Inference for observations of integrated diffusion processes.

Susanne Ditlevsen
Department of Biostatistics

Michael Sørensen
Department of Statistics and Operations Research

University of Copenhagen, Denmark

Abstract

Estimation of parameters in diffusion models is usually based on observations of the process at discrete time points. Here we investigate estimation when a sample of discrete observations is not available, but, instead, observations of a running integral of the process with respect to some weight function. This type of observations is, for example, obtained when a realization of the process is observed after passage through an electronic filter. Another example is provided by the ice-core data on oxygen isotopes used to investigate paleo-temperatures. Finally, such data play a role in connection with the stochastic volatility models of finance. The integrated process is no longer a Markov process which render the use of martingale estimating functions difficult. Therefore, a generalization of the martingale estimating functions, namely the prediction-based estimating functions, is applied to estimate parameters in the underlying diffusion process. The estimators are shown to be consistent and asymptotically normal. The method is applied to inference based on integrated data from Ornstein-Uhlenbeck processes and from the CIR-model for both of which an explicit estimating function can be found.

Key words: asymptotic normality, CIR-model, consistency, estimating equation, ice-core data, non-Markovian process, Ornstein-Uhlenbeck process, prediction based estimating functions, stochastic differential equation, quasi-likelihood.

1 Introduction

In the present paper we study statistical inference for observations of integrated diffusions. In several cases, a sample of observations at discrete time points of a diffusion process is not available, but, for example, a realization of the process has been observed after passage through an electronic filter. Another example is provided by the ice-core records from Greenland. The isotope ratio $^{18}O/^{16}O$ in the ice, measured as an average in pieces of ice, each piece representing a time interval with time increasing as a function of the depth, is a proxy for paleo-temperatures. The variation of the paleo-temperature can be modelled by a stochastic differential equation, and it is natural to model the ice-core data as an integrated diffusion process, see Ditlevsen, Ditlevsen and Andersen (2002). Integrated processes also play an important role in connection with the so-called realized stochastic volatility in finance, see Andersen and Bollerslev (1998), Genon-Catalot et al. (1999), Gloter (1999b), Sørensen (2000), Barndorff-Nielsen and Shepard (2001) and Andersen, Bollerslev, Diebold and Labys (2001).

Martingale estimating functions are a useful tool for statistical inference based on discretely sampled diffusions, see e.g. Bibby and Sørensen (1995, 1997, 2001), Pedersen (2000), and Sørensen (1997) and references therein. However, integrated diffusion processes are not Markov processes, for which reason there are no natural or easily calculated martingales on which to base a class of estimating functions. Therefore we will apply the prediction-based estimating functions that were introduced in Sørensen (2000) as a tool for drawing statistical inference about non-Markovian models and in other situations where no martingale is readily available. These estimating functions are generalizations of the martingale estimating functions. It is shown that the method of prediction-based estimating functions under mild regularity conditions provides a satisfactory solution to the inference problem investigated here. The conditions ensure existence, consistency and asymptotic normality of the estimators.

Other approaches to inference for integrated diffusions were presented in Gloter (1998, 1999a). The first paper considers the integrated Ornstein-Uhlenbeck process, and compare the Whittle estimator, which in this case is efficient, to the estimator obtained from Rydén's split data maximum pseudo-likelihood estimator, see Rydén (1994). In Gloter (1999a) minimum contrast estimators are considered that are consistent when the length of the sampling interval goes to zero as the number of observations goes to infinity.

In Section 2 the model of integrated diffusions is presented, and prediction-based estimating functions are briefly presented and applied to solve the inference problem. A way of finding the necessary moments of the integrated process is derived, and as examples prediction-based estimating functions are found for the integrated Ornstein-Uhlenbeck process and for the integrated CIR-model.

In Section 3 the optimal prediction-based estimating function is derived, and we find a way of deriving moments of order α, α being a non-negative integer, provided that these moments exist. A formula is given so that if we know an analytic expression or can simulate the moments in the underlying process, the calculation of the moments needed in order to find the optimal prediction-based estimating function is easily programmable. The optimal prediction-based estimating functions in the previous examples are discussed. This yields, in both examples, explicit estimating functions.

In Section 4 asymptotic results about the estimating functions and their estimators are proved under weak regularity conditions using the properties of the underlying process.

2 Integrated diffusions and prediction-based estimating functions

Consider the one-dimensional diffusion

$$dX_t = b(X_t;\theta)dt + \sigma(X_t;\theta)dW_t \quad , \qquad X_0 \sim \mu_\theta$$

where θ is an unknown p-dimensional parameter belonging to the parameter space $\Theta \subseteq \mathbb{R}^p$ and W is a one-dimensional standard Wiener process. We assume that X_0 is independent of W, that the stochastic differential equation has a unique weak solution, and that X is an ergodic, stationary diffusion with invariant measure μ_θ.

Suppose that a sample of observations at discrete time points is not available, but, instead, a running integral of the process with respect to some weight function. Specifically, suppose the interval of observation $[0,T]$ is subdivided into n smaller intervals of length $\Delta = T/n$, and let ν be a probability measure on the interval $[0,\Delta]$. We shall consider observations of the form

$$Y_i = \int_0^\Delta X_{(i-1)\Delta+s}\, d\nu(s) \quad ; \quad i = 1,\ldots,n. \tag{2.1}$$

Typically, ν will have a density φ with respect to the Lebesgue measure on $[0,\Delta]$, in which case

$$Y_i = \int_{(i-1)\Delta}^{i\Delta} X_s \varphi(s-(i-1)\Delta)ds \quad ; \quad i = 1,\ldots,n. \tag{2.2}$$

If our observations are obtained by integrating uniformly over the time axis, ν is simply the uniform distribution on $[0,\Delta]$ with $\varphi = 1/\Delta$, and we get the more simple observations

$$Y_i = \frac{1}{\Delta}\int_{(i-1)\Delta}^{i\Delta} X_s\, ds.$$

Note that since X is stationary, the distribution of X_t is the same for all intervals $[0,\Delta),\ldots,[(n-1)\Delta,n\Delta)$, and thus $\{Y_i\}$ is stationary.

We solve the problem of estimating the parameter θ in the underlying process X by applying the method of prediction-based estimating functions introduced in Sørensen (2000). In the following we will briefly outline the method of prediction-based estimating functions. Assume that $f_j, j=1,\ldots,N$, are one-dimensional functions such that $E_\theta(f_j(Y_i)^2)<\infty$ for all $\theta\in\Theta$. We denote the expectation when θ is the true parameter value by $E_\theta(\cdot)$. Let $h_{jk}, j=1,\ldots,N, k=1,\ldots,q_j$ be functions from $\mathbb{R}^r$ into $\mathbb{R}$, and define (for $i\geq r+1$) random variables by $Z_{jk}^{(i-1)} = h_{jk}(Y_{i-1},Y_{i-2},\ldots,Y_{i-r})$. We assume that $E_\theta((Z_{jk}^{(i-1)})^2)<\infty$ for all $\theta\in\Theta$, and let $\mathcal{P}_{i-1,j}$ denote the subspace of the space of square integrable random variables spanned by $1, Z_{j1}^{(i-1)},\ldots,Z_{jq_j}^{(i-1)}$. Finally, we make the natural assumption that $1, Z_{j1}^{(r)},\ldots,Z_{jq_j}^{(r)}$ are linearly independent. The space $\mathcal{P}_{i-1,j}$ can be interpreted as a set of predictors of $f_j(Y_i)$ based on $Y_{i-r},\ldots,Y_{i-1}$. We write the elements of $\mathcal{P}_{i-1,j}$ in the form $a_0+a^TZ_j^{(i-1)}$, where $a^T=(a_1,\ldots,a_{q_j})$ and $Z_j^{(i-1)}=(Z_{j1}^{(i-1)},\ldots,Z_{jq_j}^{(i-1)})^T$ are q_j-dimensional vectors. We denote transposition by ^{T}. We will study the estimating function

$$G_n(\theta)=\sum_{i=r+1}^{n}\sum_{j=1}^{N}\Pi_j^{(i-1)}(\theta)\left[f_j(Y_i)-\hat{\pi}_j^{(i-1)}(\theta)\right] \tag{2.3}$$

where Y_i is of the form (2.2), $\Pi_j^{(i-1)}(\theta)$ is a p-dimensional stochastic vector, the coordinates of which belong to $\mathcal{P}_{i-1,j}$, and $\hat{\pi}_j^{(i-1)}(\theta)$ is the minimum mean square error predictor of $f_j(Y_i)$ in $\mathcal{P}_{i-1,j}$.

When θ is the true parameter value, we define $C_j(\theta)$ as the covariance matrix of $Z_j^{(r)}$ and $b_j(\theta)=\Big(\mathrm{Cov}_\theta(Z_{j1}^{(r)},f_j(Y_{r+1})),\ldots,\mathrm{Cov}_\theta(Z_{jq_j}^{(r)},f_j(Y_{r+1}))\Big)^T$. Then we have

$$\hat{\pi}_j^{(i-1)}(\theta)=\hat{a}_{j0}(\theta)+\hat{a}_j(\theta)^TZ_j^{(i-1)}$$

where

$$\hat{a}_j(\theta)=C_j(\theta)^{-1}b_j(\theta) \tag{2.4}$$

and

$$\hat{a}_{j0}(\theta)=E_\theta(f_j(Y_1))-\hat{a}_j(\theta)^TE_\theta(Z_j^{(r)}). \tag{2.5}$$

If, for instance, we take $f_j(y)=y^{m_j}$ and $Z_{jk}^{(i-1)}=Y_{i-k}^{m_j}$, $k=1,\ldots,r$, for some positive integer m_j, we need to calculate the moments $E_\theta(Y_1^{m_j})$ and $E_\theta((Y_1Y_k)^{m_j})$ for $k=1,\ldots,r$. Once we have these moments, the coefficients $\hat{a}_{j0},\ldots,\hat{a}_{jr}$ can easily be found by means of the Durbin-Levinson algorithm, see Brockwell and Davis (1991). For many diffusions there exist $K>0$ and $\lambda>0$ such that $\left|\mathrm{Cov}_\theta(Y_1^m,Y_{r+1}^m)\right|\leq Ke^{-\lambda(r-1)}$, see Section 4. Therefore r will usually not need to be chosen particularly large.

Presumably $f_1(y) = y$ and $f_2(y) = y^2$ with $Z_{jk}^{(i-1)} = Y_{i-k}$, $k = 1, \ldots, r, j = 1, 2$ and $Z_{2k}^{(i-1)} = Y_{i+r-k}^2$, $k = r+1, \ldots, 2r$, will in many cases be a reasonable choice. In this case the minimum mean square error predictor of $f_1(Y_i)$ can be found as described above, while the predictor of $f_2(Y_i)$ can be found by applying the two-dimensional Durbin-Levinson algorithm to the process (Y_i, Y_i^2).

The necessary moments can in all these cases be found from the mixed moments of the process X. First we assume that $E_\theta(|X_0|^m) < \infty$. Then

$$\begin{aligned} E_\theta(Y_1^m) &= E_\theta\left(\left(\int_0^\Delta X_s \varphi(s) ds\right)^m\right) \\ &= E_\theta\left(\int_0^\Delta \cdots \int_0^\Delta X_{s_1} \cdots X_{s_m} \varphi(s_1) \cdots \varphi(s_m) ds_1 \cdots ds_m\right) \\ &= \int_0^\Delta \cdots \int_0^\Delta E_\theta(X_{s_1} \cdots X_{s_m}) \varphi(s_1) \cdots \varphi(s_m) ds_1 \cdots ds_m, \end{aligned}$$

where we have used Fubini's theorem. Specifically we get

$$E_\theta(Y_1) = \int_0^\Delta E_\theta(X_s)\varphi(s)ds = E_\theta(X_0) \int_0^\Delta \varphi(s)ds = E_\theta(X_0).$$

When $\varphi(t)$ is a constant, we see that

$$E_\theta(Y_1^m) = \frac{m!}{\Delta^m} \int_0^\Delta \int_0^{s_1} \cdots \int_0^{s_{m-1}} E_\theta(X_{s_1} \cdots X_{s_m}) ds_m \cdots ds_2 ds_1.$$

Here we have used that $E_\theta(X_{s_1} \cdots X_{s_m})$ does not depend on the ordering of $s_1, \ldots, s_m$ so that it is enough to integrate over the region where $0 \leq s_m \leq \cdots \leq s_1 \leq \Delta$. The factor $m!$ appears because $s_1, \ldots, s_m$ can be ordered in $m!$ different ways. In a similar way we obtain that when $E_\theta(|X_0|^{m_1+m_2}) < \infty$,

$$\begin{aligned} &E_\theta\left(Y_1^{m_1} Y_k^{m_2}\right) = \\ &\int_0^\Delta \cdots \int_0^\Delta E_\theta\left(X_{s_1} \cdots X_{s_{m_1}} X_{((k-1)\Delta + s_{(m_1+1)})} \cdots X_{((k-1)\Delta + s_{(m_1+m_2)})}\right) \\ &\qquad\qquad \cdot \varphi(s_1) \cdots \varphi(s_{(m_1+m_2)}) ds_1 \cdots ds_{(m_1+m_2)}. \end{aligned}$$

Example 2.1 For diffusion models where the eigenfunctions of the generator are polynomials it is possible to find all moments of type $E(X_{t_1} \cdots X_{t_m})$, see e.g. Sørensen (2000).

Consider the Ornstein-Uhlenbeck process given by

$$dX_t = -\beta X_t dt + \sigma dW_t.$$

This process is ergodic, and its stationary distribution is the normal distribution with expectation 0 and variance $\sigma^2/(2\beta)$, provided that $\beta > 0$. We have that

$$E_\theta(X_t | X_0 = x_0) = x_0 e^{-\beta t},$$

and

$$E_\theta(X_0X_t) = \sigma^2(2\beta)^{-1}e^{-\beta t}.$$

This implies that

$$E_\theta(Y_1^2) = \sigma^2\beta^{-3}\Delta^{-2}(\beta\Delta - 1 + e^{-\beta\Delta})$$

and

$$E_\theta(Y_1Y_k) = \tfrac{1}{2}\sigma^2\beta^{-3}\Delta^{-2}(1-e^{\beta\Delta})^2e^{-k\beta\Delta}$$

for $k > 1$.

For $f_1(y) = y$ and $Z_1^{(i-1)} = Y_{i-1}$ (i.e. $q_1 = 1, r = 1$), we get

$$\hat{\pi}_1^{(i-1)}(\theta) = \frac{(1-e^{-\beta\Delta})^2}{2(\beta\Delta - 1 + e^{-\beta\Delta})}Y_{i-1}.$$

Note that in this case the predictor $\hat{\pi}_1^{(i-1)}(\theta)$ depends on β only. This is also true for $r > 1$ when $f_1(y) = y$. Thus σ cannot be estimated by means of a linear estimating function; we would need for instance the function $f_2(y) = y^2$ and to include the squared observations Y_{i-k}^2 in $Z_2^{(i-1)}$.

Example 2.2 Another particular example is the model given by

$$dX_t = -\theta(X_t - \alpha)dt + \sigma\sqrt{X_t}\,dW_t.$$

This process is ergodic and its stationary distribution is the Gamma distribution with shape parameter $2\theta\alpha\sigma^{-2}$ and scale parameter $2\theta\sigma^{-2}$ provided that $\theta > 0$, $\alpha > 0$, $\sigma > 0$, and $2\theta\alpha \geq \sigma^2$. The process has many applications. It is, for instance, used in mathematical finance to model short term interest rates, see Cox, Ingersoll and Ross (1985), and Feller (1951) proposed it as a model for population growth. Recently it was used to model nitrous oxide emission from soil by Pedersen (2000).

All moments of the type $E(X_{t_1}\cdots X_{t_m})$ can be calculated by means of formulae in Sørensen (2000). In particular, $E(X_0) = \alpha, E(X_0^2) = \alpha(\alpha + \sigma^2/(2\theta))$, and

$$E(X_0X_t) = \alpha^2 + \alpha\sigma^2(2\theta)^{-1}e^{-\theta t}.$$

Thus

$$E(Y_1^2) = \alpha^2 + \alpha\sigma^2\theta^{-3}\Delta^{-2}(e^{-\theta\Delta} - 1 + \theta\Delta)$$

and

$$E(Y_1Y_k) = \alpha^2 + \frac{1}{2}\alpha\sigma^2\theta^{-3}\Delta^{-2}(e^{\theta\Delta} - 1)^2e^{-k\theta\Delta}$$

for $k > 1$. It is therefore possible to calculate explicitly the prediction-based estimating functions discussed above.

3 The optimal prediction-based inference for integrated diffusions

In this section we derive the optimal choice of the weight $\Pi^{(i-1)}(\theta)$ in (2.3) using the results and notation in Sørensen (2000). Optimality is in the sense of the theory of estimating functions, see Godambe and Heyde (1988) and Heyde (1997). The optimal member of a class of estimating functions is the one that provides the most efficient estimator. This estimator is sometimes called a quasi-likelihood estimator.

In Sørensen (2000) it was shown that the optimal estimating function of the type (2.3) is given by

$$G_n^*(\theta) = A_n^*(\theta) \sum_{i=r+1}^{n} H^{(i)}(\theta), \tag{3.1}$$

where

$$H^{(i)}(\theta) = \begin{pmatrix} \left(f_1(Y_i) - \hat{\pi}_1^{(i-1)}(\theta)\right) \\ Z_{11}^{(i-1)} \left(f_1(Y_i) - \hat{\pi}_1^{(i-1)}(\theta)\right) \\ \vdots \\ Z_{1q_1}^{(i-1)} \left(f_1(Y_i) - \hat{\pi}_1^{(i-1)}(\theta)\right) \\ \vdots \\ \vdots \\ \left(f_N(Y_i) - \hat{\pi}_N^{(i-1)}(\theta)\right) \\ Z_{N1}^{(i-1)} \left(f_N(Y_i) - \hat{\pi}_N^{(i-1)}(\theta)\right) \\ \vdots \\ Z_{Nq_N}^{(i-1)} \left(f_N(Y_i) - \hat{\pi}_N^{(i-1)}(\theta)\right) \end{pmatrix}, \tag{3.2}$$

and where

$$A_n^*(\theta) = U(\theta)^T \bar{M}_n(\theta)^{-1},$$

with

$$\bar{M}_n(\theta) = E_\theta\left(H^{(r+1)}(\theta) H^{(r+1)}(\theta)^T\right) + \sum_{k=1}^{n-r-1} \frac{(n-r-k)}{(n-r)} \left[E_\theta\left(H^{(r+1)}(\theta) H^{(r+1+k)}(\theta)^T\right) + E_\theta\left(H^{(r+1+k)}(\theta) H^{(r+1)}(\theta)^T\right)\right] \tag{3.3}$$

and

$$U(\theta) = \bar{C}(\theta) \partial_{\theta^T} \hat{a}(\theta). \tag{3.4}$$

Here $\bar{C}(\theta)$ =diag$\{\tilde{C}_1(\theta), \ldots, \tilde{C}_N(\theta)\}$ with

$$\tilde{C}_j(\theta) = [E_\theta\{Z_{jk}^{(r-1)} Z_{jl}^{(r-1)}\}]_{k,l=0,\ldots,q_j},$$

where $Z_{j0}^{(r-1)} = 1$, and

$$\hat{a}(\theta) = \left(\hat{a}_{10}(\theta), \hat{a}_1(\theta)^T, \ldots, \hat{a}_{N0}(\theta), \hat{a}_N(\theta)^T\right)^T, \tag{3.5}$$

where $\hat{a}_{j0}(\theta)$ and $\hat{a}_j(\theta)$ are given by (2.4) and (2.5). A sufficient condition for (3.1) to be optimal is that the matrix $\partial_{\theta^T}\hat{a}(\theta)$ has full rank, and that the functions $1, f_1, \ldots, f_N$ are linearly independent on the support of the conditional distribution of Y_n given $Y_1, \ldots, Y_{n-1}$. In particular, the latter condition implies that the matrix $\bar{M}_n(\theta)$ is invertible.

In Section 4 we shall see that for many diffusion models there exist $K > 0$ and $\lambda > 0$ such that the absolute values of all entries in the expectation matrices in the sum in (3.3) are dominated by $K\mathrm{e}^{-\lambda(k-r-1)}$ when $k > r$. Therefore, the sum in (3.3) can in practice often be truncated so that fewer moments need to be calculated.

Natural choices for $f_j(y)$ and $Z_{jk}^{(i-1)}$ would be $f_j(y) = y^{\alpha_{j0}}$ and $Z_{jk}^{(i-1)} = Y_{i-l_{jk}}^{\alpha_{jk}}$, where α_{j0} and α_{jk} are such that $E_\theta[Y^{4\alpha}]$ exists with $\alpha = \max\{\alpha_{10}, \ldots, \alpha_{Nq_N}\}$. Note that it is enough that $E_\theta[Y^{2\alpha}]$ exists for a prediction-based estimating function to be well-defined. The more strict condition is for the optimal prediction-based estimating function to exist. From now on we assume that f_j and $Z_{jk}^{(i-1)}$ have the form just indicated. For simplicity we assume α_{j0} and α_{jk} are integers. In order to calculate (3.3), we then need higher order moments of the form $E_\theta[Y_1^{k_1} Y_{t_1}^{k_2} Y_{t_2}^{k_3} Y_{t_3}^{k_4}]$, where $1 \leq t_1 \leq t_2 \leq t_3$ and where k_i, $i = 1, \ldots, 4$ are non-negative integers such that $(k_1 + k_2 + k_3 + k_4) \leq 4\alpha$. We will express these moments in terms of the moments of X_t, which will usually either be known or possible to determine by simulation.

Define

$$\psi(v, u, s, r; \theta) = E_\theta[X_{v_1} \cdots X_{v_{k_1}} X_{u_1} \cdots X_{u_{k_2}} X_{s_1} \cdots X_{s_{k_3}} X_{r_1} \cdots X_{r_{k_4}}],$$

where $v = (v_1, \ldots, v_{k_1})$, $u = (u_1, \ldots, u_{k_2})$, $s = (s_1, \ldots, s_{k_3})$, $r = (r_1, \ldots, r_{k_4})$,

$$\phi(x, k, t, \Delta) = \varphi(x_1 - (t-1)\Delta) \cdots \varphi(x_k - (t-1)\Delta),$$

where k is an integer and $x = (x_1, \ldots, x_k)$,

$$\Phi(v, u, s, r, t_1, t_2, t_3, \Delta) = \phi(v, k_1, 1, \Delta)\, \phi(u, k_2, t_1, \Delta)\, \phi(s, k_3, t_2, \Delta)\, \phi(r, k_4, t_3, \Delta),$$

and

$$\begin{aligned}
A_1 &= [0\,,\,\Delta]^{k_1}\\
A_2 &= [(t_1-1)\Delta\,,\,t_1\Delta]^{k_2}\\
A_3 &= [(t_2-1)\Delta\,,\,t_2\Delta]^{k_3}\\
A_4 &= [(t_3-1)\Delta\,,\,t_3\Delta]^{k_4}\\
T_1 &= \{(v_1,\cdots,v_{k_1}) : 0\le v_1\le\cdots\le v_{k_1}\le\Delta\}\\
T_2 &= \{(u_1,\cdots,u_{k_2}) : (t_1-1)\Delta\le u_1\le\cdots\le u_{k_2}\le t_1\Delta\}\\
T_3 &= \{(s_1,\cdots,s_{k_3}) : (t_2-1)\Delta\le s_1\le\cdots\le s_{k_3}\le t_2\Delta\}\\
T_4 &= \{(r_1,\cdots,r_{k_4}) : (t_3-1)\Delta\le r_1\le\cdots\le r_{k_4}\le t_3\Delta\}\\
B &= (A_1\cap T_1)\times(A_2\cap T_2)\times(A_3\cap T_3)\times(A_4\cap T_4).
\end{aligned}$$

In the same way as in Section 2 we get

$$E_\theta\left[Y_1^{k_1}Y_{t_1}^{k_2}Y_{t_2}^{k_3}Y_{t_3}^{k_4}\right] = \int_{A_1\times A_2\times A_3\times A_4}\psi(v,u,s,r;\theta)\,\Phi(v,u,s,r,t_1,t_2,t_3,\Delta)\,d\mathbf{t},$$

where $d\mathbf{t} = dr_{k_4}\cdots dr_1\,ds_{k_3}\cdots ds_1\,du_{k_2}\cdots du_1\,dv_{k_1}\cdots dv_1$. Thus we need the mixed moments of the process X of order up to $(k_1+k_2+k_3+k_4)$. These depend on the distance in time between the variables X_{t_i} appearing in the expression for the moment, and care has to be taken when different variables are integrated over the same interval when the order of the integration variables changes. When $\varphi(t) = 1/\Delta$, this can be solved in the following way. Assume that $1 < t_1 < t_2 < t_3$. Arguments of symmetry yield that

$$E_\theta[Y_1^{k_1}Y_{t_1}^{k_2}Y_{t_2}^{k_3}Y_{t_3}^{k_4}] = \frac{k_1!k_2!k_3!k_4!}{\Delta^{(k_1+k_2+k_3+k_4)}}\int_B\psi(v,u,s,r;\theta)\,d\mathbf{t} \tag{3.6}$$

The factor $k_1!$ appears because $v_1,\cdots,v_{k_1}$ can be ordered in $k_1!$ different ways. The arguments for the other factors are similar.

Example 3.1 *(Example 2.1 continued)* We will now find the optimal prediction-based estimating function for the integrated Ornstein-Uhlenbeck process with $N=1, q_1=r=1, f_1(y)=y$, and $Z_{11}^{(i-1)} = Y_{i-1}$. We have

$$E_\theta(Y_1^{k_1}Y_{t_1}^{k_2}Y_{t_2}^{k_3}) = 0$$

for $k_1+k_2+k_3 = 3$ and $1\le t_1\le t_2$, because all moments of an odd order are zero. Moreover

$$E_\theta(X_vX_uX_sX_t) = \tfrac{1}{4}\beta^{-2}\sigma^4\left(2e^{-(s-u)\beta}+e^{(s-u)\beta}\right)e^{-(t-v)\beta}$$

for $v\le u\le s\le t$. Thus, using (3.6),

$E_\theta(Y_1Y_{t_1}Y_{t_2}Y_{t_3}) =$

$$\tfrac{1}{4}\beta^{-2}\sigma^4\Delta^{-4}\int_0^{\Delta}\int_{g_1(v)}^{t_1\Delta}\int_{g_2(u)}^{t_2\Delta}\int_{g_3(s)}^{t_3\Delta}\left(2e^{-(s-u)\beta}+e^{(s-u)\beta}\right)e^{-(t-v)\beta}dt\,ds\,du\,dv$$

for $1 \leq t_1 \leq t_2 \leq t_3$, where $g_i(z) = z$ if $t_i = t_{i-1}$ and $g_i(z) = (t_i - 1)\Delta$ if $t_i > t_{i-1}$ with $t_0 = 1$ $(i = 1, 2, 3)$. Finally we end up with the following optimal prediction-based estimating function:

$$G_n^*(\theta) = \sum_{i=2}^{n}\left(Y_{i-1}Y_i - Y_{i-1}^2\frac{(1-e^{-\beta\Delta})^2}{2(\beta\Delta - 1 + e^{-\beta\Delta})}\right).$$

Example 3.2 *(Example 2.2 continued)* Consider again the CIR-model. In order to find the optimal prediction-based estimating function with $N = 1$, $f_1(y) = y$, and $Z_{1k}^{(i-1)} = Y_{i-k}$, $k = 1, \ldots, r$, we need moments of the form $E_\theta(Y_1 Y_{t_1} Y_{t_2})$ and $E_\theta(Y_1 Y_{t_1} Y_{t_2} Y_{t_3})$, $(1 \leq t_1 \leq t_2 \leq t_3)$. By the formulae above, these can be obtained by integration of moments of the form $E_\theta(X_{t_1} X_{t_2} X_{t_3})$ and $E_\theta(X_{t_1} X_{t_1} X_{t_2} X_{t_3})$, for which explicit and easily integrable expressions are known, see e.g. Sørensen (2000). As the resulting expressions are rather long, they are omitted.

4 Asymptotic results

In this section we give asymptotic results for our estimating functions and the corresponding estimators when our observations are integrated diffusions, based on general results in Sørensen (1999) and (2000). To do this we need to study which properties the process Y inherits from the underlying diffusion process X. The integrated process Y is not a Markov process, but mixing properties and moment conditions satisfied by X are preserved, which is what we will use in this section.

We begin with a result on the asymptotic behaviour of an estimating function of the general form

$$G_n(\theta) = A_n(\theta)\sum_{i=r}^{n} H^{(i)}(\theta), \tag{4.1}$$

where $\{A_n(\theta)\}$ is a sequence of $p \times \sum_{j=1}^{N}(q_j + 1)$-matrices, and where $H^{(i)}(\theta)$ is given by (3.2).

Theorem 4.1 *Suppose the diffusion process X is stationary and α-mixing with mixing coefficients $\alpha_t(\theta)$, $t > 0$, and that there exists a $\delta > 0$ such that*

$$\sum_{k=1}^{\infty}\alpha_{k\Delta}(\theta)^{\delta/(2+\delta)} < \infty \tag{4.2}$$

and

$$E_\theta\left(\left|H^{(r)}(\theta)_{jk}\right|^{2+\delta}\right) < \infty, \quad j = 1, \ldots, N, \; k = 0, \ldots, q_j. \tag{4.3}$$

Then as $n \to \infty$,

$$\bar{M}_n(\theta) \to M(\theta), \tag{4.4}$$

where $\bar{M}_n(\theta)$ *is given by* (3.3) *and where*

$$\begin{aligned} M(\theta) = {} & E_\theta\left(H^{(r)}(\theta)H^{(r)}(\theta)^T\right) + \\ & \sum_{k=1}^{\infty}\left\{E_\theta\left(H^{(r)}(\theta)H^{(r+k)}(\theta)^T\right) + E_\theta\left(H^{(r+k)}(\theta)H^{(r)}(\theta)^T\right)\right\}. \end{aligned} \tag{4.5}$$

Assume, moreover, that $A_n(\theta) \to A(\theta)$ *as* $n \to \infty$. *Then as* $n \to \infty$,

$$n^{-1}\mathrm{Var}_\theta\left(G_n(\theta)\right) \to V(\theta) = A(\theta)M(\theta)A(\theta)^T, \tag{4.6}$$

and

$$\frac{1}{\sqrt{n}}G_n(\theta) \to N\left(0, V(\theta)\right) \tag{4.7}$$

in distribution, provided that the matrix $A(\theta)$ *is such that* $A(\theta)M(\theta)A(\theta)^T$ *is strictly positive definite.*

Proof: First note that it follows that the process Y is stationary and α-mixing with mixing coefficients $\alpha_k^Y(\theta)$, satisfying $\alpha_k^Y(\theta) \leq \alpha_{(k-1)\Delta}(\theta)$, $k = 2, 3, \ldots$. This is because the σ-algebra generated by Y_i, $i = 1, \ldots, n$ is contained in the σ-algebra generated by X_u, $0 \leq u \leq n\Delta$, and the σ-algebra generated by Y_i, $i = n, n+1, \ldots$ is contained in the σ-algebra generated by X_u, $u \geq (n-1)\Delta$. Next note that since $H^{(i)}(\theta)$ is a function of $Y_{i-r}, \ldots, Y_i$, the process $H^{(i)}(\theta)$, $i = r+1, r+2, \ldots$ is α-mixing with mixing coefficients $\alpha_k^H(\theta)$, satisfying that $\alpha_k^H(\theta) \leq \alpha_{(k-r-1)\Delta}(\theta)$, $k = r+2, \ldots$, and hence (4.2) holds with $\alpha_{k\Delta}(\theta)$ replaced by $\alpha_k^H(\theta)$.

To prove asymptotic normality, it is enough to consider the one-dimensional process $v^T G_n(\theta)$ for every $v \in \mathbb{R}^p \backslash \{0\}$ (Cramér-Wold device). Hence the theorem follows from a classical central limit result by Ibragimov, see e.g. Theorem 1 in Section 1.5 of Doukhan (1994).

□

For the one-dimensional, ergodic diffusion process X there are a number of relatively simple criteria ensuring α-mixing with exponentially decreasing mixing coefficients for which (4.2) is obviously satisfied. If, for instance, the spectrum of the generator of X has a discrete spectrum then the process is α-mixing. If λ_1 denotes the smallest non-zero eigenvalue, then the mixing coefficients satisfy

$$\alpha_t(\theta_0) \leq e^{-t\lambda_1},$$

see Doukhan (1994, p. 112). Thus X is geometrically α-mixing.

The diffusion processes considered in Examples 2.1 and 2.2 both have a discrete spectrum with $\lambda_1 = \beta$ and $\lambda_1 = \theta$, respectively.

Doukhan (1994) gives other criteria for geometrical mixing too; see also Hansen and Scheinkman (1995) and Veretennikov (1997). Rather general criteria for geometric α-mixing of diffusion processes expressed in the language of Malliavin calculus were given by Kusuoka and Yoshida (1997). We cite the following straightforward set of conditions by Genon-Catalot, Jeantheau and Larédo (2000) on the coefficients b and σ that are sufficient to ensure geometric α-mixing of X. It is presupposed that X is stationary with state space (ℓ, r) $(-\infty \leq \ell < r \leq \infty)$ and that the usual conditions on the scale measure and the speed measure hold, i.e. that

$$\int_{\ell}^{x_0} s(x)dx = \int_{x_0}^{r} s(x)dx = \infty \quad \text{and} \quad \int_{\ell}^{r} m(x)dx < \infty,$$

where

$$s(x) = \exp\left(-2\int_{x_0}^{x} \frac{b(u)}{\sigma^2(u)}du\right) \quad \text{and} \quad m(x) = \frac{1}{\sigma^2(x)s(x)},$$

and where $x_0 \in (\ell, r)$.

Condition 4.2
(i) The function b is continuously differentiable and σ is twice continuously differentiable on (ℓ, r), $\sigma(x) > 0$ for all $x \in (\ell, r)$, and there exists a constant $K > 0$ such that $|b(x)| \leq K(1 + |x|)$ and $\sigma^2(x) \leq K(1 + x^2)$ for all $x \in (\ell, r)$.

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

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

This condition in fact implies more than geometric α-mixing, it actually ensures geometric ρ-mixing, which again implies the exponential bounds on certain moments mentioned in Sections 2 and 3. Specifically, there exist $K > 0$ and $\lambda > 0$ such that if Z_1 is measurable with respect to the σ-algebra generated by X_s, $s \leq t_1$, and Z_2 is measurable with respect to the σ-algebra generated by X_s, $s \geq t_2$, $t_1 < t_2$, then $|\text{Cov}(Z_1, Z_2)| \leq Ke^{-\lambda(t_2 - t_1)}\text{Var}(Z_1)\text{Var}(Z_2)$.

Weak conditions ensuring polynomial α-mixing were given by Veretennikov (1988).

The following lemma can be used to check the moment condition (4.3) in Theorem 4.1.

Lemma 4.3 *Suppose $f_j(y) = y^{\alpha_{j0}}$ and $Z_{jk}^{(i-1)} = Y_{i-l_{jk}}^{\alpha_{jk}}$ with $\alpha_{jk}, l_{jk} \geq 1$ $(j = 1, \ldots, N,\ k = 0, \ldots, q_j)$. If $E_\theta\left(|X_0|^{4\alpha+\epsilon}\right) < \infty$ for an $\epsilon > 0$, where $\alpha = \max\{\alpha_{10}, \ldots, \alpha_{Nq_N}\}$, then (4.3) holds with $\delta = \epsilon/(2\alpha)$.*

Proof: It is enough to check that $E_\theta\left(|Y_i^{\kappa_1}Y_1^{\kappa_2}|^{2+\delta}\right) < \infty$ for $1 \leq i \leq r$ and $\kappa_1, \kappa_2 \in \{\alpha_{10}, \ldots, \alpha_{Nq_N}\}$, and by Cauchy-Schwartz' inequality this is the case if $E_\theta\left(|Y_1|^{2\alpha(2+\delta)}\right) < \infty$. Finally, by Jensen's inequality, Fubini's theorem and the stationarity of X

$$
\begin{aligned}
E_\theta\left(|Y_1|^{2\alpha(2+\delta)}\right) &= E_\theta\left(\left|\int_0^\Delta X_u \nu(du)\right|^{2\alpha(2+\delta)}\right) \\
&\leq \int_0^\Delta E_\theta\left(|X_u|^{2\alpha(2+\delta)}\right)\nu(du) = E_\theta\left(|X_0|^{2\alpha(2+\delta)}\right) < \infty.
\end{aligned}
$$

□

For more general choices of $f_j(y)$ and $Z_{jk}^{(i-1)}$, the existence of the relevant moments must be checked.

The following result about existence, consistency and asymptotic normality of our estimators can now be proved exactly as the similar result in Sørensen (2000).

Theorem 4.4 *Let θ_0 denote the true value of the parameter vector. Suppose the conditions of Theorem 4.1 hold for θ in a neighbourhood $\widetilde{\Theta}$ of θ_0 and that*

(1) The vector $\hat{a}(\theta)$ given by (3.5) and the matrix $A_n(\theta)$ are twice continuously differentiable with respect to θ,

(2) The matrices $\partial_{\theta^T}\hat{a}(\theta_0)$ and $A(\theta_0)$ have rank p,

(3) The matrices $A_n(\theta)$, $\partial_{\theta_i}A_n(\theta)$ and $\partial_{\theta_i}\partial_{\theta_j}A_n(\theta)$ converge to $A(\theta)$, $\partial_{\theta_i}A(\theta)$ and $\partial_{\theta_i}\partial_{\theta_j}A(\theta)$, respectively, uniformly for $\theta \in \widetilde{\Theta}$.

Then for every $n \geq r$, an estimator $\hat{\theta}_n$ exists that solves the estimating equation $G_n(\hat{\theta}_n) = 0$ with a probability tending to one as $n \to \infty$. Moreover,

$$\hat{\theta}_n \to \theta_0 \tag{4.8}$$

in probability and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0\,,\, D(\theta_0)^{-1}V(\theta_0)(D(\theta_0)^{-1})^T\right) \tag{4.9}$$

as $n \to \infty$ with $D(\theta_0) = A(\theta_0)U(\theta_0)$, where $U(\theta_0)$ is given by (3.4).

5 Conclusion

We have demonstrated that the problem of statistical inference for integrated diffusions can be solved in a satisfactory and readily implementable way by means of prediction-based estimating functions. We have derived optimal estimating functions and have shown that under mild regularity conditions the estimators have the usual theoretical properties of consistency and asymptotic normality.

We have, moreover, considered some of the problems encountered when the method is implemented in practice. In particular, we have demonstrated that the calculation of moments needed in order to find the optimal prediction-based estimating function is easily programmable when an analytic expression is known for the moments of the underlying diffusion process or when we can obtain the moments of the diffusion by numerical simulation. Two examples were considered in which analytic expressions for these moments are available.

Acknowledgement.

The research of Michael Sørensen was supported by MaPhySto - Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation, and by the European Commission through the Human Potential Programme.

References

Andersen, T.G. and Bollerslev, T. (1998): Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review* **39,** 885–905.

Andersen, T.G., Bollerslev, T., Diebold, F.X. and Labys, P. (2001): The distribution of realized exchange rate volatility. *J. Amer. Statist. Ass.* **96.** Forthcoming.

Barndorff-Nielsen, O.E. and Shepard, N. (2001): Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in financial econometrics (with discussion). *J. R. Statist. Soc.* **B 63,** 167–241.

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

Bibby, B.M. and Sørensen, M. (1997): A hyperbolic diffusion model for stock prices. *Finance and Stochastics* **1,** 25–41.

Bibby, B.M. and Sørensen, M. (2001): Simplified estimating functions for diffusion models with a high-dimensional parameter. *Scand. J. Statist.* **28,** 99–112.

Brockwell, P.J. and Davis, R.A. (1991): *Time Series: Theory and Methods.* Springer, New York.

Cox, J.C., Ingersoll, J.E. and Ross, S.A. (1985): A theory of the term structure of interest rates. *Econometrica* **53,** 385–407.

Ditlevsen, P.D., Ditlevsen, S. and Andersen, K.K. (2002): The fast climate fluctuations during the stadial and interstadial climate states. To appear in *Journal* *of Glaciology.*

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

Feller, W. (1951): Diffusion processes in genetics. In Neyman, J. (ed.): *Proc. Second Berkeley Symposium on Mathematical Statistics and Probability*, pp. 227–246. University of California Press, Berkeley.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999): Parameter estimation for discretely observed stochastic volatility models. *Bernoulli* **5,** 855–872.

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

Gloter, A. (1998): Parameter estimation for a discrete sampling of an integrated Ornstein-Uhlenbeck process. Preprint, Univ. of Marne-la-Vallée, 14/98. To appear in *Statistics.*

Gloter, A. (1999a): Parameter estimation for a discretely observed integrated diffusion process. Preprint, Univ. of Marne-la-Vallée, 13/99.

Gloter, A. (1999b): Parameter estimation for a hidden diffusion. Preprint, Univ. of Marne-la-Vallée, 20/99.

Godambe, V.P. and Heyde, C.C. (1988): Quasi likelihood and optimal estimation. *International Statistical Review* **55,** 231–244.

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

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

Kusuoka, S. and Yoshida, N. (1997). Malliavin calculus, geometric mixing, and expansion of diffusion functionals. To appear in *Probability Theory and Related Fields* **116**.

Pedersen, A.R. (2000): Estimating the nitrous oxide emission rate from the soil surface by means of a diffusion model. *Scand. J. Statist.* **27,** 385–403.

Rydén, T. (1994): Consistent and asymptotically normal parameter estimates for hidden Markov models. *Ann. Statist.* **22,** 1884–1895.

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

Sørensen, M. (1999): On asymptotics of estimating functions. *Brazilian Journal of Probability and Statistics* **13,** 111–136.

Sørensen, M. (2000): Prediction-based estimating functions. *Econometrics Journal* **3,** 123–147.

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

Veretennikov, A.Yu. (1997). On polynomial mixing bounds for stochastic differential equations. *Stochastic Processes and their Applications* **70,** 115 – 127.