Prediction-based Estimating Functions for Diffusion Processes with Measurement Noise

Kim Nolsøe[*], Jan Nygaard Nielsen[†], and Henrik Madsen[‡]

May 22, 2000

Abstract

The prediction-based estimating functions proposed by (Sørensen, 1999) are generalized to facilitate parameter estimation in discretely observed stochastic differential equations, where the observations are corrupted by additive white noise. The new class of estimating functions has most of the nice properties of martingale estimating functions. However, they may be applied when no obvious or easily calculated martingales exist. Simple expressions are derived for the optimal estimating functions when the classes of generalized prediction-based estimating functions are defined by a finite-dimensional space of predictors. Only unconditional moments are needed for this class of estimating functions, so a considerably smaller amount of simulation is needed compared to other classes of estimating functions based on conditional moments. Particular attention is devoted to the Cox-Ingersoll-Ross model and stochastic volatility models. Using Monte Carlo simulation the small-sample properties are examined and the method is compared to other estimating functions.

KEY WORDS: Martingale estimating functions, Monte Carlo simulation, nonlinear filtering, prediction-based estimating functions, stochastic volatility models, stochastic differential equations.

JEL classification: C13

Mathematics Subject Classification (1991): 62M20, 62P05, 93E11

[*]Andersen Consulting E-mail: kim.nolsoe@ac.com

[†]E-mail: jnn@imm.dtu.dk

[‡]E-mail: hm@imm.dtu.dk

Contents

1 Introduction

Until recently the only feasible solution to the parameter estimation problem in discretely, partially observed stochastic differential equations (SDEs), where the measurements are contaminated with additive Gaussian white noise, has been to apply the Kalman-Bucy filter (Kalman and Bucy, 1961) for linear (in the narrow-sense) systems to compute the likelihood function. The filter is based on the evolution of the conditional moments of the underlying state variables, which is assumed to be given by SDEs. For nonlinear systems, ordinary differential equations describing the evolution of the conditional moments are obtained by Taylor expansions of functions of the drift and diffusion functions. For nonlinear systems, the extended Kalman filter (EKF) may be applied provided that the diffusion function does not depend on the process. Otherwise higher order filters must be applied (Jazwinski, 1970; Maybeck, 1982), see (Nielsen, Vestergaard and Madsen, 2000) for a recent application of second order filters. The EKF is particularly well-suited for handling a nonlinear measurement equation that describes the functions of the underlying state variables that are measured in noise. Nielsen, Madsen and Melgaard (2000) proposes a multivariate generalization of a transformation due to (Baadsgaard, Nielsen, Spliid, Madsen and Preisel, 1997) such that the EKF may be applied for a special class of SDEs, eventhough the diffusion function depends on the process.

The explicit treatment of measurement noise makes it possible to distinguish between process noise, i.e. the noise typically described by a Wiener process, that affects the future behavior of the process, and the measurement noise, which in technical and physical applications is merely due to uncertainty in the measurement device and in e.g. financial applications is due to rounding off prices, asynchronous trading, bid-ask spreads and other market imperfections. One-step ahead prediction errors are provided by these filters such that quasi-likelihood estimates may be obtained using a Prediction Error Decomposition (PED) (Schweppe, 1965) under the assumption that the prediction errors are for instance Gaussian distributed. Ljung and Caines (1979) provides an analysis of the properties of the estimators. In the systems and control literature this method is called a Prediction Error Method (PEM), see e.g. (Ljung, 1987). It relies on QML theory for the parameter estimation problem. However, the nonlinear filters are based on Taylor expansions in a way that makes explicit analysis of the validity of the approximations infeasible. The validity and the performance of the nonlinear filter may, to some extent, be tested using model validation tools.

Recent developments of the theory of estimating functions for discretely, partially observed SDEs in Sørensen (1999) makes explicit analysis of the properties of the estimators possible in the general case, where only some of the states are discretely observed without measurement noise. The other classes of EFs mentioned below cannot handle unobserved, latent processes. Although the Prediction-based Estimating Functions (PEFs) attributable to (Sørensen, 1999) allow for estimating parameters in the unobserved processes, if any, it does not explicitly treat the problem of estimating the states of the (un)observed processes at the sampling times, because this approach, which is also adopted in this paper, relies on unconditional moments as opposed to conditional moments. This constitutes a fundamental difference between the PEFs and the filtering methods.

The general theory of Estimating Functions (EFs) dates back to (Godambe, 1960), see also (McLeish and Small, 1988; Godambe, 1991; Heyde, 1997). However, the development of EFs for discretely observed SDEs is of a more recent date. The Martingale Estimating Functions (MEFs) from the linear family for discretely observed SDEs developed by (Bibby and Sørensen, 1995) are inspired by the properties of the pseudo-score function, i.e. the score function obtained by discretizing the continuous-time likelihood function (Liptser and Shiryayev, 1977) provided that the diffusion function does not depend on the parameter. Requiring that the EF be a martingale implies that the asymptotic properties may be obtained without letting the time between measurements tend to zero. Unfortunately it also implies that the EFs involve conditional moments and that the optimal EFs involve derivatives of these moments with respect to the parameters, which, most often, must be computed by simulation, see (Kloeden and Platen, 1995) for some approximate methods. If the diffusion function depends on the unknown parameter other classes of EFs should be used, e.g. the MEFs from the quadratic family attributable to Bibby and Sørensen (1996), see also (Bibby and Sørensen, 1997), that also requires computing the third and fourth order conditional moments. Kessler (2000) proposes a new class of simple EFs that provides explicit expressions for the estimators of the parameters in univariate SDEs. These EFs can only be used to estimate parameters appearing in the stationary density, because it is based on unconditional moments. However, the martingale property is lost and asymptotically efficient estimators are not available. Kessler and Sørensen (1999) proposes another class of MEFs that are based on eigenfunctions of the generator associated with the SDE, see also (Bibby and Sørensen, 1998), which utilizes a combination of the latter two methods. A review is provided in (Sørensen, 1997).

The PEFs proposed by Sørensen (1999) and the new class of Prediction-based Estimating Functions with Measurement noise (PEFMs) proposed in the present paper are based on predictors of functions of the observed process. Particular attention is given to classes of PEFMs given by a finite-dimensional space of predictors. For this class a simple expression for the optimal estimating functions (in the sense of fixed sample optimality) is available. Yet the presence of measurement noise makes the computation of the unconditional moments more involved. PEFMs only involve the unconditional moments as opposed to MEFs, where the conditional moments and, in some cases, their derivatives are required.

Other available methods are the Generalized Method-of-Moments (Hansen, 1982; Chan, Karolyi, Longstaff and Sanders, 1992), Simulated Method-of-Moments (Duffie and Singleton, 1993), Indirect Inference (Gourieroux, Monfort and Renault, 1993) and Efficient Method-of-Moments (Gallant and Tauchen, 1996; Gallant and Long, 1997). Only EMM can handle unobserved states, see (Gallant, Hsieh and Tauchen, 1997; Andersen and Lund, 1997) for applications to stochastic volatility models, but none of these methods allow for measurement noise explicitly. This also holds for the nonparametric methods proposed by (Aït-Sahalia, 1996*a*; Aït-Sahalia, 1996*b*; Stanton, 1997; Jiang and Knight, 1997; Bak, 1998) and compared using Monte Carlo simulation by (Chapman and Pearson, 1998; Jiang and Knight, 1999). Pritsker (1998) analyzes the power of the tests proposed by (Aït-Sahalia, 1996*b*). An overview of parameter estimation methods for discretely observed SDEs is given in (Nielsen, Madsen and Young, 1999).

In Section 2 the modelling framework is put forth. The proposed PEFMs will be presented in Section 3, where the particular problems involved in allowing for measurement noise will be discussed. A simple expression for the optimal estimating functions in the sense of (Heyde, 1997, Theorem 2.1) is presented in Section 4. Some applications will be given in Section 5, where particular attention is devoted to the CIR model (Cox, Ingersoll and Ross, 1985) and stochastic volatility models; a class of SDEs that are used extensively in mathematical finance, where one of the processes is not directly observed. In Section 6 the properties of the proposed method, simple and explicit estimating functions and a nonlinear filter used in combination with a QML method are studied using Monte Carlo simulation. Finally, Section 7 concludes.

2 The model

Consider a one-dimensional diffusion $X = (X_t)_{t\geq 0}$ defined on the state space $\mathcal{S} \subseteq \mathbb{R}$ satisfying the stochastic differential equations

$$dX_t = b(X_t; \boldsymbol{\theta})dt + \sigma(X_t; \boldsymbol{\theta})dW_t; \quad X_0 = x, \tag{1}$$

indexed by $\boldsymbol{\theta}$, where $\boldsymbol{\theta}$ belongs to $\boldsymbol{\Theta}$, an open subset of $\mathbb{R}^p$; $(W)_{t\geq 0}$ is the standard Wiener process; b and σ are known $\mathbb{R}$-valued functions defined on $\mathcal{S} \times \boldsymbol{\Theta}$, which are assumed to be smooth enough to ensure, for every $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, the uniqueness in law of the solution to (1).

Let $s(x;\boldsymbol{\theta})$ denote the density of the scale measure

$$s(x;\boldsymbol{\theta}) = \exp\left(-\int_0^x \frac{2b(y;\boldsymbol{\theta})}{\sigma^2(y;\boldsymbol{\theta})}dy\right). \tag{2}$$

CONDITION 2.1. The following hold for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$

$$\int_0^\infty s(x;\boldsymbol{\theta})dx = \int_{-\infty}^0 s(x;\boldsymbol{\theta})dx = \infty \tag{3}$$

and

$$\int_{-\infty}^\infty [s(x;\boldsymbol{\theta})\sigma^2(x;\boldsymbol{\theta})]^{-1}dx = A(\boldsymbol{\theta}) < \infty. \tag{4}$$

Under these assumptions X is ergodic, and with respect to the Lebesgue measure its stationary density is $x \mapsto [A(\boldsymbol{\theta})s(x;\boldsymbol{\theta})\sigma^2(x;\boldsymbol{\theta})]^{-1}$.

The differential operator L defined by

$$L = b(x;\boldsymbol{\theta})\frac{\partial}{\partial x} + \frac{1}{2}\sigma^2(x;\boldsymbol{\theta})\frac{\partial^2}{\partial x^2} \tag{5}$$

for all twice differentiable functions is called the *generator* of the SDE (1). A twice continuously differentiable function $\varphi(x;\boldsymbol{\theta})$ is called an *eigenfunction* for L with eigenvalue $\lambda(\boldsymbol{\theta})$ if it satisfies

$$L\varphi(x;\boldsymbol{\theta}) = -\lambda(\boldsymbol{\theta})\varphi(x;\boldsymbol{\theta}) \tag{6}$$

for all x in the state space $\mathcal{S}$.

A discretized trajectory $(Y_{t_i})_{0\leq i\leq n}$ with $t_i = i\Delta$ is assumed to be given by

$$Y_{t_i} = h(X_{t_i}) + \varepsilon_{t_i}, \tag{7}$$

where Δ is the sampling time, n the number of measurements, h a specified function and $(\varepsilon_{t_i})_{0\leq i\leq n}$ is the Gaussian white noise process $N(0,\sigma_\varepsilon^2)$ that accounts for the measurement noise.

For computational reasons the following is assumed to hold.

CONDITION 2.2. h is a polynomial in x, i.e. $h(x) = x^m + x^{m-1} + \ldots + 1$, $m \in \mathbb{N}$.

3 Prediction-based Estimating Functions

Let $\mathcal{F}_{t_i} = \sigma\{Y_{t_1},\ldots,Y_{t_i}\}$ denote the σ-algebra generated by the first i measurements $\mathbf{Y}_{t_i} = (Y_{t_1},\ldots,Y_{t_i})$ for $i = 1,\ldots,n$, where n denotes the number of measurements. Assume initially that $\sigma_\varepsilon^2 = 0$ and $h(x) = x$, i.e. $Y_{t_i} \equiv X_{t_i}$, in order to introduce the Sørensen (1999) framework. Shortly, we will generalize to the more general framework obtained for $\sigma_\varepsilon^2 > 0$ and a measurement function h satisfying Condition 2.2. Let f_j, $j = 1,\ldots,J$, be one-dimensional functions defined on the state space $\mathcal{S}$ satisfying $\mathrm{E}_{\boldsymbol{\theta}}[f_j(Y_{t_i})^2] < \infty$ for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, where $\mathrm{E}_{\boldsymbol{\theta}}$ denotes the expectation operator when $\boldsymbol{\theta}$ is the true parameter value. Let $\mathcal{H}_i^{\boldsymbol{\theta}}$ denote the L^2-space of square integrable $\mathcal{F}_{t_i}$-measurable one-dimensional stochastic variables and let $\mathcal{P}_{i,j}^{\boldsymbol{\theta}}$, $j = 1,\ldots,J$, be closed linear subspaces of $\mathcal{H}_i^{\boldsymbol{\theta}}$. As in (Sørensen, 1999) a subspace can be interpreted as a set of predictors of $f_j(Y_{t_{i+1}})$ given the previous measurements $\mathbf{Y}_{t_i}$. Let $\mathbf{\Pi}_j^{(i-1)}(\boldsymbol{\theta}) = \left(\pi_{1,j}^{(i-1)}(\boldsymbol{\theta}), \ldots, \pi_{p,j}^{(i-1)}(\boldsymbol{\theta})\right)^T$ be a p-dimensional stochastic vector with coordinates belonging to $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$. Sørensen (1999) introduces the following class of prediction-based estimating functions

$$\mathbf{G}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \sum_{j=1}^{J} \mathbf{\Pi}_j^{(i-1)}(\boldsymbol{\theta}) \left[f_j(Y_{t_i}) - \hat{\pi}_j^{(i-1)}(\boldsymbol{\theta}) \right], \tag{8}$$

where $\hat{\pi}_j^{(i-1)}(\boldsymbol{\theta})$ is the Minimum Mean Square Error (MMSE) linear predictor of $f_j(Y_{t_i})$ on $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$ and $\mathbf{G}_n(\boldsymbol{\theta})$ is a p-dimensional vector, i.e. $f_j(Y_{t_i}) - \hat{\pi}_j^{(i-1)}(\boldsymbol{\theta})$ is the i'th prediction error. It is well-known that $\hat{\pi}_j^{(i-1)}(\boldsymbol{\theta})$ is the orthogonal projection of $f_j(Y_{t_{i-1}})$ on $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$ with respect to the inner product in $\mathcal{H}_i^{\boldsymbol{\theta}}$, and it is uniquely determined by the projection equations

$$\mathrm{E}_{\boldsymbol{\theta}}[\boldsymbol{\pi}(f_j(Y_{t_i}) - \hat{\pi}_j^{(i-1)}(\boldsymbol{\theta}))] = \mathbf{0} \tag{9}$$

for all $\boldsymbol{\pi} \in \mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$. It follows from (9) that (8) provides an unbiased estimating function.

In the remainder of the paper we shall only be concerned with prediction-based estimating functions where each of the sets $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$ is finite-dimensional. Thus it is assumed that $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$ is spanned by $\mathbf{Z}_j^{(i-1)} = \left(Z_{j1}^{(i-1)}, \ldots, Z_{jq_{ij}}^{(i-1)}\right)$, where the functions $Z_{jk}^{(i-1)} = \phi_{jk}^{(i)}(Y_{t_1}, \ldots, Y_{t_i})$ for $k = 1, \ldots, q_{ij}$ are linearly independent in $\mathcal{H}_{i-1}^{\boldsymbol{\theta}}$. It is assumed that $Z_{j0}^{(i-1)} \equiv 1$ in order to ensure that the MMSE predictor of $f_j(Y_{t_i})$ in $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$ is unbiased. From (9), it follows that the optimal linear predictor is given by

$$\hat{\pi}_j^{(i-1)}(\boldsymbol{\theta}) = \hat{a}_{j0}^{(i-1)}(\boldsymbol{\theta}) + \hat{\mathbf{a}}_j^{(i-1)}(\boldsymbol{\theta})^T \mathbf{Z}_j^{(i-1)}, \tag{10}$$

where

$$\hat{\mathbf{a}}_j^{(i-1)}(\boldsymbol{\theta}) = \mathbf{C}_{i-1,j}(\boldsymbol{\theta})^{-1} \mathbf{b}_j^{(i-1)}(\boldsymbol{\theta}), \tag{11}$$

and

$$\hat{a}_{j0}^{(i-1)}(\boldsymbol{\theta}) = \mathrm{E}_{\boldsymbol{\theta}}[f_j(Y_{t_i})] - \hat{\mathbf{a}}_j^{(i-1)}(\boldsymbol{\theta})^T \, \mathrm{E}_{\boldsymbol{\theta}}[\mathbf{Z}_j^{(i-1)}]. \tag{12}$$

Here $\mathbf{C}_{i-1,j}(\boldsymbol{\theta})$ denotes the covariance matrix of $\mathbf{Z}_j^{(i-1)}$ when $\boldsymbol{\theta}$ is the true parameter value, and

$$\mathbf{b}_j^{(i-1)}(\boldsymbol{\theta}) = \left(\mathrm{Cov}_{\boldsymbol{\theta}}(Z_{j1}^{(i-1)}, f_j(Y_{t_i})), \ldots, \mathrm{Cov}_{\boldsymbol{\theta}}(Z_{jq_{ij}}^{(i-1)}, f_j(Y_{t_i}))\right)^T. \tag{13}$$

It follows that the prediction-based estimating functions proposed by (Sørensen, 1999) can be computed provided that the covariances in $\mathbf{b}_j^{(i-1)}(\boldsymbol{\theta})$ and $\mathbf{C}_{i-1,j}(\boldsymbol{\theta})$ can be computed. Since $\hat{\mathbf{\Pi}}_j^{(i-1)}(\boldsymbol{\theta})$ is completely characterized by a second order moment representation of the stochastic vector

$$\left(f_j(Y_{t_i}), Z_{j1}^{(i-1)}, \ldots, Z_{jq_{ij}}^{(i-1)}\right),$$

only parameters appearing in these moments for at least one j can be estimated using (8). For computational tractability this imposes some restrictions on the choice of the functions f_j and $\phi_{jk}^{(i)}$. Often simple polynomials in Y_{t_i} will be sufficient. There is no available theory for the optimal choice of the functions

f_j and $\phi_{jk}^{(i)}$ due to the lack of a properly defined optimality criterion, but the choice must be guided by the (subset of the) parameters that should be estimated.

Let us now turn to the general case, where $\sigma_\varepsilon^2 > 0$ and h should only satisfy Condition 2.2. The projection results described above still holds, but the computations will be more complicated due to the presence of the noise term $(\varepsilon_{t_i})_{1\leq i\leq n}$. It is assumed in (7) that $(\varepsilon_{t_i})_{1\leq i\leq n}$ is a Gaussian white noise process such that a second order moment representation is still sufficient. As it will be illustrated in Section 5 a large number of unconditional moments need be computed either explicitly or by simulation. Explicit expressions are derived in Section 4. Simulation methods are not covered in this paper, although it is noted that it is much easier to simulate unconditional moments than conditional moments, see e.g. (Kloeden and Platen, 1995).

REMARK 3.1. Hansen and Scheinkman (1995) shows that the aliasing problem does not exist for *irreversible* SDEs without measurement noise ($\sigma_\varepsilon^2 = 0$). We shall not go into the problem of uniquely identifying models of the type (1) from the discretized trajectory $(Y_{t_i})_{0\leq i\leq n}$ in the general case where $\sigma_\varepsilon^2 > 0$. ▼

4 Optimal estimating functions

In this section explicit expressions for the optimal prediction-based estimating functions with measurement noise will be given. The notation is as in (Sørensen, 1999). The presentation in Section 4.1 follows (Sørensen, 1999) that also relies on optimality results from (Heyde, 1997). Section 4.2 contains new results regarding explicit expressions for unconditional moments of both $(X_t)_{t\geq 0}$ and $(Y_{t_i})_{1\leq i\leq n}$ and their interrelations.

4.1 Optimal estimating functions

Introducing a slightly more compact notation the l'th coordinate of $\mathbf{\Pi}_j^{(i-1)}(\boldsymbol{\theta})$ in (8) is given by

$$\pi_{l,j}^{(i-1)}(\boldsymbol{\theta}) = \sum_{k=0}^{q_{ij}} a_{ljk}^{(i)} Z_{jk}^{(i-1)} \tag{14}$$

with $Z_{j0}^{(i-1)} = 1$. Writing the weights $a_{ljk}^{(i)}$ in $p \times \sum_{j=1}^{J}(q_{ij}+1)$-matrices as

$$\mathbf{A}^{(i)}(\boldsymbol{\theta}) = \begin{pmatrix} a_{110}^{(i)}(\boldsymbol{\theta}) & \cdots & a_{11q_{i1}}^{(i)}(\boldsymbol{\theta}) & \cdots & a_{1J0}^{(i)}(\boldsymbol{\theta}) & \cdots & a_{1Jq_{iJ}}^{(i)}(\boldsymbol{\theta}) \\ \vdots & & \vdots & & \vdots & & \vdots \\ a_{p10}^{(i)}(\boldsymbol{\theta}) & \cdots & a_{p1q_{i1}}^{(i)}(\boldsymbol{\theta}) & \cdots & a_{pJ0}^{(i)}(\boldsymbol{\theta}) & \cdots & a_{pJq_{iJ}}^{(i)}(\boldsymbol{\theta}) \end{pmatrix} \tag{15}$$

for $i = 1, \dots, n$, and defining $\sum_{j=1}^{J}(q_{ij}+1)$-dimensional vectors by

$$\Phi^{(i)}(\boldsymbol{\theta}) = \begin{pmatrix} Z_{10}^{(i-1)}[f_j(Y_{t_i}) - \hat{\pi}_1^{(i-1)}(\boldsymbol{\theta})] \\ \vdots \\ Z_{1q_{ij}}^{(i-1)}[f_j(Y_{t_i}) - \hat{\pi}_1^{(i-1)}(\boldsymbol{\theta})] \\ \vdots \\ \vdots \\ Z_{J0}^{(i-1)}[f_J(Y_{t_i}) - \hat{\pi}_J^{(i-1)}(\boldsymbol{\theta})] \\ \vdots \\ Z_{Jq_{iJ}}^{(i-1)}[f_J(Y_{t_i}) - \hat{\pi}_J^{(i-1)}(\boldsymbol{\theta})] \end{pmatrix} \tag{16}$$

allows us to write the estimating function $\mathbf{G}_n(\boldsymbol{\theta})$ in (8) as

$$\mathbf{G}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} \mathbf{A}^{(i)}(\boldsymbol{\theta})\Phi^{(i)}(\boldsymbol{\theta}). \tag{17}$$

The summation may be avoided by defining the $p \times \left(n\sum_{j=1}^{J}(q_{ij}+1)\right)$-matrix

$$\mathbf{D}_n(\boldsymbol{\theta}) = \left(\mathbf{A}^{(1)}(\boldsymbol{\theta}) \dots \mathbf{A}^{(n)}(\boldsymbol{\theta})\right), \tag{18}$$

and the $\left(n\sum_{j=1}^{J}(q_{ij}+1)\right)$-dimensional vector

$$\mathbf{K}_n(\boldsymbol{\theta})^T = \left(\Phi^{(1)}(\boldsymbol{\theta})^T, \dots, \Phi^{(n)}(\boldsymbol{\theta})^T\right), \tag{19}$$

i.e. the estimating function (17) takes the simple form

$$\mathbf{G}_n(\boldsymbol{\theta}) = \mathbf{D}_n(\boldsymbol{\theta})\mathbf{K}_n(\boldsymbol{\theta}) \tag{20}$$

For the estimating function $\mathbf{G}_n(\boldsymbol{\theta}) = (G_{n,1}(\boldsymbol{\theta}), \dots, G_{n,p}(\boldsymbol{\theta}))^T$ define the partial derivatives

$$\partial_{\boldsymbol{\theta}^T}\mathbf{G}_n(\boldsymbol{\theta}) = \begin{pmatrix} \partial_{\theta_1}G_{n,1}(\boldsymbol{\theta}) & \dots & \partial_{\theta_p}G_{n,1}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \partial_{\theta_1}G_{n,p}(\boldsymbol{\theta}) & \dots & \partial_{\theta_p}G_{n,p}(\boldsymbol{\theta}), \end{pmatrix} \tag{21}$$

where, say, $\partial_{\theta_1}G_{n,2}(\boldsymbol{\theta}) = \frac{\partial}{\partial\theta_1}G_{n,2}(\boldsymbol{\theta})$, and

$$\partial_{\boldsymbol{\theta}}\mathbf{G}_n(\boldsymbol{\theta})^T = (\partial_{\boldsymbol{\theta}^T}\mathbf{G}_n(\boldsymbol{\theta}))^T. \tag{22}$$

The concept of fixed sample optimality ($\mathcal{O}_F$-optimality) is defined in (Heyde, 1997), where it is shown that an estimating function $\mathbf{G}_n^*(\boldsymbol{\theta})$ is $\mathcal{O}_F$-optimal if and only if

$$\mathrm{E}_{\boldsymbol{\theta}}[\partial_{\boldsymbol{\theta}^T}\mathbf{G}_n(\boldsymbol{\theta})]^{-1}\,\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{G}_n(\boldsymbol{\theta})\mathbf{G}_n^*(\boldsymbol{\theta})^T] = \mathrm{E}_{\boldsymbol{\theta}}[\partial_{\boldsymbol{\theta}^T}\mathbf{G}_n^*(\boldsymbol{\theta})]^{-1}\,\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{G}_n^*(\boldsymbol{\theta})\mathbf{G}_n^*(\boldsymbol{\theta})^T] \tag{23}$$

for all $\mathbf{G}_n(\boldsymbol{\theta})$ of the form (20). According to (Heyde, 1997, p. 14-15) this is obtained when

$$\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{G}_n(\boldsymbol{\theta})\mathbf{G}_n^*(\boldsymbol{\theta})^T] = -\,\mathrm{E}_{\boldsymbol{\theta}}[\partial_{\boldsymbol{\theta}^T}\mathbf{G}_n(\boldsymbol{\theta})]. \tag{24}$$

This result leads to the following theorem from (Heyde, 1997, Theorem 2.1).

THEOREM 4.1. Suppose that for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the covariance matrix of $\mathbf{K}_n(\boldsymbol{\theta})$ is invertible and $\mathrm{E}_{\boldsymbol{\theta}^T}[\mathbf{K}(\boldsymbol{\theta})]$ has rank p. Then the estimating function

$$\mathbf{G}_n^*(\boldsymbol{\theta}) = \mathbf{D}_n^*(\boldsymbol{\theta})\mathbf{K}_n^*(\boldsymbol{\theta}), \tag{25}$$

where

$$\mathbf{D}_n^*(\boldsymbol{\theta}) = -\,\mathrm{E}_{\boldsymbol{\theta}}[\partial_{\boldsymbol{\theta}}\mathbf{K}_n^*(\boldsymbol{\theta})^T]\left(\mathrm{E}_{\boldsymbol{\theta}}[\mathbf{K}_n(\boldsymbol{\theta})\mathbf{K}_n(\boldsymbol{\theta})^T]\right)^{-1} \tag{26}$$

is optimal within the class of estimating functions of the form (20) for which $\mathbf{D}_n(\boldsymbol{\theta})$ has rank p.

Proof. Follows immediately by inserting (25) and (26) in (24). ■

Assuming that $(Y_{t_i})_{1\leq i\leq n}$ is stationary and that the sets $\mathcal{P}_{i-1,j}^{\boldsymbol{\theta}}$ is spanned by $Z_{jk}^{(i-1)} = \phi_{jk}(Y_{t_{i-1}},\ldots,Y_{t_{i-q}})$, $k = 0,\ldots,q_j$, where h_{jk} (now independent of i as is q_j) is a function from $\mathbb{R}^q$ to $\mathbb{R}$, $q \in \mathbb{N}$, much simpler results for the optimal estimating function can be obtained. This choice of $Z_{jk}^{(i-1)}$ being given as a function of q lagged measurements implies that $Z_{jk}^{(i-1)}$ is only well defined for $i \geq q+1$. In this case, it holds that $\mathbf{A}^{(q+1)}(\boldsymbol{\theta}) = \ldots = \mathbf{A}^{(n)}(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})$ such that (17) takes a much simpler form, namely

$$\mathbf{G}_n(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})\sum_{i=q+1}^{n}\Phi^{(i)}(\boldsymbol{\theta}). \tag{27}$$

Under this assumption the $\mathcal{O}_F$-optimal weights are given in the next theorem.

THEOREM 4.2. The optimal weights in (27) are given by

$$\mathbf{A}^*(\boldsymbol{\theta}) = \mathbf{U}(\boldsymbol{\theta})^T\mathbf{M}_n(\boldsymbol{\theta})^{-1}, \tag{28}$$

where

$$\mathbf{U}(\boldsymbol{\theta}) = -\,\mathrm{E}[\partial_{\boldsymbol{\theta}^T}\Phi^{(i)}(\boldsymbol{\theta})] = -\begin{pmatrix} \partial_{\theta_1}\Phi_1^{(i)}(\boldsymbol{\theta}) & \cdots & \partial_{\theta_p}\Phi_1^{(i)}(\boldsymbol{\theta}) \\ \vdots & & \vdots \\ \partial_{\theta_1}\Phi_{q+1}^{(i)}(\boldsymbol{\theta}) & \cdots & \partial_{\theta_p}\Phi_{q+1}^{(i)}(\boldsymbol{\theta}) \end{pmatrix} \tag{29}$$

and

$$\begin{aligned} \mathbf{M}_n(\boldsymbol{\theta}) \;=\; & \mathrm{E}[\Phi^{(i)}(\boldsymbol{\theta})\Phi^{(i)}(\boldsymbol{\theta})^T] + \sum_{k=1}^{n-q}\frac{n-q-k+1}{n-q+1}\times \\ & \left(\mathrm{E}[\Phi^{(i)}(\boldsymbol{\theta})\Phi^{(i+k)}(\boldsymbol{\theta})^T] + \mathrm{E}[\Phi^{(i+k)}(\boldsymbol{\theta})\Phi^{(i)}(\boldsymbol{\theta})^T]\right). \end{aligned} \tag{30}$$

Proof. See (Sørensen, 1999, Proposition 3.2). ■

REMARK 4.1. Conditions for invertibility of $\mathbf{M}_n$ is given in (Sørensen, 1999, Proposition 3.2). ▼

REMARK 4.2. Eq. (30) is more conveniently expressed as

$$\begin{aligned} \mathbf{M}_n(\boldsymbol{\theta}) \;=\; & \mathrm{E}[\Phi^{(i)}(\boldsymbol{\theta})\Phi^{(i)}(\boldsymbol{\theta})^T] + \sum_{k=1}^{n-q}\frac{n-q-k+1}{n-q+1}\times \\ & \left(\mathrm{E}[\Phi^{(i)}(\boldsymbol{\theta})\Phi^{(i+k)}(\boldsymbol{\theta})^T] + \mathrm{E}[\Phi^{(i)}(\boldsymbol{\theta})\Phi^{(i+k)}(\boldsymbol{\theta})^T]^T\right), \end{aligned} \tag{31}$$

when implementing the method on a computer. ▼

4.2 Computing unconditional moments

In most applications the functions f_j and ϕ_{jk} will be polynomials in the measurements at different time instants. Thus in order to determine the unconditional mixed moments constituting $\mathrm{E}[\Phi^{(i)}(\boldsymbol{\theta})\Phi^{(i)}(\boldsymbol{\theta})^T]$ expressions of the form

$$\mathrm{E}[Y_{t_1}^{j_1} Y_{t_2}^{j_2} \cdots Y_{t_m}^{j_m}] \tag{32}$$

for $m \in \mathbb{N}$, $j_1, \ldots, j_m \in \mathbb{N}_0^m$ and $t_1 < t_2 < \ldots < t_m$ being positive sampling times must be determined. Using (7) this problem reduces to determining explicit expressions for moments on the form $\mathrm{E}[X_{t_1}^{j_1} \cdots X_{t_m}^{j_m}]$ provided that the following condition holds.

CONDITION 4.1. Polynomials are eigenfunctions to the generator (5) of the diffusion process (1), i.e. the eigenfunctions

$$\varphi_i(x; \boldsymbol{\theta}) = \sum_{j=0}^{i} \gamma_{ij}(\boldsymbol{\theta}) x^j \tag{33}$$

satisfies the equation

$$L\varphi_i(x; \boldsymbol{\theta}) = -\lambda_i \varphi_i(x; \boldsymbol{\theta}) \tag{34}$$

for $i = 1, \ldots, m$.

REMARK 4.3. The constants $\gamma_{ij}(\boldsymbol{\theta})$ are computed from (33). See Section 5.1 for an example. ▼

LEMMA 4.1. Under Condition 4.1, it holds that

$$\exp(-\lambda_i t) \sum_{j=0}^{i} \gamma_{ij}(\boldsymbol{\theta}) x^j = \sum_{j=0}^{i} \gamma_{ij}(\boldsymbol{\theta}) \sum_{k=0}^{j} \nu_{jk}(t; \boldsymbol{\theta}) x^k; \quad i = 1, \ldots, m \tag{35}$$

from which the constants $\nu_{jk}(t; \boldsymbol{\theta})$ are determined.

Proof. Taking the conditional expectation on both sides of (33) yields

$$\mathrm{E}[\varphi_i(X_t; \boldsymbol{\theta}) | X_0 = x] = \sum_{j=0}^{i} \gamma_{ij}(\boldsymbol{\theta}) \, \mathrm{E}[X_t^j | X_0 = x]. \tag{36}$$

Under weak regularity conditions

$$\mathrm{E}[\varphi_i(X_t; \boldsymbol{\theta}) | X_0 = x] = \exp(-\lambda_i t) \varphi_i(x; \boldsymbol{\theta}), \tag{37}$$

see (Kessler and Sørensen, 1999, Section 5). Inserting (37) in (36) yields

$$\exp(-\lambda_i t) \varphi_i(x; \boldsymbol{\theta}) = \sum_{j=0}^{i} \gamma_{ij}(\boldsymbol{\theta}) \, \mathrm{E}[X_t^j | X_0 = x]. \tag{38}$$

Under Condition 4.1, Eq. (38) can be expressed as

$$\exp(-\lambda_i t) \varphi_i(x; \boldsymbol{\theta}) = \sum_{j=0}^{i} \gamma_{ij}(\boldsymbol{\theta}) \sum_{k=0}^{j} \nu_{jk}(t; \boldsymbol{\theta}) x^k \tag{39}$$

Inserting (33) in (39) completes the proof. ■

REMARK 4.4. It follows immediately from (38)–(39) that

$$\mathrm{E}[X_t^j|X_0 = x] = \sum_{k=0}^{j} \nu_{jk}(t;\boldsymbol{\theta})x^k; \quad j \in \mathbb{N} \tag{40}$$

▼

This leads to the main result given in the next theorem.

THEOREM 4.3. Assume that the diffusion $(X_t)_{t\geq 0}$ solves (1). For $m \in \mathbb{N}, j_1, j_2, \ldots, j_m \in \mathbb{N}_0^m$ and the sampling times $t_1 < t_2 < \ldots < t_m$, it holds that

$$\begin{aligned}\mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m}] &= \sum_{i_m=0}^{j_m} \nu_{j_m,i_m}(t_m - t_{m-1};\boldsymbol{\theta}) \times \cdots \times \sum_{i_3=0}^{j_3+i_4} \nu_{(j_3+i_4),i_3}(t_3 - t_2;\boldsymbol{\theta}) \\ &\quad \times \sum_{i_2=0}^{j_2+i_3} \nu_{(j_2+i_3),i_2}(t_2 - t_1;\boldsymbol{\theta})\, \mathrm{E}[X_{t_1}^{j_1+i_2}]\end{aligned} \tag{41}$$

Proof. The proof is made by induction. $\mathrm{E}[X_{t_1}^{j_1}]$ obviously fulfills (41).
Assume that

$$\begin{aligned}\mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m}] &= \sum_{i_m=0}^{j_m} \nu_{j_m,i_m}(t_m - t_{m-1};\boldsymbol{\theta}) \times \cdots \times \sum_{i_3=0}^{j_3+i_4} \nu_{(j_3+i_4),i_3}(t_3 - t_2;\boldsymbol{\theta}) \\ &\quad \times \sum_{i_2=0}^{j_2+i_3} \nu_{(j_2+i_3),i_2}(t_2 - t_1;\boldsymbol{\theta})\, \mathrm{E}[X_{t_1}^{j_1+i_2}]\end{aligned}$$

Next let us prove that

$$\begin{aligned}\mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m} X_{t_{m+1}}^{j_{m+1}}] &= \sum_{i_{m+1}=0}^{j_{m+1}} \nu_{j_{m+1},i_{m+1}}(t_{m+1} - t_m;\boldsymbol{\theta}) \times \cdots \times \sum_{i_3=0}^{j_3+i_4} \nu_{(j_3+i_4),i_3}(t_3 - t_2;\boldsymbol{\theta}) \\ &\quad \times \sum_{i_2=0}^{j_2+i_3} \nu_{(j_2+i_3),i_2}(t_2 - t_1;\boldsymbol{\theta})\, \mathrm{E}[X_{t_1}^{j_1+i_2}]\end{aligned}$$

By conditioning on $X_{t_m}^{j_m}$, we get

$$\begin{aligned}\mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_{m+1}}^{j_{m+1}}] &= \mathrm{E}[\mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m} X_{t_{m+1}}^{j_{m+1}}|X_{t_m}^{j_m}]] \\ &= \mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m}\, \mathrm{E}[X_{t_{m+1}}^{j_{m+1}}|X_{t_m}^{j_m}]] \\ &= \mathrm{E}[X_{t_1}^{j_1}\cdots X_{t_m}^{j_m} \sum_{i_{m+1}=0}^{j_{m+1}} \nu_{j_{m+1},i_{m+1}}(t_{m+1} - t_m;\boldsymbol{\theta}) X_{t_m}^{i_{m+1}}] \\ &= \sum_{i_{m+1}=0}^{j_{m+1}} \nu_{j_{m+1},i_{m+1}}(t_{m+1} - t_m;\boldsymbol{\theta}) \times \mathrm{E}[X_{t_1}^{j_1} X_{t_2}^{j_2}\cdots X_{t_m}^{j_m+i_{m+1}}]\end{aligned}$$

According to the assumption by induction we obtain

$$
\begin{aligned}
\mathrm{E}[X_{t_1}^{j_1} \cdots X_{t_{m+1}}^{j_{m+1}}] \quad = \quad & \sum_{i_{m+1}=0}^{j_{m+1}} \nu_{j_{m+1}, i_{m+1}}(t_{m+1} - t_m; \boldsymbol{\theta}) \\
& \times \sum_{i_m=0}^{j_m} \nu_{j_m, i_m}(t_m - t_{m-1}; \boldsymbol{\theta}) \times \cdots \\
& \times \sum_{i_3=0}^{j_3+i_4} \nu_{(j_3+i_4), i_3}(t_3 - t_2; \boldsymbol{\theta}) \\
& \times \sum_{i_2=0}^{j_2+i_3} \nu_{(j_2+i_3), i_2}(t_2 - t_1; \boldsymbol{\theta}) \, \mathrm{E}[X_{t_1}^{j_1+i_2}],
\end{aligned}
$$

which completes the proof. ■

REMARK 4.5. The result in Theorem 4.3 is a generalization of a result in (Sørensen, 1999). ▼

5 Applications

This section considers two examples of using PEFMs, namely the Cox-Ingersoll-Ross mode (Cox et al., 1985) and stochastic volatility models.

5.1 The CIR model

Consider the CIR model which is used extensively in mathematical finance as a model for e.g. spot interest rates

$$dX_t = \alpha(\beta - X_t)dt + \sigma\sqrt{X_t}dW_t; \quad X_0 = x, \tag{42}$$

where $\boldsymbol{\theta} = (\alpha, \beta, \sigma) \in (0, \infty)^3$ such that the process $(X_t)_{t \geq 0}$ is ergodic. It is well-known that X_t has a non-central chi-square distribution and that X (the stationary case) has a gamma-distribution $\Gamma\left(\frac{2\alpha\beta}{\sigma^2}, \frac{\sigma^2}{2\alpha}\right)$. It follows that the stationary mean and variance are given by $\mathrm{E}[X] = \beta$ and $\mathrm{V}[X] = \frac{\beta\sigma^2}{2\alpha}$, respectively. The higher order moments satisfy the recursive relation

$$\mathrm{E}[X^m] = \left(\beta + \frac{(m-1)\sigma^2}{2\alpha}\right) \mathrm{E}[X^{m-1}] \tag{43}$$

for $m \geq 2$. The spectrum (set of eigenvalues) is $\Lambda_{\boldsymbol{\theta}} = \{j\alpha \colon j \in \mathbb{N}_0\}$ with corresponding eigenfunctions $\phi_j(x; \boldsymbol{\theta}) = L_j^{(\eta)}(2\alpha x\sigma^{-2})$, where $L_j^{(\eta)}$ is the jth order Laguerre polynomial with parameter $\eta = 2\alpha\beta\sigma^{-2} - 1$ (Karlin and Taylor, 1981).

A discretized trajectory $(Y_{t_i})_{0 \leq i \leq n}$ with $t_i = i\Delta$ is assumed to be given by

$$Y_{t_i} = X_{t_i} + \varepsilon_{t_i}. \tag{44}$$

For notational simplicity, we assume that only one parameter is to be estimated, i.e. $p = 1$.

Let $f(y) = y^2$, $J = 1$, and let the space $\mathcal{P}^{\boldsymbol{\theta}}_{i-1}$ on which the linear projection is made be spanned by $\mathbf{Z}^{(i-1)} = \left(Z^{(i-1)}_1, \ldots, Z^{(i-1)}_5\right)$, where $Z^{(i-1)}_k = \phi_k\left(Y_{t_i}, Y_{t_{i-1}}, \ldots, Y_{t_{i-5}}\right) = Y^2_{t_{i-k}}$. The dimension of the space $\mathcal{P}^{\boldsymbol{\theta}}_{i-1}$ is chosen arbitrarily to be equal to 5.

Assuming stationarity, the optimal (MMSE) predictor of $f(Y_{t_i})$ on $\mathcal{P}^{\boldsymbol{\theta}}_{i-1}$ is given by a special case of (10), i.e.

$$\hat{\pi}^{(i-1)} = \hat{a}_0 + \hat{\mathbf{a}}(\boldsymbol{\theta})^T \mathbf{Z}^{(i-1)} = \hat{a}_0 + \hat{a}_1 Y^2_{t_{i-1}} + \ldots + \hat{a}_5 Y^2_{t_{i-5}}$$

such that the projection errors are given by

$$Y^2_{t_i} - \hat{a}_0 - \hat{a}_1 Y^2_{t_{i-1}} - \ldots - \hat{a}_5 Y^2_{t_{i-5}}$$

and (16) takes the form

$$\Phi^{(i)} = \begin{bmatrix} Y^2_{t_i} - \hat{a}_0 - \hat{a}_1 Y^2_{t_{i-1}} - \ldots - \hat{a}_5 Y^2_{t_{i-5}} \\ Y^2_{t_{i-1}}(Y^2_{t_i} - \hat{a}_0 - \hat{a}_1 Y^2_{t_{i-1}} - \ldots - \hat{a}_5 Y^2_{t_{i-5}}) \\ \vdots \\ Y^2_{t_{i-5}}(Y^2_{t_i} - \hat{a}_0 - \hat{a}_1 Y^2_{t_{i-1}} - \ldots - \hat{a}_5 Y^2_{t_{i-5}}) \end{bmatrix}.$$

In order to compute the optimal weights $\mathbf{A}^*(\boldsymbol{\theta})$ given by (28) and the optimal PEFM in (27), it is necessary to compute the matrix of partial derivatives $\mathbf{U}(\boldsymbol{\theta})$ in (29) and the matrix $\mathrm{M}_n(\boldsymbol{\theta})$ in (30). The latter consists of (mixed) moments of Y_{t_i} that may be computed as follows. It is possible to derive a general expression that relates the simple unconditional moments of Y_{t_i} and X_{t_i}.

LEMMA 5.1. For the discretized trajectory $(Y_{t_i})_{1 \le i \le n}$ given by (44), it holds that

$$\mathrm{E}[Y^{2m}_{t_i}] = \sum_{j=0}^{2m} \delta_j \, \mathrm{E}[X^{2m-j}_{t_i}] \tag{45}$$

with

$$\delta_j = \begin{cases} \binom{2m}{j} \sigma_\varepsilon^j \prod\limits_{k=1}^{j/2} (2k-1) & \text{for} \quad j = 0, 2, 4, \ldots \\ 0 & \text{for} \quad j = 1, 3, 5, \ldots \end{cases} \tag{46}$$

for $m \in \mathbb{N}$.

Proof. The binomial formula yields the result

$$\mathrm{E}[Y^{2m}_{t_i}] = \mathrm{E}[(X_{t_i} + \varepsilon_{t_i})^{2m}] = \sum_{j=1}^{2m} \binom{2m}{j} \mathrm{E}[X^{2m-j}_{t_i} \varepsilon^j_{t_i}] \tag{47}$$

As $(\varepsilon_{t_i})_{1 \le i \le n}$ is Gaussian white noise, it holds that

$$\mathrm{E}[X^{2i_1+1}_{t_i} \varepsilon^{2i_2+1}_{t_i}] = 0$$

for $i_1, i_2 \in \mathbb{N}_0^2$. As $\mathrm{E}[\varepsilon^{2i+1}_t] = 0$ and $\mathrm{E}[\varepsilon^{2j}_t] = \sigma_\varepsilon^{2j} \prod_{k=1}^{j} (2k-1)$ for $j \in \mathbb{N}_0$ the definition of the constants $\delta_j \in \mathbb{N}$ in (46), and hence (45), follows. ■

From the lemma, it follows that e.g.

$$
\begin{aligned}
\mathrm{E}[Y_{t_i}^2] &= \mathrm{E}[X_{t_i}^2] + \sigma_\varepsilon^2 \\
\mathrm{E}[Y_{t_i}^4] &= \mathrm{E}[X_{t_i}^4] + 6\,\mathrm{E}[X_{t_i}^2]\sigma_\varepsilon^2 + 3\sigma_\varepsilon^4 \\
\mathrm{E}[Y_{t_i}^6] &= \mathrm{E}[X_{t_i}^6] + 15\,\mathrm{E}[X_{t_i}^4]\sigma_\varepsilon^2 + 45\,\mathrm{E}[X_{t_i}^2]\sigma_\varepsilon^4 + 15\sigma_\varepsilon^6 \\
\mathrm{E}[Y_{t_i}^8] &= \mathrm{E}[X_{t_i}^8] + 28\,\mathrm{E}[X_{t_i}^6]\sigma_\varepsilon^2 + 210\,\mathrm{E}[X_{t_i}^4]\sigma_\varepsilon^4 + 420\,\mathrm{E}[X_{t_i}^2]\sigma_\varepsilon^6 + 105\sigma_\varepsilon^8
\end{aligned}
$$

Furthermore, it holds that

$$
\begin{aligned}
\mathrm{V}[Y_{t_i}^2] &= \mathrm{E}[X_{t_i}^4] + 4\,\mathrm{E}[X_{t_i}^2]\sigma_\varepsilon^2 + 2\sigma_\varepsilon^4 - \mathrm{E}[X_{t_i}^2]^2 \\
\mathrm{Cov}[Y_{t_1}^2, Y_{t_2}^2] &= \mathrm{E}[(Y_{t_1}^2 - \mathrm{E}[Y_{t_1}^2])(Y_{t_2}^2 - \mathrm{E}[Y_{t_2}^2])] = \mathrm{Cov}[X_{t_1}^2, X_{t_2}^2];\ t_1 \neq t_2,
\end{aligned}
$$

Using these relations higher order unconditional moments of $(Y_{t_i})_{1\le i\le n}$ may be expressed in terms of the unconditional moments of $(X_{t_i})_{1\le i\le n}$ given by (43). The same applies for mixed moments of $(Y_{t_i})_{1\le i\le n}$, e.g.

$$
\begin{aligned}
\mathrm{E}[Y_{t_1}^2 Y_{t_2}^2 Y_{t_3}^2 Y_{t_4}^2] &= \mathrm{E}[(X_{t_1}+\varepsilon_{t_1})^2 (X_{t_2}+\varepsilon_{t_2})^2 (X_{t_3}+\varepsilon_{t_3})^2 (X_{t_4}+\varepsilon_{t_4})^2] \\
&= \mathrm{E}[X_{t_1}^2 X_{t_2}^2 X_{t_3}^2 X_{t_4}^2] + \sigma_\varepsilon^2\,\mathrm{E}[X_{t_1}^2 X_{t_2}^2 X_{t_3}^2] \\
&\quad + \sigma_\varepsilon^2\,\mathrm{E}[X_{t_1}^2 X_{t_2}^2 X_{t_4}^2] + \sigma_\varepsilon^2\,\mathrm{E}[X_{t_1}^2 X_{t_3}^2 X_{t_4}^2] + \sigma_\varepsilon^4\,\mathrm{E}[X_{t_2}^2 X_{t_3}^2] \\
&\quad + \sigma_\varepsilon^4\,\mathrm{E}[X_{t_2}^2 X_{t_4}^2] + \sigma_\varepsilon^4\,\mathrm{E}[X_{t_3}^2 X_{t_4}^2] + 4\sigma_\varepsilon^6\,\mathrm{E}[X_{t_1}^2] + \sigma_\varepsilon^8
\end{aligned}
$$

Theorem 4.3 provides expressions for the mixed moments $\mathrm{E}[X_{t_1}^2 X_{t_2}^2 X_{t_3}^2 X_{t_4}^2]$ in terms of $\mathrm{E}[X_t^m]$, where the latter is given by (43).

EXAMPLE 5.1. Let us now try to determine expressions of $\nu_{20}(t;\boldsymbol{\theta})$, $\nu_{21}(t;\boldsymbol{\theta})$ and $\nu_{22}(t;\boldsymbol{\theta})$ by applying (35). To ease the notation $\nu_{jk}(t;\boldsymbol{\theta})$ will be written as ν_{jk} and $\gamma_{jk}(\boldsymbol{\theta})$ as γ_{jk}. Inserting $i = 2$ in (35) yields

$$
\begin{aligned}
\exp(-\lambda_2 t)\cdot(\gamma_{20} + \gamma_{21}x + \gamma_{22}x^2) &= \gamma_{20}\nu_{00} + \gamma_{21}(\nu_{10} + \nu_{11}x) \\
&\quad + \gamma_{22}(\nu_{20} + \nu_{21}x + \nu_{22}x^2)
\end{aligned}
$$

Collecting terms in x yields

$$
\begin{aligned}
\gamma_{20}\exp(-\lambda_2 t) &= \gamma_{20}\nu_{00} + \gamma_{21}\nu_{10} + \gamma_{22}\nu_{20} \\
\gamma_{21}\exp(-\lambda_2 t) &= \gamma_{21}\nu_{11} + \gamma_{22}\nu_{21} \\
\gamma_{22}\exp(-\lambda_2 t) &= \gamma_{22}\nu_{22}.
\end{aligned}
$$

These equations are then solved with respect to ν_{2j} for $j = 1, 2, 3$, which yields

$$\nu_{20} = \frac{\gamma_{20}}{\gamma_{22}}(\exp(-\lambda_2 t) - \nu_{00}) - \frac{\gamma_{21}}{\gamma_{22}}\nu_{10} \tag{48a}$$

$$\nu_{21} = \frac{\gamma_{21}}{\gamma_{22}}(\exp(-\lambda_2 t) - \nu_{11}) \tag{48b}$$

$$\nu_{22} = \exp(-\lambda_2 t) \tag{48c}$$

Inserting the solutions for ν_{00}, ν_{10} and ν_{11} obtained by solving $L\varphi_2(x;\boldsymbol{\theta}) = -\lambda_2\varphi_2(x;\boldsymbol{\theta})$ and $L\varphi_1(x;\boldsymbol{\theta}) = -\lambda_1\varphi_1(x;\boldsymbol{\theta})$ in (48) yields

$$
\begin{aligned}
\nu_{20} &= \frac{\gamma_{20}}{\gamma_{22}}(\exp(-\lambda_2 t) - 1) - \frac{\gamma_{21}}{\gamma_{22}}\frac{\gamma_{10}}{\gamma_{11}}(\exp(-\lambda_1 t) - 1) \\
\nu_{21} &= \frac{\gamma_{21}}{\gamma_{22}}(\exp(-\lambda_2 t) - \exp(-\lambda_1 t)) \\
\nu_{22} &= \exp(-\lambda_2 t)
\end{aligned}
$$

♦

EXAMPLE 5.2. Under Condition 4.1, it follows that the second eigenfunction $\phi_2(x;\boldsymbol{\theta})$ is given by (up to a multiplicative constant)

$$\phi_2(x;\boldsymbol{\theta}) = \gamma_{20} + \gamma_{21}x + \gamma_{22}x^2 \tag{49}$$

where the constants γ_{2i} for $i = 0, 1, 2$ are determined as follows. For the CIR model, Eq. (34) takes the form

$$\alpha(\beta - x)(\gamma_{21} + 2\gamma_{22}) + \sigma^2 x\gamma_{22} = -\lambda_2(\gamma_{20} + \gamma_{21}x + \gamma_{22}x^2)$$

Making the arbitrary choice $\gamma_{22} = 1$ and matching coefficients yields

$$\gamma_{21} = -\frac{2\alpha\beta + \sigma^2}{\alpha} \text{ and } \gamma_{20} = \frac{\beta(2\alpha\beta + \sigma^2)}{2\alpha}$$

such that

$$\phi_2(x;\boldsymbol{\theta}) = \frac{\beta(2\alpha\beta + \sigma^2)}{2\alpha} - \frac{2\alpha\beta + \sigma^2}{\alpha}x + x^2$$

♦

It is clear that even for a small sample a huge amount of mixed moments should be computed in order to determine the optimal PEFM given by (27). These computations may be carried out using a symbolic mathematical package, say, Maple[1] or the moments may be computed by simulation. In both cases, it is convenient to introduce an algebra to handle all the special cases, where some of the indices in the mixed moment $\mathrm{E}[X_{t_1}^2 X_{t_2}^2 X_{t_3}^2 X_{t_4}^2]$ coincides. Depending on the (autocorrelation and mixing) properties of the data, it may also be possible to simplify and reduce the number of mixed moments. See (Nolsøe, 1999) for further details.

5.2 Stochastic Volatility Models

Sørensen (1999) considers the simple Stochastic Volatility (SV) model

$$dX_t = \sqrt{v_t}dW_t \tag{50a}$$
$$dv_t = b(v_t;\boldsymbol{\theta})dt + c(v_t;\boldsymbol{\theta})dB_t \tag{50b}$$

where $(W_t)_{t\geq 0}$ and $(B_t)_{t\geq 0}$ are assumed to be two independent standard Wiener processes.

Consider the differences $Z_{t_i} = X_{t_i} - X_{t_{i-1}}$ that turns out to be returns between sampling instants provided that X_{t_i} is the logarithm of a stock price. Since

$$Z_{t_i} = \int_{t_{i-1}}^{t_i} \sqrt{v_t}dW_t, \tag{51}$$

it follows, as in (Sørensen, 1999), that the process $Z = (Z_{t_i})_{0\leq i\leq n}$ is stationary, that the Z_{t_i}'s are uncorrelated, but not independent, and that

$$Z_{t_i} = \sqrt{S_{t_i}}\xi_{t_i}, \tag{52}$$

[1] Some programs may be obtained from the authors on request.

where

$$S_{t_i} = \int_{t_{i-1}}^{t_i} v_t dt \tag{53}$$

and where the ξ_{t_i}'s are independent, identically standard normal stochastic variables, and independent of $S = (S_{t_i})_{0 \le i \le n}$.

Introducing the measurement equation

$$Y_{t_i} = Z_{t_i} + \varepsilon_{t_i}, \tag{54}$$

it follows that the moments of Y_t may be expressed in terms of the moments of v_t, Z_t and ε_t at the sampling instants, i.e.

$$\begin{aligned} E[(Z_t+\varepsilon_t)^2] &= E[Z_t^2+\varepsilon_t^2+2Z_t\varepsilon_t] = E[v_t]+\sigma_\varepsilon^2 && (55)\\ \text{Var}[(Z_t+\varepsilon_t)^2] &= E[(Z_t+\varepsilon_t)^4]-E[(Z_t+\varepsilon_t)^2]^2 \\ &= E[Z_t^4]+6E[Z_t^2]\sigma_\varepsilon^2+3\sigma_\varepsilon^4-(E[Z_t^2]+\sigma_\varepsilon^2)^2 \\ &= E[Z_t^4]+4E[Z_t^2]\sigma_\varepsilon^2+2\sigma_\varepsilon^4-E[Z_t^2]^2 && (56) \end{aligned}$$

Under the assumption that (50b) is the CIR-model (42) the expressions in Lemma 5.1 may be used to compute these moments (by substituting X_t for Z_t). For other models, see (Sørensen, 1999).

REMARK 5.1. Allowing for correlated Wiener processes would make it possible to model the socalled "leverage" effect, , i.e. the fact that large upward moves in equity markets typically have smaller volatility impacts than large downwards moves of the same magnitude. It would be particularly interesting to model the leverage effect introduced by (Black, 1976; Christie, 1982) dynamically, because (Engle and Lee, 1999) suggests that this effect is only a temporary behavior in the stock market. ▼

Alternatively, the Efficient Method-of-Moments (EMM) method (Gallant and Tauchen, 1996) is applied in (Gallant et al., 1997; Gallant, Hsu and Tauchen, 1998). A Bayesian approach based on a Markov Chain Monte Carlo (MCMC) methodology is proposed by (Eraker, 1998). Neither of these methods allow for measurement noise. However, another estimation method for this class of models based on a second order nonlinear filter is proposed by (Nielsen, Vestergaard and Madsen, 2000).

Additional theoretical results on SV-models may be found in e.g. the series of papers (Genon-Catalot, Jeantheau and Larédo, 1998*a*; Genon-Catalot, Jeantheau and Larédo, 1998*b*; Genon-Catalot, Jeantheau and Larédo, 1999).

6 Monte Carlo studies

In this section the properties of the proposed PEFMs will be analyzed and compared to the properties of the simple and explicit estimating functions (Kessler, 2000), and a nonlinear filter used in combination with a QML method using Monte Carlo simulation. Discrete measurements of the CIR model are used, i.e. the model is given by (42)+(44).

6.1 A nonlinear filter

The general continuous-discrete time nonlinear filtering problem is described in (Jazwinski, 1970; Maybeck, 1982) with recent applications in finance given in (Nielsen, Vestergaard and Madsen, 1999; Nielsen, Vestergaard and Madsen, 2000). Briefly the general idea is to infer information about the unobserved states X_{t_i} from the measurements Y_{t_i} for $i = 1, \dots, n$ using two sets of equations: A propagation set describing the evolution of the states between the sampling times and an update set that updates the estimates of the states at the sampling times t_i. Let $\mathcal{Y}_{t_i}$ denote the information set provided by the measurements up to and including time t_i. The transition densities $p(X_{t_i}|X_{t_{i-1}};\boldsymbol{\theta})$ can, at least in principle, be found as the solution to the Chapman-Kolmogorov forward equation, and the conditional density $p(X_t|\mathcal{Y}_{t_{i-1}};\boldsymbol{\theta})$ may then be found from

$$p(X_t|\mathcal{Y}_{t_{i-1}};\boldsymbol{\theta}) = \int_{\mathcal{S}} p(X_t|X_{t_{i-1}};\boldsymbol{\theta})p(X_{t_{i-1}}|Y_{t_{i-1}};\boldsymbol{\theta})dX_{t_{i-1}} \text{ for } t \in [t_{i-1}, t_i), \tag{57}$$

where $p(X_{t_i}|Y_{t_{i-1}};\boldsymbol{\theta})$ is the conditional density for the previous measurement update that follows from Bayes' rule

$$p(X_{t_i}|Y_{t_i};\boldsymbol{\theta}) = \frac{p(Y_{t_i}|X_{t_i}, Y_{t_{i-1}};\boldsymbol{\theta})p(X_{t_i}|X_{t_{i-1}};\boldsymbol{\theta})}{p(Y_{t_i}|Y_{t_{i-1}};\boldsymbol{\theta})}. \tag{58}$$

Assuming that the measurement noise is Gaussian, the first numerator term may be simplified to $p(Y_{t_i}|X_{t_i}, Y_{t_{i-1}};\boldsymbol{\theta}) = p(Y_{t_i}|X_{t_i};\boldsymbol{\theta})$, where the latter is Gaussian. The denominator is given by

$$p(Y_{t_i}|Y_{t_{i-1}};\boldsymbol{\theta}) = \int_{\mathcal{S}} p(Y_{t_i}|X_{t_i};\boldsymbol{\theta})p(X_{t_i}|Y_{t_i};\boldsymbol{\theta})dX_{t_i} \tag{59}$$

Unfortunately, an explicit expression for the conditional density $p(X_t|\mathcal{Y}_{t_{i-1}};\boldsymbol{\theta})$ given in (57) cannot be obtained for the CIR model (42), which implies that an approximate solution to the continuous-discrete time filtering problem given by (57)–(59) must be found. Motivated by the linear Kalman filter (Kalman and Bucy, 1961; Harvey, 1989) that is based on the first two conditional moments, i.e. the conditional mean and variance, the exact time propagation of the states are approximated by two Ordinary Differential Equations (ODEs) for the first two conditional moments of the conditional density, see e.g. (Maybeck, 1982). However, for the CIR model (42) it turns out that these approximate ODEs coincide with the true ODEs[2]. Indeed these ODEs may be solved explicitly, i.e.

$$\begin{aligned} \mathrm{E}[X_{t_i}|X_{t_{i-1}}] &= \beta + (X_{t_{i-1}} - \beta)e^{-\alpha\Delta} && (60a) \\ \mathrm{V}[X_{t_i}|X_{t_{i-1}}] &= \frac{\sigma^2\beta}{2\alpha} + \frac{\sigma^2}{\alpha}(X_{t_{i-1}} - \beta)e^{-\alpha\Delta} - \frac{\sigma^2}{\alpha}\left(X_{t_{i-1}} - \frac{\beta}{2}\right)e^{-2\alpha\Delta} && (60b) \end{aligned}$$

with $\Delta = t_i - t_{i-1}$. The approximate updating equations are given by

$$\begin{aligned} \mathrm{E}[X_{t_i}|X_{t_i}] &= \mathrm{E}[X_{t_i}|X_{t_{i-1}}] + K_{t_i}(Y_{t_i} - \mathrm{E}[X_{t_i}|X_{t_{i-1}}]) && (61a) \\ \mathrm{V}[X_{t_i}|X_{t_i}] &= (1 - K_{t_i})\,\mathrm{V}[X_{t_i}|X_{t_{i-1}}], && (61b) \end{aligned}$$

where the Kalman gain K_{t_i} is

$$K_{t_i} = \frac{\mathrm{V}[X_{t_i}|X_{t_{i-1}}]}{\mathrm{V}[X_{t_i}|X_{t_{i-1}}] + \sigma_\varepsilon^2} \tag{62}$$

The equations (60)–(62) constitute the modified truncated second order filter.

QML estimates of the parameter $\boldsymbol{\theta}$ are readily obtained by assuming that the one-step ahead prediction errors $Y_{t_i} - \mathrm{E}[X_{t_i}|X_{t_{i-1}}]$ in (61a) are Gaussian. Given that the first two conditional moments are correctly specified by the filter, it follows immediately from (Bollerslev and Wooldridge, 1992) that the QML estimates attain the same nice properties as ML estimates.

[2]This statement holds for the socalled truncated second order filter that ignores all central moments of X_t of higher order than two.

REMARK 6.1. It is noted that the assumption inherent to the applied filter is that the conditional mean and variance provide an adequate description of the transition density. This assumption only holds for narrow-sense linear SDEs with a Gaussian density. ▼

6.2 Simple and explicit estimating functions

Kessler (2000) proposes the class of Simple and Explicit Estimating Functions (SEEF) given by

$$\mathbf{F}_n(\boldsymbol{\theta}) = \sum_{i=1}^{n} L\mathbf{g}(X_{t_i}), \tag{63}$$

where $\mathbf{g}$ belongs to the set of twice continuously differentiable functions on $\mathcal{S}$, and the generator L is defined in (5). This class may be used provided that the sampling interval Δ is "large" such that the autocorrelation between subsequent measurements may effectively be neglected, and that the measurements are not observed in noise. Under some regularity conditions, the methodology provides consistent and asymptotically normal, but not asymptotically efficient, estimates of parameters in the stationary density. It is based on the unconditional moments as the PEFs proposed by (Sørensen, 1999).

Inserting $\mathbf{g}(x) = (x, x^2)^T$ in (63) and solving $\mathbf{F}_n(\boldsymbol{\theta}) = \mathbf{0}$ for α and σ^2 yields the following explicit estimators for the CIR model (42), i.e.

$$\hat{\alpha}_n = \frac{\sigma^2 \sum_{i=1}^{n} X_{t_i}}{2\sum_{i=1}^{n} X_{t_i}^2 - 2\beta \sum_{i=1}^{n} X_{t_i}} \quad \text{and} \quad \hat{\sigma}_n^2 = \frac{2\alpha \sum_{i=1}^{n} X_{t_i}^2 - 2\alpha\beta \sum_{i=1}^{n} X_{t_i}}{\sum_{i=1}^{n} X_{t_i}}, \tag{64}$$

where the estimator for each parameter is obtained under the assumption that the remaining parameters are fixed at their true values. In the following two simulation studies the estimators (64) will be used on both the simulated states X_{i_i} and the measurements Y_{t_i}.

6.3 Simulation studies

Two Monte Carlo simulation studies are reported in this section. In each study 20 stochastically independent time series each consisting of n=500 measurements have been simulated. The Milstein discretization scheme (Kloeden and Platen, 1995) is used to obtain a numerical solution to (42) and measurement noise is subsequently added to the simulated states. For computational convenience, one parameter is estimated at a time while the others are fixed at their true value. The nuisance parameter σ_ε^2 is also fixed. Due to the severe computational requirements of PEFM only estimates of α and σ^2 are provided in both studies. The results obtained in Section 5.1 have been used extensively.

In the first study the parameter vector

$$\boldsymbol{\theta}^T = (\alpha, \beta, \sigma^2, \sigma_\varepsilon^2) = (0.08, 0.25, 0.0073, (0.015)^2)$$

is used. The sampling time is Δ=0.1. The results reported in Table 1 are obtained. The mean and standard deviation of the 20 samples of each parameter is listed. All the methods underestimate α, but almost the same level of efficiency is obtained. This, in particular, holds for the filtering/QML method. In our experience filters tend to underestimate the *speed-of-adjustment* parameter α when it is "small" compared to the sampling time Δ, see e.g. (Baadsgaard, 1996; Nielsen, Vestergaard and Madsen, 2000). On the other hand, the filtering/QML method outperforms the EF-based methods regarding the estimate of σ^2. It is noted that the SEEF method provides remarkably poor estimates of σ^2. When the SEEF method is applied to the states (*SEEFx*) only 10 out of the 20 estimates are positive, while 14 estimates are positive when the SEEF method is applied to the measurements (*SEEFy*)[3]. It appears that Δ is too small to allow for the autocorrelation between subsequent measurement to be neglected.

	True	PEFM	SEEFx	SEEFy	Filter
α	0.08	0.0743 (0.0145)	0.0725 (0.0148)	0.0756 (0.0145)	0.0546 (0.0144)
σ^2	0.0073	0.0066 (0.0011)	0.0120 (0.0082)	0.0124 (0.0091)	0.0074 (0.0013)

Table 1: Estimation results from the first simulation study. The estimates provided are obtained as the mean of 20 independent samples each consisting of 500 measurements from the CIR model (42). The standard deviations are given in parenthesis. The true values are $(\alpha, \beta, \sigma^2, \sigma_\varepsilon^2)$ = (0.08, 0.25, 0.0073, $(0.015)^2$).

In the second study the parameter vector

$$\boldsymbol{\theta}^T = (\alpha, \beta, \sigma^2, \sigma_\varepsilon^2) = (0.5, 0.25, 0.0225, 0.01)$$

is used. The sampling time is Δ=0.1. The results reported in Table 2 are obtained. In this study the filtering/QML method and the PEFM method provide almost identical results. The filtering estimates are slightly less biased, and the efficiency of the estimate of σ^2 is higher. Thus the value of α in this study is sufficiently large compared to the sampling time Δ to avoid the afore-mentioned problems with the filtering/QML method. Similar problems have not occurred with the PEFM method.

	True	PEFM	Filter
α	0.5	0.4911 (0.0553)	0.4959 (0.0584)
σ^2	0.0225	0.01873 (0.007322)	0.01973 (0.003349)

Table 2: Estimation results from the second simulation study. The estimates provided are obtained as the mean of 20 independent samples each consisting of 500 measurements from the CIR model (42). The standard deviations are given in parenthesis. The true values are $(\alpha, \beta, \sigma^2, \sigma_\varepsilon^2)$ = (0.5, 0.25, 0.0225, 0.01).

7 Conclusion

A generalization of the prediction-based estimating functions (PEFs) proposed by (Sørensen, 1999) that allows for measurement noise has been proposed. The method is based on unconditional (mixed) moments and new explicit results have been obtained for computing these moments. The CIR model is used to illustrate some of the moment calculations inherent to the method; an explicit expression for the relation between the measurements and the underlying states is derived. The method has been presented for univariate diffusion processes, but it may readily be generalized to multifactor models.

An obvious advantage of PEFM is that optimal estimators may be derived and their asymptotic properties are well-established. An equally obvious disadvantage is the huge number of unconditional and mixed

[3]These results are obtained simply by substituting Y_{t_i} for X_{t_i} in (64).

moments that need be computed even for univariate SDEs. From a theoretical point of view it is possible to generalize the PEFM method to cope with multivariate SDEs, but they may be very difficult to implement, unless suboptimal weights are used. Nonlinear filtering methods are well-suited for multivariate SDEs, but the approximations that are inherent to the ODEs for the conditional moments are very difficult to assess such that the properties of the QML-estimates are difficult to establish. In our experience the bias imposed by these moment approximations are, however, negligible in most applications.

References

Aït-Sahalia, Y. (1996*a*), 'Nonparametric Pricing of Interest Rate Derivative Securities', *Econometrica* **64**, 527–560.

Aït-Sahalia, Y. (1996*b*), 'Testing Continuous-time Models of the Spot Interest Rate', *Review of Financial Studies* **9**, 385–426.

Andersen, T. G. and Lund, J. (1997), 'Estimating continuous time stochastic volatility models of the short term interest rate', *Journal of Econometrics* **77**, 343–378.

Baadsgaard, M., Nielsen, J. N., Spliid, H., Madsen, H. and Preisel, M. (1997), Estimation in stochastic differential equations with a state dependent diffusion term, *in* Y. Sawaragi and S. Sagara, eds, 'SYSID '97 - 11th IFAC Symposium on System Identification', IFAC.

Baadsgaard, M. T. (1996), Estimation in Stochastic Differential Equations, Master's thesis, Department of Mathematical Modelling, Lyngby, Denmark.

Bak, J. (1998), Nonparametric methods in finance, Master's thesis, Institute of Mathematical Modelling, Technical University of Denmark.

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. (1996), 'On Estimation for Discretely Observed Diffusions: A Review', *Theory of Stochastic Processes* **2**(18), 49–56.

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

Bibby, B. M. and Sørensen, M. (1998), Simplified estimating functions for diffusion models with a high-dimensional parameter, Preprint 1998-10, Department of Theoretical Statistics, University of Copenhagen.

Black, F. (1976), Studies of stock price volatility changes, *in* 'Proceedings of the 1976 Meeting of the American Statistical Association', pp. 177–181.

Bollerslev, T. and Wooldridge, J. M. (1992), 'Quasi maximum likelihood estimation and inference in dynamic models with time-varying covariances', *Econometric Reviews* **11**(2), 143–172.

Chan, K. C., Karolyi, G. A., Longstaff, F. A. and Sanders, A. B. (1992), 'An empirical comparison of alternative models of the short-term interest rate', *Journal of Finance* **47**, 1209–1227.

Chapman, D. A. and Pearson, N. D. (1998), Is the short rate drift actually nonlinear?, Technical report, Finance Department, Graduate School of Business, The University of Texas at Austin.

Christie, A. A. (1982), 'The Stochastic Behaviour of Common Stock Variances: Value, Leverage and Interest Rate Effects', *Journal of Financial Econometrics* **10**, 407–432.

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

Duffie, D. and Singleton, K. J. (1993), 'Simulated moments estimation of markov models of asset prices', *Econometrica* **61**, 929–952.

Engle, R. F. and Lee, G. G. J. (1999), A Permanent and Transitory Component Model of Stock Return Volatility, *in* R. Engle and H. White, eds, 'Cointegration, Causality, and Forecasting: A Festschrift in Honor of Clive W. J. Granger', Oxford University Press, pp. 475–497.

Eraker, B. (1998), MCMC Analysis of Diffusion Models with Applications to Finance. Manuscript, NHH, Bergen.

Gallant, A., Hsu, C.-T. and Tauchen, G. (1998), Using High/Low Data to Calibrate Volatility Diffusions and Extract the Forward Integrated Variance. Manuscript, Department of Economics, Duke University.

Gallant, A. R., Hsieh, D. and Tauchen, G. E. (1997), 'Estimation of stochastic volatility models with diagnostics', *Journal of Econometrics* **81**, 159–192.

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. E. (1996), 'Which Moments to Match?', *Econometric Theory* **12**(4), 657–681.

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1998*a*), 'Limit theorems for discretely observed stochastic volatility models', *Bernoulli* **4**, 283–303.

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

Genon-Catalot, V., Jeantheau, T. and Larédo, C. (1999), 'Parameter estimation for discretely observed stochastic volatility models', *Bernoulli* **5**(5), 855–872.

Godambe, V. P. (1960), 'An optimum property of regular maximum-likelihood estimation', *Annals of Mathematical Statistics* **31**, 1208–1211.

Godambe, V. P. E. (1991), *Estimating Functions*, Oxford Science Publications, Oxford.

Gourieroux, C., Monfort, A. and Renault, E. (1993), 'Indirect inference', *Journal of Applied Econometrics* **8**, 85–118.

Hansen, L. P. (1982), 'Large sample properties of generalized method of moments', *Econometrica* **50**, 1029–1054.

Hansen, L. P. and Scheinkman, J. A. (1995), 'Back to the future: Generating moment implications for continuous-time markov processes', *Econometrica* **63**(4), 767–804.

Harvey, A. C. (1989), *Forecasting, Structural Models and the Kalman Filter*, Cambridge University Press, New York.

Heyde, C. C. (1997), *Quasi-Likelihood And Its Applications*, Springer Series in Statistics, Springer, New York.

Jazwinski, A. H. (1970), *Stochastic Processes and Filtering Theory*, Academic Press, New York.

Jiang, G. J. and Knight, J. L. (1997), 'A nonparametric approach to the estimation of diffusion process - with an application to a short-term interest rate model', *Econometric Theory* **13**, 615–645.

Jiang, G. J. and Knight, J. L. (1999), 'Finite sample comparison of alternative estimators of Itô diffusion processes - a monte carlo study', *Journal of Computational Finance* **2**(3).

Kalman, R. E. and Bucy, R. S. (1961), 'New results in linear filtering and prediction problems', *Journal of Basic Engineering, Transactions ASMA, Series D* **83**, 95–108.

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

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

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

Kloeden, P. E. and Platen, E. (1995), *Numerical Solutions of Stochastic Differential Equations, Second Edition*, Springer-Verlag, Heidelberg.

Liptser, R. S. and Shiryayev, A. N. (1977), *Statistics of Random Processes I, II*, Springer-Verlag, New York.

Ljung, L. (1987), *System Identification: Theory for the User*, Prentice-Hall, New York.

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

Maybeck, P. S. (1982), *Stochastic Models, Estimation and Control*, Academic Press, London.

McLeish, D. L. and Small, C. G. (1988), *The Theory and Applications of Statistical Inference Functions*, Vol. 44 of *Lecture Notes in Statistics*, Springer, New York.

Nielsen, J. N., Madsen, H. and Melgaard, H. (2000), Estimating parameters in discretely, partially observed stochastic differential equations, Technical Report 2000-xx, Dept. of Math. Modelling, Tech. Univ. of Denmark, Lyngby, Denmark.

Nielsen, J. N., Madsen, H. and Young, P. (1999), Parameter estimation in stochastic differential equations: An overview, *in* 'Proceedings of The 14th IFAC World Congress', IFAC, Elsevier Science, pp. 289–294.

Nielsen, J. N., Vestergaard, M. and Madsen, H. (1999), Nonlinear filtering of univariate stochastic volatility models, *in* H.-F. Chen, D.-Z. Cheng and J.-F. Zhang, eds, 'Proceedings of the 14th IFAC World Congress', Vol. M, IFAC, Elsevier Science, pp. 123–128.

Nielsen, J. N., Vestergaard, M. and Madsen, H. (2000), 'Estimation in continuous-time stochastic volatility models using nonlinear filters', *International Journal of Theoretical and Applied Finance* **3**(2), 1–30.

Nolsøe, K. (1999), Estimating functions and applications, Master's thesis, Institute of Mathematical Modelling, Technical University of Denmark.

Pritsker, M. (1998), 'Nonparametric density estimation and tests of continuous time interest rate models', *The Review of Financial Studies* **11**(3), 449–487.

Schweppe, F. (1965), 'Evaluation of likelihood function for Gaussian signals', *IEEE Transactions on Information Theory* **11**, 61–70.

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

Sørensen, M. (1999), Prediction-Based Estimating Functions, Preprint 1999-5, Department of Theoretical Statistics, University of Copenhagen.

Stanton, R. (1997), 'A Nonparametric Model of Term Structure Dynamics and the Market Price of Interest Rate Risk', *Journal of Finance* **52**(5), 1973–2002.