ISSN 1398-599X MPS-RR 1999-14 www.maphysto.dk

Prediction-Based Estimating Functions.

Michael Sørensen
Department of Theoretical Statistics
University of Copenhagen
and MaPhySto*

Abstract

A generalization of martingale estimating functions is presented which is useful when there are no natural or easily calculated martingales that can be used to construct a class of martingale estimating functions. It is demonstrated that the new class of estimating functions has most of the nice properties of martingale estimating functions. Particular attention is given to classes of prediction-based estimating functions given by a finite-dimensional space of predictors. For such a class, a simple expression is found for the optimal estimating function. This particular type of prediction-based estimating functions only involve unconditional moments, in contrast to the martingale estimating functions where conditional moments are required. Thus for applications to discretely observed continuous time models, a considerably smaller amount of simulation is, in general, needed for these than for martingale estimating functions. This is also true of the optimal prediction-based estimating functions. Conditions are given that ensure the existence, consistency and asymptotic normality of the corresponding estimators. The new method is applied to inference for diffusion compartment models, sums of Ornstein-Uhlenbeck type processes, and stochastic volatility models. Stochastic volatility models are studied in considerable detail. It is demonstrated that for inference about the models in Hull and White (1988) and Chesney and Scott (1989), an explicit optimal prediction-based estimating function can be found so that no simulations are needed.

Key words: Asymptotic normality, consistency, diffusion compartment model, diffusion processes, discrete time observation of continuous time models, linear predictors, martingale estimating functions, mixing, optimal estimating functions, stochastic differential equation, stochastic volatility model, stock prices, sum of Ornstein-Uhlenbeck type processes, quasi-likelihood.

*MaPhySto - Centre for Mathematical Physics and Stochastics, funded by a grant from The Danish National Research Foundation

1 Introduction

Martingale estimating functions are useful for drawing statistical inference about diffusion models from discrete time data, see Bibby and Sørensen (1995, 1997), Sørensen (1997) and Kessler and Sørensen (1998). For some models, such as stochastic volatility models (see Example 1.1), there either is no natural martingale on which to base a class of estimating functions, or the natural martingales are too complicated to be useful in practice. One way around this problem is to use simple, explicit estimating functions of the type proposed by Kessler (1996). Another idea is studied in the present paper: A generalization of the martingale estimating functions is proposed that is shown to have most of the nice properties of the martingale estimating functions. These new estimating functions are based on predictors of functions of the observed process. Particular attention is given to classes of prediction-based estimating functions given by a finite-dimensional space of predictors. For such a class, a simple expression is found for the optimal estimating function. This particular type of prediction-based estimating functions only involve unconditional moments, in contrast to the martingale estimating functions where conditional moments are required. Thus for applications to discretely observed continuous time models, a considerably smaller amount of simulation is, in general, needed for these new estimating functions than for martingale estimating functions. This is also true of the optimal prediction-based estimating functions where the calculation of the optimal weights require extra computation. The prediction-based estimating functions are closely related to the method of prediction error estimation that is used in the stochastic control literature, see e.g. Ljung and Caines (1979), see Section 2. In this paper we consider application of the new method to diffusion compartment models, sums of Ornstein-Uhlenbeck processes, and stochastic volatility models. Obviously, the method will be useful for more general hidden Markov models too.

In Section 2, the prediction based estimating functions are presented, while the optimal estimating functions based on a finite-dimensional predictor space is derived in Section 3.

In Section 4, the use of prediction-based estimating functions for inference based on discrete time data about stochastic volatility models of the diffusion type is studied in considerable detail. It is discussed how to calculate optimal prediction-based estimating functions. In particular, it is demonstrated that for inference about the Hull and White (1988) model, an explicit optimal prediction-based estimating function can be found so that no simulations are needed. For the Chesney and Scott (1989) model too, an explicit optimal estimating function can be found. In general a certain amount of simulation is necessary. Prediction-based estimating functions can also be applied to inference about more general stochastic volatility models, for instance models with a leverage effect or the models proposed by Barndorff-Nielsen and Shephard (1998).

Other methods have previously been proposed for stochastic volatility models. One is indirect inference or the so-called efficient method of moments (which is actually only efficient under rather strong regularity conditions), see Gourieroux, Montfort, and Renault (1993), Gallant and Tauchen (1996), and Gallant and Long (1997). For an interesting application of the efficient method of moments technology to a stochastic volatility model, see Andersen and Lund (1998). These indirect inference methods are based on very extensive simulations. Probably the methods proposed in this paper require considerably less computer time. A different type of estimators was proposed by Genon-Catalot, Jeantheau and Larédo (1998b) based in limit results (where the time between observations goes to zero) in Genon-Catalot, Jeantheau and Larédo (1998a). This estimation method is simpler than the indirect inference methods and that of prediction-based estimation functions, but is probably less efficient, it can be biased if the time between observations is large, and only parameters appearing in the invariant measure can be estimated. In Genon-Catalot, Jeantheau and Larédo (1998c), estimators based on empirical moments were proposed that are consistent without the assumption that the time between observations goes to zero. Ruiz (1994) proposed a pseudo-likelihood method based on a Gaussain approximation that allowed her to apply the Kalman filter. An estimation method based on nonlinear filters was proposed by Nielsen, Vestergaard and Madsen (1999), who in particular studied the use of a second order filter.

In Section 5, two other type of models are considered briefly: diffusion compartment models and sums of Ornstein-Uhlenbeck type processes. In Section 6, asymptotic results about the estimating functions and the estimators are proved using results for mixing stochastic processes. In particular, conditions are given that ensure the existence, consistency and asymptotic normality of the estimators. For many stochastic volatility models it turns out that a result with simpler conditions hold.

We conclude this introduction by considering three examples of processes for which prediction-based estimating functions seem to be useful.

Example 1.1 *The Simple Stochastic Volatility Model.* Consider the model

$$\begin{aligned} dX_t &= \sqrt{v_t} dW_t \\ dv_t &= b(v_t;\theta)dt + c(v_t;\theta)dB_t, \end{aligned} \tag{1.1}$$

where W and B are independent standard Wiener processes. We assume that v is an ergodic, positive diffusion with invariant measure μ_θ, and that $v_0 \sim \mu_\theta$ and is independent of B. This type of model is used in mathematical finance, where X is, for instance, a model for the logarithm of the price of a stock. More complicated models are sometimes used, where X also has a drift dependent on v, or where the two Wiener processes are dependent (leverage). The model (1.1) is, however, sufficiently complex to illustrate the usefulness of the prediction-based estimating functions and some problems met when applying them. The estimating functions proposed in this paper can also be applied to the more complex models, but for these more simulation is needed than for the simple stochastic volatility model (1.1).

Suppose the process X has been observed at discrete time points $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$. The process X is not a Markov process, so it is not clear what martingale estimating function to use. Also likelihood inference is usually out of the question because it is impossible to find an explicit expression for the likelihood function.

In view of the structure of the model, it is natural to base the statistical inference on the differences $Y_i = X_{i\Delta} - X_{(i-1)\Delta}$. When X is the logarithm of a stock price, the Y_i-s are the (continuous-time) returns between the observation times. However, martingale estimating functions based on $f(Y_i) - E_\theta(f(Y_i)|Y_{i-1}, \ldots, Y_1)$ is not a feasible approach because of the problems involved in calculating the conditional expectation, analytically as well as numerically. In Section 4 we shall show that prediction-based estimating functions can be applied to this type of models.

□

Example 1.2 *Sums of Ornstein-Uhlenbeck type processes.* Consider the model given by

$$Y_t = \sum_{i=1}^{m} X_t^{(i)}, \tag{1.2}$$

where $X^{(i)}$, $i = 1, \ldots, m$, are independent, stationary Ornstein-Uhlenbeck-type processes satisfying the stochastic differential equations

$$dX_t^{(i)} = -\gamma_i X_t^{(i)} dt + dZ_t^{(i)}, i = 1, \ldots, m. \tag{1.3}$$

Here $\gamma_i > 0$, $i = 1, \ldots, m$, and $Z^{(i)}$, $i = 1, \ldots, m$, are independent (and possibly quite different) Lévy processes the distributions of which can also depend on unknown parameters. The process $X^{(i)}$ can be expressed as

$$X_t^{(i)} = e^{-\gamma_i t} X_0^{(i)} + \int_0^t e^{-\gamma_i (t-s)} dZ_s^{(i)}. \tag{1.4}$$

This model type was introduced and studied in Barndorff-Nielsen, Jensen and Sørensen (1998). A process Y of this simple type has properties that make it, for instance, useful as a model for the velocity in a turbulent fluid, see Barndorff-Nielsen, Jensen and Sørensen (1990, 1993). In such an application the $X^{(i)}$-s describe what happens at different time-scales, where the dependence structure is very dissimilar. The process type is also useful as a model for other phenomena, where there the dependence structure varies considerably over time-scales. In finance such a situation could be caused by groups of investors with different time horizons.

As in the previous example the model is non-Markovian, and when the data are observations at discrete time points, $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$, there is neither a natural martingale estimating function nor an explicit likelihood function to use for inference about the parameters of the model. We shall return to this model type in Section 5 and see how prediction-based estimating function can be applied.

□

Example 1.3 Consider the *diffusion compartment model* with d compartments defined by the stochastic differential equation

$$dX_t = A(\theta)X_t dt + \sigma(X_t;\theta)dW_t, \;\; X_0 = x_0, \tag{1.5}$$

where the jth coordinate, $X_t^{(j)}$, of the d-dimensional process X_t is the amount of material in the jth compartment at time t. The ijth non-diagonal entry of the $d \times d$-matrix $A(\theta)$, the compartmental matrix, is interpreted as the flow of material from compartment j to compartment i relative to the amount of material in compartment j. The ith diagonal element of $A(\theta)$ is usually negative and is interpreted as the flow out of compartment i relative to the amount of material in it. The process W is a d_1-dimensional standard Wiener process, and $\sigma(x;\theta)$ is a $d \times d_1$-matrix satisfying that $\sigma(x;\theta)\sigma(x;\theta)^T$ is invertible. Here and later, M^T denotes the transpose of the matrix M. The parameter θ is assumed to belong to an open subset $\Theta \subseteq \mathbb{R}^p$ such that the equation (1.5) has a unique solution for all $\theta \in \Theta$. Usually the data are observations at discrete time points of the amount of material in only a subset of the compartments, $X_t^{(1)}, \ldots, X^{(m)}$ ($m < d$) say. Let Y denote the vector of observed compartment amounts. The compartmental diffusion models of this type were introduced and studied by Bibby (1995b), who has also studied the more restrictive situation where σ is constant so that the model is Gaussian. In the Gaussian case the likelihood function is tractable, see Bibby (1994, 1995a), but in the non-Gaussian case the fact that the observed process Y is not Gaussian can cause problems. Bibby (1995b) proposed an algorithm similar to the EM-algorithm, but based on martingale estimating functions, for estimation. In Section 5 we shall consider inference for diffusion compartment models by means of prediction-based estimating functions.

□

2 Prediction-based estimating functions.

Suppose as a model for the data $Y_1, Y_2, \ldots, Y_n$ that they are observations from a stochastic process model indexed by a p-dimensional parameter $\theta \in \Theta$ ($\Theta \subseteq \mathbb{R}^p$). The model could be a continuous time model, and in that case the observation time points need not be equidistant.

Assume that f_j, $j = 1, \ldots, N$, are one-dimensional functions, defined on the state space of Y, such that $E_\theta(f_j(Y_i)^2) < \infty$ for all $\theta \in \Theta$ and for $i = 1, \ldots, n$. We denote the expectation when θ is the true parameter value by $E_\theta(\cdot)$. Let $\mathcal{F}_i$ be the σ-algebra generated by $Y_1, Y_2, \ldots, Y_i$, let $\mathcal{H}_i^\theta$ denote the L^2-space of square integrable $\mathcal{F}_i$-measurable one-dimensional random variables when θ is the true parameter value, and let $\mathcal{P}_{i,j}^\theta$, $j = 1, \cdots, N$ be closed linear subspaces of $\mathcal{H}_i^\theta$. A subspace $\mathcal{P}_{i,j}^\theta$ can be interpreted as a set of square integrable predictors of $f_j(Y_{i+1})$ given $Y_1, Y_2, \ldots, Y_i$. In this paper we shall study estimating functions of the form

$$G_n(\theta) = \sum_{i=1}^{n} \sum_{j=1}^{N} \Pi_j^{(i-1)}(\theta)[f_j(Y_i) - \hat{\pi}_j^{(i-1)}(\theta)], \tag{2.1}$$

where $\Pi_j^{(i-1)}(\theta) = (\pi_{1,j}^{(i-1)}(\theta), \ldots, \pi_{p,j}^{(i-1)}(\theta))^T$ is a p-dimensional stochastic vector, the coordinates of which belong to $\mathcal{P}_{i-1,j}^\theta$, and where $\hat{\pi}_j^{(i-1)}(\theta)$ is the minimum mean square error predictor of $f_j(Y_i)$ in $\mathcal{P}_{i-1,j}^\theta$. As is well-known, $\hat{\pi}_j^{(i-1)}(\theta)$ is the orthogonal projection of $f_j(Y_i)$ on $\mathcal{P}_{i-1,j}^\theta$ with respect to the inner product in $\mathcal{H}_i^\theta$. The projection exists and is uniquely determined by the normal equations

$$E_\theta\left(\pi[f_j(Y_i) - \hat{\pi}_j^{(i-1)}(\theta)]\right) = 0 \tag{2.2}$$

for all $\pi \in \mathcal{P}_{i-1,j}^\theta$. It follows from (2.2) that $G_n(\theta)$ is an unbiased estimating function. We will refer to estimating functions of the type (2.1) as *prediction-based estimating functions.*

An important and well-studied particular type of prediction-based estimating functions are the *martingale estimating functions.* These are obtained when $\mathcal{P}_{i-1,j}^\theta = \mathcal{H}_{i-1}^\theta$. In this case, the projection $\hat{\pi}_j^{(i-1)}(\theta)$ is the conditional expectation $E_\theta(f_j(Y_i)|Y_1, Y_2, \ldots, Y_{i-1})$, so that $G_n(\theta)$ is a θ-martingale. In some models it is, however, very difficult to find $E_\theta(f_j(Y_i)|Y_1, Y_2, \ldots, Y_{i-1})$, even numerically. In such cases other prediction-based estimating functions can be useful.

We shall be particularly interested in prediction-based estimating functions where each of the sets $\mathcal{P}_{i-1,j}^\theta$ is finite-dimensional. In this case a sufficiently explicit expression can be given for the projection to allow a study of optimal estimation and asymptotic properties of estimators. Thus we shall from now on assume that $\mathcal{P}_{i-1,j}^\theta$ is spanned by $Z_{j0}^{(i-1)}, Z_{j1}^{(i-1)}, \ldots, Z_{jq_{ij}}^{(i-1)}$ of the form $Z_{jk}^{(i-1)} = h_{jk}^{(i)}(Y_1, Y_2, \ldots, Y_{i-1})$, $k = 1, \ldots, q_{ij}$, which are linearly independent in $\mathcal{H}_{i-1}^\theta$. Note that we assume that we can choose the functions $h_{jk}^{(i)}$ to be independent of θ. To ensure that the minimum mean square error predictor of $f_j(Y_i)$ in $\mathcal{P}_{i-1,j}^\theta$ is unbiased, we assume that $Z_{j0}^{(i-1)}$ is constantly equal to 1. In this case, the predictors in $\mathcal{P}_{i-1,j}^\theta$ have the form $a_0 + a^T Z_j^{(i-1)}$, where $a^T = (a_1, \ldots, a_{q_{ij}})$ and $Z_j^{(i-1)} = (Z_{j1}^{(i-1)}, \ldots, Z_{jq_{ij}}^{(i-1)})^T$. By the normal equations (2.2), the minimum mean square error predictor of $f_j(Y_i)$ in $\mathcal{P}_{i-1,j}^\theta$ is given by

$$\hat{\pi}_j^{(i-1)}(\theta) = \hat{a}_{j0}^{(i-1)}(\theta) + \hat{a}_j^{(i-1)}(\theta)^T Z_j^{(i-1)}, \tag{2.3}$$

where

$$\hat{a}_j^{(i-1)}(\theta) = C_{i-1,j}(\theta)^{-1} b_j^{(i-1)}(\theta) \tag{2.4}$$

and

$$\hat{a}_{j0}^{(i-1)}(\theta) = E_\theta \left(f_j(Y_i)\right) - \hat{a}_j^{(i-1)}(\theta)^T E_\theta \left(Z_j^{(i-1)}\right). \tag{2.5}$$

Here $C_{i-1,j}(\theta)$ denotes the covariance matrix of $Z_j^{(i-1)}$ when θ is the true parameter value, while

$$b_j^{(i-1)}(\theta) = \left(\mathrm{Cov}_\theta\left(Z_{j1}^{(i-1)}, f_j(Y_i)\right), \ldots, \mathrm{Cov}_\theta\left(Z_{jq_{ij}}^{(i-1)}, f_j(Y_i)\right)\right)^T. \tag{2.6}$$

Thus a prediction-based estimating function can be calculated provided only that we can calculate the covariances in $C_{i-1,j}(\theta)$ and $b_j^{(i-1)}(\theta)$. Since $\hat{\pi}_j^{(i-1)}(\theta)$ depends exclusively on the first and second order moments of the random vector $\left(f_j(Y_i), Z_{j1}^{(i-1)}, \ldots, Z_{jq_{ij}}^{(i-1)}\right)$, only parameters appearing in these moments for at least one j can be estimated using (2.1). This is intuitively obvious and indeed follows from conditions given in Section 6. Of course, other f_j-s and/or other predictor spaces can be used if it is required to estimate other parameters.

In most models the moments needed must be calculated numerically, typically by simulations. That is usually not a problem because unconditional moments are much easier to determine by simulation than the conditional moments appearing in martingale estimating functions. In a number of models it is possible to calculate the required moments explicitly for certain choices of the f_j-s; see Section 4.

Example 2.1 Suppose that the observations are one-dimensional, that $N = 1$ ($j = 1$ is suppressed in the notation in this example), that $f(x) = x^2$, and that

$$\mathcal{P}_{i-1}^\theta = \left\{a_0 + a_1 Y_{i-1}^2 + \cdots + a_q Y_{i-q}^2 \,|\, a_j \in \mathrm{I\!R},\, j = 0, 1, \ldots, q\right\}. \tag{2.7}$$

Then $C_{i-1}(\theta)$ is the covariance matrix of the stochastic vector $Z^{(i-1)} = \left(Y_{i-1}^2, \ldots, Y_{i-q}^2\right)^T$, and $b^{(i-1)}(\theta) = \left(\mathrm{Cov}_\theta\left(Y_i^2, Y_{i-1}^2\right), \ldots, \mathrm{Cov}_\theta\left(Y_i^2, Y_{i-q}^2\right)\right)^T$. If moreover the process Y is stationary, $C_{i-1}(\theta)$ and $b^{(i-1)}(\theta)$ do not depend on i, so

$$\hat{\pi}^{(i-1)}(\theta) = \hat{a}_0(\theta) + \hat{a}(\theta)^T Z^{(i-1)},$$

where $\hat{a}(\theta)$ is given by (2.4), but does not depend on i, and where

$$\hat{a}_0(\theta) = E_\theta\left(Y_1^2\right)\left(1 - [\hat{a}(\theta)_1 + \cdots + \hat{a}(\theta)_q]\right). \tag{2.8}$$

With this choice of predictor space, the sum in (2.1) can only be for $i = q+1, \ldots, n$. The problem can, of course, be avoided by choosing $q_i = i - 1$ in (2.7). When the volatility process is ρ-mixing, q need not be very large because in that case the coefficients $\hat{a}(\theta)_i$ decreases exponentially with i. □

To briefly indicate the connection of prediction based estimating functions to prediction error estimation, consider the contrast function

$$K(\theta) = \sum_{i=1}^{n} \left(f(Y_i) - \breve{\pi}^{(i-1)}(\theta) \right)^2$$

where f and $\breve{\pi}^{(i-1)}(\theta)$ are as in (2.1) and (2.3) with $N = 1$, and where θ is one-dimensional. Under weak differentiability conditions, an estimator $\hat{\theta}$ that minimizes $K(\theta)$ will also solve the estimating equation $K'(\hat{\theta}) = 0$, where K' denotes the derivative of K. The estimating function $K'(\theta)$ is of the type (2.1), but it will, in general, not be the optimal estimating function based on f and $\breve{\pi}^{(i-1)}(\theta)$. In the following section, we shall see how to find the optimal prediction-based estimating function.

Finally note that a non-optimal prediction-based estimating function is obtained by differentiating the logarithm of the pseudo-likelihood function obtained by pretending that the process $\{f(Y_i)\}$ is Gaussian with the correct first and second order moments and multiplying the Gaussian conditional densities of $f(Y_i)$ given $(f(Y_{i-1}), \ldots, f(Y_{i-q}))$ for $i = q+1, \ldots, n$. The set of predictors is then spanned by 1 and $Z_k^{(i-1)} = f(Y_{i-k})$, $k = 1, \ldots, q$.

3 Optimal estimation based on linear predictors

In this section, we will give results on how to find the optimal estimating function in a class of prediction-based estimating functions given by finite-dimensional predictor-sets, $\mathcal{P}_{ij}^{\theta}$, $i = 1, \ldots, n$, $j = 1, \ldots, N$. The notation is as in the previous section. It is, however, convenient to use a more compact notation too.

The ℓth coordinate of the vector $\Pi_j^{(i-1)}(\theta)$ in (2.1) has the form

$$\pi_{\ell,j}^{(i-1)}(\theta) = \sum_{k=0}^{q_{ij}} a_{\ell jk}^{(i)}(\theta) Z_{jk}^{(i-1)},$$

where, as earlier, $Z_{j0}^{(i-1)} = 1$. Define $p \times \sum_{j=1}^{N}(q_{ij}+1)$-matrices by

$$A^{(i)}(\theta) = \tag{3.1}$$

$$\begin{pmatrix} a_{110}^{(i)}(\theta) & \cdots & a_{11q_{i1}}^{(i)}(\theta) & \cdots & \cdots & a_{1N0}^{(i)}(\theta) & \cdots & a_{1Nq_{iN}}^{(i)}(\theta) \\ \vdots & & \vdots & & & \vdots & & \vdots \\ a_{p10}^{(i)}(\theta) & \cdots & a_{p1q_{i1}}^{(i)}(\theta) & \cdots & \cdots & a_{pN0}^{(i)}(\theta) & \cdots & a_{pNq_{iN}}^{(i)}(\theta) \end{pmatrix},$$

$i = 1, \ldots, n$, and $\sum_{j=1}^{N}(q_{ij}+1)$-dimensional vectors by

$$H^{(i)}(\theta) = \begin{pmatrix} Z_{10}^{(i-1)}\left[f_1(Y_i) - \hat{\pi}_1^{(i-1)}(\theta)\right] \\ \vdots \\ Z_{1q_{i1}}^{(i-1)}\left[f_1(Y_i) - \hat{\pi}_1^{(i-1)}(\theta)\right] \\ \vdots \\ \vdots \\ Z_{N0}^{(i-1)}\left[f_N(Y_i) - \hat{\pi}_N^{(i-1)}(\theta)\right] \\ \vdots \\ Z_{Nq_{iN}}^{(i-1)}\left[f_N(Y_i) - \hat{\pi}_N^{(i-1)}(\theta)\right] \end{pmatrix}, \quad i = 1, \ldots, n. \tag{3.2}$$

Then the estimating function $G_n(\theta)$ can be written as

$$G_n(\theta) = \sum_{i=1}^{n} A^{(i)}(\theta) H^{(i)}(\theta). \tag{3.3}$$

If we, moreover, define the $p \times \left(n\sum_{j=1}^{N}(q_{ij}+1)\right)$-matrix $D_n(\theta) = \left[A^{(1)}(\theta) \ldots A^{(n)}(\theta)\right]$, and the $\left(n\sum_{j=1}^{N}(q_{ij}+1)\right)$-dimensional vector $K_n(\theta)^T = \left(H^{(1)}(\theta)^T, \ldots, H^{(n)}(\theta)^T\right)$, then

$$G_n(\theta) = D_n(\theta) K_n(\theta). \tag{3.4}$$

Proposition 3.1 *Suppose that for all $\theta \in \Theta$ the covariance matrix of $K_n(\theta)$ is invertible and $E_\theta(\partial_{\theta^T} K_n(\theta))$ has rank p. Then the estimating function*

$$G_n^*(\theta) = D_n^*(\theta) K_n(\theta), \tag{3.5}$$

where

$$D_n^*(\theta) = -E_\theta\left(\partial_\theta K_n(\theta)^T\right)\left(E_\theta\left(K_n(\theta)K_n(\theta)^T\right)\right)^{-1}, \tag{3.6}$$

is optimal within the class of estimating functions of the form (3.4) *for which $D_n(\theta)$ has rank p. The estimating function $G_n^*(\theta)$ satisfies the second Bartlett-identity with Godambe-information*

$$E_\theta(\partial_\theta K_n(\theta)^T)(E_\theta(K_n(\theta)K_n(\theta)^T))^{-1}E_\theta(\partial_{\theta^T}K_n(\theta))$$

.

Proof: By Theorem 2.1 in Heyde (1997), G^* is optimal if and only if

$$\begin{aligned} & E_\theta\left(\partial_{\theta^T} G_n(\theta)\right)^{-1} E_\theta\left(G_n(\theta) G_n^*(\theta)^T\right) \\ = \; & E_\theta\left(\partial_{\theta^T} G_n^*(\theta)\right)^{-1} E_\theta\left(G_n^*(\theta) G_n^*(\theta)^T\right) \end{aligned} \tag{3.7}$$

for all G of the form (3.4). This is the case when

$$E_\theta\left(G_n(\theta)G_n^*(\theta)^T\right) = -E_\theta\left(\partial_{\theta^T} G_n(\theta)\right) \tag{3.8}$$

for all G of the form (3.4), which obviously holds when $D_n^*(\theta)$ is given by (3.6), because

$$E_\theta\left(G_n(\theta)G_n^*(\theta)^T\right) = D_n(\theta)E_\theta(K_n(\theta)K_n(\theta)^T))D_n^*(\theta)^T$$

and

$$E_\theta\left(\partial_{\theta^T}G_n(\theta)\right) = D_n(\theta)E_\theta(\partial_{\theta^T}K_n(\theta)).$$

□

This result is not terribly interesting in practice when the dimension of the covariance matrix of $K_n(\theta)$ is large, which it is when n is large. Simpler and more useful results can be given when the process Y is *stationary*, provided that we choose as the basis of $\mathcal{P}_{i-1,j}^\theta$ vectors of the form $Z_{jk}^{(i-1)} = h_{jk}(Y_{i-1},\ldots,Y_{i-s})$, $k = 0,\ldots,q_j$, where h_{jk} is a function (independent of i) from $\mathbb{R}^s$ into $\mathbb{R}$, and where $s \in \mathbb{N}$. As usual $Z_{j0}^{(i-1)} = 1$. Note that q_j+1, the dimension of $\mathcal{P}_{i-1,j}^\theta$, does not depend on i. For some values of j and k the function $h_{jk}(x_1,\ldots,x_s)$ will usually depend only on a subset of $\{x_1,\ldots,x_s\}$. Note that the $Z_{jk}^{(i-1)}$-s are well-defined only when $i \geq s+1$, and that the process $(Y_i, Z_{11}^{(i-1)},\ldots,Z_{Nq_N}^{(i-1)})$ is stationary.

In view of these considerations, we consider estimating functions of the form (2.1) except that the sum starts at $i = s+1$, and that the $p \times \sum_{j=1}^N (q_j+1)$-matrix $A^{(i)}(\theta)$ defined by (3.1) is equal to a fixed matrix $A(\theta)$ for all i. Obviously,

$$\hat{\pi}_j^{(i-1)}(\theta) = \hat{a}_{j0}(\theta) + \hat{a}_j(\theta)^T Z_j^{(i-1)} \tag{3.9}$$

where $\hat{a}_{j0}(\theta)$ and $\hat{a}_j(\theta)$ do not depend on i. Specifically,

$$\hat{a}_j(\theta) = C_j(\theta)^{-1}b_j(\theta) \tag{3.10}$$

where $C_j(\theta)$ is the covariance matrix of $Z_j^{(s)} = \left(Z_{j1}^{(s)},\ldots,Z_{jq_j}^{(s)}\right)^T$, while the k-th entry of the q_j-dimensional vector $b_j(\theta)$ is $\mathrm{Cov}_\theta\,(Z_{jk}^{(s)},\, f_j(Y_{s+1}))$. The j-th coordinate of the N-dimensional vector $\hat{a}_0(\theta)$ is given by (2.5) with for instance $i = s+1$. In the following we will use $r = s+1$.

We can write estimating functions of the type just described as

$$G_n(\theta) = A(\theta)\sum_{i=r}^n H^{(i)}(\theta), \tag{3.11}$$

where the $\sum_{j=1}^N (q_j+1)$-dimensional vectors $H^{(i)}(\theta)$, $i = r,\ldots,n$, are given by (3.2). The process $H^{(i)}(\theta)$ is stationary, cf. (3.9). We assume that

$$p \leq \sum_{j=1}^N (q_j+1)$$

and define the $\left(\sum_{j=1}^N (q_j+1)\right) \times p$-matrix

$$U(\theta) = \bar{C}(\theta)\partial_{\theta^T}\hat{a}(\theta), \tag{3.12}$$

where $\bar{C}(\theta) = \operatorname{diag}\left(\tilde{C}_1(\theta), \ldots, \tilde{C}_N(\theta)\right)$ with

$$\tilde{C}_j(\theta) = \left\{ E_\theta \left(Z_{jk}^{(r-1)} Z_{j\ell}^{(r-1)} \right) \right\}_{k,\ell=0,\ldots,q_j}$$

(as usual $Z_{j0}^{(r-1)} = 1$) and

$$\hat{a}(\theta)^T = (\hat{a}_{10}(\theta), \hat{a}_{11}(\theta), \ldots, \hat{a}_{1q_1}(\theta), \ldots, \hat{a}_{N0}(\theta), \ldots \hat{a}_{Nq_N}(\theta))^T. \quad (3.13)$$

Note that $\tilde{C}_j(\theta)$ can be related to the matrix $C_j(\theta)$ that was used earlier by

$$\tilde{C}_j(\theta) = \begin{pmatrix} 0 & \cdots & 0 \\ \vdots & C_j(\theta) & \\ 0 & & \end{pmatrix} + E_\theta \left(\tilde{Z}_j^{(r-1)} \right)^{\otimes 2},$$

where $v^{\otimes 2} = vv^T$ for a vector v. Here $\tilde{Z}_j^{(r-1)} = \left(Z_{j0}^{(r-1)}, Z_{j1}^{(r-1)}, \ldots, Z_{jq_j}^{(r-1)} \right)^T$.

Proposition 3.2 *Suppose that for all $\theta \in \Theta$ the matrix $\partial_{\theta^T} \hat{a}(\theta)$ has rank p and that $1, f_1, \ldots, f_N$ are linearly independent on the support of the conditional distribution of Y_n given $\underline{Y}_{n-1} = (Y_1, \ldots, Y_{n-1})$. Then the matrix*

$$\bar{M}_n(\theta) = E_\theta \left(H^{(r)}(\theta) H^{(r)}(\theta)^T \right) + \quad (3.14)$$

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

is invertible, and the estimating function

$$G_n^*(\theta) = A^*(\theta; n) \sum_{i=r}^{n} H^{(i)}(\theta), \quad (3.15)$$

where

$$A^*(\theta; n) = U(\theta)^T \bar{M}_n(\theta)^{-1}, \quad (3.16)$$

is optimal within the class of estimating functions of the form (3.11) for which $A(\theta)$ has rank p. The estimating function $G_n^*(\theta)$ satisfies the second Bartlett-identity with Godambe-information

$$U(\theta)^T \bar{M}_n(\theta)^{-1} U(\theta). \quad (3.17)$$

Proof: Using the stationarity of the process $\{H^{(i)}(\theta) : i = r, r+1, \ldots\}$ it is not difficult to see that the covariance matrix of the random vector $\sum_{i=r}^n H^{(i)}(\theta)$ is $(n-r+1)\bar{M}_n(\theta)$ given by (3.14). If this covariance matrix is not strictly positive definite, there exists a non-trivial linear combination of the coordinates of the random vector $\sum_{i=r}^n H^{(i)}(\theta)$ which is identically equal to zero. This implies that there exists a j, $j = 1$ say, such that

$$f_1(Y_n) = \sum_{j=2}^{N} c_j(\underline{Y}_{n-1}) f_j(Y_n) + d(\underline{Y}_{n-1}),$$

but this contradicts the assumption made about $f_1, \ldots, f_N$. Hence the matrix $\bar{M}_n(\theta)$ is invertible.

Now, it follows from (3.9) that $U(\theta) = -E_\theta\left(\partial_{\theta^T} H^{(r)}(\theta)\right)$, so

$$E_\theta\left(G_n(\theta)G_n^*(\theta)^T\right) = (n-r+1)A(\theta)\bar{M}_n(\theta)A^*(\theta;n)^T$$

and

$$E_\theta\left(\partial_{\theta^T} G_n(\theta)\right) = (n-r+1)A(\theta)U(\theta).$$

Since the coordinates of $\tilde{Z}_j^{(r-1)}$ form a basis of $\mathcal{P}_{r-1,j}^\theta$, the matrices $\tilde{C}_j(\theta)$ are strictly positive definite. Hence so is $\bar{C}(\theta)$, and thus the matrices $U(\theta)$ and $A^*(\theta;n)$ both have rank p if and only if $\partial_{\theta^T}\hat{a}(\theta)$ has rank p. Now the proposition follows in complete analogy with the proof of Proposition 3.1.

□

Note that in Proposition 3.2 it was in particular assumed that the conditional distribution of Y_n given $\underline{Y}_{n-1}$ is not concentrated in one point. This follows from the condition on $f_1, \ldots, f_N$. Note also that the condition on $f_1, \ldots, f_N$ is satisfied for a stochastic volatility model provided that $1, f_1, \ldots, f_N$ are linearly independent on $\mathbb{R}$ and that the volatility process is non-degenerate conditionally on the past, i.e. it is a proper stochastic volatility model. For instance, $f_1(y), \ldots, f_N(y)$ could be different powers of y.

When $p = \sum_{j=1}^N (q_j+1)$ the theorem is rather empty, because in that case $A^*(\theta;n)$ is invertible and thus does not influence the estimator. In this uncommon case, $A^*(\theta;n)$ only ensures that the second Bartlett-identity holds.

It follows from (3.10) that

$$\partial_{\theta_k}\hat{a}_j(\theta) = C_j(\theta)^{-1}\left(\partial_{\theta_k} b_j(\theta) - \left(\partial_{\theta_k} C_j(\theta)\right)\hat{a}_j(\theta)\right), \tag{3.18}$$

while $\partial_{\theta^T}\hat{a}_{j0}(\theta)$ is finally found using (2.5). Thus if we know $C_j(\theta)$, $b_j(\theta)$, $E_\theta(Z_j^{(r-1)})$, $E_\theta(f_j(Y_r))$, $j = 1, \ldots, N$, their derivatives with respect to θ, and the moments appearing in (3.14), we can calculate the optimal prediction-based estimating function. Note that only moments and derivatives of moments are needed. Note also that $C_j(\theta)$, $b_j(\theta)$, $E_\theta(Z_j^{(r-1)})$, and $E_\theta(f_j(Y_r))$ were needed earlier to find the predictor $\hat{\pi}_j^{(i-1)}(\theta)$, so the only new requirements here are the derivatives and the moments in (3.14).

If the process Y is sufficiently mixing (see the Section 6), the matrix $\bar{M}_n(\theta)$ converges to a matrix $M(\theta)$ as $n \to \infty$. Asymptotically, is does not matter whether we use $U(\theta)^T\bar{M}_n(\theta)^{-1}$ or $U(\theta)^T M(\theta)^{-1}$ to define the optimal estimating function. The asymptotic variance of the two estimators will be the same, see Section 6.

4 Stochastic volatility models

In this section we will consider application of prediction-based estimating functions to inference for the stochastic volatility model in Example 1.1 when the data are $X_0, X_\Delta, X_{2\Delta}, \ldots, X_{n\Delta}$. As mentioned in Section 1, it is natural to base the statistical inference on the differences $Y_i = X_{i\Delta} - X_{(i-1)\Delta}$, $i = 1, \ldots, n$. Since

$$Y_i = \int_{(i-1)\Delta}^{i\Delta} \sqrt{v_t} dW_t \tag{4.1}$$

it follows that the process $\{Y_i\}$ is stationary, that the Y_i-s are uncorrelated, but not independent, and that

$$Y_i = \sqrt{S_i} Z_i, \tag{4.2}$$

where

$$S_i = \int_{(i-1)\Delta}^{i\Delta} v_t dt \tag{4.3}$$

and where the Z_i-s are independent, identically standard normal distributed random variables, and independent of $\{S_i\}$.

First we study the class of estimating functions given in Example 2.1, i.e. estimating functions of the form

$$G_n(\theta) = \sum_{i=q+1}^{n} \Pi^{(i-1)}(\theta) \left[Y_i^2 - \hat{a}_0(\theta) - \hat{a}_1(\theta) Y_{i-1}^2 - \cdots - \hat{a}_q(\theta) Y_{i-q}^2 \right] \tag{4.4}$$

with $\Pi^{(i-1)}(\theta) = A(\theta) \tilde{Z}^{(i-1)}$, where $A(\theta)$ is a $p \times (q+1)$-matrix and $\tilde{Z}^{(i-1)} = \left(1, Y_{i-1}^2, \ldots, Y_{i-q}^2\right)$.

In order that the minimum mean square error predictor of Y_i^2 in $\mathcal{P}_{i-1}^{\theta}$ is defined, we must assume that $E_\theta(Y_i^4) < \infty$, and in order to find it, i.e. to find $\hat{a}_0(\theta), \hat{a}_1(\theta), \cdots, \hat{a}_q(\theta)$, cf. (3.9), we must calculate $E_\theta(Y_i^2)$, $\mathrm{Var}_\theta\,(Y_i^2)$ and $\mathrm{Cov}_\theta\,(Y_i^2, Y_{i+j}^2)$, $j = 1, \ldots, q$. It is easy to see (by Jensen's inequality) that $E_\theta(v_t^{\beta/2}) < \infty$ implies $E_\theta(Y_i^\beta) < \infty$ for $\beta \geq 2$. For $\beta \leq 2$, $E_\theta(v_t) < \infty$ implies that $E_\theta(Y_i^\beta) < \infty$. In the following we shall see that we can actually calculate the moments of the observations that we need from moments of the volatility process.

Obviously, $E_\theta(Y_i^2) = E_\theta(S_1)$, $\mathrm{Var}_\theta\,(Y_i^2) = 3\mathrm{Var}_\theta\,(S_1) + 2E_\theta(S_1)^2$, and $\mathrm{Cov}_\theta\,(Y_i^2, Y_{i+j}^2) = \mathrm{Cov}_\theta\,(S_1, S_{1+j})$. Define

$$\xi(\theta) = E_\theta(v_t) \tag{4.5}$$

$$\omega(\theta) = \mathrm{Var}_\theta(v_t) \tag{4.6}$$

$$r(u;\theta) = \mathrm{Cov}_\theta(v_t, v_{t+u})/\omega(\theta). \tag{4.7}$$

It is not difficult to see that

$$
\begin{aligned}
E_\theta\left(Y_n^2\right) &= \Delta\xi(\theta) && (4.8)\\
\mathrm{Var}_\theta\left(Y_n^2\right) &= 6\omega(\theta)R^*(\Delta;\theta)+2\Delta^2\xi(\theta)^2 && (4.9)\\
\mathrm{Cov}_\theta\left(Y_n^2,Y_{n+i}^2\right) &= \omega(\theta)\left[R^*(\Delta(i+1);\theta)\right. && (4.10)\\
&\qquad \left.-\,2R^*(\Delta i;\theta)+R^*(\Delta(i-1);\theta)\right],
\end{aligned}
$$

where

$$R^*(t;\theta)=\int_0^t\int_0^s r(u;\theta)\,du\,ds;$$

see Barndorff-Nielsen and Shephard (1998). The formula (4.10) is mainly useful when a simple explicit expression for $R^*(t;\theta)$ is available. It is less useful when $R^*(t;\theta)$ must be found numerically because it expresses a possibly small number as a difference between large quantities. For numerical calculations it is perhaps more useful that

$$\mathrm{Cov}_\theta\left(Y_n^2,Y_{n+i}^2\right)=\omega(\theta)\int_{(i-1)\Delta}^{i\Delta}\int_s^{s+\Delta} r(u;\theta)\,du\,ds, \qquad (4.11)$$

which follows by easy calculations. In the following, we will use the notation $\zeta(i;\theta)=\mathrm{Cov}_\theta\left(Y_n^2,Y_{n+i}^2\right)$.

If one of the sub-optimal prediction-based estimating functions is sufficient for the purpose we have in mind (for instance, if we just need a $\sqrt{n}$-consistent estimator), we can find an explicit estimating function provided only that we can explicitly calculate the first and second order moments of the volatility process. This is simple if the volatility process has a linear drift (mean reversion), under mild regularity conditions on the diffusion coefficient.

In order to find the optimal choice of the matrix $A(\theta)$ in the expression for $\Pi^{(i-1)}(\theta)$ in (4.4), i.e. the optimal matrix $A^*(\theta;n)$ in Proposition 3.2, we also need to calculate the covariances $E_\theta\left(H^{(r)}(\theta)_j H^{(r+i)}(\theta)_k\right)$, where $H^{(i)}(\theta)_j$ is the j-th coordinate of

$$H^{(i)}(\theta)=\begin{pmatrix} Y_i^2-\hat a_0(\theta)-\hat a_1(\theta)Y_{i-1}^2-\cdots-\hat a_q(\theta)Y_{i-q}^2 \\[2ex] Y_{i-1}^2\left[Y_i^2-\hat a_0(\theta)-\hat a_1(\theta)Y_{i-1}^2-\cdots-\hat a_q(\theta)Y_{i-q}^2\right] \\ \vdots \\ Y_{i-q}^2\left[Y_i^2-\hat a_0(\theta)-\hat a_1(\theta)Y_{i-1}^2-\cdots-\hat a_q(\theta)Y_{i-q}^2\right]\end{pmatrix}, \qquad (4.12)$$

$i=1,\ldots,n$. For these covariances to exist, we must assume that $E_\theta(Y_i^8)<\infty$. Straightforward calculations show that if $j\neq 0$ and $k\neq 0$, then

$$
\begin{aligned}
E_\theta\left(H^{(r)}(\theta)_j H^{(r+i)}(\theta)_k\right) &= \sum_{\nu=0}^q\sum_{\kappa=0}^q b_\nu(\theta)b_\kappa(\theta)E_\theta\left(Y_{r-\nu}^2Y_{r+i-\kappa}^2Y_{r-j}^2Y_{r+i-k}^2\right)\\
&+\hat a_0(\theta)\sum_{\kappa=0}^q b_\kappa(\theta)\left[E_\theta\left(Y_{r+i-\kappa}^2Y_{r-j}^2Y_{r+i-k}^2\right)+E_\theta\left(Y_{r-\kappa}^2Y_{r-j}^2Y_{r+i-k}^2\right)\right]\\
&+\hat a_0(\theta)^2\left[\zeta(|j-k+i|;\theta)+\Delta^2\xi(\theta)^2\right], \qquad (4.13)
\end{aligned}
$$

where

$$b_\kappa(\theta) = \begin{cases} -1 & \text{for } \kappa = 0 \\ \hat{a}_\kappa(\theta) & \text{for } \kappa = 1, \ldots, q. \end{cases}$$

When $k = 0$ and $j \neq 0$,

$$\begin{aligned} E_\theta\left(H^{(r)}(\theta)_j H^{(r+i)}(\theta)_0\right) &= \sum_{\nu=0}^{q}\sum_{\kappa=0}^{q} b_\nu(\theta) b_\kappa(\theta) E_\theta\left(Y_{r-\nu}^2 Y_{r+i-\kappa}^2 Y_{r-j}^2\right) \\ &+ \hat{a}_0(\theta) \sum_{\kappa=0}^{q} b_\kappa(\theta)\left[\zeta(|j-\kappa+i|;\theta) + \zeta(|j-\kappa|;\theta)\right] \\ &- \hat{a}_0(\theta)^2 \Delta\xi(\theta), \end{aligned} \tag{4.14}$$

and when $j = k = 0$,

$$\begin{aligned} &E_\theta\left(H^{(r)}(\theta)_0 H^{(r+i)}(\theta)_0\right) \\ &= \sum_{\nu=0}^{q}\sum_{\kappa=0}^{q} b_\nu(\theta) b_\kappa(\theta) \zeta(|\nu-\kappa+i|;\theta). \end{aligned} \tag{4.15}$$

A number of times we have used the identity $\sum_{\nu=0}^{q} b_\nu(\theta) = -\hat{a}_0(\theta)/(\Delta\xi(\theta))$, cf. (2.8). The expression for $E_\theta\left(H^{(r)}(\theta)_0 H^{(r+i)}(\theta)_j\right)$ $(j \neq 0)$ is obtained from (4.14) by interchanging r and $r+i$ in the first line and changing the sign of i in the second line. Another expression for $E_\theta\left(H^{(r)}(\theta)_j H^{(r+i)}(\theta)_k\right)$ with $j \neq 0$ and $k \neq 0$ is

$$\sum_{\nu=0}^{q}\sum_{\kappa=0}^{q} b_\nu(\theta) b_\kappa(\theta) E_\theta\left[Y_{r-j}^2 Y_{r+i-k}^2 \left(Y_{r-\nu}^2 - \Delta\xi(\theta)\right)\left(Y_{r+i-\kappa}^2 - \Delta\xi(\theta)\right)\right].$$

A similar expression can be found when $j = 0$ or $k = 0$. If, for instance, $j = 0$, the random variable Y_{r-j}^2 should simply be removed under the expectation sign. Analogously when $k = 0$.

We see that apart from the moments we already have considered, we need moments of the form $E_\theta\left(Y_i^2 Y_j^2 Y_1^2\right)$ and $E_\theta\left(Y_i^2 Y_j^2 Y_k^2 Y_1^2\right)$, where we can assume that $i \geq j \geq k$. The stationarity of the process Y implies that we can always take the smallest index to equal one. Obviously the value of these moments depends on whether some of the indices coincide.

Define

$$\begin{aligned} \varphi(s,t,u;\theta) &= E_\theta\left(v_s v_t v_u\right) \\ \psi(s,t,u,z;\theta) &= E_\theta\left(v_s v_t v_u v_z\right) \end{aligned}$$

and

$$\begin{aligned} T_3 &= \{(s,t,u) : s > t > u\} \\ T_4 &= \{(s,t,u,z) : s > t > u > z\} \\ B_3 &= [(i-1)\Delta, i\Delta] \times [(j-1)\Delta, j\Delta] \times [0,\Delta] \ \cap \ T_3 \\ B_4 &= [(i-1)\Delta, i\Delta] \times [(j-1)\Delta, j\Delta] \times [(k-1)\Delta, k\Delta] \times [0,\Delta] \ \cap \ T_4. \end{aligned}$$

Moreover, for a positive integer λ, define $c_\lambda = \lambda! \prod_{i=1}^{\lambda}(2i-1)$. If λ indices coincide

$$E_\theta\left(Y_i^2 Y_j^2 Y_1^2\right) = c_\lambda \int_{B_3} \varphi(s,t,u;\theta)dsdtdu. \tag{4.16}$$

If the three indices are different, $\lambda = 1$ and $c_1 = 1$. For $E_\theta\left(Y_i^2 Y_j^2 Y_k^2 Y_1^2\right)$ the situation is a bit more complicated. If all indices are different ($\lambda = 1$) or if there is only one set of coinciding indices of size $\lambda > 1$, then

$$E_\theta\left(Y_i^2 Y_j^2 Y_k^2 Y_1^2\right) = c_\lambda \int_{B_4} \psi(s,t,u,z;\theta)dsdtdudz. \tag{4.17}$$

If there are two pairs of indices, each of which has the same value (with a different value for the two pairs), then

$$E_\theta\left(Y_i^2 Y_j^2 Y_k^2 Y_1^2\right) = c_2^2 \int_{B_4} \psi(s,t,u,z;\theta)dsdtdudz. \tag{4.18}$$

When there are no coinciding indices these formulae are straightforward since, for instance, $E_\theta\left(Y_i^2 Y_j^2 Y_1^2\right) = E_\theta\left(S_i S_j S_1\right) E_\theta\left(Z_1^2\right)^3$, where $E_\theta\left(Z_1^2\right) = 1$. In this case, the integral in (4.16) is over the whole set $[(i-1)\Delta, i\Delta] \times [(j-1)\Delta, j\Delta] \times [0,\Delta]$. If there are two indices that are identical, then $E_\theta\left(Y_i^2 Y_j^2 Y_1^2\right) = E_\theta\left(S_i S_j S_1\right) E_\theta\left(Z_1^4\right)^2 E_\theta\left(Z_1^2\right)$, where $E_\theta\left(Z_1^4\right) = 3$. Here it is not necessary to integrate over the whole box $[(i-1)\Delta, i\Delta] \times [(j-1)\Delta, j\Delta] \times [0,\Delta]$ because the value of the function $\varphi(s,t,u;\theta)$ does not change when the arguments s,t,u are permuted. Thus it is enough to integrate over the set B_3 and then multiply by $f_2^{-1} = 2$, where f_2 is the ratio of the volume of the set B_3 to the volume of the whole box over which we should integrate. The formulae for the other cases are obtained similarly. The constant c_λ is simply the product of $E_\theta\left(Z_1^{2\lambda}\right) = \prod_{i=1}^{\lambda}(2i-1)$ and $f_\lambda^{-1} = \lambda!$, where f_λ is the ratio of the volume of the set over which we actually integrate (B_3 or B_4) to the volume of the whole box over which we should integrate.

To finish we only need to find the moments in (4.5) – (4.7) and in the functions φ and ψ. The marginal moments (4.5) and (4.6) are relatively easy to calculate from the marginal distribution of v_t, for which a simple expression exists in terms of the drift and diffusion coefficients of the volatility process. The other moments require knowledge of the dependence structure of the volatility process, and for most models these moments must be obtained by simulation. The extent of the simulations is considerably smaller than that needed for martingale estimating functions because only unconditional moments must be calculated, which can be obtained from a single long trajectory.

There are, however, models for which it is possible to give explicit expressions for the moments needed because certain conditional moments can be explicitly found. Assume for instance that the eigenfunctions of the generator of the volatility process are polynomials. Recall that the generator is the differential operator

$$L_\theta = b(x;\theta)\frac{d}{dt} + \tfrac{1}{2}c^2(x;\theta)\frac{d^2}{dx^2}\ , \tag{4.19}$$

and that a function h is called an eigenfunction for L_θ if there exists a positive number λ (called the eigenvalue) such that $L_\theta h = -\lambda h$. Suppose that

$$h_i(x;\theta) = \sum_{j=0}^{i} \rho_{ij}(\theta) x^j, \quad i = 1, 2, 3, \tag{4.20}$$

where $\rho_{ii}(\theta) \neq 0$, are eigenfunction for L_θ with eigenvalues $\lambda_1(\theta), \lambda_1(\theta), \lambda_3(\theta)$. Under weak regularity conditions, see e.g. Kessler and Sørensen (1998),

$$E_\theta(h_i(v_t;\theta)|v_0 = x) = e^{-\lambda_i(\theta)t} h_i(x;\theta).$$

Thus inserting v_t for x and taking the conditional expectation given v_0 on both sides of (4.20) yields

$$e^{-\lambda_i(\theta)t} h_i(v_0;\theta) = \sum_{j=0}^{i} \rho_{ij}(\theta) \pi_j(x,t;\theta) \quad i = 1, 2, 3,$$

where $\pi_k(x,t;\theta) = E_\theta(v_t^k | v_0 = x)$. In particular, $\pi_0(x,t;\theta) = 1$. From this linear equation it is easy to get explicit expressions for the conditional moments $\pi_i(x,t;\theta)$:

$$\pi_i(x,t;\theta) = \sum_{j=0}^{i} \nu_{ij}(t;\theta) x^j,$$

where

$$\begin{aligned}
\nu_{10}(t;\theta) &= \rho_{11}(\theta)^{-1} \rho_{10}(\theta) \left(e^{-\lambda_1(\theta)t} - 1\right) \\
\nu_{11}(t;\theta) &= e^{-\lambda_1(\theta)t} \\
\\
\nu_{20}(t;\theta) &= \rho_{22}(\theta)^{-1} \Big[\rho_{20}(\theta) \left(e^{-\lambda_2(\theta)t} - 1\right) \\
&\qquad - \rho_{21}(\theta)\rho_{11}(\theta)^{-1}\rho_{10}(\theta) \left(e^{-\lambda_1(\theta)t} - 1\right)\Big] \\
\nu_{21}(t;\theta) &= \rho_{22}(\theta)^{-1} \rho_{21}(\theta) \left(e^{-\lambda_2(\theta)t} - e^{-\lambda_1(\theta)t}\right) \\
\nu_{22}(t;\theta) &= e^{-\lambda_2(\theta)t} \\
\\
\nu_{30}(t;\theta) &= \rho_{33}(\theta)^{-1} \Big\{\rho_{30}(\theta) \left(e^{-\lambda_3(\theta)t} - 1\right) \\
&\qquad - \rho_{32}(\theta)\rho_{22}(\theta)^{-1}\rho_{20}(\theta) \left(e^{-\lambda_2(\theta)t} - 1\right) \\
&\qquad + \left[\rho_{32}(\theta)\rho_{22}(\theta)^{-1}\rho_{21}(\theta) - \rho_{31}(\theta)\right] \rho_{11}(\theta)^{-1} \\
&\qquad\qquad \times \rho_{10}(\theta) \left(e^{-\lambda_1(\theta)t} - 1\right)\Big\} \\
\nu_{31}(t;\theta) &= \rho_{33}(\theta)^{-1} \Big\{\rho_{31}(\theta) \left(e^{-\lambda_3(\theta)t} - e^{-\lambda_1(\theta)t}\right) \\
&\qquad - \rho_{32}(\theta)\rho_{22}(\theta)^{-1}\rho_{21}(\theta) \left(e^{-\lambda_2(\theta)t} - e^{-\lambda_1(\theta)t}\right)\Big\} \\
\nu_{32}(t;\theta) &= \rho_{33}(\theta)^{-1} \rho_{32}(\theta) \left(e^{-\lambda_3(\theta)t} - e^{-\lambda_2(\theta)t}\right) \\
\nu_{33}(t;\theta) &= e^{-\lambda_3(\theta)t}.
\end{aligned}$$

Now, repeated iterations of conditional expectations show that for $s > t > u > z$ (which is enough)

$$E_\theta\left(v_s v_t\right) = \tag{4.21}$$

$$\left(E_\theta(v_t^2) + \rho_{11}(\theta)^{-1}\rho_{10}(\theta)\xi(\theta)\right)e^{-\lambda_1(\theta)(s-t)} - \rho_{11}(\theta)^{-1}\rho_{10}(\theta)\xi(\theta),$$

$$\begin{aligned}\varphi(s,t,u;\theta) = \nu_{10}(s-t;\theta)\sum_{j=0}^{1}\nu_{1j}(t-u;\theta)E_\theta(v_t^{j+1}) \\ + \nu_{11}(s-t;\theta)\sum_{j=0}^{2}\nu_{2j}(t-u;\theta)E_\theta(v_t^{j+1}),\end{aligned} \tag{4.22}$$

and

$$\psi(s,t,u,z;\theta) = \tag{4.23}$$

$$\begin{aligned}&[\nu_{10}(s-t;\theta)\nu_{10}(t-u;\theta) + \nu_{11}(s-t;\theta)\nu_{20}(t-u;\theta)] \\ &\qquad \times \sum_{j=0}^{1}\nu_{1j}(u-z;\theta)E_\theta(v_t^{j+1}) \\ &+ [\nu_{10}(s-t;\theta)\nu_{11}(t-u;\theta) + \nu_{11}(s-t;\theta)\nu_{21}(t-u;\theta)] \\ &\qquad \times \sum_{j=0}^{2}\nu_{2j}(u-z;\theta)E_\theta(v_t^{j+1}) \\ &+ \nu_{11}(s-t;\theta)\nu_{22}(t-u;\theta)\sum_{j=0}^{3}\nu_{3j}(u-z;\theta)E_\theta(v_t^{j+1}).\end{aligned}$$

The marginal moments $E_\theta(v_t^2)$, $E_\theta(v_t^3)$, and $E_\theta(v_t^4)$ can be found from the stationary distribution for which an explicit expression is available in terms of the drift and the diffusion coefficient.

Example 4.1 Consider the volatility process given by

$$dv_t = -\theta(v_t - \alpha)dt + \sigma\sqrt{v_t}dB_t. \tag{4.24}$$

This stochastic volatility model was proposed by Hull and White (1988), and was also considered by Heston (1993).

The process (4.24) 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 $2\theta\alpha \geq \sigma^2$. Thus

$$E_{\alpha,\theta,\sigma}\left(v_t^i\right) = \left(\alpha + \frac{(i-1)\sigma^2}{2\theta}\right)E_{\alpha,\theta,\sigma}\left(v_t^{i-1}\right),\ i = 1, 2, \ldots.$$

In particular, $\xi(\alpha,\theta,\sigma) = \alpha$, $E_{\alpha,\theta,\sigma}(v_t^2) = \alpha(\alpha + \sigma^2/(2\theta))$, $E_{\alpha,\theta,\sigma}(v_t^3) = \alpha(\alpha + \sigma^2/(2\theta))(\alpha + \sigma^2/\theta)$, $E_{\alpha,\theta,\sigma}(v_t^4) = \alpha(\alpha + \sigma^2/(2\theta))(\alpha + \sigma^2/\theta)(\alpha + 3\sigma^2/(2\theta))$, and $\mathrm{Var}_{\alpha,\theta,\sigma}(v_t) = \alpha\sigma^2/(2\theta)$.

The eigenfunctions of the generator are Laguerre polynomials evaluated at $2\theta x\sigma^{-2}$, see e.g. Karlin and Taylor (1981). Specifically, the i-th eigenfunction is the i-th order polynomial with coefficients

$$\rho_{ij}(\alpha,\theta,\sigma) = \frac{(-2\theta\sigma^{-2})^j}{j!}\begin{pmatrix} i+2\alpha\theta\sigma^{-2}-1 \\ i-j \end{pmatrix},$$

$j = 0,\ldots,i$, $i = 1,2,\ldots$. The corresponding eigenvalue is $i\theta$. It follows that $r(u;\alpha,\theta,\sigma) = e^{-\theta u}$. Explicit expressions for the functions $\varphi(s,t,u;\alpha,\theta,\sigma)$ and $\psi(s,t,u,z;\alpha,\theta,\sigma)$ too follow from (4.22) and (4.23).

□

Example 4.2 Consider the volatility process given by $v_t = \exp(U_t)$, where U is a stationary Gaussian Ornstein-Uhlenbeck process, $dU_t = -\theta(U_t-\alpha)dt + \sigma dB_t$ with $\theta > 0$ (Wiggins (1987); Chesney and Scott (1989); Melino and Turnbull (1990)). The model can be obtained as a limit of the EGARCH(1,1) model, see Nelson (1990).

Also for this model the moments necessary to find the optimal prediction-based estimating function can be calculated explicitly, because U is Gaussian. For instance,

$$\psi(s,t,u,z;\theta,\alpha,\sigma) = E_{\theta,\alpha,\sigma}\left[\exp\left(U_s+U_t+U_u+U_z\right)\right],$$

which is the Laplace transform of a known Gaussian distribution.

□

It is important to have alternatives to the simple estimating function discussed so far in this section. It might, for instance, be the case that the moments needed in the optimal estimating function do not exist. An alternative could be estimating functions of the form

$$G_n(\theta) = \sum_{i=q+1}^{n} \Pi^{(i-1)}(\theta)\left\{|Y_i|^\gamma - \hat{a}_0(\theta) - \hat{a}_1(\theta)|Y_{i-1}|^\gamma - \cdots - \hat{a}_q(\theta)|Y_{i-q}|^\gamma\right\} \tag{4.25}$$

with $\Pi^{(i-1)}(\theta) = A(\theta)\tilde{Z}^{(i-1)}$ and $\tilde{Z}^{(i-1)} = (1,|Y_{i-1}|^\gamma,\ldots,|Y_{i-q}|^\gamma)$. Here γ is some suitably chosen positive real number. If, for instance, $\gamma = \frac{1}{2}$ we need to assume only that $E_\theta(Y_i^2) < \infty$ for the optimal estimating function of the type (4.25) to exist. The price is that it is not as easy to calculate the moments needed.

For the estimating function (4.25) the analysis goes much like the analysis above with Y_i^2 replaced by $|Y_i|^\gamma$. The problem is that the expression (4.2) for Y_i fits estimating functions with $\gamma = 2$ particularly well. In order to find the minimum mean square error predictor of Y_i^2 in $\mathcal{P}_{i-1}^\theta$, we need $E_\theta(|Y_i|^\gamma)$, $\mathrm{Var}_\theta\left(|Y_i|^\gamma\right)$ and $\mathrm{Cov}_\theta\left(|Y_i|^\gamma,|Y_{i+j}|^\gamma\right)$, $j = 1,\ldots,q$. Obviously,

$$\begin{aligned} E_\theta(|Y_i|^\gamma) &= E_\theta(S_1^{\gamma/2})E_\theta(|Z_1|^\gamma) \\ &= E_\theta(S_1^{\gamma/2})2^{\gamma/2},\ ((\gamma+1)/2)/\sqrt{\pi}, \end{aligned} \tag{4.26}$$

$$
\begin{aligned}
E_\theta\left(|Y_i|^{2\gamma}\right) &= E_\theta(S_1^\gamma)E_\theta(|Z_1|^{2\gamma}) \\
&= E_\theta(S_1^\gamma)2^\gamma ,\ (\gamma+1/2)/\sqrt{\pi},
\end{aligned}
\tag{4.27}
$$

and

$$
E_\theta(|Y_i|^\gamma|Y_{i+j}|^\gamma) = E_\theta(S_1^{\gamma/2}S_{1+j}^{\gamma/2})2^\gamma ,\ ((\gamma+1)/2)^2/\pi,
\tag{4.28}
$$

where , denotes the gamma-function. Unfortunately, there is not in general a simple way of relating the moments $E_\theta(S_1^\beta)$ to the moments of the volatility process v when β is not an integer. Therefore the non-integer moments of S_1 must be found by simulation, which on the other hand is not difficult. If $\gamma = 1$, we can of course find $E_\theta(S_1^\gamma)$ by using the results derived earlier.

In order to find the optimal choice of the matrix $A(\theta)$ in the expression for $\Pi^{(i-1)}(\theta)$ in (4.25), we need to calculate the covariances $E_\theta\left(H^{(r)}(\theta)_j H^{(r+i)}(\theta)_k\right)$. Calculations similar to those made earlier in this section show that what we further need are moments of the form $E_\theta\left(|Y_i|^\gamma|Y_j|^\gamma|Y_k|^\gamma|Y_1|^\gamma\right)$ and $E_\theta\left(|Y_i|^\gamma|Y_j|^\gamma|Y_1|^\gamma\right)$ $(i \geq j \geq k)$. As above these moments can be related to moments of a standard normal distribution and the moments $E_\theta(S_i^{\gamma/2}S_j^{\gamma/2}S_k^{\gamma/2}S_1^{\gamma/2})$ and $E_\theta(S_i^{\gamma/2}S_j^{\gamma 2}S_1^{\gamma/2})$. In a few cases where some indices coincide and where $\gamma = 1$ or $\gamma = \frac{1}{2}$, these moments can be calculated by results given earlier, but in most cases they must be found by simulation. As the details are rather similar to those for the case $\gamma = 2$ they are omitted.

If computing time is important, it might be an idea to sacrifice some efficiency by replacing S_i by $\Delta v_{(i-1)\Delta}$ when calculating the moments used in the optimal choice of the vectors $A_j(\theta)$. For instance, $E_\theta(S_i^{\gamma/2}S_j^{\gamma 2}S_1^{\gamma/2})$ could be approximated by $\Delta^{3\gamma/2}E_\theta\left(v_{(i-1)\Delta}^{\gamma/2}v_{(j-1)\Delta}^{\gamma/2}v_0^{\gamma/2}\right)$.

5 Other models

In this section we will briefly consider two other types of models, the diffusion compartment models and sums of Ornstein-Uhlenbeck-type processes.

5.1 Diffusion compartment models

First we consider the two compartment model $X_t = (X_t^{(1)}, X_t^{(2)})^T$ given by (1.5). The data are $Y_1, \ldots, Y_n$, where $Y_i = X_{i\Delta}^{(1)}$. A natural choice of prediction-based estimating function is one with $N = 2$, $f_1(y) = y$, and $f_2(y) = y^2$. It might simplify matters to choose different spaces of predictors for each of the two functions f_j. Here, however, we specify a single space of predictors by the choice $Z_j^{(i-1)} = (Y_{i-1}, \ldots, Y_{i-q}, Y_{i-1}^2, \ldots, Y_{i-q}^2)^T$, $j = 1, 2$.

To find the optimal estimating function, we need moments of the form $E_\theta\left(Y_{i_1}^{\kappa_1}Y_{i_2}^{\kappa_2}Y_{i_3}^{\kappa_3}Y_{i_4}^{\kappa_4}\right)$ with $\kappa_j \in \{0,1,2\}$. Some of the indices i_j

might coincide. Under weak regularity conditions on σ, the conditional moments $\mu_t(x;\theta) = E_\theta(X_t|X_0 = x)$ can be found explicitly, because $\mu_t(x;\theta)$ satisfies the differential equation

$$\frac{d\mu_t(x;\theta)}{dt} = A(\theta)\mu_t(x;\theta),$$

the solution of which can be expressed in terms of a matrix exponential function. For a particular model, Bibby (1994) gave an explicit expression for the solution. If also the marginal second moment of Y_i is known explicitly, the moments $Y_{i_1}Y_{i_2}$ can be found. If higher order marginal moments are known, we can also find a few other mixed moments. Except for very special models, the other moments must be found numerically.

5.2 Sums of Ornstein-Uhlenbeck-type processes

For a model given by (1.2), one could try a prediction-based estimating function like then one considered in the previous subsection. Also for this type of models the marginal moments and the moments of the form $Y_{i_1}Y_{i_2}$ can be calculated explicitly, whereas the rest must be found numerically. For the discrete time versions of these models (see Barndorff-Nielsen, Jensen and Sørensen (1998)), the necessary moments can, as least in principle, be calculated explicitly.

The idea behind models that are sums of Ornstein-Uhlenbeck type processes is, however, that the processes in the sum (1.2) represent what happens at different time scales, and the estimating function ought to be chosen in accordance with this. One way of doing so is to use the same functions f_j as in subsection 5.1, but to use a number of different spaces of predictors for each of them. If $m = 2$, two spaces of predictors could, for instance, be specified by the choices $Z_1^{i-1} = (Y_{i-1}, \ldots, Y_{i-q_1}, Y_{i-1}^2, \ldots, Y_{i-q_1}^2)$ and $Z_2^{i-1} = (Y_{i-q_2}, \ldots, Y_{i-q_3}, Y_{i-q_2}^2, \ldots, Y_{i-q_3}^2)$ for appropriate choices of q_1, q_2 and q_3 $(q_2 < q_3)$.

6 Asymptotic results

In this section we give asymptotic results for prediction-based estimating functions and the corresponding estimators when the observed process Y is stationary. We will consider estimating function of the general form

$$G_n(\theta) = A_n(\theta)\sum_{i=r}^{n} H^{(i)}(\theta), \tag{6.1}$$

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) with Z-s of the type defined after the proof of Proposition 3.1 and with $\hat{\pi}$-s given by (3.9).

The asymptotic results are based on results for mixing stochastic processes. An extensive review of such results are given in the book by

Doukhan (1994). A very useful review is also contained in Genon-Catalot, Jeantheau and Larédo (1998c).

Theorem 6.1 *Suppose Y is stationary and α-mixing with mixing coefficients $\alpha_k(\theta)$, $k = 1, 2, \ldots$, and that there exists a $\delta > 0$ such that*

$$\sum_{k=1}^{\infty} \alpha_k(\theta)^{\delta/(2+\delta)} < \infty \tag{6.2}$$

and

$$E_{\theta_0}\left(\left| H^{(r)}(\theta)_j \right|^{2+\delta} \right) < \infty, \quad j = 0, \ldots, q. \tag{6.3}$$

Then as $n \to \infty$,

$$\bar{M}_n(\theta) \to M(\theta), \tag{6.4}$$

where $\bar{M}_n(\theta)$ is given by (3.14) and where

$$M(\theta) = E_\theta\left(H^{(r)}(\theta) H^{(r)}(\theta)^T \right) + \sum_{k=1}^{\infty} \left[E_\theta\left(H^{(r)}(\theta) H^{(r+k)}(\theta)^T \right) + E_\theta\left(H^{(r+k)}(\theta) H^{(r)}(\theta)^T \right) \right]. \tag{6.5}$$

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

$$n^{-1}\mathrm{Var}_\theta\left(G_n(\theta)\right) \to V(\theta) = A(\theta) M(\theta) A(\theta)^T, \tag{6.6}$$

and

$$\frac{1}{\sqrt{n}} G_n(\theta) \to N\left(0, V(\theta)\right) \tag{6.7}$$

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

Proof: Next note that since $H^{(i)}(\theta)$ is a function of $Y_{i-s}, \ldots, Y_i$ (see the definition of $Z_{jk}^{(i-1)}$ after Proposition 3.1), it follows that the process $H^{(i)}(\theta)$, $i = s+1, s+2, \ldots$ is α-mixing with mixing coefficients $\alpha_{k+s}(\theta)$, $k = 1, 2, \ldots$. 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 Theorem 1 in Section 1.5 of Doukhan (1994).

□

In several models, the sequence of mixing coefficients $\alpha_k(\theta)$ decreases exponentially so that (6.2) is automatically satisfied. This is, for instance, often the case for stochastic volatility models, see Subsection 6.1 below. A weaker condition ensuring that a central limit theorem holds can be found in Doukhan, Massart and Rio (1994).

It follows from Theorem 6.1 that also the optimal estimating function $G_n^*(\theta)$ given by (3.15) and (3.16) is asymptotically normal, i.e.

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

in distribution as $n \to \infty$, under the conditions of the theorem, provided that the limit matrix $A^*(\theta) = \lim_{n\to\infty} U(\theta)^T \bar{M}_n(\theta)^{-1}$ is finite and that the limit covariance matrix

$$V^*(\theta) = A^*(\theta)M(\theta)A^*(\theta)^T = A^*(\theta)U(\theta)$$

is strictly positive definite. A sufficient condition for this is that the limit covariance matrix $M(\theta)$ is strictly positive definite and that the matrix $\partial_{\theta^T}\hat{a}(\theta)$ has full rank (rank p). The vector $\hat{a}(\theta)$ is given by (3.13). In this case $A^*(\theta) = U(\theta)^T M(\theta)^{-1}$ and $V^*(\theta) = U(\theta)^T M(\theta)^{-1} U(\theta)$. Note that the estimating function where $A_n(\theta)$ is equal to $U(\theta)^T M(\theta)^{-1}$ for all n is optimal too.

Theorem 6.1 implies a result about the estimator obtained from an estimating function $G_n(\theta)$ of the form (6.1), in particular $G_n^*(\theta)$. In the following θ_0 denotes the true value of the parameter vector.

Theorem 6.2 *Suppose the conditions of Theorem 6.1 hold for θ in a neighbourhood $\widetilde{\Theta}$ of θ_0 and that*

(1) The vector $\hat{a}(\theta)$ given by (3.13) 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,

$$\hat{\theta}_n \to \theta_0 \tag{6.8}$$

in probability and

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{\mathcal{D}} N\left(0\,,\ D(\theta_0)^{-1}V(\theta_0)(D(\theta_0)^{-1})^T\right) \tag{6.9}$$

as $n \to \infty$ with $D(\theta_0) = A(\theta_0)U(\theta_0)$, where $U(\theta_0)$ is given by (3.12).

Proof: The theorem follows from Corollary 2.7 and Theorem 2.8 in Sørensen (1998) with $\bar{\theta} = \theta_0$. To simplify the notation we will only consider the case where θ is one-dimensional. The general case is proved in exactly the same way, apart from notational complications and the fact that it must be checked that the matrix $D(\theta_0)$ is invertible, which is straightforward under the conditions imposed (cf. (3.12)).

Define $M_n = \{\theta \in \Theta : |\theta - \theta_0| \leq \alpha/\sqrt{n}\}$ $(\alpha > 0)$. Conditions 2.6 (i) and (iv) in Sørensen (1998) are obviously satisfied. Condition (v) follows because (3) implies that there exists $n_0 \in \mathbb{N}$ such that when $n \geq n_0$, the absolute value of the coordinates of the vectors $A_n(\theta)$, $\partial_\theta A_n(\theta)$ and $\partial_\theta^2 A_n(\theta)$ are bounded by some $M > 0$ for all $\theta \in \tilde{\Theta}$. Hence for $n \geq n_0$

$$\begin{aligned}\sup_{\theta \in M_n} n^{-1}|G_n(\theta)| \;&\leq\; M \sum_{j=1}^{N} \sum_{k=1}^{q_j} \Bigg(|n^{-1} F_n(\theta_0)_{jk}| \\ &\quad + \sup_{\theta \in M_n} |\hat{a}_{jk}(\theta) - \hat{a}_{jk}(\theta_0)| \sum_{\ell=1}^{q_j} n^{-1} \sum_{i=r}^{n} |Z_{jk}^{(i-1)} Z_{j\ell}^{(i-1)}| \Bigg),\end{aligned}$$

where $F_n(\theta) = \sum_{i=r}^n H^{(i)}(\theta)$. Now use that $\hat{a}_{jk}(\theta)$ is continuous, that $n^{-1}F_n(\theta_0) \to 0$, and that $n^{-1}\sum_{i=r}^n |Z_{jk}^{(i-1)} Z_{j\ell}^{(i-1)}| \to E_\theta \left(Z_{jk}^{(s)} Z_{j\ell}^{(s)} \right)$, which is finite.

To check Condition (ii) in Sørensen (1998), note that

$$\begin{aligned} &|n^{-1}\partial_\theta G_n(\theta) + D(\theta_0)| \qquad\qquad (6.10)\\ &\quad\leq\; |\partial_\theta A_n(\theta) n^{-1} F_n(\theta)| + |(A_n(\theta) - A(\theta_0)) n^{-1} \partial_\theta F_n(\theta)| \\ &\qquad + |A(\theta_0) n^{-1} (\partial_\theta F_n(\theta) - \partial_\theta F_n(\theta_0))| + |A(\theta_0)(n^{-1}\partial_\theta F_n(\theta_0) + U(\theta_0))|.\end{aligned}$$

The first term on the right hand side is treated exactly as $n^{-1}G_n(\theta)$ in the previous condition (with $A_n(\theta)$ replaced by $\partial_\theta A_n(\theta)$), and the last term goes to zero because $n^{-1}\partial_\theta F_n(\theta_0) \to -U(\theta_0)$). That the supremum over M_n of the second term goes to zero follows because $A_n(\theta) \to A(\theta_0)$ uniformly in $\tilde{\Theta}$, while the coordinates of $n^{-1}\partial_\theta F_n(\theta)$ on M_1 can be dominated by a convergent sequence (because $\partial_\theta \hat{a}(\theta)$ is continuous). To finally treat the third term, we use that

$$\begin{aligned}&|A(\theta_0) n^{-1} (\partial_\theta F_n(\theta) - \partial_\theta F_n(\theta_0))| \\ &\quad \leq M \sum_{jk} \sup_{\theta \in M_n} |\partial_\theta \hat{a}_{jk}(\theta) - \partial_\theta \hat{a}_{jk}(\theta_0)| \sum_{\ell} n^{-1} \sum_{i=r}^{n} |Z_{jk}^{(i-1)} Z_{j\ell}^{(i-1)}|,\end{aligned}$$

which goes to zero because $\partial_\theta \hat{a}_{jk}(\theta)$ is continuous.

For Condition (iii), we use that

$$\begin{aligned}\partial_\theta^2 G_n(\theta) &= \partial_\theta^2 A_n(\theta) n^{-1} F_n(\theta) \\ &\quad + 2\partial_\theta A_n(\theta) n^{-1} \partial_\theta F_n(\theta) + A_n(\theta) n^{-1} \partial_\theta^2 F_n(\theta).\end{aligned}$$

The first term is treated as $n^{-1}G_n(\theta)$ was treated when checking Condition (v) (with $A_n(\theta)$ replaced by $\partial_\theta^2 A_n(\theta)$). The second term is dominated by an expression like the sum of the last three terms in (6.10) with $A_n(\theta)$ and $A(\theta)$ replaced by $\partial_\theta A_n(\theta)$ and $\partial_\theta A(\theta)$, respectively, and the sum is treated in the same way. The last term is dominated by a similar expression, now with $\partial_\theta \hat{a}$ replaced by $\partial_\theta^2 \hat{a}$.

□

Note that if $A_n(\theta)$ does not depend on n, condition (3) of Theorem 6.2 is trivially satisfied.

For the optimal estimating function, $V^*(\theta_0) = A^*(\theta_0)U(\theta_0)$, so the asymptotic variance of the optimal estimator is simply $V^*(\theta_0)^{-1}$. This is no surprise because the second Bartlett identity is satisfied for the optimal estimating function.

6.1 Stochastic volatility models

For the stochastic volatility model given by (1.1), α-mixing of the volatility process v implies α-mixing of the observed return process Y.

Lemma 6.3 *Suppose the volatility process v is α-mixing with mixing coefficients $\alpha_t(\theta)$, $t > 0$. Then Y is α-mixing with the mixing coefficients $\tilde{\alpha}_k(\theta) \leq \alpha_k(\theta)$, $k = 1, 2 \ldots$.*

Proof: Let $\mathcal{F}^v$ denote the σ-algebra generated by the volatility process v. Then for Borel subsets M_1 and M_2 of $\mathbb{R}^t$ and $\mathbb{R}^{\ell+1}$, respectively,

$$
\begin{aligned}
& |P\left((Y_1, \ldots, Y_t) \in M_1\right) P\left((Y_{t+k}, \ldots, Y_{t+k+\ell}) \in M_2\right) - \\
& - \quad P\left(\{(Y_1, \ldots, Y_t) \in M_1\} \cap \{(Y_{t+k}, \ldots, Y_{t+k+\ell}) \in M_2\}\right)| \\
& = \quad \left| E\left[P\left(\left(\sqrt{S_1}Z_1, \ldots, \sqrt{S_t}Z_t\right) \in M_1 \middle| \mathcal{F}^v\right)\right]\right. \\
& \qquad \times E\left[P\left(\left(\sqrt{S_{t+k}}Z_{t+k}, \ldots, \sqrt{S_{t+k+\ell}}Z_{t+k+\ell}\right) \in M_2 \middle| \mathcal{F}^v\right)\right] - \\
& \qquad - E\left[P\left(\left\{\left(\sqrt{S_1}Z_1, \ldots, \sqrt{S_t}Z_t\right) \in M_1\right\}\right.\right. \\
& \qquad \cap \left.\left.\left.\left\{\left(\sqrt{S_{t+k}}Z_{t+k}, \ldots, \sqrt{S_{t+k+\ell}}Z_{t+k+\ell}\right) \in M_2\right\} \middle| \mathcal{F}^v\right)\right]\right| \\
& = \quad \left|E\left[f_1\left(S_1, \ldots, S_t\right)\right] E\left[f_2\left(S_{t+k}, \ldots, S_{t+k+\ell}\right)\right]\right. \\
& \qquad \left. - E\left[f_1\left(S_1, \ldots, S_t\right) f_2\left(S_{t+k}, \ldots, S_{t+k+\ell}\right)\right]\right| \\
& = \quad \left|\mathrm{Cov}\left(f_1\left(S_1, \ldots, S_t\right), f_2\left(S_{t+k}, \ldots, S_{t+k+\ell}\right)\right)\right| \leq \alpha_k(\theta_0)
\end{aligned}
$$

where

$$
f_1\left(S_1, \ldots, S_t\right) = P\left(\left(\sqrt{S_1}Z_1, \ldots, \sqrt{S_t}Z_t\right) \in M_1 \middle| S_1, \ldots, S_t\right)
$$

and

$$
\begin{aligned}
& f_2\left(S_{t+k}, \ldots, S_{t+k+\ell}\right) \\
& \quad = \quad P\left(\left(\sqrt{S_{t+k}}Z_{t+k}, \ldots, \sqrt{S_{t+k+\ell}}Z_{t+k+\ell}\right) \in M_1 \middle| S_{t+k}, \ldots, S_{t+k+\ell}\right),
\end{aligned}
$$

and where we have used that the Z_is are mutually independent and independent of $\mathcal{F}^v$. The last inequality follows from (1') in Doukhan (1994, p. 3) because $0 \leq f_i \leq 1$, $i = 1, 2$.

□

A more general result that covers hidden Markov models was given independently by Genon-Catalot, Jeantheau and Larédo (1998c).

For the one-dimensional, ergodic diffusion process v there are a number of relatively simple criteria ensuring α-mixing with exponentially decreasing mixing coefficients. If, for instance, the spectrum of the generator of v (cf. (4.19)) 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 v is geometrically mixing, and the condition (6.2) in Theorem 6.1 is satisfied.

The volatility process (4.24) has a discrete spectrum with $\lambda_1 = \theta$, see Example 4.1. For this example a first numerical experiment has been carried out. For $\alpha = 2$, $\theta = 0.5$, $\sigma = 1$ and $\Delta = 1$ a sample of 1000 observation were simulated. The parameters were estimated using a non-optimal prediction-based estimating function of the type considered in Example 2.1 with $q = 20$. The estimates were $\hat{\alpha} = 2.01$, $\hat{\theta} = 0.64$ and $\hat{\sigma} = 1.30$. It is not surprising that α seems to be the easiest parameter to estimate. A serious simulation study is planned.

Doukhan (1994) gives other criteria for geometrical mixing too; see also Genon-Catalot, Jeantheau and Larédo (1998c). Rather general criteria for α-mixing of diffusion processes expressed in the language of Malliavin calculus were given by Kusuoka and Yoshida (1997).

Acknowledgement.

I am grateful to Valentine Genon-Catalot, Thierry Jeantheau and Catherine Laredo for fruitful and inspiring discussions, for proposing Example 4.2, and for their critical comments on an earlier version of this paper. Thanks are also due to Anders Nielsen for assistance with the numerical experiment.

References

Andersen, T.G. and Lund, J. (1998): Stochastic volatility and mean drift in the short rate diffusion: Sources of steepness, level and curvature in the yield curve. Preprint.

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1990): Parametric modelling of turbulence. *Phil. Trans. R. Soc. Lond. A* **332**, 439 – 455.

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1993): A statistical model for the streamwise component of a turbulent velocity field. *Annales Geophysicae* **11**, 99–103.

Barndorff-Nielsen, O.E., Jensen, J.L. and Sørensen, M. (1998): Some stationary processes in discrete and continuous time. *Adv. Appl. Prob.*

30, 989 – 1007.

Barndorff-Nielsen, O.E. and Shephard, N. (1998): Aggregation and model construction for volatility models. Centre for Analytical Finance, Working Paper No. 10.

Bibby, B.M. (1994): A two-compartment model with additive white noise. Research Report No. 290, Dept. Theor. Statistics, Univ. of Aarhus.

Bibby, B.M. (1995a): Analysis of a tracer experiment based on compartmental diffusion models. Research Report No. 303, Dept. Theor. Statistics, Univ. of Aarhus.

Bibby, B.M. (1995b): On estimation in compartmental diffusion models. Research Report No. 305, Dept. Theor. Statistics, Univ. of Aarhus.

Bibby, B.M. and Sørensen, M. (1995): Martingale estimating 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.

Chesney, M. and Scott, L. (1989): Pricing European currency options: A comparison of the modified Black-Scholes model and a random variance model. *J. Financial Quantitative Anal.*, **24**, 267 – 289.

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

Doukhan, P., Massart, P. and Rio, E. (1994): The functional central limit theorem for strongly mixing processes. *Ann. Inst. Henri Poincaré*, **30**, 63 – 82.

Gallant, A.R. and Long, J.R. (1997): Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* **84**, 125 – 141.

Gallant, A.R. and Tauchen, G. (1996): Which moments to match? *Econometric Theory* **12**, 657 – 681.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998a): Limit theorems for discretely observed stochastic volatility models. *Bernoulli* **4**, 283 – 303.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998b): Parameter estimation for discretely observed stochastic volatility models. To appear in *Bernoulli* **5**.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998c): Stochastic volatility models as hidden Markov models and statistical applications. Preprint, Université de Marne-la-Vallée.

Godambe, V.P. (1960): An optimum property of regular maximum likelihood estimation. *Ann. Math. Statist.* **31**, 1208–1212.

Godambe, V.P. (1985): The foundations of finite sample estimation in stochastic processes. *Biometrika* **72**, 419–428.

Godambe, V.P. and Heyde, C.C. (1987): Quasi likelihood and optimal estimation. *Int. Statist. Rev.* **55**, 231–244.

Gourieroux, C., Montfort, A. and Renault, E. (1993): Indirect inference. *Journal of Applied Econometrics* **8**, S85 – S118.

Heston, S.L. (1993): A closed-form solution for options with stochastic volatility with applications to bond and currency options. *Review of Financial Studies* **6**, 327–343.

Heyde, C.C. (1988): Fixed sample and asymptotic optimality for classes of estimating functions. *Contemporary Mathematics* **80**, 241–247.

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

Hull, J. and White, A. (1988): An analysis of the bias in option pricing caused by a stochastic volatility. *Advances in Futures and Options Research* **3**, 29–61.

Karlin, S. and Taylor, H.M. (1981): *A Second Course in Stochastic Processes.* Academic Press, Orlando.

Kessler, M. (1996): Simple and explicit estimating functions for a discretely observed diffusion process. Research Report No. 336, Dept. Theor. Statist., Univ. of Aarhus. To appear in *Scand. J. Statist.*

Kessler, M. and Sørensen, M. (1998): Estimating equations based on eigenfunctions for a discretely observed diffusion process. Research Report No. 332, Dept. Theor. Statist., Univ. of Aarhus. To appear in *Bernoulli.*

Kusuoka, S. and Yoshida, N. (1997): Malliavin calculus, strong mixing, and expansion of diffusion functionals. Preprint.

Ljung, L. and Caines, P.E. (1979): Asymptotic normality of predictor error estimators for approximate system models. *Stochastics*, **3**, 29 – 46.

Melino, A. and Turnbull, S.M. (1990): Pricing foreign currency options with stochastic volatility. *Journal of Econometrics*, **45**, 239 – 265.

Nelson, D.B. (1990): ARCH models as diffusion approximations. *Journal of Econometrics*, **45**, 7 – 38.

Nielsen, J.N., Vestergaard, M. and Madsen, H. (1999): Estimation in continuous-time stochastic volatility models using nonlinear filters. To appear in *Int. J. of Theoretical and Applied Finance*.

Ruiz, E. (1994): Quasi-maximum likelihood estimation of stochastic volatility models. *Journal of Econometrics*, **63**, 289 – 306.

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, 305–325.

Sørensen, M. (1998): On asymptotics of estimating functions. Preprint No. 6, Dept. of Theor. Statist. Univ. of Copenhagen.

Wiggins, J.B. (1987): Option values under stochastic volatility: Theory and empirical estimates. *Journal of Financial Economics*, **19**, 351 – 372.