PERGAMON

Annual Reviews in Control 24 (2000) 83–94

Annual Reviews in Control

www.elsevier.com/locate/arcontrol

PARAMETER ESTIMATION IN STOCHASTIC DIFFERENTIAL EQUATIONS: AN OVERVIEW

Jan Nygaard Nielsen* Henrik Madsen*
Peter C. Young**

* *Department of Mathematical Modelling, The Technical University of Denmark, DK-2800 Lyngby, Denmark, Fax: +45 4588 1397, Email: {jnn,hm}@imm.dtu.dk*
** *Control and Systems Group, CRES/IEBS, Lancaster University, Lancaster LA1 4YQ, U.K., Fax: +44 1524 593985, Email: p.young@lancaster.ac.uk*

Abstract: This paper presents an overview of the progress of research on parameter estimation methods for stochastic differential equations (mostly in the sense of Itô calculus) over the period 1981-99. These are considered both without measurement noise and with measurement noise, where the discretely observed stochastic differential equations are embedded in a continuous-discrete time state space model. Every attempt has been made to include results from other scientific disciplines. Maximum likelihood estimation of parameters in nonlinear stochastic differential equations is in general not possible due to the unavailability of closed form expressions for the transition and stationary probability density functions of the states. However, major developments are classified according to their approximation to the "true" maximum likelihood solution as opposed to a historical order of presentation.

Keywords: Brownian motion, continuous-time systems, estimation theory, nonlinear filters, non-Gaussian processes, statistical inference, stochastic modelling.

1. INTRODUCTION

Following the survey paper (Young, 1981), a remarkable development in methods for parameter estimation in stochastic systems has taken place in a variety of fields with only limited cross fertilization of ideas. The objective of this paper is to provide a survey of new and existing methods for parameter estimation in discretely observed stochastic differential equation (SDEs), where the latter are embedded in a continuous-discrete time state space model.

Mathematical modelling of dynamical systems in continuous-time has recently received much attention in diverse fields, see e.g. (Young, 1984; Unbehauen and Rao, 1990; Kloeden and Platen, 1995), where the latter focuses on Itô stochastic differential equations. The use of continuous-time stochastic models is also advocated in (Bergstrom, 1990; Bohlin and Graebe, 1995; Unbehauen and Rao, 1997). Following (Åström, 1970) little work has been reported on Itô SDEs in the control literature, whereas a partial list of applications include biological waste treatment (Unny and Karmeshu, 1983), communications (Horstkempe and Lefever, 1984), structural mechanics (Hennig and Grunwald, 1984), climatology (Madsen, 1985), heat dynamics of buildings (Madsen and Holst, 1995; Nielsen and Madsen, 1996), hydrology (Jacobsen *et al.*, 1996), and finance (Nielsen *et al.*, 1999; Nielsen *et al.*, 2000). Yet it is evident that there are numerous open research problems with respect to model structure identification and related topics.

PII: S1367-5788(00)00005-5

SDEs provide a means of combining the hallmarks of grey box modelling in system identification (Ljung, 1987; Söderström and Stoica, 1989), namely by combining a priori knowledge about the system and statistical methods for parameter estimation and model validation. However, the formal definition of SDEs in the system identification community e.g. (Söderström *et al.*, 1997; Haverkamp *et al.*, 1997) differs from the definition in other fields (Protter, 1990; Karatzas and Shreve, 1996) in a way that makes genuine probabilistic methods difficult to apply. It is, for instance, essential that the distribution of the solution of an Itô SDE is obtained from the Kolmogorov partial differential equation, which in turn forms the basis for e.g. exact filtering and maximum likelihood estimation.

Parameter estimation in nonlinear SDEs driven by Wiener processes using discrete-time measurements is an inherently difficult problem to which an increasing number of solutions exist in the literature. Itô SDEs imply that estimation methods in addition to e.g. (Young, 1981; Ljung, 1987; Söderström and Stoica, 1989; Sagara and Zhao, 1990; Unbehauen and Rao, 1990; Söderström *et al.*, 1991; Söderström *et al.*, 1997; Haverkamp *et al.*, 1997) are called for.

There is increasing evidence of both a theoretical and an empirical nature that the level of the (process) noise depends on the state variables in a variety of applications, which represents a challenge to the estimation methods listed above. This multiplicative process noise is an important reason for considering SDEs more rigorously in the Itô sense [1]. An alternative formulation is stochastic integration in the Stratonovitch sense, which adheres to the classical rules of calculus, yet as argued in (Kloeden and Platen, 1995) it is unsuitable for state and parameter estimation.

The remainder of the paper is organized as follows: Section 2 describes the mathematical framework. In Section 3 the generalized method of moments (GMM), the efficient method of moments (EMM) and indirect inference (II) will be discussed. Section 4 considers methods that approximate the probability density functions and, if applicable, the likelihood function by Monte Carlo simulation. Section 5 considers new methods for conventional, stochastic state space models and a linear observation space formulation in which the model is defined in terms of continuous-time, Transfer Function (TF) models. Section 6 considers nonlinear filtering methods. Finally, Section 7 summarizes.

[1] In discrete-time, a plethora of Generalized AutoRegressive with Conditional Heteroscedasticity (GARCH) models have been proposed to describe such phenomena, see e.g. (Bollerslev *et al.*, 1992; Engle, 1995; Rossi, 1996).

2. STOCHASTIC DIFFERENTIAL EQUATIONS

Let the time evolution of the states of a dynamical system $\mathbf{X}_t \in \mathbb{R}^d$, $t_0 \leq t \leq T$, be described by the solution to the stochastic differential equation in the Itô sense

$$d\mathbf{X}_t = \mathbf{b}(t, \mathbf{X}_t; \boldsymbol{\theta})dt + \boldsymbol{\sigma}(t, \mathbf{X}_t; \boldsymbol{\theta})d\mathbf{W}_t \quad (1)$$

or, written componentwise as,

$$dX_t^i = b^i(t, \mathbf{X}_t; \boldsymbol{\theta})dt + \sum_{j=1}^{m} \sigma^{ij}(t, \mathbf{X}_t; \boldsymbol{\theta})dW_t^j \quad (2)$$

for $j = 1, \ldots, d$, where $\mathbf{X}_{t_0}$ is a stochastic initial condition satisfying $E[\|\mathbf{X}_{t_0}\|^2] < \infty$. The drift function $\mathbf{b}: [t_0, T] \times \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^d$ accounts for the evolution in mean of the state variables in the interval $[t, t + dt)$. The evolution of the covariance of the states in the same time interval, which is defined in terms of the dispersion matrix $\boldsymbol{\sigma}: [t_0, T] \times \mathbb{R}^d \times \mathbb{R}^p \mapsto \mathbb{R}^{d \times m}$ is described by the diffusion matrix

$$\boldsymbol{\Sigma}(t, \mathbf{X}_t; \boldsymbol{\theta}) \equiv \boldsymbol{\sigma}(t, \mathbf{X}_t; \boldsymbol{\theta})\boldsymbol{\sigma}(t, \mathbf{X}_t; \boldsymbol{\theta})^T. \quad (3)$$

The infinitesimal characteristics $\mathbf{b}$ and $\boldsymbol{\sigma}$ are nonlinear and time-varying, assumptions which reflect the idea that systems will, in general, be both dynamically nonlinear and nonstationary. It is also assumed throughout that they satisfy sufficient regularity (Lipschitz and bounded growth) conditions to ensure the existence and uniqueness of strong solutions to (1), see e.g. (Øksendal, 1995). The process noise is modelled as a standard Wiener process $\mathbf{W}_t = (W_t^1, \ldots, W_t^m)^T$. The parameter vector $\boldsymbol{\theta}$ may be restricted to a subset $\boldsymbol{\Theta}$ of $\mathbb{R}^p$. Such parameter restrictions may readily be applied to the model, because it is formulated in continuous-time.

It is convenient to think of $\mathbf{X}_t$ as the output of a system described by $\mathbf{b}$ and $\boldsymbol{\sigma}$ with inputs $\mathbf{W}_t$ and $\mathbf{X}_{t_0}$, see Figure 1.

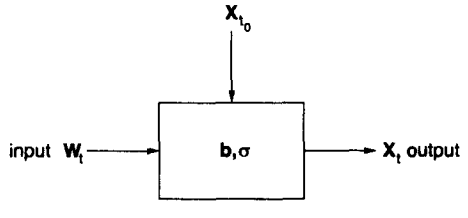

Fig. 1. A schematic of a stochastic differential equation with drift $\mathbf{b}$ and diffusion $\boldsymbol{\sigma}$ driven by the Wiener process $\mathbf{W}_t$ and the initial condition $\mathbf{X}_{t_0}$.

Remark 1. The specification in (1) does not allow for time delays as in socalled *stochastic functional equations (stochastic differential delay equations and stochastic hereditary equations).*

Discrete observations $\mathbf{Y}_{t_k}$ ($l = \dim(\mathbf{Y}_{t_k})$) of the SDE (1) are obtained at the sampling instants $t_1 < t_2 < \cdots < t_N = T$, where N denotes the number of observations, and T is the final time in the time interval $[t_0, T]$ ($t_0 < t_1$). The measurement equation is

$$\mathbf{Y}_{t_k} = \mathbf{h}(t_k, \mathbf{X}_{t_k}; \boldsymbol{\theta}) + \mathbf{e}_{t_k}; \quad k = 1, \ldots, N \tag{4}$$

where $\mathbf{h}(t_k, \mathbf{x}_{t_k}; \boldsymbol{\theta})$ is a nonlinear function, which is assumed to be continuously differentiable with respect to $\mathbf{x}_{t_k}$, and $\{\mathbf{e}\}$ is a zero mean Gaussian white noise process with covariance $\boldsymbol{\Sigma}^e_{t_k}$. The stochastic entities $\mathbf{X}_{t_0}$, $\{\mathbf{W}\}$ and $\{\mathbf{e}\}$ are assumed to be mutually independent for all t and t_k.

Remark 2. The methods described in Sections 3–4 do not allow for measurement errors, so (4) will be excluded from the discussion until Section 5.

Denote the transition probability density function (pdf) of the state vector by

$$p(\mathbf{x}_t|\mathbf{x}_\tau; \boldsymbol{\theta}) = p_{\mathbf{x}_t|\mathbf{x}_\tau}(\mathbf{x}|\xi) = p(t, \mathbf{x}; \tau, \xi, \boldsymbol{\theta}) \tag{5}$$

for $t > \tau$, i.e. loosely speaking the probability that the process is in state $\mathbf{x}_t$ at time t provided that it was in state $\mathbf{x}_\tau$ at time τ. It may be shown that $p(t, \mathbf{x}; \tau, \xi, \boldsymbol{\theta})$ satisfies the Kolmogorov forward equation

$$\frac{\partial p}{\partial t} = \mathcal{L}(p) \tag{6}$$

with the initial condition $\lim_{t \to \tau} p_{\mathbf{x}_t|\mathbf{x}_\tau}(\mathbf{x}|\xi) = \delta(\mathbf{x} - \xi)$ (Dirac's delta function) and the forward diffusion operator $\mathcal{L}(p)$ defined by

$$\mathcal{L}(p) = -\sum_{i=1}^{n} \frac{\partial(pb^i)}{\partial x^i} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \frac{\partial^2(p\Sigma^{ij})}{\partial x^i \partial x^j}, \tag{7}$$

where b^i and Σ^{ij} denote the elements of $\mathbf{b}$ and $\boldsymbol{\Sigma}$ evaluated at $(t, \mathbf{x}; \boldsymbol{\theta})$, and p is short for $p(t, \mathbf{x}; \tau, \xi, \boldsymbol{\theta})$. Assuming stationarity, it holds that $p(t, \mathbf{x}; \tau, \xi, \boldsymbol{\theta}) = p(t-\tau, \mathbf{x}; 0, \xi, \boldsymbol{\theta})$ or $p(\tau, \mathbf{x}; \xi, \boldsymbol{\theta})$ for short. The stationary pdf solves $\mathcal{L}(p) = 0$ provided that it exists.

With discrete observations the stationary pdf is denoted by $p(\Delta_k, \mathbf{x}_{t_k}; \mathbf{x}_{t_{k-1}}, \boldsymbol{\theta})$, where $\Delta_k = t_k - t_{k-1}$ denote the sampling instants, such that the likelihood function of the parameter given the observations $\mathcal{X}^N = (\mathbf{x}_{t_1}, \mathbf{x}_{t_2}, \ldots, \mathbf{x}_{t_N})^T$ follows

$$L_N(\boldsymbol{\theta}) = \left[\prod_{k=1}^{N} p(\Delta_k, \mathbf{x}_{t_k}; \mathbf{x}_{t_{k-1}}, \boldsymbol{\theta})\right] p(\mathbf{x}_0; \boldsymbol{\theta}) \tag{8}$$

where $p(\mathbf{x}_0; \boldsymbol{\theta})$ is the pdf of the initial condition. This may be a Dirac delta function $\delta(\mathbf{x}_0)$ or a specified pdf. Of course, this pdf follows from the specification of the initial condition in (1).

Assuming that the states are observed discretely (without measurement noise), Maximum Likelihood (ML) estimates are given by

$$\boldsymbol{\theta}_{\mathrm{ML}} = \arg\max_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} L_N(\boldsymbol{\theta}) \tag{9}$$

Following standard likelihood theory the *score function* (or *score vector*) is given by

$$\mathbf{S}_N(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}} \ln L_N(\boldsymbol{\theta}) \tag{10}$$

such that the ML estimates may be found by solving the *estimating equation* $\mathbf{S}_N(\boldsymbol{\theta}) = \mathbf{0}$. As might be expected closed form solutions to (6) are rarely available – except for the linear case, which implies that ML estimates cannot be obtained explicitly.

2.1 *Discretization schemes*

Some estimation methods involve numerical solutions of the SDE (1), so some comments and formulae regarding numerical discretization schemes for SDEs will be made here for easy reference, see also (Kloeden and Platen, 1995) and the references therein. A numerical solution of the SDE (1) is obtained by iteratively solving a stochastic difference equation obtained by computing an Itô-Taylor expansion of the drift and diffusion functions $\mathbf{b}$ and $\boldsymbol{\sigma}$.

Assume, for simplicity, that N equidistant observations are available, i.e. $\Delta_k = \Delta$. Let $\delta = \Delta/K$ denote the length of the discretization time step, where $K > 1$ is the number of time steps in each interval $[t_{k-1}, t_k]$ for $k = 1, \ldots, N$. Furthermore, introduce $\tau_{k-1,i} = t_{k-1} + i\delta$ for $i = 0, \ldots, K$, and let the stochastic process $\{\mathbf{Z}\}$ be a discrete-time approximation of $\{\mathbf{X}\}$. For the SDE (1) the ν'th component of the Euler discretization scheme is given by the stochastic difference equation

$$Z^\nu_{\tau_{k-1,i}} = Z^\nu_{\tau_{k-1,i-1}} + b^\nu(\tau_{k-1,i-1}, \mathbf{Z}_{\tau_{k-1,i-1}}; \boldsymbol{\theta})\delta + \sum_{j=1}^{m} \sigma^{\nu j}(\tau_{k-1,i-1}, \mathbf{Z}_{\tau_{k-1,i-1}}; \boldsymbol{\theta})\delta W^j_{\tau_{k-1,i}} \tag{11}$$

for $i = 0, \ldots, K$ with the initial condition $\mathbf{Z}_{\tau_{k-1,0}} = \mathbf{x}_{t_{k-1}}$ and $\delta W^j_{\tau_{k-1,i}} = W^j_{\tau_{k-1,i}} - W^j_{\tau_{k-1,i-1}}$ is the $N(0, \delta)$ distributed increment of the jth component of the m-dimensional standard Wiener process $\mathbf{W}_t$.

Kloeden and Platen (1995) contains a vast range of discretization schemes for Itô SDEs driven by Wiener processes, but there is a need for more research work on numerical properties such

as stability, convergence and the magnitude of approximation error of these schemes. There is also a need for schemes allowing for variable step sizes.

3. METHODS OF MOMENTS

An important class of widely used estimation methods is method of moments, because they do not rely upon distributional assumptions. In this section the Generalized Method of Moments (GMM) attributable to (Hansen, 1982), see also (Ogaki, 1993), will be discussed prior to considering some more efficient generalizations and implementations.

It may be shown that other estimation methods such as ordinary least squares (OLS), nonlinear Instrumental Variable (IV) estimation, Two-Stage Least Squares (2SLS) and, under some regularity conditions, maximum likelihood, may be viewed as special cases of GMM (and subsequently EMM and SMM). Hamilton (1994, Section 14.2) discusses the technical details of these special cases. GMM was originally developed for discrete-time stochastic models, yet it may be applied to SDEs by computing moment conditions from a discretized version of the SDE (Chan *et al.*, 1992).

3.1 *The generalized method of moments*

The main advantage of the GMM method is that it requires specification only of certain moment conditions rather than the full density. This can also be a drawback, because then GMM does not make efficient use of all the information in the sample, which may lead to a loss of efficiency. Furthermore the parameter estimates depend on the choice of the moment conditions and the theory does not provide any guidance regarding the choice of the moment conditions. A lot of research work in parallel has been reported in the system parameter estimation literature eg. (Young, 1981; Söderström and Stoica, 1989; Young, 1993) that is clearly related to GMM and provides various extensions, in particular in terms of recursive implementations, see eg. (Ljung and Söderström, 1983; Young, 1984; Unbehauen and Rao, 1987). However, to our knowledge, these relationships have not been explored further in the literature. Let $\mathbf{X}_{t_k}$, $k = 1, \ldots, N$ denote the observations and let $\mathbf{f}(\mathbf{X}_k; \boldsymbol{\theta})$ denote a r-dimensional vector function ($r \geq p$) satisfying

$$E[\mathbf{f}(\mathbf{X}_{t_k}; \boldsymbol{\theta})] = \mathbf{0} \tag{12}$$

which consists of r moment conditions of the discretized SDE [2]. The r moment conditions that may include instrumental variables must be chosen (subjectively) by the modeller. See (Bates and White, 1988; Hall, 1993) and Section 3.2 herein for a discussion on instrument selection.

By the law of large numbers $E[\mathbf{f}(\mathbf{X}_{t_k}; \boldsymbol{\theta})]$ may be estimated by

$$\mathbf{g}(\boldsymbol{\theta}; \mathcal{X}^N) = \frac{1}{N} \sum_{k=1}^{N} \mathbf{f}(\mathbf{X}_{t_k}; \boldsymbol{\theta}) \tag{13}$$

where it is duly noted that $\mathbf{g}: \mathbb{R}^p \mapsto \mathbb{R}^r$ such that the GMM estimate $\hat{\boldsymbol{\theta}}$ solves

$$\mathbf{g}(\hat{\boldsymbol{\theta}}; \mathcal{X}^N) = \mathbf{0} \tag{14}$$

The GMM method provides an estimate $\hat{\boldsymbol{\theta}}$ by minimizing a quadratic function

$$Q_N(\boldsymbol{\theta}; \mathcal{X}^N) = [\mathbf{g}(\boldsymbol{\theta}; \mathcal{X}^N)]^T \boldsymbol{\Sigma}_N [\mathbf{g}(\boldsymbol{\theta}; \mathcal{X}^N)] \tag{15}$$

where $\boldsymbol{\Sigma}_N$ is a $(r \times r)$ positive semidefinite weight matrix. By using $\boldsymbol{\Sigma}_N = \mathbf{S}_N^{-1}$, where $\mathbf{S}_N$ denotes the covariance matrix of the moment restrictions, the estimate of $\boldsymbol{\theta}$ is said to be efficient, although, unfortunately, the efficiency depends on the choice of the moment conditions $\mathbf{f}(\mathbf{X}_{t_k}; \boldsymbol{\theta})$. Newey and West (1987) provides consistent estimates of $\boldsymbol{\Sigma}_N$ in the presence of heteroscedasticity and autocorrelation in the moment conditions. However, the modeller may be willing to sacrifice asymptotic efficiency in exchange for not having to specify completely the nature of the heterescedasticity and/or autocorrelation. See (Andrews, 1991) for a nonparametric approach.

Hansen (1982) shows that the GMM estimate obtained by solving

$$\left\{ \partial_{\boldsymbol{\theta}^T} \mathbf{g}(\boldsymbol{\theta}; \mathcal{X}^N) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \right\}^T (\hat{\mathbf{S}}^N)^{-1} \mathbf{g}(\hat{\boldsymbol{\theta}}; \mathcal{X}_N) = 0$$

is consistent and asymptotic normal, i.e.

$$\sqrt{N}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \to \mathrm{N}(\mathbf{0}, \mathbf{V}) \tag{16}$$

where $\mathbf{V} = (\mathbf{D}\mathbf{S}^{-1}\mathbf{D})^{-1}$ and $\mathbf{D}$ is estimated by

$$\hat{\mathbf{D}}_N = \partial_{\boldsymbol{\theta}^T} \mathbf{g}(\boldsymbol{\theta}; \mathcal{X}^N) \Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} \tag{17}$$

The GMM method also provides Wald, Lagrange multiplier and Likelihood Ratio (LR) type statistics for hypothesis and diagnostics testing (Hansen, 1982). Considering the number of well-known estimation methods that GMM encompasses, it is a very useful method and it is

[2] In the special case $r \to \infty$ such that all moments of order less than or equal to r are contained in $\mathbf{f}(\mathbf{X}_{t_i}; \boldsymbol{\theta})$, GMM may be considered as a ML method.

easy to implement. However, it has a number of drawbacks: *i*) The parameter estimates depend on the subjective choice of moment conditions, and there is no theory available for selecting neither optimal moment conditions nor instruments; and *ii*) The needed discretization of the SDE (1) introduces biases in the parameter estimates that are difficult to quantify and thus to correct for, see (Baadsgaard, 1996) for some Monte Carlo experiments. The amount of bias clearly depends on the choice of discretization scheme and sampling time in a complicated way that it is arguably not worth exploring further.

Hansen and Scheinkman (1995) considers a generalization to more general continuous-time Markov processes, where two sets of moment conditions based on (7) are obtained. Thus the methods for parameter estimation and statistical inference inherent to GMM may be used, although the issue of statistical efficiency is not formally addressed in (Hansen and Scheinkman, 1995).

3.2 *The efficient method of moments*

The Efficient Method of Moments (EMM) developed by (Gallant and Tauchen, 1996) and elaborated in a series of papers (Andersen and Lund, 1997; Gallant *et al.*, 1997; Gallant and Tauchen, 1999) is a natural progression from GMM that alleviates some of the problems listed above - except that it cannot handle neither measurement noise nor nonstationary data. A computer implementation is described in (Gallant and Tauchen, 1995). Obviously the EMM method is also based on matching moments, but the crucial moment conditions are obtained from the score vector of an auxiliary discrete-time model (see details later). This construction also facilitates detailed model specification tests that are based directly on each individual moment condition, see e.g. (Andersen *et al.*, 1998).

The first step in EMM is to fit an auxiliary model (which is often called a score generator) to the observed data. The model is characterized by the cpdf $p(\mathbf{x}|\mathcal{X}^N;\boldsymbol{\psi})$, where $\mathcal{X}^N$ contains lagged values of the process $\mathbf{X}_N, \mathbf{X}_{N-1}, \ldots$ (and possibly explanatory variables) and $\boldsymbol{\psi} \in \mathbb{R}^q$ is a q-dimensional parameter vector ($p \leq q$). Not all the elements of $\mathbf{X}_N$ need be observed.

To fix ideas, assume, for a moment, that $d = 1$. The score generator may consist of any order of, say, an ARMA-model for the conditional mean μ_t and any order of, say, a GARCH model (Bollerslev, 1986) for the conditional variance h_t. Compute the normalized variables $z_t = (x_t - \mu_t)/\sqrt{h_t}$ and express the pdf $p(x|\mathcal{X}^N;\boldsymbol{\psi})$ in terms of the semi-nonparametric (SNP) density estimator attributable to (Gallant and Tauchen, 1989), i.e.

$$p(x_t|\mathcal{X}^N;\boldsymbol{\psi}) = \frac{[P_K(z_t,x_t)]^2}{\int [P_K(u,x_t)]^2\phi(u)du}\frac{\phi(z_t)}{\sqrt{h_t}} \quad (18)$$

where $\phi(\cdot)$ is the standard normal density and the Hermite polynomials

$$P_K(z_t,x_t) = \sum_{i=0}^{K_z}\left(\sum_{j=0}^{K_x} a_{ij}(x_t)^i\right)(z_t)^i \quad (19)$$

with $(K_x, K_z) \in \mathbb{N} \times \mathbb{N}$ should account for any deviance from the normal density. To achieve identifiability, the parameter a_{00} is set to 1.

The standard information criteria such as BIC, AIC, YIC and Hannan-Quinn may be used to identify K_z, K_x, and the order of the ARMA and GARCH models. Given a properly identified pdf p, quasi-maximum likelihood (QML) estimates of $\boldsymbol{\psi}$ may be obtained and the first step of EMM is completed.

The expected value of the score vector (10) now provides the moment conditions for simulated method of moments estimation of the continuous-time model. Let $\{\tilde{\mathbf{X}}_t(\boldsymbol{\theta}), \tilde{\mathcal{X}}_t(\boldsymbol{\theta})\}_{t=1}^M$ denote a sample of size M simulated from a discretized version of (1) using the parameter vector $\boldsymbol{\theta}$, where $M \gg N$. The EMM estimator is then defined by

$$\hat{\boldsymbol{\theta}}_N = \arg\min_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \mathbf{m}_M(\boldsymbol{\theta},\hat{\boldsymbol{\psi}})\boldsymbol{\Sigma}_N\mathbf{m}_M(\boldsymbol{\theta},\hat{\boldsymbol{\psi}})^T \quad (20)$$

where $\mathbf{m}_M(\boldsymbol{\theta},\hat{\boldsymbol{\psi}})$ is the expected value of the score vector, evaluated by Monte Carlo integration at the QML estimate of the score generator parameter $\hat{\boldsymbol{\psi}}$, i.e.

$$\mathbf{m}_M(\boldsymbol{\theta},\hat{\boldsymbol{\psi}}) = \frac{1}{M}\sum_{k=1}^{M}\partial_{\boldsymbol{\psi}}\ln \mathbf{p}(\tilde{\mathbf{x}}_t,(\boldsymbol{\theta})|\tilde{\mathcal{X}}^{t_k}(\boldsymbol{\theta});\hat{\boldsymbol{\psi}})$$

As before the weight matrix $\boldsymbol{\Sigma}_N$ is a consistent estimator of the inverse asymptotic covariance matrix of the auxiliary score vector. It may be estimated by

$$\hat{\mathbf{S}}_N = \frac{1}{N}\mathbf{m}_M(\boldsymbol{\theta},\hat{\boldsymbol{\psi}})\mathbf{m}_M(\boldsymbol{\theta},\hat{\boldsymbol{\psi}})^T \quad (21)$$

with $\boldsymbol{\Sigma}_N = \mathbf{S}_N^{-1}$ under the assumption that there is no autocorrelation in the expected likelihood scores. Gallant and Tauchen (1996) presents alternate formulae in the presence of autocorrelation. This construction is very similar to GMM, although the moment conditions are obtained in another way. However, the covariance matrix is independent of $\boldsymbol{\theta}$ such that the usual iterative computation of it in GMM is avoided. This reduces the computational burden considerably.

The asymptotic properties of the EMM estimator (20) is summarized in (Gallant and Tauchen, 1999): If the score generator encompasses the true generating process (DGP), then the QML estimates become sufficient statistics and EMM is fully efficient. Gallant and Tauchen (1999) performs an analysis of the relative efficiency of EMM compared to other method of moments using the Marron and Wand (1992) test suite of scale mixtures of normals and concludes that the relative efficiency of EMM is uniformly higher. The small-sample properties of EMM are studied by (Chumacero, 1997; Ng and Michaelides, 1997; Andersen *et al.*, 1999) using Monte Carlo simulation. Their work support the conjecture that EMM is more efficient than other method of moments, but apparently no one has performed a comparative study with ML in those models, where ML is feasible. This is an interesting subject for further study that might clarify the concept of an "efficient" method in this context.

The Indirect Inference (II) method attributable to (Gourieroux *et al.*, 1993; Smith, 1993) is closely related to EMM, yet it is computationally more demanding, because the QML estimates of the score generator must be computed repeatedly. EMM requires only one computation of the QML estimates such that the score generator may be more elaborate. EMM and II both encompass a Simulated Method of Moments (SMM) developed by (Duffie and Singleton, 1993). Duffie and Glynn (1996) proposes a method similar to the infinitesimal generator-based method due to (Hansen and Scheinkman, 1995) that is based on random sampling instants. Yet both these methods cannot handle unobserved states.

4. LIKELIHOOD-BASED METHODS

In this section three likelihood-based methods are considered. The first method consists of a discretization of the likelihood function that, under some technical conditions, follows from an assumption of continuous observations being available. Secondly, a likelihood function is derived for a discretized version of the SDE (1), where the discretization time step δ is equal to the sampling interval Δ. The extension proposed by (Pedersen, 1995*b*) assumes that $\delta \ll \Delta$.

Assume initially that the diffusion function is known, i.e. $\boldsymbol{\sigma}(t, \mathbf{X}_t; \boldsymbol{\theta}) = \boldsymbol{\sigma}(t, \mathbf{X}_t)$. Under some technical conditions (Liptser and Shiryayev, 1977) the log-likelihood function for $\boldsymbol{\theta}$ based on continuous observations of $\mathbf{X}_t$ in the time interval $[t_0, T]$ can be written in terms of the integrals

$$l_T^c(\boldsymbol{\theta}) = \int_{t_0}^{T} \mathbf{b}(s, \mathbf{X}_s; \boldsymbol{\theta})^T \boldsymbol{\Sigma}(s, \mathbf{X}_s)^{-1} d\mathbf{X}_s - \frac{1}{2} \int_{t_0}^{T} \mathbf{b}(s, \mathbf{X}_s; \boldsymbol{\theta})^T \boldsymbol{\Sigma}(s, \mathbf{X}_s)^{-1} \mathbf{b}(s, \mathbf{X}_s; \boldsymbol{\theta}) ds \quad (22)$$

The usual approximation of these integrals by finite Itô and Riemann sums, respectively, leads to the approximate log-likelihood function for $\boldsymbol{\theta}$ based on discrete observations

$$\tilde{l}_T(\boldsymbol{\theta}) = \sum_{k=1}^{N} \boldsymbol{\Lambda}(t_{k-1}, \mathbf{X}_{t_{k-1}}; \boldsymbol{\theta})(\mathbf{X}_{t_k} - \mathbf{X}_{t_{k-1}}) - \frac{1}{2} \sum_{k=1}^{N} \boldsymbol{\Lambda}(t_{k-1}, \mathbf{X}_{t_{k-1}}; \boldsymbol{\theta}) \mathbf{b}(t_{k-1}, \mathbf{X}_{t_{k-1}}; \boldsymbol{\theta}) \times (t_k - t_{k-1}) \quad (23)$$

with

$$\boldsymbol{\Lambda}(\cdot, \cdot; \cdot) = \mathbf{b}(t_{k-1}, \mathbf{X}_{t_{k-1}}; \boldsymbol{\theta})^T \boldsymbol{\Sigma}(t_{k-1}, \mathbf{X}_{t_{k-1}})^{-1}.$$

This method can also be used if $\boldsymbol{\theta}$ admits a partition $\boldsymbol{\theta} = (\boldsymbol{\theta}_1^T \boldsymbol{\theta}_2^T)^T$, $\boldsymbol{\theta}_1$ being a subset of $\mathbb{R}^{p-1}$, such that $\mathbf{b}(t, \mathbf{X}_t; \boldsymbol{\theta})$ depends only on $\boldsymbol{\theta}_1$ and $\boldsymbol{\sigma}(t, \mathbf{X}_t; \boldsymbol{\theta})$ is known up to the scale factor θ_2, i.e. $\boldsymbol{\sigma}(t, \mathbf{X}_t; \boldsymbol{\theta}) = \theta_2 \tilde{\boldsymbol{\sigma}}(t, \mathbf{X}_t)$, because $\tilde{l}_T(\boldsymbol{\theta})$ essentially only depends on $\boldsymbol{\theta}_1$, and $(\theta_2)^2$ can be estimated by a quadratic-variation-like formula (Florens-Zmirou, 1989). In cases where the diffusion function depends on $\boldsymbol{\theta}$ in a more general way, Hutton and Nelson (1986) shows that the discretized score function $\tilde{S}_T(\boldsymbol{\theta}) = \partial_{\boldsymbol{\theta}} \tilde{l}_T^c(\boldsymbol{\theta})$ can, under certain regularity conditions, still be used to obtain an estimator for $\boldsymbol{\theta}$ with some nice properties, see also (Pedersen, 1993). It is, however, not a ML-estimator. The last approach is also feasible if $\boldsymbol{\Sigma}(t_{k-1}, \mathbf{X}_{t_{k-1}})^{-1}$ is not regular provided that the Moore-Penrose pseudo-inverse of $\boldsymbol{\Sigma}(t_{k-1}, \mathbf{X}_{t_{k-1}})$ is used. Whereas Dacunha-Castelle and Florens-Zmirou (1986) shows consistency for the estimates obtained from continuous observations (22) for $N \to \infty$ (irrespective of the value of Δ), inconsistency of the estimates in the case of time-equidistant observations obtained from (23) are shown in (Florens-Zmirou, 1989). In general, it holds that estimation methods originating from the theory of continuous observations yield strongly biased estimators, unless $\max_{1 \le k \le N} \Delta_k$ is very "small", because the drift and diffusion functions only provide an accurate description of the process $\{\mathbf{X}\}$ on an infinitesimal time scale.

An alternative to discretizing the likelihood function for continuous observations is to derive the likelihood of a discretized version of the SDE (1). This approach has been proposed by (Pedersen, 1995*b*) and shortly thereafter by (Santa-Clara, 1995). Pedersen (1995*b*) proposes a method that

is based on simulating the SDE using the Euler scheme (11) with a discretization time step orders of magnitude smaller than the sampling time ($\delta \ll \Delta$). This method is called an *Approximate ML (AML)* method.

The basic idea is to split the time interval between two subsequent observations into a number of subintervals, K, and construct a sequence of pdfs that converges to the true pdf for $K \to \infty$ by averaging over J simulations. To fix ideas, assume that $d = 1$. Assume further that equidistant observations are available $\Delta_k = \Delta$ for $k = 1, \ldots, N$ and that the discretization time step δ satisfies $\Delta = K\delta$. Let $Z_i^{(j)}$ denote the value of the pseudo-observation at time $\tau_{k-1,i} = t_{k-1} + i\delta$ in the jth simulation for $i = 1, \ldots, K$ and $j = 1, \ldots, J$ using the Euler scheme (11). Introduce the sequence of approximate pdfs $\{p^{(K)}(\tau_{k-1,i-1}, z_{i-1}, \tau_{k-1,i}, z_i; \boldsymbol{\theta})\}_{K=1}^{\infty}$ as an approximation to the true pdf. For a sufficiently small δ, it holds that $Z_1^{(j)} \sim N(x_{t_{k-1}} + b(t_{k-1}, x_{t_{k-1}}; \boldsymbol{\theta})\delta, \sigma^2(t_{k-1}, x_{t_{k-1}}; \boldsymbol{\theta})\delta)$ using the Euler scheme (11), i.e. that $p^{(1)}$ is Gaussian. This does, however, not hold for $p^{(K)}$ for $K \geq 2$. If the Euler scheme is used again to compute the next pseudo-observation $Z_2^{(j)}$ the Chapman-Kolmogorov equation yields

$$p^{(2)}(\tau_2, x_{t_k}, \xi_2; \boldsymbol{\theta}) = \int_{\mathcal{S}} p^{(1)}(\tau_2, \xi_1, \xi_2; \boldsymbol{\theta}) \times p^{(1)}(\tau_1, x_{t_{k-1}}, \xi_1; \boldsymbol{\theta}) d\xi_1 \quad (24)$$

where the intermediate pseudo-observation Ξ_1 (capital ξ_1) is integrated out. If this scheme is repeated $K - 1$ times, the transition pdf can be approximated by $E[p^{(1)}(\tau_{K-1}, Z_{\tau_{K-1}}, x_{t_k}; \boldsymbol{\theta})]$. The idea attributable to (Pedersen, 1995*b*; Pedersen, 1995*a*) is to use Monte Carlo methods to compute the expectation, i.e. a $(K-1)$-dimensional integral, by simulating J sample paths of the pseudo-observations and then average over these sample paths to get an approximation of $p^{(K)}(\Delta, x_{t_{k-1}}, x_{t_k}; \boldsymbol{\theta})$. The conditional (on x_0) likelihood function (8) then readily follows. However, this method is very time-consuming due to the Monte Carlo-based integration. Variance-reduction methods may be applied to minimize the Monte Carlo variance using weak discretization schemes, see eg. (Kloeden and Platen, 1995; Fishman, 1996; Fang and Wang, 1994) and the references therein, but this has not been considered in the literature.

Bergstrom (1990) proposes a methodology, where the diffusion function σ is assumed to be constant during the sampling intervals such that σ may be moved outside the stochastic integral. This implies that the pdf is Gaussian. An application is given by (Nowman, 1997) for a univariate SDE.

5. OTHER APPROACHES

In this section, we review briefly two other approaches to the estimation of stochastic ODE models: the first, which is set within the conventional, stochastic state space formulation of the problem and relates closely to the methodology based on Itô calculus discussed in previous sections of the paper; and the second, applied to an alternative linear, observation space formulation, in which the model is defined in terms of continuous-time, transfer function models. These methods do not allow for both process noise and measurement noise, see Section 6 for a treatment of the general setting.

Early research of the estimation of stochastic ODE models utilised the Extended Kalman Filter (EKF) approach (Jazwinski, 1970; Gelb, 1974; Young and Beck, 1974), in which both the state variables and any unknown parameters, jointly considered as the elements of an 'extended' state vector, are estimated simultaneously using 'relinearization' of the filter-estimation equations about the latest recursive estimates at each recursive update, see (Ljung, 1978; Ljung, 1979) for an analysis of the properties of the estimated parameters. In parameter estimation terms, however, a more satisfactory approach to the parameter/state estimation problem is the ML method proposed by (Stepner and Mehra, 1973; Åström and Kallström, 1973). This is reviewed in (Young, 1981; Young, 1984), who discusses other related references using similarly motivated methodology.

Since this ML approach is closely related to the methods discussed previously, it will not be discussed in detail here. Basically, however, it involves numerical optimization of the state space model parameters and associated hyperparameters (i.e. noise variances and covariances) based on a cost function of the general form, see (Young, 1984, Chapter 9):

$$J = \sum_{k=1}^{N} \left\{ \|\hat{\mathbf{e}}_k\|^2_{\mathbf{P}^{-1}_{t_k|t_{k-1}}} + \ln\det \mathbf{P}_{t_k|t_{k-1}} \right\} + \|\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}_0\|_{\mathbf{P}_{\theta}^{-1}(0)} \quad (25)$$

where $\hat{\mathbf{e}}_k$ is the innovation (recursive residual); $\mathbf{P}_{t_k|t_{k-1}}$ is the covariance matrix generated by the associated continuous-discrete Kalman Filter (KF), or its nonlinear equivalent, at the kth sampling instant; $\hat{\boldsymbol{\theta}}_0$ is the a priori estimate of the model parameter vector ; and $\mathbf{P}_{\theta}(0)$ is its associated error-covariance matrix. This latter term reveals an alternative Bayesian interpretation of the method that allows for the introduction of a priori information on the parameters. This ML approach exploits the 'prediction error decomposition' (PED) method of Schweppe (1965) to

generate the likelihood function from the KF equations, see e.g. (Harvey, 1989; Young, 1998*c*; Young, 1998*b*) and Section 6 herein.

Within the linear, automatic control context, continuous-time models are often formulated in terms of continuous-time Laplace transform transfer functions and the estimation of such models has received a lot of interest since the early nineteen sixties, see (Young, 1981; Sinha and Rao, 1991) and the many references therein. Considering, for simplicity, the single input-single output case, this model is normally written in the following form:

$$x(t) = \frac{B(s)}{A(s)}u(t-\tau) \tag{26}$$

$$y(t) = x(t) + \xi(t) \tag{27}$$

Here $x(t)$ and $u(t)$ denote, respectively, the deterministic output and input signals of the system; $y(t)$ is the observed output, which is assumed to be contaminated by a noise signal $\xi(t)$; τ is a pure time (transport) delay affecting the input signal; $A(s)$ and $B(s)$ are polynomials in the Laplace operator s of the following form,

$$A(s) = s^n + a_1 s^{n-1} + a_2 s^{n-2} + \ldots + a_n \tag{28}$$

$$B(s) = b_0 s^m + b_1 s^{m-1} + b_2 s^{m-2} + \ldots + b_m \tag{29}$$

where n and m can take on any positive integer values; the transfer function (TF) does not need to be proper (i.e. there is no restriction that $m < n$); and the pure time delay of τ time units is introduced to reflect the fact that real systems are often characterized by pure transport delays (note that in the state space situation, this would necessitate the use of a differential-delay form of the state equations).

For the purposes of stochastic estimation e.g. (Young *et al.*, 1991; Young, 1996), the noise signal $\xi(t)$ is assumed to be a zero mean stochastic disturbance with rational spectral density, which can be described by the following TF model,

$$\xi(t) = \frac{D(s)}{C(s)}e(t) \tag{30}$$

where,

$$C(s) = s^p + c_1 s^{p-1} + c_2 s^{p-2} + \ldots + c_p \tag{31}$$

$$D(s) = s^q + d_1 s^{q-1} + d_2 s^{q-2} + \ldots + d_q \tag{32}$$

while $e(t)$ is a zero mean, serially uncorrelated and normally distributed random variable (white noise) with variance σ^2. This model is the s operator equivalent of the well known discrete-time ARMA process; however, it is often sufficient to consider only the simpler equivalent of the AR process, i.e. when $D(s) = 1$.

The complete model obtained in the above manner can be written in the form,

$$y(t) = \frac{B(s)}{A(s)}u(t) + \frac{D(s)}{C(s)}e(t) \tag{33}$$

and estimates of the model parameters can be obtained by standard numerical optimisation of the associated Gaussian Likelihood function. An attractive alternative approach is obtained by noting that $\hat{e}_k$ can be written in the alternative form,

$$\hat{e}_k = \frac{\hat{C}(s)}{\hat{D}(s)\hat{A}(s)}\left(\hat{A}(s)y_k - \hat{B}(s)u_k\right) = \hat{A}(s)y_k^* - \hat{B}(s)u_k^* \tag{34}$$

where y_k^* and u_k^* are sampled values of the continuous-time "prefiltered" variables obtained in the following manner,

$$y^*(t) = \frac{\hat{C}(s)}{\hat{D}(s)\hat{A}(s)}y(t) \tag{35}$$

$$u^*(t) = \frac{\hat{C}(s)}{\hat{D}(s)\hat{A}(s)}u(t) \tag{36}$$

Equation (34) is now linear-in-the-parameters of the deterministic part of TF model from u_k to y_k, so that it is possible to estimate these parameters by an iterative or recursive/iterative relaxation optimization process based on discrete-time data sampled from the above continuous-time operations and involving the use of linear least squares or instrumental variable (IV) algorithms at each iteration. The continuous-time prefilters and the auxiliary model in the IV case are updated adaptively at each iteration, as described in (Young, 1984; Young, 1996; Young and Jakeman, 1980). In the case of non-uniformly sampled $y(t)$, these filtering operations are implemented in a continuous-discrete KF form to allow for interpolation. The inclusion of the prefilters in (35)–(36) is not only justified statistically but it also nicely solves the problem of time-derivative measurement by providing the physically realizable filtered time derivatives of the input and output signals as by-products of these prefiltering operations. This relaxation solution is justified theoretically by the discrete-time analysis of Pierce (1972), which can be extended to this continuous-time situation.

The IV solution is particularly robust since it can yield sub-optimal solutions that do not rely on restrictive assumptions on the nature of $\xi(t)$ provided the deterministic input signal $u(t)$ is independent of $\xi(t)$. The simplest and most robust solution of this type is the Simplified Refined Instrumental Variable (SRIV) method, see e.g. (Young *et al.*, 1991; Young, 1996), where it

is assumed for convenience that the $\xi(t)$ is white noise, so that the adaptive prefilters are simply of the form $1/A(s)$. Although this is a sub-optimal solution, the parameter estimates are consistent and relatively efficient. Consequently, it provides a very practical method of identifying and estimating continuous-time models of stochastic systems and has advantages over alternative, more ad hoc methods (e.g. most of the methods described in (Sinha and Rao, 1991)) since the prefilters are selected on a more objective statistical basis, thus improving the efficiency of the estimates. Recent examples of its practical application in terms of databased mechanistic modelling (DBM) are given in (Young, 1998*a*; Young, 1998*b*; Young, 1998*c*; Young *et al.*, 1999). The DBM approach allows for state-dependent parameters such that nonlinear stochastic systems may be modelled.

6. NONLINEAR FILTERING

In this section the continuous-discrete nonlinear filtering problem for the state space model (1)+(4) is considered. This framework allows for both process and measurement noise compared to the methods surveyed in the previous sections. Emphasis is placed on parameter estimation.

If $\mathbf{b}$, $\mathbf{h}$ are linear in the states $\mathbf{X}_t$, and $\boldsymbol{\sigma}$ does not depend on $\mathbf{X}_t$, the SDE (1) is linear in the narrow-sense and an explicit solution may be found, see eg. (Arnold, 1974; Kloeden and Platen, 1995). The Kalman-Bucy filter (Kalman and Bucy, 1961) provides an exact solution to the filtering problem. The filter also provides recursive residuals such that a Gaussian likelihood function follows from a Prediction Error Decomposition (PED) approach (Schweppe, 1965). A maximum likelihood method for direct estimation of embedded parameters in SDEs is proposed in (Madsen, 1985; Graebe, 1990; Madsen and Melgaard, 1991; Melgaard and Madsen, 1993; Bohlin and Graebe, 1995) based on the EKF and the PED. The EKF supplies state estimates and the recursive residuals, where the latter are used to compute QML estimates of the parameters using the Gaussian likelihood function implied by the PED. This approach may also be used for nonlinear systems. For SDEs with a state-dependent diffusion function higher order filters are needed (Maybeck, 1982), see, however, the discussion of a particular transformation below. This also holds for linear SDEs in the wide-sense, i.e. SDEs where both the drift and the diffusion functions are linear in the states. An approach based on the iterated EKF (Jazwinski, 1970) is given in (Melgaard and Madsen, 1993).

Nielsen and Madsen (2000) introduces a generalization of the transformation proposed by (Baadsgaard *et al.*, 1997) such that the method described in (Graebe, 1990; Melgaard and Madsen, 1993) may be applied to a restricted class of SDEs with a state-dependent diffusion term.

7. DISCUSSIONS AND CONCLUSIONS

The overview provided here focuses on estimation methods for discretely observed Itô stochastic differential equations. It indicates that only nonlinear filtering methods provide an approximate solution to the full state and parameter estimation problem posed in (1) and (4) using nonlinear filters and a PED approach to obtain ML-estimates. The method-of-moments GMM can be used only for models with fully observed states with no measurement noise, where EMM/II allows for partially observed states. GMM is fairly easy to implement, but being based on a discretized version of the SDE, it is not an optimal choice. If it is possible to derive explicit expressions for the moment resctrictions, GMM will lead to unbiased and efficient estimators. However, in such simple cases, it is most likely the true ML solution is also feasible. EMM is more efficient than other methods of moments, and it relies on the test methodology that is an inherent part of GMM also. However, for both methods, the efficiency loss will inevitably lead to low power of these tests.

It seems evident that the most general and useful approach today is nonlinear filtering augmented by the prediction error decomposition. Unfortunately, it is difficult to assess the model error introduced by the approximate filtering equations in terms of biasedness and efficiency loss. It is, however, possible to test for the correct model, see e.g. (Bak *et al.*, 1998).

8. REFERENCES

Andersen, T. G. and J. Lund (1997). Estimating continuous time stochastic volatility models of the short term interest rate. *Journal of Econometrics* **77**, 343–378.

Andersen, T. G., H-J Chung and B. E. Sørensen (1999). Efficient method of moments estimation of a stochastic volatility model: a Monte Carlo study. *Journal of Econometrics* **91**(1), 61–87.

Andersen, T. G., L. Benzoni and J. Lund (1998). Estimating jump-diffusions for equity returns. Manuscript, Finance Department, Kellogg Graduate School of Management, Northwestern University.

Andrews, D. W. K. (1991). Heteroskedasticity and autocorrelation consistent covariance matrix estimation. *Econometrica* **59**, 817–858.

Arnold, L. (1974). *Stochastic differential equations.* Wiley. New York.

Åström, K. J. (1970). *Introduction to Stochastic Control Theory.* Academic Press. New York.

Åström, K. J. and C. G. Kallström (1973). Application of system identification techniques to the determination of ship dynamics. In: *Identification and System Parameter Estimation* (P. Eykhoff, Ed.). pp. 415–424. North Holland/American Elsevier.

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

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

Bak, J., H. Madsen and H. Aa. Nielsen (1998). Goodness of fit of stochastic differential equations. In: *Symposium i Anvendt Statistik* (P. Linde and A. Holm, Eds.).

Bates, C. and H. White (1988). Efficient instrumental variables estimation of systems of implicit heterogeneous nonlinear dynamic equations with nonspherical errors. In: *Dynamic Econometric Modeling* (W. A. Barnett, E. R. Berndt and H. White, Eds.). Cambridge University Press. Cambridge, England.

Bergstrom, A. R. (1990). *Continuous-time Econometric Modelling.* Oxford University Press. Oxford.

Bohlin, T. and S. F. Graebe (1995). Issues in nonlinear stochastic grey-box identification. *International Journal of Adaptive Control and Signal Processing* **9**(6), 465–490.

Bollerslev, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* **31**, 307–327.

Bollerslev, T., R. Y. Chou and K. F. Kroner (1992). ARCH models in finance: A review of the theory and evidence. *Journal of Econometrics* **52**, 5–59.

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

Chumacero, R. (1997). Finite sample properties of the efficient method of moments. Manuscript, Department of Economics, University of Chile.

Dacunha-Castelle, D. and D. Florens-Zmirou (1986). Estimation of the coefficients of a diffusion from discrete observations. *Stochastics* **19**, 263–284.

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

Duffie, D. and P. Glynn (1996). Estimation of continuous-time markov processes sampled at random time intervals. Manuscript, Graduate School of Business, Stanford University.

Engle, R. F. (1995). *ARCH Selected Readings.* Cambridge University Press. Cambridge.

Fang, K. and Y. Wang (1994). *Number-Theoretic Methods in Statistics.* Chapman and Hall. London.

Fishman, G. (1996). *Monte Carlo: Concepts, Algorithms, and Applications.* Springer-Verlag. New York.

Florens-Zmirou, D. (1989). Approximate discrete-time schemes for statistics of diffusion processes.. *Statistics* **20**, 547–557.

Gallant, A. R. and G. E. Tauchen (1989). Semi-nonparametric estimation of conditionally constrained heterogenous processes: Asset pricing applications. *Econometrica* **57**, 1091–1120.

Gallant, A. R. and G. E. Tauchen (1995). A program for nonparametric time series analysis, version 8.4, user's guide. *Working paper, University of North Carolina.*

Gallant, A. R. and G. E. Tauchen (1996). Which Moments to Match?. *Econometric Theory* **12**(4), 657–681.

Gallant, A. R. and G. Tauchen (1999). The Relative Efficiency of Method of Moments Estimators. *Journal of Econometrics* **92**, 149–172.

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

Gelb, A. (1974). *Applied Optimal Estimation.* MIT Press for The Analytic Sciences Corporation. Cambridge, Mass. (ed.).

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

Graebe, S. F. (1990). IDKIT: A Software Guide for Gray Box Identification, User's Guide, Version RT-1.0. Technical report. TRITA-REG 9004, Automatic Control, Royal Institute of Technology, Stockholm, Sweden.

Hall, A. (1993). Some aspects of the generalized method of moments estimation. In: *Handbook of Statisitics, Vol 11.* (C. R. Rao G. S. Maddala and H.D. Vinod, Eds.). Vol. 11. Chap. 15. North Holland, Amsterdam.

Hamilton, J. D. (1994). *Time Series Analysis.* Princeton University Press. Princeton, New Jersey.

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

Hansen, L. P. and J. A. Scheinkman (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.

Haverkamp, B. R. J., M. Verhaegen, C. T. Chou and R. Johansson (1997). Continuous-time subspace model identification method using Laguerre filtering. In: *SYSID' 97 - 11th IFAC Symposium on System Identification* (Y. Sawaragi and S. Sagara, Eds.). IFAC.

Hennig, K. and G. Grunwald (1984). Treatment of flow induced pendulum oscillations. *Kernenergie* **27**, 286–292.

Horstkempe, W. and R. Lefever (1984). *Noise Induced Transitions*. Springer. New York.

Hutton, J. E. and P. I. Nelson (1986). Quasi-likelihood estimation for semimartingales. *Stochastic Processes and their Applications* **22**, 245–257.

Jacobsen, J. L., H. Madsen and P. Harremoës (1996). Modelling the transient impact of rain events on the oxygen content of a small creek. *Water, Science and Technology* **33**(2), 177–187.

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

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

Karatzas, I. and S. E. Shreve (1996). *Brownian Motion and Stochastic Calculus, Second Edition*. Springer-Verlag. New York.

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

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

Ljung, L. (1978). Convergence analysis of parametric identification methods. *IEEE Transactions on Automatic Control* **23**(5), 770–783.

Ljung, L. (1979). Asymptotic behavior of the extended kalman filter as a parameter estimator for linear systems. *IEEE Transactions on Automatic Control* **24**(1), 36–50.

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

Ljung, L. and T. Söderström (1983). *Theory and practice of recursive identification*. MIT Press. Cambridge, Massachusetts.

Madsen, H. (1985). Statistically determined dynamical models for climate processes. PhD thesis. IMSOR, DTH. Lyngby.

Madsen, H. and H. Melgaard (1991). The mathematical and numerical methods used in CTLSM - a program for ML-estimation in stochastic, continuous time dynamical models.. Technical report. Institute of Mathematical Statistics and Operations Research.

Madsen, H. and J. Holst (1995). Estimation of continuous-time models for the heat dynamics of a building. *Energy and Buildings* **22**, 67–79.

Marron, J.S. and M.P. Wand (1992). Exact mean integrated squared error. *The Annals of Statistics* **20**, 712–736.

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

Melgaard, H. and H. Madsen (1993). CTLSM version 2.6 - a program for parameter estimation in stochastic differential equations.. Technical report. IMSOR. No. 1.

Newey, W. J. and K. D. West (1987). A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**(3), 703–708.

Ng, S. and A. Michaelides (1997). Estimating the rational expectations model of speculative storage: a Monte Carlo comparison of three simulation estimators. Manuscript. Department of Economics, Boston College.

Nielsen, J. N. and H. Madsen (1996). Modeling heat dynamics using thermal networks. In: *System Identification Competition*. Chap. 13. Joint Research Centre, Ispra, Italy.

Nielsen, J. N. and H. Madsen (2000). Applying the EKF to Stochastic Differential Equations with Level Effects. Submitted.

Nielsen, J. N., M. Vestergaard and H. Madsen (1999). Nonlinear filtering of univariate stochastic volatility models. In: *Proceedings of the 14th IFAC World Congress*. Vol. M. IFAC. Elsevier Science. pp. 123–128.

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

Nowman, K. B. (1997). Gaussian estimates of single-factor continuous time models of the term structure of interest rates. *Journal of Finance* **52**(4), 1695–1706.

Ogaki, M. (1993). Generalized method of moments: Econometric applications. In: *Handbook of Statistics, Vol. 11*. Elsevier Publishers.

Øksendal, B. (1995). *Stochastic Differential Equations, 4th Edition*. Springer-Verlag. Heidelberg.

Pedersen, A. R. (1993). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. Technical Report 264. Department of Theoretical Statistics, University of Aarhus.

Pedersen, A. R. (1995*a*). Consistency and asymptotic normality of an approximate maximum likelihood estimator for discretely observed diffusion processes. *Bernoulli* **1**, 257–279.

Pedersen, A. R. (1995*b*). A new approach to maximum likelihood estimation for stochastic differential equations based on discrete observations. *Scandinavian Journal of Statistics* **22**, 55–71.

Pierce, D. A. (1972). Least squares estimation in dynamic disturbance time-series models. *Biometrika* **5**(9), 73–78.

Protter, P. (1990). *Stochastic Integration and Differential Equations.* Springer-Verlag. New York.

Rossi, P. E. (1996). *Modelling Stock Market Volatility - Bridging the Gap to Continuous Time.* Academic Press. San Diego.

Sagara, S. and Z. Y. Zhao (1990). Numerical integration approach to on-line identification of continuous-time systems. *Automatica* **26**(1), 63–74.

Santa-Clara, P. (1995). Simulated likelihood estimation of diffusion with application to the short term interest rate. Technical report. Anderson Graduate School of Management, University of California at Los Angeles.

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

Sinha, N. K. and G. P. Rao (1991). *Identification of Continuous-time Systems.* Kluwer. Dordrecht.

Smith, A. A. (1993). Estimating nonlinear time series models using simulated vector autoregressions. *Journal of Applied Econometrics* **8**, S63–S84.

Söderström, T. and P. Stoica (1989). *System Identification.* Prentice Hall. London.

Söderström, T., H. Fan, M. Mossberg and B. Carlsson (1997). A bias-compensation scheme for estimating continuous time AR process parameters. In: *SYSID' 97 - 11th IFAC Symposium on System Identification* (Y. Sawaragi and S. Sagara, Eds.). IFAC.

Söderström, T., P. Stoica and B. Friedlander (1991). An indirect prediction error method for system identification. *Automatica* **27**(1), 183–188.

Stepner, D. E. and R. K. Mehra (1973). Maximum likelihood identification and optimal input design for identifying aircraft stability and control derivatives. NASA Rep. No. NASA CR-2200.

Unbehauen, H. and G. P. Rao (1987). *Identification of continuous systems.* North-Holland. Amsterdam.

Unbehauen, H. and G. P. Rao (1990). Continuous-time approaches to system identification - a survey. *Automatica* **26**(1), 23–35.

Unbehauen, H. and G. P. Rao (1997). Identification of continuous-time systems: A tutorial. In: *SYSID' 97 - 11th IFAC Symposium on System Identification* (Y. Sawaragi and S. Sagara, Eds.). IFAC.

Unny, T. E. and Karmeshu (1983). Stochastic nature of outputs from conceptual reservoir model cascades. *Journal of Hydrology* **68**, 161–180.

Young, P. (1981). Parameter estimation for continuous-time models - a survey. *Automatica* **17**, 23–39.

Young, P. C. (1984). *Recursive Estimation and Time-Series Analysis.* Springer-Verlag. Berlin.

Young, P. C. (1993). Time Variable and State Dependent Parameter Modelling of Nonstationary and Nonlinear Time Series. In: *Developments in Time Series* (T. Subba Rao, Ed.). Chapman and Hall: London. pp. 374–413.

Young, P. C. (1996). Identification, estimation and control of continuous-time and delta operator systems. In: *Identification in Engineering Systems* (M. I. Friswell and J. E. Mottershead, Eds.). pp. 1–17. University of Wales: Swansea.

Young, P. C. (1998*a*). Data-based mechanistic modelling, generalised sensitivity and dominant mode analysis. *Computer Physics Communications* **1**, 1–17.

Young, P. C. (1998*b*). Data-based mechanistic modelling of environmental, ecological, economic and engineering systems. *Environmental Modelling and Software* **1**, 105–122.

Young, P. C. (1998*c*). Data-based mechanistic modellling of engineering systems. *Journal of Vibration and Control* **4**, 5–28.

Young, P. C., A. Chotai and W. Tych (1991). Identification, estimation and control of continuous-time systems described by delta operator models. In: *Identification in Continuous-time Systems* (N. K. Sinha and G. P. Rao, Eds.). pp. 363–418. Kluwer: Dordrecht.

Young, P. C. and A. J. Jakeman (1980). Refined instrumental variable methods of recursive time-series analysis, part III: Extensions. *International Journal of Control* **3**, 741–764.

Young, P. C. and M. B. Beck (1974). The modelling and control of water quality in a river system. *Automatica* **10**, 455–468.

Young, P. C., L. Price, D. Berckmans and K. Janssens (1999). Recent development in the modelling and control of climate and ventilation in agricultural buildings. In press, *Computers in Agriculture.*