Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood versus Extended Kalman Filter and Itô-Taylor Expansion

Hermann SINGER

This article compares several estimation methods for nonlinear stochastic differential equations with discrete time measurements. The likelihood function is computed by Monte Carlo simulations of the transition probability (simulated maximum likelihood SML) using kernel density estimators and functional integrals and by using the extended Kalman filter (EKF and second-order nonlinear filter SNF). The relation with a local linearization method is discussed. A simulation study for a diffusion process in a double well potential (Ginzburg–Landau equation) shows that, for large sampling intervals, the SML methods lead to better estimation results than the likelihood approach via EKF and SNF. A second study using a nonlinear diffusion coefficient (generalized Cox–Ingersoll–Ross model) demonstrates that the EKF type estimators may serve as efficient alternatives to simple maximum quasi-likelihood approaches and Monte Carlo methods.

Key Words: Discrete measurements; Importance sampling; Nonlinear systems; Maximum likelihood estimation; Monte Carlo simulation; Stochastic differential equations.

1. INTRODUCTION

Nonlinear dynamical systems have been a focus of interest since the advent of "deterministic chaos," "catastrophy theory," "synergetics," and "dissipative systems" (see Schuster 1984; Arnold 1986; Ozaki 1985) which swapped from physics (Haken 1977) into other fields such as economics (Zhang 1991; Feichtinger 1992; Puu 1992) and sociology (Weidlich and Haag 1983).

This article attempts to derive and compare parameter estimation methods for nonlinear stochastic differential equations (SDE). To meet requirements of empirical data analysis, only discrete time measurements are assumed (sampling), but the modeling of the dynamics is specified in physical continuous time. These models are well known in econometrics and are considered as models of choice by some authors, since they permit an optimal

Hermann Singer is Professor, FernUniversität Hagen, Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany (E-mail: Hermann.Singer@FernUni-Hagen.de).

Journal of Computational and Graphical Statistics, Volume 11, Number 4, Pages 972–995
DOI: 10.1198/106186002808

usage of restrictions in the dynamical specification at a micro level, that is, for very small extrapolation to the future. By contrast, the dynamical connection between discrete times of measurement may be very intricate because of the cumulated interactions in this time span (exact discrete model; see Bergstrom 1990, Singer 1995, and the references cited therein).

Maximum likelihood (ML), least squares, generalized method of moments (GMM), and minimum χ^2 estimation of nonlinear differential or difference systems with continuous or discrete measurements have been treated by Basawa and Prakasa Rao (1980), Kitagawa (1987, 1996), Lo (1988), Campillo and Le Gland (1989), Florens-Zmirou (1989), Genon-Catalot (1990), Brockwell and Hyndman (1992), Singer (1992), Komaki (1993), Hansen and Scheinkman (1995), Wymer (1995), Tanizaki and Mariano (1995), Gallant and Long (1997), Overbeck and Rydén (1997), Elerian, Chib and Shephard (2001), and others.

Recently, Shoji and Ozaki (1997, 1998) compared the performance of discrete time schemes for stochastic differential equations using the Euler scheme and Itô-Taylor expansions of the vector field with a discretized likelihood ratio approach (discretized continuous sampling) and the generalized method of moments (GMM). The discretization intervals were small, however ($\Delta t = 0.05, \ldots, 0.2$).

In this article we choose larger sampling intervals, $\Delta t = 0.5, \ldots, 2$, in order to address the needs of econometrics and the social sciences. Furthermore, an estimation method is proposed which is asymptotically unbiased for arbitrary sampling interval (simulation of the transition probability). It is shown how the Itô-Taylor expansion of Shoji and Ozaki corresponds to a second-order nonlinear filter SNF (Jazwinski 1970).

The article is organized as follows: Section 2 introduces the continuous-discrete state space model. Section 3 derives and compares the likelihood-based methods

1. extended Kalman filter (EKF),
2. second order EKF (SNF),
3. local linearization with Itô-formula (Shoji and Ozaki),
4. simulated transition probability (kernel density method), and
5. simulated transition probability (functional integral method)

with each other.

Section 4 presents a simulation study that uses an SDE with nonlinear drift coefficient (fourth-order potential; Ginzburg–Landau equation), and compares the performance of the several estimation methods with each other. Section 5 gives a second simulation study that addresses the generalized Cox–Ingersoll–Ross model (GCIR) used in the financial modeling of interest rates (nonlinear diffusion coefficient).

2. NONLINEAR CONTINUOUS/DISCRETE STATE SPACE MODELS

We discuss the nonlinear *continuous/discrete state space model*

$$dy(t) = f(y(t), t, \psi)dt + g(y(t), t, \psi)dW(t), \tag{2.1}$$

where discrete measurements z_i are taken at times $\{t_0, t_1, \ldots, t_T\}$ and $t_0 \leq t \leq t_T$ according to

$$z_i = h(y(t_i), t_i, \psi) + \epsilon_i. \tag{2.2}$$

In state Equation (2.1), $W(t)$ denotes a r-dimensional Wiener process and the state is described by the p-dimensional state vector $y(t)$. It fulfills a system of stochastic differential equations in the sense of Itô (see Arnold 1974) with fixed initial condition $y(t_0)$. The functions $f : \mathsf{R}^p \times \mathsf{R} \times \mathsf{R}^u \to \mathsf{R}^p$ and $g : \mathsf{R}^p \times \mathsf{R} \times \mathsf{R}^u \to \mathsf{R}^p \times \mathsf{R}^r$ are called drift and diffusion coefficients, respectively. In measurement Equation (2.2), $\epsilon_i \sim N(0, R(t_i, \psi))$ is a k-dimensional discrete time white noise process (measurement error). Parametric estimation is based on the u-dimensional parameter vector ψ. The key quantity for the computation of the likelihood function is the transition probability $p(y, t|x, s)$ which is a solution of the Fokker–Planck equation

$$\frac{\partial p(y,t|x,s)}{\partial t} = -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t,\psi)p(y,t|x,s)] + \frac{1}{2}\sum_{ij}\frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t,\psi)p(y,t|x,s)] \quad (2.3)$$

subject to the initial condition $p(y, s|x, s) = \delta(y - x)$ (Dirac delta function). The diffusion matrix is given by $\Omega = gg' : \mathsf{R}^p \times \mathsf{R} \times \mathsf{R}^u \to \mathsf{R}^p \times \mathsf{R}^p$. Under certain technical conditions the solution of (2.3) is the conditional density of $y(t)$ given $y(s) = x$ (see, e.g., Wong and Hajek 1985, chap. 4). Furthermore, in this article we assume direct noiseless measurements of $y(t_i)$, that is, $h(y(t), t, \psi) = y(t)$ and $R(t_i, \psi) = 0$.

Extensions to nonlinear noisy measurements are given in Gordon, Salmond, and Smith (1993); Kitagawa (1987, 1996); Hürzeler and Künsch (1998); and Singer (2000).

In order to model exogenous influences, f and g are assumed to depend on deterministic regressor variables $x(t) : \mathsf{R} \to \mathsf{R}^q$, that is, $f(y, t, \psi) = f(y, t, x(t), \psi)$ and so on. For notational simplicity, the dependence on the $x(t)$ will be suppressed.

3. COMPUTATION OF THE LIKELIHOOD FUNCTION

In order to compute the likelihood function of system (2.1)–(2.2), we can express the probability distribution of states $y(t_0), \ldots, y(t_T)$ in terms of solutions of Fokker–Planck Equation (2.3). Using the Markov property of $y(t)$ we obtain

$$p(y_T, \ldots, y_1|y_0; \psi) = \prod_{i=0}^{T-1} p(y_{i+1}|y_i; \psi), \quad (3.1)$$

where $p(y_T, \ldots, y_1|y_0; \psi)$ is the joint distribution of measurements conditional on $y(t_0) = y_0$ (see Lo 1988) and $p(y_{i+1}|y_i; \psi) := p(y_{i+1}, t_{i+1}|y_i, t_i; \psi)$ is the transition probability density. Defining the likelihood function as $L_\psi(y) := p(y_T, \ldots, y_1|y_0; \psi)$ and the ML estimator as $\hat{\psi} := \arg\max_\psi L_\psi(y)$ we must solve Fokker–Planck Equation (2.3) repeatedly in a nonlinear optimization algorithm. Only in the case of a linear vector field f and state independent diffusion coefficient g do we obtain a Gaussian density, but otherwise complicated functions arise. In some special cases analytical solutions have been derived. For example, in the case of linear $f(y) = \mu y$ and $g(y) = \sigma y^{\alpha/2}$, which is the well-known constant elasticity of variance (CEV) diffusion process used in option pricing (see Feller 1951; Cox and Ross 1976), an analytical solution has been derived by Feller involving Bessel functions.

In the general multivariate case, we cannot hope to obtain analytical solutions and must resort to approximations and numerical procedures for (2.3) (matrix continued-fractions, finite differences, Monte Carlo methods, and so on; see Risken 1989; Press, Teukolsky, Vetterling, and Flannery 1992; Ames 1992; Kloeden and Platen 1992). Alternatively, the prediction error identification method (Ljung and Söderström 1983), the extended Kalman filter EKF or other linearization methods (e.g., Shoji and Ozaki 1997) lead to approximations of the conditional density in terms of conditional Gauss distributions (see Sections 3.1–3.4). These approximations work well when the sampling intervals $\Delta t_i = t_{i+1} - t_i$ are not too large in comparision with the dynamics as specified in f and g (see Section 3.4). On the other hand, time series and panel data often involve large sampling intervals which are fixed by the design of the study. Therefore, as in the linear case, the likelihood should be computed exactly for the times of measurement t_i (exact discrete model).

3.1 EXTENDED KALMAN FILTER

The extended Kalman filter (EKF) is a well-known extension of the linear Kalman filter to nonlinear systems (see Jazwinski 1970). Here we use a version which is applicable to the continuous/discrete state space model.

Denoting the conditional mean and covariance matrix as $y(t|t_i) = E[y(t)|Z^i]$ and $P(t|t_i) = \mathrm{var}[y(t)|Z^i]$, where $Z^i = \{z_i, \ldots, z_0\}$ are the measurements up to time t_i, these fulfil the exact equations ($t_i \le t \le t_{i+1}$)

$$\dot{y}(t|t_i) = E[f|Z^i] \tag{3.2}$$

$$\dot{P}(t|t_i) = \mathrm{cov}(f, y|Z^i) + \mathrm{cov}(y, f|Z^i) + E[\Omega|Z^i] \tag{3.3}$$

between measurements. These are not differential equations, however, because the expectation values require knowledge of the conditional distribution $p(y, t|Z^i)$.

Expanding the drift f and the diffusion matrix Ω around the estimate $y(t|t_i)$ up to first order, and inserting this into (3.2)–(3.3) leads to the coupled approximate moment equations (Jazwinski 1970, chap. 9)

$$\dot{y}(t|t_i) = f(y(t|t_i), t) \tag{3.4}$$

$$\dot{P}(t|t_i) = f_y(y(t|t_i), t)P(t|t_i) + P(t|t_i)f_y'(y(t|t_i), t) + \Omega(y(t|t_i), t), \tag{3.5}$$

where the Jacobian of the drift is defined by

$$(f_y)_{jk} := f_{j,k} := \frac{\partial f_j(y,t)}{\partial y_k}; \quad j,k = 1, \ldots, p \tag{3.6}$$

3.2 SECOND-ORDER EXTENDED KALMAN FILTER

Expanding the drift f and the diffusion matrix Ω up to second order around the estimate $y(t|t_i)$ and inserting this into (3.2)–(3.3) leads to the coupled approximate moment equations (second-order nonlinear filter SNF; Jazwinski 1970, chap. 9)

$$\dot{y}(t|t_i) = f(y(t|t_i), t) + \frac{1}{2} f_{yy}(y(t|t_i), t) * P(t|t_i) \tag{3.7}$$

$$\begin{aligned}\dot{P}(t|t_i) &= f_y(y(t|t_i),t)P(t|t_i) + P(t|t_i)f_y'(y(t|t_i),t)\Omega(y(t|t_i),t) \\ &\quad + \frac{1}{2}\Omega_{yy}(y(t|t_i),t) * P(t|t_i), \end{aligned} \tag{3.8}$$

where the Jacobians are defined by

$$(f_y)_{jk} := f_{j,k} := \frac{\partial f_j(y,t)}{\partial y_k}; \quad j,k = 1,\ldots,p \tag{3.9}$$

$$(f_{yy})_{jkl} := f_{j,kl} := \frac{\partial^2 f_j(y,t)}{\partial y_k \partial y_l}; \quad j,k,l = 1,\ldots,p \tag{3.10}$$

$$(\Omega_{yy})_{jklm} := \Omega_{jk,lm} := \frac{\partial^2 \Omega_{jk}(y,t)}{\partial y_l \partial y_m}; \quad j,k,l,m = 1,\ldots,p. \tag{3.11}$$

The second-order terms $f_{yy} * P(t|t_i)$ and $\Omega_{yy} * P(t|t_i)$ explicitly read

$$(f_{yy} * P(t|t_i))_j = \sum_{kl} (f_{yy})_{jkl} P(t|t_i)_{kl} = \mathrm{tr}[(f_{yy})_j P(t|t_i)], \tag{3.12}$$

and

$$(\Omega_{yy} * P(t|t_i))_{jk} = \sum_{lm} (\Omega_{yy})_{jklm} P(t|t_i)_{lm} = \mathrm{tr}[(\Omega_{yy})_{jk} P(t|t_i)]. \tag{3.13}$$

At times of measurement t_i, the conditional mean and covariance are changed by the EKF measurement update. In the present context, where the state is measured directly without noise, this leads to the initial conditions $y(t_i|t_i) = y_i, P(t_i|t_i) = 0$. In the usual EKF, the second-order term $f_{yy} * P(t|t_i)$ in the mean equation and $\Omega_{yy} * P(t|t_i)$ in the variance equation are neglected. Moment equations including the second-order terms define the second-order nonlinear filter SNF (Jazwinski 1970, chap. 9). In both cases the likelihood function is computed recursively using the prediction error decomposition (Schweppe 1965)

$$L_\psi(z) = \prod_{i=0}^{T-1} |2\pi\Gamma_{i+1|i}|^{-1/2} \exp\left\{-\frac{1}{2}\nu_{i+1}'\Gamma_{i+1|i}^{-1}\nu_{i+1}\right\}, \tag{3.14}$$

$$\nu_{i+1} = z_{i+1} - h(y(t_{i+1}|t_i), t_{i+1}), \tag{3.15}$$

$$\Gamma_{i+1|i} = H_{i+1}P(t_{i+1}|t_i)H_{i+1}' + R_{i+1}, \tag{3.16}$$

and $H_{i+1} = h_y(y(t_{i+1}|t_i), t_{i+1}) : k \times p$ is the Jacobian of the output function $h(y,t)$.

3.3 LOCAL LINEARIZATION METHOD OF SHOJI AND OZAKI

There is a close relation to the local linearization method of Shoji and Ozaki (1997, 1998) where the Itô formula was used to expand the vector field $f(y,t)$. We obtain

$$\begin{aligned} f(y(t),t) - f(y(t_i),t_i) &= \int_{t_i}^t f_y(y(s),s)dy(s) + \int_{t_i}^t \frac{1}{2} f_{yy}(y(s),s) * \Omega(y(s),s)ds \\ &\quad + \int_{t_i}^t f_t(y(s),s)ds. \end{aligned} \tag{3.17}$$

Freezing the coefficients at (y_i, t_i) and using a state independent diffusion coefficient $\Omega(s)$ Shoji and Ozaki obtained the linearized SDE ($t_i \le t \le t_{i+1}$)

$$dy(t) = [f_y(y_i,t_i)(y(t)-y_i) + f(y_i,t_i) + f_t(y_i,t_i)(t-t_i) + \frac{1}{2} f_{yy}(y_i,t_i) * \Omega(t_i)(t-t_i)]dt + g(t)dW(t).$$

The corresponding moment equations are

$$\begin{aligned} \dot{y}(t|t_i) &= f_y(y_i,t_i)(y(t|t_i)-y_i) + f(y_i,t_i) + f_t(y_i,t_i))(t-t_i) \\ &\quad + \frac{1}{2} f_{yy}(y_i,t_i) * \Omega(t_i)(t-t_i) \qquad (3.18) \\ \dot{P}(t|t_i) &= f_y(y_i,t_i)P(t|t_i) + P(t|t_i)f_y'(y_i,t_i) + \Omega(t_i). \qquad (3.19) \end{aligned}$$

By contrast to the EKF moment Equations (3.7)–(3.8), which is a system of nonlinear differential equations, the Jacobians are evaluated once at the measurements (y_i, t_i) and the differential equations are linear and not coupled. In both cases, the differential equations were discretized on a grid with spacing $\delta t = (t_{i+1} - t_i)/J_i$ and numerically solved with the Euler method.

The similarity with the SNF of Section 3.2 is seen from a Taylor expansion of the drift around (y_i, t_i) up to second order leading to

$$\begin{aligned} f(y(t),t) &= f(y_i,t_i) + f_y(y_i,t_i)(y(t)-y_i) + f_t(y_i,t_i)(t-t_i) \\ &\quad + \frac{1}{2} f_{yy}(y_i,t_i) * (y(t)-y_i)(y(t)-y_i)' \\ &\quad + O(||y(t)-y_i||^3) + O((t-t_i)^2). \qquad (3.20) \end{aligned}$$

Inserting this into (3.2) and replacing $E[(y(t)-y_i)(y(t)-y_i)'|Z^i] \approx \Omega(t_i)(t-t_i)$ yields the first moment equation of the Shoji–Ozaki approach. By contrast, the SNF results by Taylor expansion of f around the filtered state $y(t|t_i)$. Of course, this is not surprising since the Itô formula is based on Taylor expansion as well.

3.4 VALIDITY OF THE LINEARIZATION METHODS

As mentioned earlier, the EKF and also the Shoji–Ozaki method rest on the assumption, that the transition density $p(y_{i+1}|y_i;\psi)$ is Gaussian and the moments $y(t|t_i) = E[y(t)|Z^i]$ and $P(t|t_i) = \text{var}[y(t)|Z^i] := M_2$ can be computed by solving approximate moment equations. If p is Gaussian, the third moment $M_3 = E[(y(t) - y(t|t_i))^3|Z^i]$ must vanish and the fourth moment will factorize as $M_4 = 3M_2^2$ in the scalar case. Since these moments fulfill the equations

$$\begin{aligned} \dot{M}_3(t|t_i) &= 3E[f(y(t),t)\,((y(t)-y(t|t_i))^2 - M_2)|Z^i] \\ &\quad + 3E[\Omega(y(t),t)\,(y(t)-y(t|t_i))|Z^i] \qquad (3.21) \\ \dot{M}_4(t|t_i) &= 4E[f(y(t),t)\,((y(t)-y(t|t_i))^3 - M_3)|Z^i] \\ &\quad + 6E[\Omega(y(t),t)\,(y(t)-y(t|t_i))^2|Z^i], \qquad (3.22) \end{aligned}$$

with initial conditions $M_3(t_i|t_i) = 0; M_4(t_i|t_i) = 3M_2(t_i|t_i)^2$, M_3 will not remain zero if f is nonlinear and $\Omega(y,t)$ depends on y. Expanding again around the conditional mean $y(t|t_i)$ up to second order, one obtains the approximate equations

$$\dot{M}_3(t|t_i) = 3f_yM_3 + \frac{3}{2}f_{yy}(M_4 - M_2^2) + 3\Omega_yM_2 + \frac{3}{2}\Omega_{yy}M_3 \tag{3.23}$$

$$\dot{M}_4(t|t_i) = 4f_yM_4 + 2f_{yy}(M_5 - M_2M_3) + 6\Omega M_2 + 6\Omega_yM_3 + 3\Omega_{yy}M_4. \tag{3.24}$$

Only in the linear case $f = Ay + b; \Omega = \text{const.}$, the solution of $\dot{M}_3 = 3AM_3$ is zero for all times and $\dot{M}_4 = 4AM_4 + 6\Omega M_2$ is solved by $M_4 = 3M_2^2$, as required by a Gaussian solution. Nonlinearities lead to deviations from the Gauss distribution which is valid only for short sampling intervals $\Delta t_i = t_{i+1} - t_i$.

Similar considerations apply to the Shoji–Ozaki method, where $\int f_y dy \approx f_y(y(t) - y(t_i))$ and $\int f_{yy}\Omega_y dt \approx f_{yy}\Omega_y(t - t_i)$. This is valid only for small $t - t_i$ and for longer intervals nonlinear terms appear in the approximate stochastic differential equation leading to deviations from Gaussianity.

For example, if $f = ay + b/2y^2; f_y = a + by$, the Itô integral is exactly solved by $\int f_y dy = a(y(t) - y(t_i)) + b/2(y(t)^2 - y(t_i)^2 - \Omega(t - t_i))$, whereas the approximation leads to $(a + by(t_i))(y(t) - y(t_i))$ Using $y^2 - x^2 = (y + x)(y - x) \approx 2x(y - x)$ for $y(t) \downarrow y(t_i)$ it is seen that the expressions coincide for $t \downarrow t_i$ by the continuity of the solution $y(t)$.

3.5 KERNEL DENSITY METHOD

We choose a solution method of Fokker–Planck Equation (2.3), which basically involves simulation of N independent trajectories $y_n(t), n = 1, \ldots, N$ between times t_i and t_{i+1}. The functional

$$p(y,t|x,s) = E[\delta(y(t) - y)|y(s) = x] \tag{3.25}$$

is approximated by the mean (N-point estimator)

$$\hat{p}(y,t|x,s) = \frac{1}{N}\sum_{n=1}^{N}\delta(y_{nxs}(t) - y), \tag{3.26}$$

where $y_{nxs}(t)$ is an ensemble of solutions of (2.1) starting at (x,s). $\hat{p}$ is an unbiased estimator of p, but its variance does not exist since integrals involving the square of the delta function $\delta(.)^2$ cannot be defined (see, e.g., Lighthill 1958). If we use a unimodal positive kernel function $K(.)$ integrating to one, such as the Gauss density $\phi(y; 0, \Sigma)$ instead of $\delta(.)$, we obtain a kernel density estimator of p (see Silverman 1986). This estimator is used for the transition density $p(y_{i+1}|y_i)$ and likelihood (3.1). Since it is necessary to evaluate $\hat{p}$ very often, the simple choice of the smoothing parameter Σ, namely

$$\Sigma(t) = (aN^{-\frac{1}{p+4}})^2\widehat{\text{var}}(y_{nxs}(t))$$

$$a = \left(\frac{4}{p+2}\right)^{\frac{1}{p+4}} \tag{3.27}$$

was used. It is only optimal for Gauss distributed data (see Silverman 1986). Here, $\widehat{\text{var}}(y_{nxs}(t))$ is the empirical covariance matrix at time t of the N trajectories starting at (x, s). The smoothing parameter is separately chosen for each evaluation of the transition density $p(y_{i+1}|y_i;\psi)$ occuring in (3.1). In the context of Monte Carlo filtering in discrete time, a similar approach was used by Hürzeler and Künsch (1998). Letting the smoothing parameter go to zero, the filtering method of Kitagawa (1996) is recovered and (3.26) is obtained. The trajectories $y_{nxs}(t)$ were simulated using the simplest stochastic Runge–Kutta scheme, the Euler–Maruyama method (see Fahrmeir 1976; Rümelin 1982; Kloeden and Platen 1992). More accurate higher order methods such as Heun's scheme could be used, but these schemes generate solutions in the sense of Stratonovich. Of course it is always possible to modify the drift such that the simulated solution coincides with the Itô solution of (2.1).

Since we want to evaluate only a functional of $y(t)$, a weak Euler scheme was used. Instead of the Gaussian increments of the Wiener process $\delta W(t) = W(t+\delta t) - W(t)$ uniform random variables with mean zero and variance δt were used leading to savings in CPU time. Alternatively, a two-point variable with the same moments could be used, too (see Kloeden and Platen 1992, part VI).

In comparision to other solution methods of the Fokker–Planck equation, the Monte Carlo approach works well in higher dimensions (p-dimensional state y) and is applicable even if the diffusion matrix $\Omega = gg'$ is singular at some points. Furthermore, $\hat{p}$ is always positive and integrates to one. Finite difference methods, for examples, can lead to negative values of the solution matrix because of approximation errors.

3.6 FUNCTIONAL INTEGRAL METHOD

An alternative representation of the transition probability is obtained by repeatedly using the Markov property and the Chapman–Kolmogorov equation

$$p(y_{i+1}|y_i) = \int p(y_{i+1}|\eta_{J_i-1})p(\eta_{J_i-1}|\eta_{J_i-2}) \\ \dots p(\eta_1|y_i)d\eta_{J_i-1}\dots d\eta_1, \qquad (3.28)$$

where the interval $[t_i, t_{i+1}]$ is divided in $J_i = \Delta t_i/\delta t$ subintervals of spacing $\delta t = \Delta t_i/J_i$ and auxiliary variables $y_i = \eta_0, \dots, \eta_{J_i} = y_{i+1}$ were introduced ($\eta_j = y(\tau_j); \tau_j = t_i + j\delta t; j = 0, \dots, J_i$). Then, for small δt the transition density can be approximated by the normal density

$$p(\eta_{j+1}|\eta_j) \quad \approx \quad \phi(\eta_{j+1}; \eta_j + f(\eta_j, \tau_j)\delta t, \Omega(\eta_j, \tau_j)\delta t). \qquad (3.29)$$

In the limit $J_i \to \infty$ we obtain the path (functional) integral representation

$$p(y_{i+1}|y_i) = \lim_{J_i\to\infty} \int \exp\left[-\frac{1}{2}\sum_{j=0}^{J_i-1}(\eta_{j+1}-\eta_j - f_j\delta t)'(\Omega_j\delta t)^{-1}(\eta_{j+1}-\eta_j-f_j\delta t)\right] \times \prod_{j=0}^{J_i-1}|2\pi\Omega_j\delta t|^{-1/2} d\eta_{J_i-1}\ldots d\eta_1 \tag{3.30}$$

$$:= \int \exp\left(-\frac{1}{2}O[y]\right) Dy(t) \tag{3.31}$$

(see Haken 1977, chap. 6.6; Risken 1989, ch. 4.4.2; Kloeden and Platen 1992, chap. 16.3; Pedersen 1995, theorem 3). The formal exponent

$$O[y] = \int_{t_i}^{t_{i+1}} [\dot{y}(t) - f(y,t)]'\Omega(y,t)^{-1}[\dot{y}(t)-f(y,t)]dt \tag{3.32}$$

is called a *Onsager-Machlup functional* and $Dy(t)$ symbolizes the integration over paths. In practical computations we do not go to the limit but choose δt small enough [a so-called ϵ version in the sense of Stratonovich (1989)]. Thus, we obtain a $(J_i - 1)$-dimensional integral (3.28), which is approximated by the mean

$$\hat{p}(y_{i+1}|y_i) = N^{-1}\sum_{n=1}^{N} p(y_{i+1}|\eta_{n,J_i-1}), \tag{3.33}$$

where η_{n,J_i-1} is the nth simulated time path $y_{n,y_i,t_i}(t)$ evaluated at time $t = t_i + (J_i - 1)\delta t$. As with the kernel density method, a weak Euler scheme (with uniform variates) was used for the simulation of the sample paths.

3.6.1 Importance Sampling

The integral representation (3.28) can be rewritten to reduce the the variance of the estimate (3.33). In general, the integral

$$E_1[g] = \int g(y)p_1(y)dy = \int g(y)\frac{p_1(y)}{p_2(y)}p_2(y)dy = E_2\left[g\frac{p_1}{p_2}\right] \tag{3.34}$$

leads to a variance reduced unbiased estimate

$$\widehat{E_1[g]} = N^{-1}\sum_{n=1}^{N} g(y_n)\frac{p_1(y_n)}{p_2(y_n)} \tag{3.35}$$

$y_n \sim p_2$ if the density p_2 is chosen appropriately. One can show that the optimal density is given by

$$p_{2,\text{opt}} = \frac{|g(y)|p_1(y)}{E_1|g(y)|} \tag{3.36}$$

and the variance of (3.35) is zero if g is positive (see Kloeden and Platen 1992, chap. 16.3). Setting $g(\eta_{J_i-1}) = p(y_{i+1}|\eta_{J_i-1})$ and $p_1 = p(\eta_{J_i-1}|\eta_{J_i-2}) \dots p(\eta_1|\eta_0)$ leads to the optimal density

$$p_{2,\text{opt}} = p(\eta_{J_i-1}|\eta_{J_i-2}, y_{i+1}) \dots p(\eta_1|\eta_0, y_{i+1}), \tag{3.37}$$

where the transition densities are conditioned on future measurements y_{i+1}. In general these are difficult to compute, but for the linear system an exact result is available which can be generalized to nonlinear systems. Thus, in linear systems, the estimate (3.35) has variance zero and the mean is exact with one trajectory. In nonlinear systems, the density p_2 leads to an estimate $\hat{p}(y_{i+1}|y_i)$ with variance > 0 but still to a variance reduction. Since in the linear case all densities are Gaussian, one obtains the conditional mean and variance ($\eta_j = y(\tau_j); \tau_j = t_i + j\delta t; j = 0, \dots, J_i - 1$)

$$\begin{aligned}
E(\eta_{j+1}|\eta_j, y_{i+1}) &= E(\eta_{j+1}|\eta_j) + F_j[y_{i+1} - E(y_{i+1}|\eta_j)] && (3.38)\\
\text{var}(\eta_{j+1}|\eta_j, y_{i+1}) &= \text{var}(\eta_{j+1}|\eta_j) - F_j \text{var}(y_{i+1}|\eta_j)]F_j' && (3.39)\\
F_j &:= \text{var}(\eta_{j+1}|\eta_j)\Phi'(t_{i+1}, \tau_{j+1})\text{var}(y_{i+1}|\eta_j)^{-1} && (3.40)
\end{aligned}$$

which characterize the density $p(\eta_{j+1}|\eta_j, y_{i+1})$. The smoother gain F_j and the update formulas are analogous to the fixed interval smoother (see Anderson and Moore 1979) and can be easily generalized to noisy measurements z_{i+1} (Singer 2000). The quantities in the update formulas can be obtained by solving the differential equations ($\tau_j \le t \le t_{i+1}$)

$$\begin{aligned}
\dot{y}(t|\tau_j) &= f(y(t|\tau_j), t);\ y(\tau_j|\tau_j) = \eta_j && (3.41)\\
\dot{P}(t|\tau_j) &= A(t)P(t|\tau_j) + P(t|\tau_j)A'(t) + \Omega(t);\ P(\tau_j|\tau_j) = 0 && (3.42)\\
\dot{\Phi}(t|\tau_{j+1}) &= A(t)\Phi(t|\tau_{j+1});\ \Phi(\tau_{j+1}|\tau_{j+1}) = I && (3.43)\\
f(y,t) &:= A(t)y + b(t) && (3.44)
\end{aligned}$$

and setting

$$\begin{aligned}
E(\eta_{j+1}|\eta_j) &= y(\tau_{j+1}|\tau_j) && (3.45)\\
E(y_{i+1}|\eta_j) &= y(t_{i+1}|\tau_j) && (3.46)\\
\text{var}(\eta_{j+1}|\eta_j) &= P(\tau_{j+1}|\tau_j) && (3.47)\\
\text{var}(y_{i+1}|\eta_j) &= P(t_{i+1}|\tau_j). && (3.48)
\end{aligned}$$

In the limit of small δt one can write

$$\begin{aligned}
E(\eta_{j+1}|\eta_j, y_{i+1}) &= \eta_j + f(\eta_j, \tau_j)\delta t + F_j[y_{i+1} - E(y_{i+1}|\eta_j)] && (3.49)\\
\text{var}(\eta_{j+1}|\eta_j, y_{i+1}) &= \Omega(\tau_j)\delta t - F_j \text{var}(y_{i+1}|\eta_j)F_j' && (3.50)
\end{aligned}$$

which can be interpreted as a correction to the drift f and the diffusion matrix Ω. Therefore the optimal density p_2 and trajectories drawn from it can be obtained by using a modified drift f_2 and diffusion coefficient Ω_2.

Since the moment equations (3.41)–(3.44) can be generalized to nonlinear systems by replacing $A(t) \to f_y(y,t)$ and $\Omega(t) \to \Omega(y,t)$ one obtains a sampling scheme where

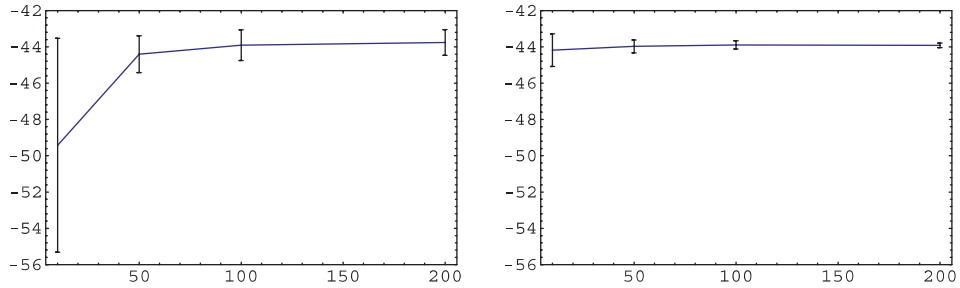

Figure 1. Convergence of the simulated likelihood (Ginzburg–Landau model). Means ± standard deviations in $M = 10$ replications. Right picture: with importance sampling. Sample size $N = 10, 50, 100, 200$.

the (sub)optimal density p_2 is implemented by means of the EKF updates and trajectories $\{\eta_{J_i-1}, \ldots, \eta_0\}_n \sim p_2$ can be simulated using f_2 and Ω_2. The convergence of the simulated likelihood is shown in Figure 1, where the model discussed in Section 4 was used. The variance of the estimates is considerably reduced in comparision to the simple estimate (3.33), but improvements of the density p_2 may be possible. Heuristically, the benchmark of the sampler is the EKF, and systems well described by it will exhibit small variances even for small sample size N.

4. SIMULATION STUDY 1: GINZBURG–LANDAU MODEL

The several estimation methods were tested in a simulation study. The model under consideration is a diffusion process in a double well potential $\Phi(y, \{\alpha, \beta, \sigma\}) = \frac{\alpha}{2}y^2 + \frac{\beta}{4}y^4$

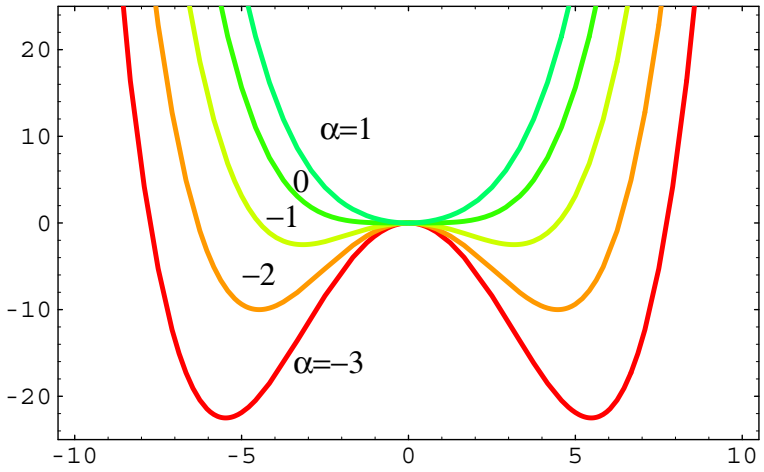

Figure 2. Potential $\Phi(y, \{\alpha, \beta, \sigma\}) = \frac{\alpha}{2}y^2 + \frac{\beta}{4}y^4$ for $\beta = 1$ and several values of $\alpha = -3, \ldots, 1$. Bifurcation of f when α becomes negative.

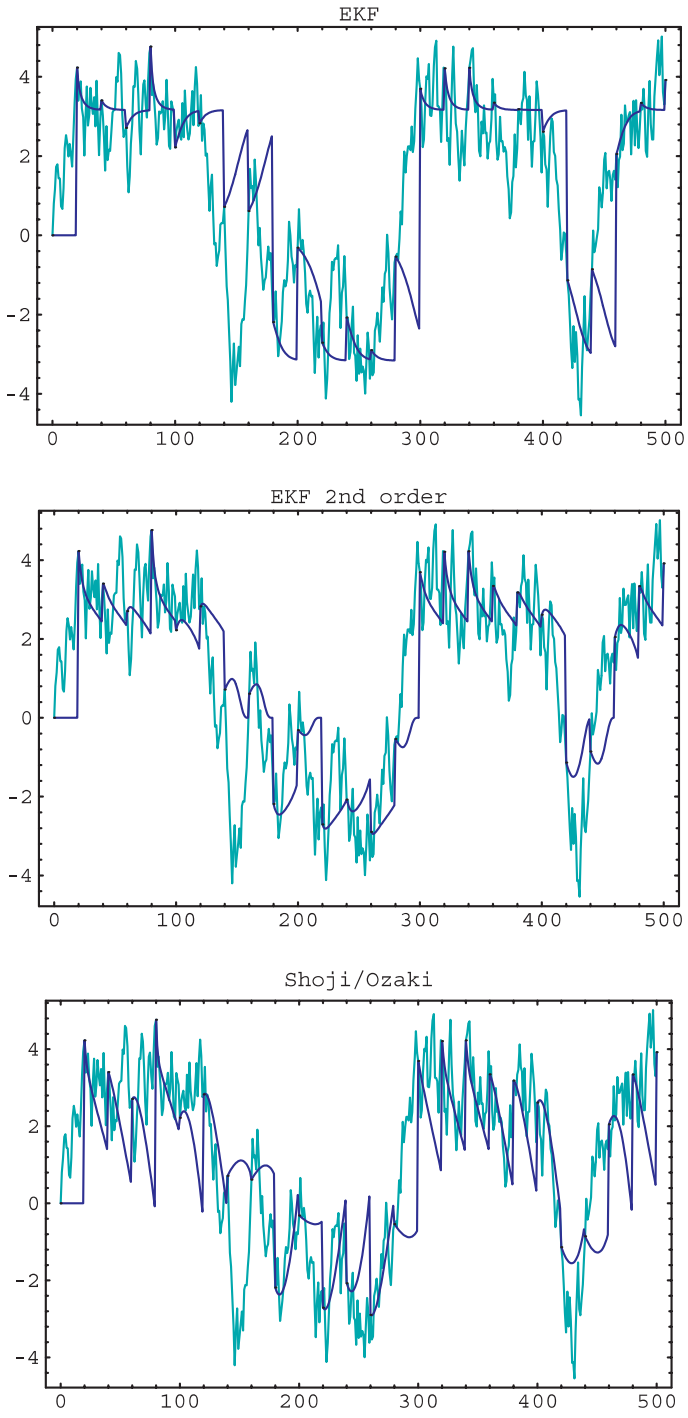

Figure 3. Diffusion process in a bimodal potential. Filtered estimates of $y(t)$ using the EKF, SNF and local linearization (true parameters inserted, sampling interval $\Delta t = 2$).

Table 1. Means and Standard Deviations of ML Estimates in $M = 100$ Replications. Comparison of EKF, SNF, and Shoji–Ozaki's local linearization at different sampling intervals $\Delta t = 2, 1, 0.5$.

		EKF			*SNF*			*Shoji–Ozaki*		
para-meter	*true values*	*mean*	*std*	*RMSE*	*mean*	*std*	*RMSE*	*mean*	*std*	*RMSE*
					Δt = 2					
α	–1.	–0.4363	0.5689	0.8009	–0.2728	0.2227	0.760503	–0.164693	0.277562	0.880215
β	0.1	0.1192	0.1545	0.15565	0.04142	0.04057	0.071258	0.0214504	0.0119278	0.079450
σ	2.	2.204	0.9013	0.92401	1.554	0.242	0.507157	1.59717	0.349	0.532984
					Δt = 1					
α	–1.	–0.5145	0.4284	0.64744	–0.6365	0.2748	0.455647	–0.461325	0.230173	0.58579
β	0.1	0.08236	0.05755	0.06020	0.07118	0.02423	0.037655	0.0450723	0.0155778	0.05709
σ	2.	1.992	0.3542	0.35427	1.804	0.1785	0.265299	1.76298	0.198182	0.30896
					Δt = 0.5					
α	–1.	–0.7608	0.2817	0.36955	–0.9744	0.3626	0.36345	–0.81903	0.285035	0.33763
β	0.1	0.09741	0.02708	0.02721	0.1002	0.02942	0.02942	0.076245	0.019856	0.030961
σ	2.	2.028	0.1617	0.16417	1.989	0.1475	0.14783	1.91593	0.131459	0.156042

with vector field $f = -\partial\Phi/\partial y$ and state independent diffusion coefficient $g = \sigma$. The SDE reads explicitly (Ginzburg–Landau equation)

$$dy \quad = \quad -[\alpha y + \beta y^3]dt + \sigma dW(t). \tag{4.1}$$

A physical picture is the strongly damped random movement of a sphere in a landscape defined by the potential Φ.

The potential Φ can exhibit a Hopf bifurcation when α becomes negative ($\beta > 0$; see Figure 2). The singularity (zero of f) at $y = 0$ forks into 3 singularities at $\{\pm\sqrt{-\alpha/\beta}, 0\}$. In the simulation study a parameter constellation of $\psi = \{\alpha, \beta, \sigma\} = \{-1, 0.1, 2\}$ corresponding to minima of Φ at $\pm\sqrt{10}$ was used. It can be easily shown that the stationary density $p_0(y) = \lim_{t\to\infty} p(t, y|x, s)$ is proportional to $\exp(-\frac{2}{\sigma^2}\Phi)$ leading to a bimodal shape with maxima at the singularities. Models of this kind occur in the context of limit cycles, phase transitions and as normal forms of nonlinear systems (Arnold 1973, 1986; Haken 1977; Holmes 1981, normal form theorem 4.4). From a more practical point of view, the model has been studied in order to guide the application of filter methods in the atmospheric and oceanic sciences (Miller, Ghil, Gauthiez 1994). In the present context the non-Gaussian density of the model is a test for the local linearization and EKF methods, which still rely on small deviations from the linear case.

The data used for estimation were $M = 100$ simulated trajectories on a grid with $J = 500, \delta t = 0.1$, but these were sampled at uniform time intervals of $\Delta t = 2$, that is, $i = 0, \ldots T = 25$. Also, smaller intervals $\Delta t = 1, 0.5$ were investigated.

The likelihood was computed using the several estimation methods and maximized by using a quasi Newton scheme with numerical score and BFGS secant updates (Dennis and Schnabel 1983). The iteration was stopped at k, if the score and the difference in the log-likelihood was small, that is, $||s_k||_\infty < \epsilon_1 = 0.001$ and $|l_k - l_{k-1}| < \epsilon_2 = 0.001$.

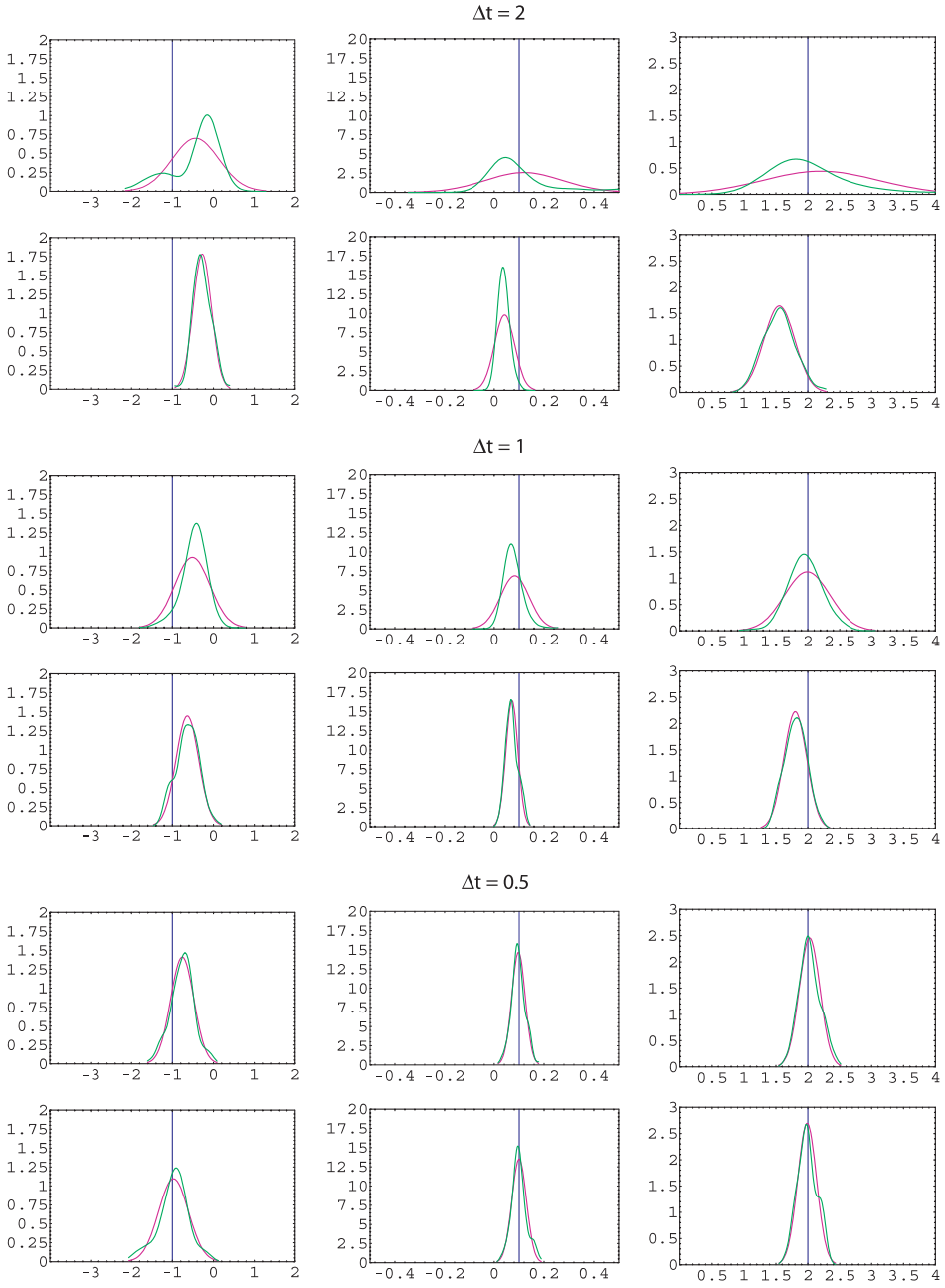

Figure 4. EKF and SNF. Empirical distribution (kernel density estimates and fitted normal distribution) of $M = 100$ ML estimates of $\{\alpha, \beta, \sigma\} = \{ -1., 0.1, 2. \}$ (from left). Rows 1 – 2: $\Delta t = 2$, row 1: EKF, row 2: SNF. Rows 3 – 4: $\Delta t = 1$, row 3: EKF, row 4: SNF. Rows 5 – 6: $\Delta t = 0.5$, row 5: EKF, row 6: SNF. True values are marked by vertical lines.

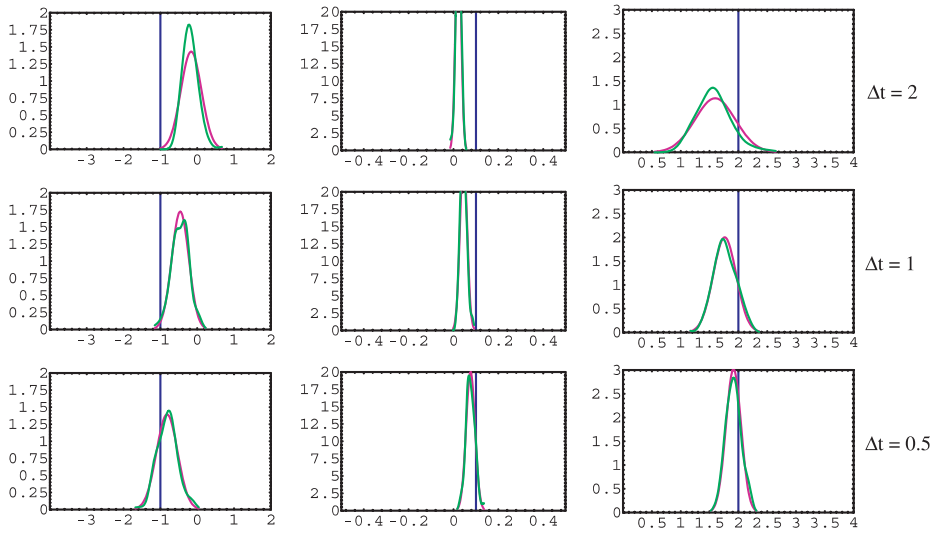

Figure 5. Local linearization (Shoji–Ozaki). Empirical distribution of $M = 100$ ML estimates of $\{\alpha, \beta, \sigma\} = \{-1., 0.1, 2.\}$ (from left). Row 1: $\Delta t = 2$, Row 2: $\Delta t = 1$, Row 3: $\Delta t = 0.5$.

4.1 EKF, SECOND-ORDER EKF, AND LOCAL LINEARIZATION

First the EKF, the second-order EKF (SNF; Taylor expansions of f and g), and the Shoji–Ozaki approach were compared using three different sampling intervals ($\Delta t = 2, 1, 0.5$). Figure 3 shows a simulated trajectory and the approximate filter solution $E[y(t)|Z^t]$ for the methods EKF, SNF, and the Shoji–Ozaki approach. At the times of measurement $t_i = i\Delta t$, $\Delta t = 2$ the solution changes discontinuously due to the new measurement information. The figures show that the second-order corrections lead to a better tracking of the trajectory $y(t)$.

The results of the simulation study are shown in Table 1 and Figures 4–5. For the sampling intervals $\Delta t = 2, 1$, the second-order filter is best in terms of the root mean square error (RMSE), but for $\Delta t = 0.5$ each of the methods leads to good results. Generally, the smaller the sampling interval, the smaller the bias and RMSE of the estimates.

4.2 KERNEL DENSITY AND FUNCTIONAL INTEGRAL APPROACH

Since the linearized estimators of the last section are asymptotically biased in the limit $T \to \infty$, Δt fixed, the asymptotically unbiased ($N \to \infty$, $T \to \infty$, Δt fixed) kernel density and functional integral approaches (simulated likelihood) were tested with the large sampling interval $\Delta t = 2$ using different sizes of the Monte Carlo sample ($N = 10, 50, 100$). The results are shown in Tables 2–4 and Figures 6–8.

In comparision with the EKFs, the estimates computed with the simulated likelihood method exhibit smaller bias and less dispersion. Without variance reduction, the size N of the Monte Carlo sample must be large enough (50 or more) in order to keep the simulation error small, but the variance reduced sampler (3.35) yields good results already for $N =$

Table 2. Kernel Density Method. Means and standard deviations of ML estimates in $M = 100$ replications (Monte Carlo sample size $N = 10, 50, 100$).

$\Delta t = 2$	True values	Mean	Std	RMSE
		$N = 10$		
α	–1.	–2.73522	1.66331	2.40366
β	0.1	0.334058	0.193477	0.303671
σ	2.	2.37004	0.634993	0.734947
		$N = 50$		
α	–1.	–1.47312	0.776728	0.909478
β	0.1	0.179503	0.102761	0.129926
σ	2.	1.81311	0.355508	0.40164
		$N = 100$		
α	–1.	–1.42286	0.720371	0.835311
β	0.1	0.158894	0.084701	0.103164
σ	2.	1.91374	0.367772	0.377752

10. Comparing the two different approaches, the functional integral method is superior, especially in the estimation of parameters α and β related to the drift.

4.3 COMPARISION OF ALL METHODS

In the case of large sampling interval $\Delta t = 2$, the estimation methods relying on a simulation of the likelihood function lead to the smallest RMSE, if the Monte Carlo sample size is large enough or variance reduction techniques are used. This is because the simulation methods take into account, that the conditional density $p(y, t|y_i)$ is bimodal, whereas the EKFs and the Shoji–Ozaki approach always assume a Gaussian distribution. This is a

Table 3. Functional Integral Method. Means and standard deviations of ML estimates in $M = 100$ replications (Monte Carlo sample size $N = 10, 50, 100$).

$\Delta t = 2$	True values	Mean	Std	RMSE
		$N = 10$		
α	–1.	–1.79274	1.03301	1.30213
β	0.1	0.200757	0.115922	0.15359
σ	2.	2.50679	0.550433	0.748208
		$N = 50$		
α	–1.	–1.11849	0.579374	0.591366
β	0.1	0.119159	0.0553919	0.0586115
σ	2.	2.00896	0.379529	0.379634
		$N = 100$		
α	–1.	–1.06157	0.555947	0.559346
β	0.1	0.110808	0.0578512	0.0588522
σ	2.	1.90572	0.375373	0.387032

Table 4. Functional Integral Method with Importance Sampling. Means and standard deviations of ML estimates in $M = 100$ replications (Monte Carlo sample size $N = 10, 50, 100$).

$\Delta t = 2$	*True values*	*Mean*	*Std*	*RMSE*
		N = 10		
α	–1.	–1.04213	0.328788	0.331476
β	0.1	0.10861	0.0377964	0.0387645
σ	2.	2.04636	0.388395	0.391153
		N = 50		
α	–1.	–0.981368	0.360238	0.36072
β	0.1	0.102277	0.0416743	0.0417364
σ	2.	1.93997	0.361415	0.366367
		N = 100		
α	–1.	–0.820112	0.312953	0.36097
β	0.1	0.0872694	0.0371133	0.039236
σ	2.	1.85711	0.345669	0.374037

good approximation only for small sampling intervals Δt_i (see Section 3.4). On the other hand, the simulation of the likelihood requires much more computation time (see Table 5), but for large sampling intervals the improvements in bias and RMSE are considerable. Computation of the importance density by smoothing requires higher computational costs but the results for $N = 10$ are better than the corresponding estimates without importance sampling ($N = 100$). The computations were performed on a PowerPC 604e with 180 MHz using MPW C and Mathematica with the MathLink communication library.

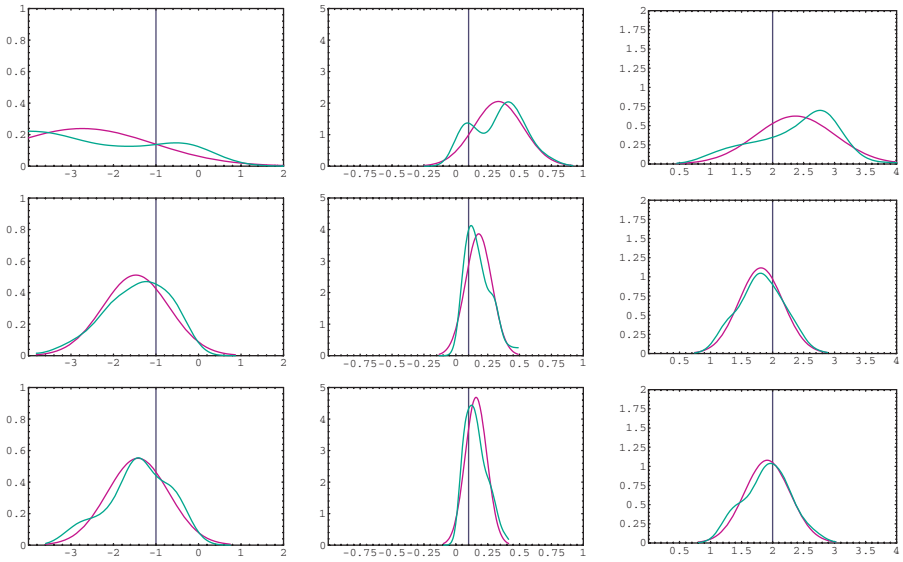

Figure 6. Kernel density method. Empirical distribution of $M = 100$ ML estimates of $\{\alpha, \beta, \sigma\} = \{ -1., 0.1, 2. \}$ (from left). Sample size $N = 10, 50, 100$ (from above).

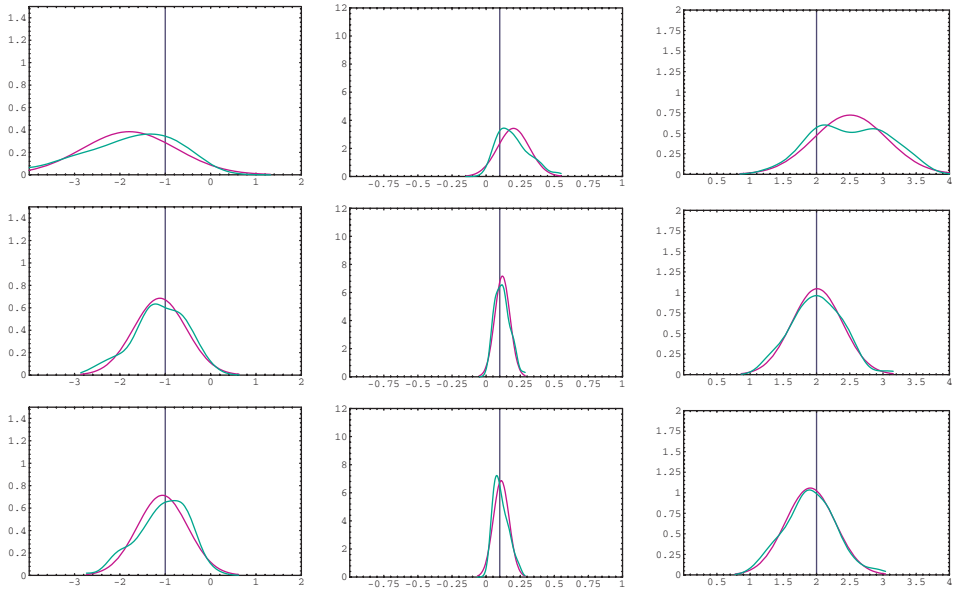

Figure 7. Functional integral method. Empirical distribution of $M = 100$ ML estimates of $\{\alpha, \beta, \sigma\} = \{ -1., 0.1, 2. \}$ (from left). Sample size $N = 10, 50, 100$ (from above).

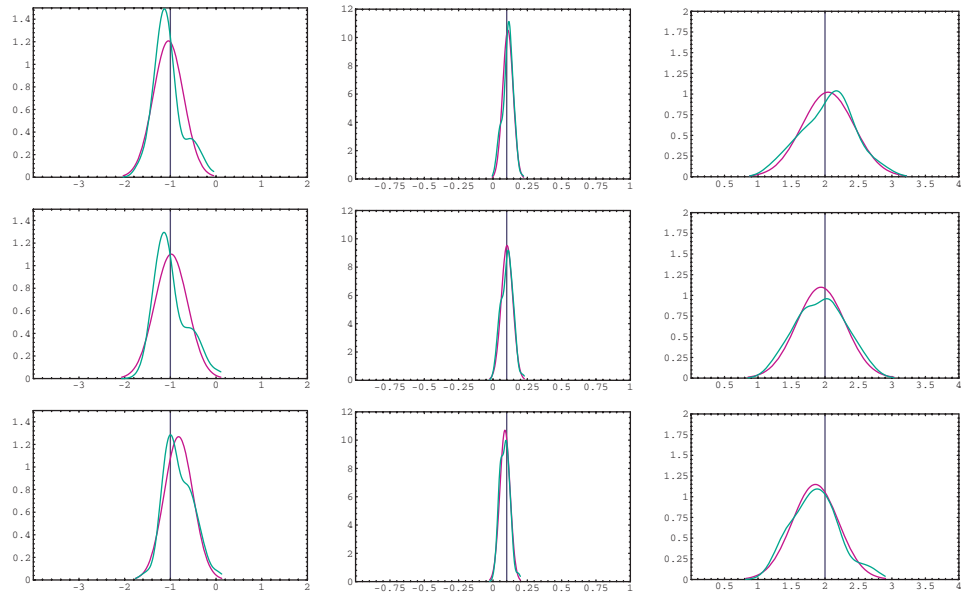

Figure 8. Functional integral method with importance sampling. Empirical distribution of $M = 100$ ML estimates of $\{\alpha, \beta, \sigma\} = \{ -1., 0.1, 2. \}$ (from left). Sample size $N = 10, 50, 100$ (from above).

Table 5. Comparision of CPU Times (seconds) for 10 Likelihood Evaluations and Varying Monte Carlo sample size *N* (Ginzburg–Landau model).

Method		*N* = *10*	*50*	*100*
EKF	1.76	–	–	–
SNF	1.95	–	–	–
FIF		1.08	1.70	2.63
FIF/IS		6.56	58.85	131.27

5. SIMULATION STUDY 2: GENERALIZED COX-INGERSOLL-ROSS MODEL

The Cox–Ingersoll–Ross (CIR) model ($\alpha = 1$) and its generalization (GCIR)

$$dr(t) \quad = \quad \mu[r(t) - r_0]dt + \sigma r(t)^{\alpha/2} dW(t), \tag{5.1}$$

were designed to model interest rates $r(t)$ which fluctuate around an equilibrium value r_0 (Cox, Ingersoll, and Ross 1985). The solution is always positive but only known explicitly for $\alpha = 1$ (square root process; see, e.g., Overbeck and Rydén 1997) or $r_0 = 0$ (CEV model; see Section 3). In the present context it serves as an example for a nonlinear diffusion coefficient. The model was estimated on the weekly three months to maturity interest rate series as published by the Federal Reserve Bank (H.15 data, series tbsm3m) from 01/08/1954 to 12/31/1999 (see Figure 9). Similar data have been estimated by the so called efficient method of moments (EMM) using the GCIR model (Andersen and Lund 1997). The ML estimates $\hat{\psi} = \{\hat{\mu}, \hat{r_0}, \hat{\alpha}, \hat{h} = \log \hat{\sigma}^2\} = \{-0.128, 6.15, 1.43, -1.91\}$ obtained by using the EKF were taken as true values and the estimation methods quasi-ML (see

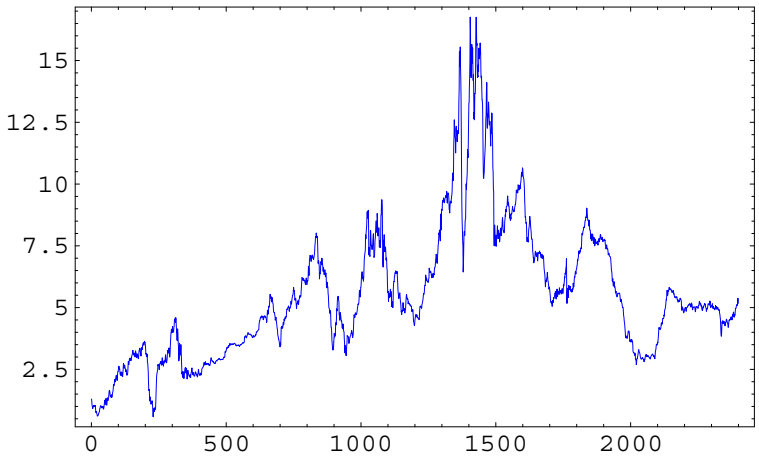

Figure 9. Weekly three months to maturity interest rate series tbsm3m (source: Federal Reserve Bank).

Table 6. Simulation Study Using the GCIR Model and Simulated Interest Rate Data ($M = 100$ replications; see text).

True values	*Mean*	*STD*	*RMSE*
	MQLE		
–0.128	–0.140027	0.0854657	0.0863078
6.15	6.62231	2.42761	2.47312
1.43	1.62785	0.313982	0.371122
–1.91	−2.26715	0.593798	0.69293
	EKF		
–0.128	–0.191776	0.0918065	0.111784
6.15	6.10848	2.17473	2.17513
1.43	1.43853	0.0726564	0.0731559
–1.91	–1.92337	0.116938	0.1177
	SNF		
–0.128	–0.191776	0.0918113	0.111789
6.15	6.10869	2.17541	2.17581
1.43	1.43863	0.072659	0.0731699
–1.91	–1.92369	0.116973	0.117771
	FIF/N = 100		
–0.128	–0.199283	0.0983863	0.121495
6.15	5.78519	1.95189	1.98569
1.43	1.43861	0.0813057	0.0817604
–1.91	–2.09144	0.129997	0.223206
	FIF/N = 200		
–0.128	–0.189293	0.09179	0.110373
6.15	6.20549	2.17114	2.17185
1.43	1.42951	0.0829616	0.082963
–1.91	–2.02065	0.130989	0.171466
	FIF/Importance Sampling/N = 10		
–0.128	–0.190444	0.0920237	0.11121
6.15	6.26824	2.38093	2.38387
1.43	1.41205	0.0718943	0.0741003
–1.91	–1.88408	0.116619	0.119465
	FIF/Importance Sampling/N = 50		
–0.128	–0.193136	0.0933753	0.113849
6.15	6.04289	2.57447	2.5767
1.43	1.42757	0.071608	0.0716494
–1.91	–1.90635	0.115705	0.115763
	FIF/Importance Sampling/N = 100		
–0.128	-0.192599	0.0918176	0.112265
6.15	6.02299	2.148	2.15175
1.43	1.44306	0.072509	0.0736752
−1.91	−1.92798	0.11644	0.117821

below), EKF, SNF, FIF and FIF/IS (with importance sampling) were tested on $M = 100$ replications. The estimation and simulations were based on a discretization scheme with $\Delta t = 1/52$ (weekly sampling interval) and $\delta t = 1/(52 * 7)$ (daily discretization interval). Since the drift function is linear, the first moment equation $\dot{y}(t|t_i) = f(y(t|t_i), t)$, $y(t|t_i) := E[r(t)|r(t_i)]$ is exact, but the variance equation $\dot{P}(t|t_i) = f_y(y(t|t_i), t)P(t|t_i) + P(t|t_i)f'_y(y(t|t_i), t) + \Omega(y(t|t_i), t)$ is approximate except for $\alpha = 1$ (CIR model). In this case $E[\Omega(r(t))|r(t_i)] = E[\sigma^2 r(t)|r(t_i)] = \sigma^2 y(t|t_i)$ (see Section 3.1). It can be shown that the score function of the EKF is equivalent to the estimating equations derived by Wefelmeyer (1996), if fourth moments are factorized by the Gaussian assumption $E[z^4] = 3E[z^2]^2$. These equations combine first and second conditional moments and are called the extended quasi-likelihood model. The resulting efficient estimator is the extendend maximum quasi-likelihood estimator (EMQLE). From these theoretical results the EKF type estimators should perform well when first- and second-order moments are sufficient to describe likelihood and score. A simpler MQLE using only the first conditional moment was computed as well (martingale estimator, see Bibby and Sorensen 1995; Overbeck and Rydén 1997). In scalar notation without measurement model the estimating equations are of the form

$$0 = \sum_{i=0}^{T-1} w_i(y_{i+1} - y(t_{i+1}|t_i)), \tag{5.2}$$

where the optimal weights are given by $w_i = y_\psi(t_{i+1}|t_i)/P(t_{i+1}|t_i)$ and $y_{i+1} := r(t_{i+1})$. The resulting MQL estimates of ψ can be interpreted as conditional least squares estimates using the prediction error (see Bibby and Sorensen 1995; Klimko and Nelson 1978). It may seem that Equation (5.2) uses the information in the second conditional moment $P(t_{i+1}|t_i)$, but as shown by Wefelmeyer (1996), the estimator does not use P, although it is contained in the definition. A second term using the squared prediction error $\nu_i := y_{i+1} - y(t_{i+1}|t_i)$ yields the combined estimation equations

$$0 = \sum_{i=0}^{T-1} w_{1i}\nu_i + w_{2i}[\nu_i^2 - P(t_{i+1}|t_i)], \tag{5.3}$$

and the resulting EQMLE is equivalent to the EKF estimates.

The results of the simulation study are presented in Table 6. It is somewhat surprising, that the EKF and SNF perform very well in terms of the RMSE, but the functional integral approach with $N = 200$ has slight advantages in the estimation of the parameter μ (mean reversion). The simple QMLE (5.2) using only the first conditional moment is outperformed by the ML methods considering also second moments (EKF) and by simulated ML. Finally, the functional integral approach with importance sampling ($N = 100$) is slightly better in relation to EKF and FIF in terms of RMSE for the mean level r_0.

6. CONCLUSION

We compared several estimation methods for the structural parameters of a sampled stochastic differential equation system. In a first simulation study using a nonlinear drift

in a double well potential, linearization methods based on the extended Kalman filter and on Itô's formula are shown to be time saving, but the estimates exhibit strong bias for large sampling intervals. The simulation based likelihood methods are preferable for large sampling intervals, but require more CPU time.

In a second simulation study adressing nonlinear diffusion coefficients, but linear drift, the EKF type estimators perform as well as the computationally more demanding Monte Carlo methods with simulated likelihood. The advantages of the latter methods may be more pronounced if the sampling interval (weekly) is larger or if stronger nonlinearities are considered. The simulation parameters were determined by a real dataset, however.

Although the simulation studies assume uniform sampling intervals, the software implementation also allows irregular sampling and missing values in some components of the measurements.

In Miller, Ghil, and Gauthiez (1994), using a double well potential, it was demonstrated that the inclusion of measurement noise leads to a degradation of performance of the EKF in tracking the system trajectories. This will also diminish the quality of the parameter estimates which are based on the prediction error decomposition and may lead to a further advantage of the simulation methods.

Present investigations include a measurement model for the kernel density and functional integral method and simulation experiments for multivariate systems such as the Lorenz model.

ACKNOWLEDGMENTS

The author thanks an associate editor and two referees for their thorough reading of the manuscript and helpful suggestions.

[Received January 2001. Revised July 2001.]

REFERENCES

Ames, W. F. (1992), *Numerical Methods for Partial Differential Equations* (3rd ed.), Boston: Academic Press.

Andersen, T. G., and Lund, J. (1997), "Estimating Continuous-Time Stochastic Volatility Models of the Short-Term Interest Rate," *Journal of Econometrics*, 77, 343–377.

Anderson, B. D. O., and Moore, J. B. (1979), *Optimal Filtering*, Englewood Cliffs, NJ: Prentice Hall.

Arnold, L. (1974), *Stochastic Differential Equations*, New York: Wiley.

Arnold, V. I. (1973), *Ordinary Differential Equations*, Cambridge, MA: MIT Press.

——— (1986), *Catastrophy Theory*, Berlin: Springer.

Basawa, I. V., and Prakasa Rao, B. L. S. (1980), *Statistical Inference for Stochastic Processes*, London: Academic Press.

Bergstrom, A. R. (1990), *Continuous Time Econometric Modelling*, Oxford: Oxford University Press.

Bibby, M., and Sorensen, M. (1995), "Martingale Estimation Functions for Discretely Observed Diffusion Processes," *Bernoulli*, 1, 1–39.

Brockwell, P. J., and Hyndman, R. J. (1992), "On Continuous Time Threshold Autoregression," *International Journal of Forecasting*, 8, 157–173.

Campillo, F., and Le Gland, F. (1989), "MLE for Partially Observed Diffusions: Direct Maximization vs. the EM Algorithm," *Stochastic Processes and their Applications*, 33, 245–274.

Cox, J. C., Ingersoll, J. E., and Ross, S. A. (1985), "A Theory of the Term Structure of Interest Rates," *Econometrica*, 53, 385–407.

Cox, J. C., and Ross, S. A. (1976), "The Valuation of Options for Alternative Stochastic Processes," *Journal of Financial Economics*, 3, 145–166.

Dennis, J. E., Jr., and Schnabel, R.B. (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Englewood Cliffs, NJ: Prentice Hall.

Elerian, O., Chib, S., and Shephard, N. (2001), "Likelihood Inference for Discretely Observed Nonlinear Diffusions," *Econometrica*, 69, 959–993.

Fahrmeir, L. (1976), "Approximation von Stochastischen Differentialgleichungen auf Digital- und Hybridrechnern," *Computing*, 16, 359–371.

Feichtinger, G. (1992), "Nonlinear Threshold Dynamics: Further Examples for Chaos in Social Sciences," in *Economic Evolution and Demographic Change*, eds. G. Haag, U. Mueller, and K. G. Troitzsch, Berlin: Springer, pp. 141–154.

Feller, W. (1951), "Two Singular Diffusion Problems," *Annals of Mathematics*, 54, 173–182.

Florens-Zmirou, D. (1989), "Approximate Discrete-Time Schemes for Statistics of Diffusion Processes," *Statistics*, 20, 547–557.

Gallant, R. A., and Long, J. R. (1997), "Estimating Stochastic Differential Equations Efficiently by Minimum Chi-Squared," *Biometrika*, 84, 125–141.

Genon-Catalot, V. (1990), "Maximum Contrast Estimation for Diffusion Processes from Discrete Observations," *Statistics*, 21, 99–116.

Gordon, N. J., Salmond, D. J., and Smith, A. F. M. (1993), "Novel Approach to Nonlinear/Non-Gaussian Bayesian State Estimation," *IEEE Transactions on Radar and Signal Procesing*, 140, 107–113.

Haken, H. (1977), *Synergetics*, Berlin: Springer.

Hansen, L. P., and Scheinkman, J. A. (1995), "Back to the Future: Generating Moment Implications for Continuous-Time Markov Processes," *Econometrica*, 63, 767–804.

Holmes, Ph. J. (1981), "Center Manifolds, Normal Forms and Bifurcations of Vector Fields," *Physica D*, 2, 449–481.

Hürzeler, M., and Künsch, H. (1998), "Monte Carlo Approximations for General State-Space Models," *Journal of Computational and Graphical Statistics*, 7, 175–193.

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

Kitagawa, G. (1987), "Non-Gaussian State Space Modeling of Nonstationary Time Series," *Journal of the American Statistical Association*, 82, 1032–1063.

——— (1996), "Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models," *Journal of Computational and Graphical Statistics*, 5, 1–25.

Klimko, L. A., and Nelson, P. I. (1978), "On Conditional Least Squares Estimation For Stochastic Processes," *The Annals of Statistics*, 6, 629–642.

Kloeden, P. E., and Platen, E. (1992), *Numerical Solution of Stochastic Differential Equations*, Berlin: Springer.

Komaki, F. (1993), "State-Space Modelling of Time Series Sampled From Continuous Processes with Pulses," *Biometrika*, 80, 417–429.

Lighthill, M. J. (1958), *Introduction to Fourier Analysis and Generalised Functions*, Cambridge, UK: Cambridge University Press.

Ljung, L., and Söderström, T. (1983), *Theory and Practice of Recursive Identification*, Cambridge, MA: MIT Press.

Lo, A. W. (1988), "Maximum Likelihood Estimation of Generalized Itô Processes with Discretely Sampled Data," *Econometric Theory*, 4, 231–247.

Miller, R. N., Ghil, M., and Gauthiez, F. (1994), "Advanced Data Assimilation in Strongly Nonlinear Dynamical

Systems," *Journal of the Atmospheric Sciences*, 51, 1037–1056.

Overbeck, L., and Rydén, T. (1997), "Estimation in the Cox-Ingersoll-Ross Model," *Econometric Theory*, 13, 430–461.

Ozaki, T. (1985), "Nonlinear Time Series and Dynamical Systems," in *Handbook of Statistics*, ed. E. J. Hannan, Amsterdam: North Holland, pp. 25–83.

Pedersen, A. R. (1995), "A New Approach to Maximum Likelihood Estimation for Stochastic Differential Equations Based on Discrete Observations," *Scandinavian Journal of Statistics*, 22, 55–71.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., and Flannery, B. P. (1992), *Numerical Recipes in C* (2nd ed.), Cambridge: Cambridge University Press.

Puu, T. (1992), "A Chaotic Process with Slow Feed Back: The Case of Business Cycles," in *Economic Evolution and Demographic Change*, eds. G. Haag, U. Mueller, and K.G. Troitzsch, New York: Springer, pp. 3–33.

Risken, H. (1989), *The Fokker-Planck Equation* (2nd ed.), Berlin: Springer.

Rümelin, W. (1982), "Numerical Treatment of Stochastic Differential Equations," *SIAM Journal of Numerical Analysis*, 19, 604–613.

Schuster, H.G. (1984), *Deterministic Chaos*, Weinheim: Physik-Verlag.

Schweppe, F. (1965), "Evaluation of Likelihood Functions for Gaussian Signals," *IEEE Transactions on Information Theory*, 11, 61–70.

Shoji, I., and Ozaki, T. (1997), "Comparative Study of Estimation Methods for Continuous Time Stochastic Processes," *Journal of Time Series Analysis*, 18, 485–506.

——— (1998), "A Statistical Method of Estimation and Simulation for Systems of Stochastic Differential Equations," *Biometrika*, 85, 240–243.

Silverman, B. W. (1986), *Density Estimation for Statistics and Data Analysis*, London: Chapman and Hall.

Singer, H. (1992), "Dynamic Structural Equations in Discrete and Continuous Time," in *Economic Evolution and Demographic Change*, eds. G. Haag, U. Mueller, and K. G. Troitzsch, Berlin: Springer, pp. 306–320.

——— (1995), "Analytical Score Function for Irregularly Sampled Continuous Time Stochastic Processes with Control Variables and Missing Values," *Econometric Theory*, 11, 721–735.

——— (2000), "Variance Reduced Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models," in *Proceedings of the Fifth International Conference on Social Science Methodology (RC33, ISA and GESIS)*, Zentralarchiv für Empirische Sozialforschung, Universität Köln.

Stratonovich, R.L. (1989), "Some Markov Methods in the Theory of Stochastic Processes in Nonlinear Dynamic Systems," in *Noise in Nonlinear Dynamic Systems*, eds. F. Moss and P. McClintock, Cambridge University Press, pp. 16–71.

Tanizaki, H., and Mariano, R.S. (1995), "Prediction, Filtering and Smoothing in Nonlinear and Non-normal Cases using Monte-Carlo Integration," in *Econometric Inference using Simulation Techniques*, eds. H.K. Van Dijk, A. Monfort, and B.W. Brown, New York: Wiley, pp. 245–261.

Wefelmeyer, W. (1996), "Quasi-likelihood Models and Optimal Inference," *The Annals of Statistics*, 24, 405–422.

Weidlich, W., and Haag, G. (1983), *Quantitative Sociology*, Berlin: Springer.

Wong, E., and Hajek, B. (1985), *Stochastic Processes in Engineering Systems*, New York: Springer.

Wymer, C. (1995), "Advances in the Estimation and Analysis of Nonlinear Differential Equation Models in Economics," in *Methods and Applications of Economic Dynamics*, eds. L. Schoonbek, E. Sterken, and S.K. Kuipers, Amsterdam: Elsevier Science.

Zhang, W. B. (1991), *Synergetic Economics*, Berlin: Springer.