{(m)} = \frac{1}{\widetilde{\xi}_i} (-1)^{i-1} \binom{m}{i}.$$

(Equivalently, for each m, the polynomium $\sum_i \beta_i^{(m)} \left(x^i - \widetilde{\xi}_i\right)$ of degree m is an eigenfunction for the generator for the CIR–process (39) corresponding to the eigenvalue $\widetilde{b}m$, see Kessler and Sørensen [5] for estimating functions built from eigenfunctions, and their Example 2.1 for the CIR–process).

Note. Because all conditional moments for the ergodic CIR–process are finite, one verifies directly that the local martingale $M^{(m)}$ satisfies $E_\theta^x \left[M^{(m)}\right]_t <$

∞ for all x and t, in particular $M^{(m)}$ is a true martingale under P_θ^x (L^2-bounded on $[0,t]$ for all t).

Turning now to the problem of estimating θ from discrete observations of X, it is clear that the model (38) belongs to case (iii) with $p=4$, $p'=2$, so we shall apply Theorems 2 and 3 for that case with $r = 1+\dim(1) = 2$. In view of the above, a simple candidate for the base (f^1, f^2) is

$$f^1(x) = x^{2-2\gamma}, \quad f^2(x) = x^{4-4\gamma}, \tag{43}$$

which trivially satisfies Assumption A. Note that f^1, f^2 both depend on θ, cf. the comment immediately preceding Proposition 1.

Using Theorem 2 with $*$ in (18) equal to 0, and listing the parameters in the order γ, σ^2, a, b one finds that

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

and eventually arrives at the estimating function

$$g_{t,\theta}(x,y) = \begin{pmatrix} -2\log x & x^{2\gamma-2}\log x \\ -2 & x^{2\gamma-2} \\ 2(3-4\gamma)x^{2\gamma-2} & -(1-2\gamma)x^{4\gamma-4} \\ 2(3-4\gamma) & -(1-2\gamma)x^{2\gamma-2} \end{pmatrix} \begin{pmatrix} y^{2-2\gamma} - \pi_{t,\theta}x^{2-2\gamma} \\ y^{4-4\gamma} - \pi_{t,\theta}x^{4-4\gamma} \end{pmatrix}, \tag{44}$$

which requires the use of (42) for $m = 1, 2$ in order to find the conditional expectations.

That $g_{t,\theta}$ given by (44) indeed satisfies the conditions (13) for small $\Delta-$optimality is most easily verified directly. Note that the asserted (Theorem 2) linear independence between the columns in h_θ, i.e. the functions comprising the rows in the $4\times 2-$matrix in (44) where rows 2 and 4 are similar, holds precisely because $\gamma \neq 1$. Note that this similarity also implies that the $4\times 2-$matrix in (44) may be replaced the simpler

$$\begin{pmatrix} -2\log x & x^{2\gamma-2}\log x \\ 1 & 0 \\ 2(3-4\gamma)x^{2\gamma-2} & -(1-2\gamma)x^{4\gamma-4} \\ 0 & x^{2\gamma-2} \end{pmatrix}$$

without affecting the solutions to the estimating equations.

For the flow $(g_{t,\theta})$ given by (44) one still needs to check the integrability assumptions from Jacobsen [2] Theorem 7.5, and the conditions on estimating flows made prior to that theorem. In the case at hand the conditions in particular amount to requiring that $E_\theta^\mu \left| g_{t,\theta}^k \left(X_0, X_t \right) \right|^K < \infty$ for all components k and moderate values of $K \in \mathbb{N}$. The problem is the appearance of powers $x^{2\gamma-2}$ and $x^{4\gamma-4}$ in the expression for $g_{t,\theta}$, which translates into negative powers $\widetilde{X}_0^{-1}$ and $\widetilde{X}_0^{-2}$ of the CIR–process $\widetilde{X}$, and of course e.g. $E_\theta^\mu \widetilde{X}_0^{-K} = E_\theta^\mu X_0^{(2\gamma-2)K} < \infty$ iff $\frac{2\widetilde{a}}{\widetilde{\sigma}^2} > K$. Thus some care should be taken before applying (44), at least it must be assumed that $\frac{2\widetilde{a}}{\widetilde{\sigma}^2}$ be suitably large

To find the optimal martingale estimating function with base f^1, f^2 given by (43), one needs (42) also for $m = 3, 4$ *and* conditional moments involving logarithms, see (7) in which the term $\pi_{t,\theta} \dot{f}_\theta$ appears. The latter moments are easy to find in terms of the conditional expectation $E_\theta \left(\log \widetilde{X}_t \middle| \widetilde{X}_0 = \widetilde{x} \right)$ for the CIR–process starting at an arbitrary level $\widetilde{x}$, but the explicit form for this is unpleasant to work with.

Whether one uses the small $\Delta-$optimal flow (44) or the optimal flow, since $d = 1$ a slight improvement in efficiency may be gained by symmetrizing, using e.g. $\frac{1}{2} \left(g_{t,\theta} \left(x, y \right) + g_{t,\theta} \left(y, x \right) \right)$ instead of (44), cf. Jacobsen [2], Proposition 6.1, and the discussion there about time reversal.

3.2 The finite-dimensional Gaussian diffusions

We consider now the $d-$dimensional diffusion

$$dX_t = \left(A + BX_t \right) \, dt + D \, dW_t \tag{45}$$

where the unknown parameters are $A \in \mathbb{R}^{d \times 1}$, $B \in \mathbb{R}^{d \times d}$ and $C := DD^T \in \mathbb{R}^{d \times d}$, with the symmetric matrix C assumed strictly positive definite. (In this subsection the symbol B is used to denote the matrix of linear drift parameters and the driving $d-$dimensional Brownian motion is denoted W instead of B). Thus

$$p' = |J| \, , \quad p = |J| + d + d^2.$$

The diffusion (45) has Gaussian transitions (for the expectation and second order momments, see (46) and (47) below) and is ergodic iff $\text{spec}(B) \subset \{ \lambda \in \mathbb{C} : \text{Re}(\lambda) < 0 \}$.

As base (of dimension $\dim(d)$) for the martingale estimating functions we shall use (f^q) for $q \in \{1,\ldots,d\} \cup J$, where

$$f^i(x) = x_i \quad (1 \le i \le d), \qquad f^{i'j'}(x) = x_{i'}x_{j'} \quad ((i',j') \in J),$$

writing $x = (x_1,\ldots,x_d)$ for a generic point in $\mathbb{R}^d$. Clearly (f^q) satisfies the conditions from Assumption A and also, as a little work shows, the conditions on $\partial_x f$, $\partial^2_{xx} f$ from Theorem 2. To proceed we need the conditional moments $\pi_{t,\theta} f^q$, conveniently collected in the vector $\pi_{t,\theta}x = (\pi_{t,\theta}x_i)_{1\le i\le d}$ and the matrix $\pi_{t,\theta}xx^T$ and known to be given by the expressions

$$\pi_{t,\theta}x = \left(e^{tB} - I_d\right) B^{-1}A + e^{tB}x \tag{46}$$

$$\pi_{t,\theta}xx^T = (\pi_{t,\theta}x)\,(\pi_{t,\theta}x)^T + \int_0^t e^{sB}Ce^{sB^T}\,ds. \tag{47}$$

(As in the previous example, notation like $\pi_{t,\theta}xx^T$ is short for $\pi_{t,\theta}f(x)$, where $f(y) = yy^T$.)

Invoking Theorem 2 with $*$ in (18) equal to 0, one eventually arrives at the following small $\Delta-$optimal estimating function $g_{t,\theta}$ with $g_{1,t,\theta} = \left(g^{i'j'}_{t,\theta}\right)_{(i',j')\in J}$ and $g_{2,t,\theta}$ split into the vector–valued component $g^A_{2,t,\theta} = \left(g^i_{t,\theta}\right)_{1\le i\le d}$ and the matrix–valued component $g^B_{2,t,\theta} = \left(g^{ij}_{t,\theta}\right)_{1\le i,j\le d}$ and $g_{1,t,\theta}$ and $g^A_{2,t,\theta}$ and $g^B_{2,t,\theta}$ given by

$$\begin{aligned}
g^{i'j'}_{1,t,\theta}(x,y) &= \Big(C^{-1}\Big[-x\,(y-\pi_{t,\theta}x)^T - (y-\pi_{t,\theta}x)\,x^T \\
&\qquad + yy^T - \pi_{t,\theta}\left(xx^T\right)\Big]C^{-1}\Big)_{i'j'}, \\
g^A_{2,t,\theta}(x,y) &= C^{-1}\,(y-\pi_{t,\theta}x), \\
g^B_{2,t,\theta}(x,y) &= C^{-1}\,(y-\pi_{t,\theta}x)\,x^T.
\end{aligned}$$

For the calculations one uses that

$$\dot{C}^T_{1,\theta}(x) \in \mathbb{R}^{|J|\times d^2}, \quad \left(\dot{C}^T_{1,\theta}(x)\right)_{i'j',ij} = \begin{cases} \delta_{i'i}\delta_{j'j} & \text{if } i \le j, \\ \delta_{i'j}\delta_{j'i} & \text{if } i > j, \end{cases}$$

$$\begin{aligned}
\left(\dot{b}^A_{2,\theta}\right)^T(x) &= I_d \in \mathbb{R}^{d\times d}, \\
\left(\dot{b}^B_{2,\theta}\right)^T(x) &= I_d \otimes x \in \mathbb{R}^{d^2\times d}.
\end{aligned}$$

Also note that

$$\left(\partial_x f(x) \quad \partial^2_{xx} f(x) R\right)^{-1} = \begin{pmatrix} I_d & 0_{d\times J} \\ P(x) & D \end{pmatrix}$$

where $D = \operatorname{diag}(d_{i'j'}) \in \mathbb{R}^{J\times J}$ with

$$d_{i'j'} = \begin{cases} 1 & \text{if } i' < j' \\ \frac{1}{2} & \text{if } i' = j', \end{cases}$$

and $P(x) \in \mathbb{R}^{J\times d}$ with

$$P_{i'j',j}(x) = -d_{i'j'}\left(\delta_{i'j} x_{j'} + \delta_{j'j} x_{i'}\right).$$

As in the previous example, the simplest way to verify the small $\Delta-$optimality is to verify directly from these expressions that the conditions (13) are satisfied.

The resulting estimating equations are not affected by multiplication from the left and/or right by C, and it is now an easy task to write down the estimators of the parameter functions

$$\mathfrak{A} := \left(e^{\Delta B} - I_d\right) B^{-1} A, \quad e^{\Delta B}, \quad \mathfrak{C} := \int_0^\Delta e^{sB} C e^{sB^T}\, ds$$

based on the observations $X_0, X_\Delta, \ldots, X_{n\Delta}$: defining

$$\overline{X}_* := \tfrac{1}{n}\sum_{i=1}^n X_{(i-1)\Delta}, \quad \overline{X}^* := \tfrac{1}{n}\sum_{i=1}^n X_{i\Delta},$$

using (46) and (47) one sees that the estimating equations obtained from g_1, g_2^A, g_2^B are equivalent to the equations

$$\sum_{i=1}^n \left(X_{i\Delta} - \mathfrak{A} - e^{\Delta B} X_{(i-1)\Delta}\right) = 0, \tag{48}$$

$$\sum_{i=1}^n \left(X_{i\Delta} - \mathfrak{A} - e^{\Delta B} X_{(i-1)\Delta}\right) X^T_{(i-1)\Delta} = 0, \tag{49}$$

$$\sum_{i=1}^n \left(X_{i\Delta} X^T_{i\Delta} - \left(\mathfrak{A} + e^{\Delta B} X_{(i-1)\Delta}\right)\left(\mathfrak{A} + e^{\Delta B} X_{(i-1)\Delta}\right)^T - \mathfrak{C}\right) = 0, \tag{50}$$

and hence

$$\hat{\mathfrak{A}} = \overline{X}^* - e^{\Delta\widehat{B}}\overline{X}_*, \tag{51}$$

$$e^{\Delta\widehat{B}} = \left(\sum_{i=1}^{n}\left(X_{i\Delta} - \overline{X}^*\right)X_{(i-1)\Delta}^T\right)\left(\sum_{i=1}^{n}\left(X_{(i-1)\Delta} - \overline{X}_*\right)\left(X_{(i-1)\Delta} - \overline{X}_*\right)^T\right)^{-1} \tag{52}$$

$$\hat{\mathfrak{C}} = \tfrac{1}{n}\sum_{i=1}^{n}\left(X_{i\Delta}X_{i\Delta}^T - Z_iZ_i^T\right) \tag{53}$$

where in the last line,

$$Z_i := \widehat{\mathfrak{A}} + e^{\Delta\widehat{B}}X_{(i-1)\Delta}.$$

Note that the equation (53) may be written

$$\hat{\mathfrak{C}} = \tfrac{1}{n}\sum_{i=1}^{n}\left(X_{i\Delta} - \hat{\mathfrak{A}} - e^{\Delta\widehat{B}}X_{(i-1)\Delta}\right)\left(X_{i\Delta} - \hat{\mathfrak{A}} - e^{\Delta\widehat{B}}X_{(i-1)\Delta}\right)^T \tag{54}$$

as is seen using that from (48) and (49) it follows that

$$\sum_{i=1}^{n}\left(X_{i\Delta} - \hat{\mathfrak{A}} - e^{\Delta\widehat{B}}X_{(i-1)\Delta}\right)Z_i^T = 0.$$

The likelihood function for observing $X_0, X_\Delta, \ldots, X_{n\Delta}$ conditionally on X_0 is

$$\prod_{i=1}^{n}\frac{1}{(2\pi)^{d/2}\left|\mathfrak{C}\right|}\exp\left(-\tfrac{1}{2}\left(X_{i\Delta} - \xi_i\right)^T\mathfrak{C}^{-1}\left(X_{i\Delta} - \xi_i\right)\right)$$

where

$$\xi_i = \pi_{\Delta,\theta}\left(X_{(i-1)\Delta}\right) = \mathfrak{A} + e^{\Delta B}X_{(i-1)\Delta}.$$

Maximizing this over $\mathfrak{A}$, $e^{\Delta B}$ and $\mathfrak{C}$ varying *freely* in $\mathbb{R}^{d\times 1}$, $\mathbb{R}^{d\times d}$ and the space of symmetric positive definite $d\times d-$ matrices yields the estimators $\hat{\mathfrak{A}}$, $e^{\Delta\widehat{B}}$ and $\hat{\mathfrak{C}}$ from (51), (52) and (53). In a forthcoming paper by Mathieu Kessler and Anders Rahbek [4] (for the model with $A = 0$), the authors study this maximum-likelihood estimator and also tackle the non-trivial problem of converting the expressions (51), (52) and (53) into estimators for A, B and C: it may of course happen with probability > 0 that the right–hand side of (52) is not the exponential of any square matrix, and even if it is, $\widehat{B}$ may not satsify the basic condition for ergodicity that $\mathrm{Re}\,(\lambda) < 0$ for all $\lambda \in \mathrm{spec}(\widehat{B})$.

As pointed out by Kessler and Rahbek, a particularly unpleasant problem is that the $\widehat{B}$ solving (52) may not even be unique, due to certain periodicities. If $\widehat{B}$ is found of course

$$\widehat{A} = \widehat{B}\left(e^{\Delta\widehat{B}} - I_d\right)^{-1} \overset{\wedge}{\mathfrak{A}}$$

(with $e^{\Delta\widehat{B}} - I_d$ non–singular iff 0 is not an eigenvalue for $\widehat{B}$, hence non-singular almost surely if only n is large enough). Finally $\widehat{C}$ should be found from (53), but again here, Kessler and Rahbek show that there need not be a unique solution.

Remark 4 *If the observations X_{t_i} are not equidistant as assumed above (where $t_i = i\Delta$), maximum-likelihood estimation if at all possible will certainly be very difficult. Also, there are no analogues of $\overset{\wedge}{\mathfrak{A}}$, $e^{\Delta\widehat{B}}$ and $\overset{\wedge}{\mathfrak{C}}$ and in order to solve the estimating equations which are obtained from (48), (49) and (50), one must rely entirely on numerics and proceed directly to find $\widehat{A}$, $\widehat{B}$ and $\widehat{C}$. A possible method is to reparametrize B through its spectrum and look for B only with d distinct eigenvalues (it may be argued that with probability one, $\widehat{B}$ will have this property) that are either real or come in complex conjugate pairs. The parametrization of B could then be in terms of these eigenvalues and a concrete choice of corresponding complex eigenvectors of unit length. At least, if the t_i are non-lattice, the uniqueness problems emphasized by Kessler and Rahbek should disappear.*

A Appendix

The following result was used in the proof of Theorem 3:

Lemma 4 *Let $A, B \in \mathbb{R}^{m\times m}$ be symmetric and positive semidefinite matrices such that $1 \leq \operatorname{rank}(A) = m' < m$ and such that the columns (or rows) of A and B jointly span all of $\mathbb{R}^m$. Let further $\mathcal{O} = \binom{\mathcal{O}_1}{\mathcal{O}_2}$ be an orthogonal $m \times m-$matrix with $\mathcal{O}_1$ comprising the first m', $\mathcal{O}_2$ the last $m - m'$ rows of $\mathcal{O}$ such that*

$$\mathcal{O}A\mathcal{O}^T = \operatorname{diag}(\lambda_1, \ldots, \lambda_{m'}, 0, \ldots, 0),$$

$\lambda_1, \ldots, \lambda_{m'} > 0$ denoting the non-zero eigenvalues for A. Then as $t \to 0$,

$$(A + tB)^{-1} = \frac{1}{t}\mathcal{O}_2^T\left(\mathcal{O}_2 B\mathcal{O}_2^T\right)^{-1}\mathcal{O}_2 + N + O(t),$$

where N is of the form

$$\mathcal{O}_1^T \left(\mathcal{O}_1 A \mathcal{O}_1^T\right)^{-1} \mathcal{O}_1 + \mathcal{O}_2^T S + S^T \mathcal{O}_2$$

for some $(m-m') \times m-$matrix S.

Proof. Assume first that $A = \operatorname{diag}(\lambda_1, \ldots, \lambda_{m'}, 0, \ldots, 0)$ (with all $\lambda_\ell > 0$) and write

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

with e.g. B_{22} the lower right $(m-m') \times (m-m')-$submatrix of B. Then

$$|A + tB| = t^{m-m'} \left(\prod_{\ell=1}^{m'} \lambda_\ell \right) |B_{22}| + O\left(t^{m-m'+1}\right),$$

Also for the subdeterminants obtained by deleting the ℓ'th row and ℓ''th column,

$$|A + tB|_{\ell\ell'} = \begin{cases} O\left(t^{m-m'-1}\right) & \text{if } \ell, \ell' > m' \\ O\left(t^{m-m'}\right) & \text{otherwise.} \end{cases}$$

It follows from this that $(A+tB)^{-1}$ is of the form

$$\frac{1}{t} \begin{pmatrix} 0 & 0 \\ 0 & M \end{pmatrix} + N + O(t)$$

and it is then easy to see that, writing

$$D = \operatorname{diag}(\lambda_1, \ldots, \lambda_{m'}) \in \mathbb{R}^{m' \times m'}$$

one has

$$(A+tB)^{-1} = \frac{1}{t} \begin{pmatrix} 0 & 0 \\ 0 & B_{22}^{-1} \end{pmatrix} + \begin{pmatrix} D^{-1} & -D^{-1} B_{12} B_{22}^{-1} \\ -B_{22}^{-1} B_{21} D^{-1} & 0 \end{pmatrix} + O(t). \tag{55}$$

For the general case, just use that

$$(A+tB)^{-1} = \mathcal{O}^T \left(\mathcal{O} A \mathcal{O}^T + t \mathcal{O} B \mathcal{O}^T\right)^{-1} \mathcal{O}$$

with $\left(\mathcal{O} A \mathcal{O}^T + t \mathcal{O} B \mathcal{O}^T\right)^{-1}$ of the form (55) and $D = \mathcal{O}_1 A \mathcal{O}_1^T$. ■

References

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

[4] Kessler, M., Rahbek, A. (2000). Inference for discretely observed multivariate homogeneous Gaussian diffusions (manuscript).

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

[6] Sørensen, M. (1998). On asymptotics of estimating functions. Preprint 6, Department of Theoretical Statistics, University of Copenhagen.