A SURVEY OF ESTIMATION METHODS FOR STOCHASTIC DIFFERENTIAL EQUATIONS

Hermann Singer
FernUniversität in Hagen *

This article gives a short review of key issues and of existing estimation methods in differential equation modeling. It covers both linear and nonlinear stochastic differential equations (SDE) and continuous-discrete statespace models including panel data.

Key Words: Ito calculus; sampling; exact discrete model; transition densities; Extended Kalman filter; Nonlinear filtering

1 INTRODUCTION

Differential equations are the continuous time analog of time series models, i.e. the state vector $Y(t)$ is a function of the real parameter t (time) and the desired time function is given in terms of time derivatives. For example, the simple **growth model**

$$dY(t)/dt = \lambda Y(t) \tag{1}$$

states, that the time change of $Y(t)$ in the interval $[t, t + dt]$ is proportional to the state at this time point. Examples are the growth of populations or the attenuation of radioactive rays in media ($\lambda < 0$). The solution of (1) with initial condition $Y(t_0)$ is given by

$$Y(t) = \exp[\lambda(t - t_0)]Y(t_0) \tag{2}$$

which can be verified by computing the derivative. In a social science context, the simple deterministic equation must be extended by stochastic equation errors which model neglected variables and misspecifications in the functional form of the differential equation. A linear specification is given by the **stochastic differential equation** (SDE)

$$dY(t)/dt = \lambda Y(t) + g\zeta(t) \tag{3}$$

*Lehrstuhl für angewandte Statistik und Methoden der empirischen Sozialforschung, D-58084 Hagen, Germany, Hermann.Singer@FernUni-Hagen.de

where $\zeta(t)$ is a zero mean **Gaussian white noise** process with autocorrelation $\gamma(t-s) = E[\zeta(t)\zeta(s)] = \delta(t-s)$ (Dirac delta function). This means, that the noise process is only autocorrelated for very short time spans. The solution of (3) is given by

$$Y(t) = \exp[\lambda(t-t_0)]Y(t_0) + \int_{t_0}^{t} \exp(\lambda(t-s))g\zeta(s)ds. \tag{4}$$

Since the continuous time white noise process is a generalized function (cf. Arnold, 1973, ch. 3), the solution is usually written by the replacement $\zeta(s)ds = dW(s)$, where $W(s)$ is the Wiener process, a continuous, but not differentiable process (cf. Arnold, ch. 4). Thus, in order to avoid derivatives of nondifferentiable processes, one writes

$$\begin{aligned} dY(t) &= \lambda Y(t)dt + gdW(t) & (5)\\ Y(t) &= \exp[\lambda(t-t_0)]Y(t_0) + \int_{t_0}^{t} \exp[\lambda(t-s)]gdW(s). & (6) \end{aligned}$$

and interpretes the first equation as an integral equation. This is the so called **Itô calculus**, which is a well defined method for the treatment of stochastic differential equations (cf. Arnold, 1973). From the explicit solution (6) it is seen that the solution process is Gaussian if the initial condition is Gaussian or constant as well.
In the context of parameter estimation it is helpful to write the solution as a time series (**exact discrete model**)

$$Y_{i+1} = \exp[\lambda(t_{i+1}-t_i)]Y_i + \int_{t_i}^{t_{i+1}} \exp[\lambda(t_{i+1}-s)]gdW(s), \tag{7}$$

$Y_i = Y(t_i)$, or more concise as

$$Y_{i+1} = \Phi(t_{i+1}, t_i)Y_i + u_i, \tag{8}$$

where Φ is the fundamental matrix, but it should be noted that the parameters of the AR(1) model are highly nonlinearly restricted (e.g. $\mathrm{Var}(u_i) = \int \Phi(t_{i+1}, s)^2 g^2 ds$).
This is the main problem for the task of parameter estimation. Software must be able to implement the required restrictions, especially in the multivariate case where time ordered matrix exponentials are involved (some references are Phillips, 1976, Hamerle et al., 1991, 1993, Singer, 1998).

2 LINEAR CONTINUOUS/DISCRETE STATE SPACE MODELS

In order to incorporate higher order derivatives and errors of measurement, the **general linear continuous/discrete state space model**

$$\begin{aligned} dY(t) &= [A(t,\psi)Y(t) + b(t,\psi)]dt + G(t,\psi)dW(t) & (9)\\ Z_i &= H(t_i,\psi)Y(t_i) + d(t_i,\psi) + \epsilon_i & (10) \end{aligned}$$

is introduced. In the state equation (9), $W(t)$ denotes a r-dimensional **Wiener process** and the state is described by the p-dimensional state vector $Y(t)$. It fulfils a system of stochastic differential equations in the sense of Itô (cf. Arnold, 1974) The measurement error $\epsilon_i \sim N(0, R(t_i, \psi)$ is discrete time white noise.
Parametric estimation is based on the u-dimensional parameter vector ψ and the time dependence t also incorporates deterministic regressor variables $x(t)$. Moreover, panel data may be treated by joining a panel index $n, n = 1, ..., N$ (cf. Singer, 1998).
The state equation (9) may be solved for the times of measurement t_i and the likelihood function can be computed recursively by means of the **Kalman filter algorithm** (Liptser and Shiryaev, 1977, 2001, Harvey and Stock, 1985, Singer, 1998). The computation proceeds in steps of time updates and measurement updates involving the conditional moments $\mu(t|t_i) = E[Y(t)|Z^i]$ and $\Sigma(t|t_i) = \mathrm{Var}[Y(t)|Z^i]$, where $Z^i = \{Z_i, ..., Z_0\}$ are the measurements up to time t_i. The time updates fulfil ($t \in [t_i, t_{i+1}]$)

$$
\begin{aligned}
(d/dt)\mu(t|t_i) &= A(t,\psi)\mu(t|t_i) + b(t,\psi) && (11)\\
(d/dt)\Sigma(t|t_i) &= A(t,\psi)\Sigma(t|t_i) + \Sigma(t|t_i)A'(t,\psi) + \Omega(t,\psi) && (12)
\end{aligned}
$$

where $\Omega(t,\psi) = GG'(t,\psi)$ is the diffusion matrix. At the measurement times, the time update (optimal prediction) is corrected by the measurement Z_{i+1} using the Bayes formula

$$
\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + K(t_{i+1}|t_i)\nu(t_{i+1}|t_i) && (13)\\
\Sigma(t_{i+1}|t_{i+1}) &= [I - K(t_{i+1}|t_i)H(t_{i+1},\psi)]\Sigma(t_{i+1}|t_i) && (14)\\
\nu(t_{i+1}|t_i) &= Z_{i+1} - Z(t_{i+1}|t_i) && (15)\\
Z(t_{i+1}|t_i) &= H(t_{i+1},\psi)\mu(t_{i+1}|t_i) + d(t_{i+1},\psi) && (16)\\
\Gamma(t_{i+1}|t_i) &= H(t_{i+1},\psi)\Sigma(t_{i+1}|t_i)H'(t_{i+1},\psi) + R(t_{i+1},\psi) && (17)\\
K(t_{i+1}|t_i) &:= \Sigma(t_{i+1}|t_i)H'(t_{i+1},\psi)\Gamma(t_{i+1}|t_i)^{-1} && (18)
\end{aligned}
$$

where $K(t_{i+1}|t_i)$ is the **Kalman gain**, $Z(t_{i+1}|t_i) = H(t_{i+1},\psi)\mu(t_{i+1}|t_i) + d(t_{i+1},\psi)$ is the optimal predictor of the measurement Z_{i+1}, $\nu(t_{i+1}|t_i)$ is the **prediction error** and $\Gamma(t_{i+1}|t_i)$ is the **prediction error covariance matrix**. Fortunately, the updated state $Y(t_{i+1})|Z^{i+1}$ is again conditionally Gaussian and one can proceed with two conditional moments. After T steps one obtains the likelihood

$$
l(\psi; Z) = \log p(Z_T, \ldots, Z_0; \psi) = \sum_{i=0}^{T-1} \log p(Z_{i+1}|Z^i; \psi)p(Z_0), \qquad (19)
$$

where the transition densities are given in terms of the Gauss distributions

$$
p(Z_{i+1}|Z^i; \psi) = \phi(\nu(t_{i+1}|t_i); 0, \Gamma(t_{i+1}|t_i)). \qquad (20)
$$

Thus the Kalman filter computes the likelihood recursively in terms of predictions, prediction errors and their conditional variance (cf. Jazwinski, 1970).
It should be noted that the system matrices may also depend on earlier measurements (i.e. $A(t) = A(t, Z^i), t_i \leq t, H(t_i) = H(t_i, Z^{i-1})$ etc., but the distributions in the filter are still conditionally Gaussian if $Y(t_0)|Z(t_0)$ is such (cf. Liptser and Shiryaev, 2001, ch.

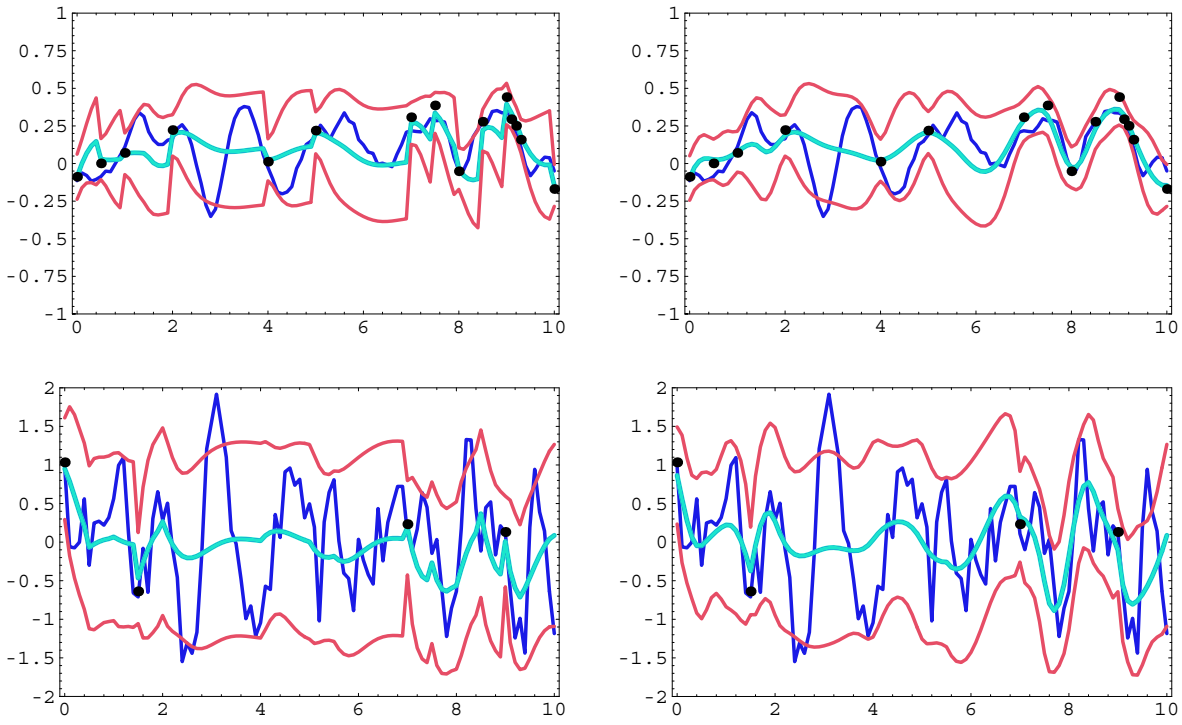

Figure 1: Linear oszillator with irregulary measured states: Filtered state (left), smoothed state (right). Measurements at $\tau_1 = \{0, .5, 1, 2, 4, 5, 7, 7.5, 8, 8.5, 9, 9.1, 9.2, 9.3, 10\}$, $\tau_2 = \{0, 1.5, 7, 9\}$.

13). Thus, although the latent states may be not Gaussian due to the Z-dependence, the conditionally Gaussian filtering scheme is still valid. For example, the diffusion matrix $G(t)$ may contain delayed prediction errors $G(t, \nu(t_i|t_{i-1}), ...)$ and thus ARCH and GARCH effects in a continuous time model.

The state space model is very general and allows the specification of models with higher order derivatives where only the first component is measured at sampling times t_i. It should be noted that no approximation whatsoever is involved. For example, the CAR(2) process $\ddot{y} + \gamma\dot{y} + \omega_0^2 y = g\zeta(t)$ has state space representation

$$d\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} := \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}dt + \begin{bmatrix} 0 & 0 \\ 0 & g \end{bmatrix}d\begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix} \tag{21}$$

$$z_i := \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \end{bmatrix} + \epsilon_i \tag{22}$$

and there is **no need to approximate differentials by finite differences** etc. The unobservable derivative $y_2 = \dot{y}$ is reconstructed by the filter as $E[y_2(t)|Z^T]$ (smoothed, filtered or predicted state depending on time t) (Singer, 1993). Approximating the derivatives by finite differences introduces unnecessary specification error and biased estimates. Figure

1 shows the true, filtered and smoothed trajectories of a irregularly sampled oscillator (cf. Singer, 1995) and 95% HPD confidence intervals.
At present it may be stated that the **linear state space model is completely solved and may be estimated efficiently by the Kalman filter algorithm**. This works with only one trajectory and/or with panel data. One only has to sum up the N likelihood contributions. References are Jones and Ackerson (1990), Singer (1995, 1998).
It may be the case that some nonlinear SEM software is able to incorporate (in some special cases) the nonlinear restrictions as created by the moment equations (11,12)(cf. Oud and Jansen, 2000), but in general special software with direct implementation of the Kalman filter (KF) is preferrable. The KF is the recursive, most direct and efficient implementation of the continuous/discrete state space model.
For example, the time dependent model of growth rates with constant part

$$A_0 = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \tag{23}$$

and a linearly changing part

$$A_1(t) = \begin{bmatrix} \alpha\, t & 0 \\ 0 & \beta\, t \end{bmatrix}. \tag{24}$$

yields a drift matrix $A(t) = A_0 + A_1(t)$. Since $A(t)$ does not commute with $A(s)$, i.e. $A(t)A(s) - A(s)A(t) \neq 0$, the state transition matrix fulfilling the equation

$$(d/dt)\Phi(t|t_i) \;=\; A(t,\psi)\Phi(t|t_i) \tag{25}$$

is not simply given by

$$\Phi(t|t_i) \;=\; \exp[\int_{t_i}^{t} A(s,\psi)ds] \tag{26}$$

but by the time ordered expression

$$\Phi(t|t_i) \;=\; \overleftarrow{T}\exp[\int_{t_i}^{t} A(s,\psi)ds] \tag{27}$$

where $\overleftarrow{T}\, A(t)A(s) = A(s)A(t); t < s$ is the Wick time ordering operator. Using the moment equations (11,12) with a respective numerical approximation (Euler- or higher order Runge-Kutta scheme) does automatically produce the time ordered expressions. This is the crucial step each candidate software package must take in order to implement the correct parameter functionals (for details see Singer, 1998).

Other issues

In economics many variables (such as gross national product) are cumulated or averaged over certain period of time. These so called **flow data** can be naturally modelled by differential equations if additional derivatives are included in the state space model which generate

integrated measurements (cf. Bergstrom, 1984, Singer, 1995). Then, the dynamics are between stocks and latent differentiated flows, but only integrated measurements are needed. Often, **missing data** occur in panel designs or data are measured at **arbitrary frequencies**. Sometimes, on may combine different time series collected at different frequencies (daily, quarterly etc.). The continuous-discrete state space model can handle all these cases easily, since the system model proceeds in continuous time but is only measured at certain irregular times t_i. Even different sampling intervals of the panel units and/or exogenous variables are possible. Thus, in principle, one could collect data at arbitrary times (no panel waves). Missing data can be treated by the Kalman filter measurement update, if missing components are canceled in the respective matrices. In fact, the discrete time measurement of a continuous time model may be viewed as a missing data problem and all nonmeasured states are reconstructed by the filter.
Individual specific random effects π_n can be treated in the state space model by writing $(n = 1, \ldots, N)$

$$\begin{aligned} d\begin{bmatrix} y_n \\ \pi_n \end{bmatrix} &= \begin{bmatrix} A & I \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_n \\ \pi_n \end{bmatrix}dt + \begin{bmatrix} G \\ 0 \end{bmatrix}dW_n(t) \\ z_{ni} &= [\,H \quad 0\,]\begin{bmatrix} y_{ni} \\ \pi_{ni} \end{bmatrix} + \epsilon_{ni}, \end{aligned} \tag{28}$$

From this one gets estimates of the covariance matrix $\mathrm{Var}(\pi_n)$ of the several components and is able to specify correlations between states and random effects $\mathrm{Cov}(y_n(t_0), \pi_n)$.

3 NONLINEAR DIFFERENTIAL EQUATIONS AND STATE SPACE MODELS

Whereas the linear case seems to be solved and implemented completely (but is not frequently applied), there are many issues and competing approaches in the nonlinear field. It is presently an area of very active research due to the growing interest in finance models. The option price model of Black and Scholes relies on a SDE model for the underlying stock variable and Merton's monograph (1990) on continuous finance has been given the field a strong 'continuous' flavor. This is in contrast to econometrics where still times series method dominate and also sociology despite the old tradition of Coleman and others.
Here we have the general system

$$dY(t) = f(Y(t), t, \psi)dt + g(Y(t), t, \psi)dW(t) \tag{29}$$

$$Z_i = h(Y(t_i), t_i, \psi) + \epsilon_i. \tag{30}$$

A strong simplification occurs when the state is completely measured at times t_i, i.e. $Z_i = Y_i = Y(t_i)$. Then, only the transition probability $p(y_{i+1}, t_{i+1}|y_i, t_i)$ must be computed in order to obtain the likelihood function

$$l(\psi; Y) = \log p(Y_T, \ldots, Y_0; \psi) = \sum_{i=0}^{T-1} \log p(Y_{i+1}|Y_i; \psi)p(Y_0). \tag{31}$$

Unfortunately, the transition probability can be computed analytically only in some special cases (including the linear), but in general approximation methods must be employed. Since the transition density fulfils a partial differential equation (PDE), the so called **Fokker-Planck equation**, approximation methods for PDE, e.g. **finite difference methods** can be used (cf. Jensen and Poulsen, 2002). .
A large class of approximations rests on linearization methods which can be applied to the exact moment equations (**extended Kalman Filter EKF**; cf. Jazwinski, 1970 and below) or directly to the nonlinear differential equation using Itô's lemma (**local linearization LL**; Shoji and Ozaki, 1997, 1998). Since the linearity is only approximate in the vicinity of a measurement or reference trajectory, the conditional Gaussian schemes are valid only for short measurement intervals $t_{i+1} - t_i$. Other linearization methods relate to the diffusion term, but are interpretable in terms of the EKF (Nowman, 1997)
Alternatively, the **Monte Carlo method** can be employed to obtain approximate transition densities (Pedersen, 1995, Andersen and Lund, 1997, Elerian et al., 2001, Singer, 2002, 2003). More recently, **Hermite expansions** of the transition density have been utilized by Aït-Sahalia, 2002). In this approach, the expansion coefficients are expressed in terms of conditional moments and computed analytically by using computer algebra programs. The computations comprise the multiple action of the backward operator $L = F^\dagger$ on polynomials. Alternatively, one can use systems of moment differential equations (Singer, 2004). It seems that this approach is most efficient both in accuracy and computing time (cf. Aït-Sahalia, loc. cit., fig. 1, Jensen and Poulsen, 2002).
Nonparametric approaches attempt to estimate the drift function $f()$ and the diffusion function $\Omega()$ without assumptions about a certain functional form. They typically involve kernel density estimates of conditiona densities (cf. Bandi and Phillips, 2001). Other approaches utilize Taylor series expansions of the drift function and estimate the derivatives (expansion coefficients) as latent states using the LL method (similarly to the SNF) (Shoji, 2002).

State space models

The continuous-discrete state space model (29) may be treated approximately by linearized moment equations, if one computes the exact evolution equations (the dependence on ψ is dropped)

$$\begin{aligned} (d/dt)\mu(t|t_i) &= E[f(Y,t)|Z^i] && (32)\\ (d/dt)\Sigma(t|t_i) &= E[f(Y,t)(Y(t)-\mu(t|t_i))|Z^i] + E[\Omega(Y,t)]. && (33) \end{aligned}$$

These are not differential equations, since they contain the conditional density $p(y,t|Z^i)$ which would already contain the complete solution of the filtering problem. Taylor expansion of f up to first order around the conditional mean $\mu(t|t_i) = E[Y(t)|Z^i]$ yields the **continuous-discrete extended Kalman filter EKF**

$$\begin{aligned} (d/dt)\mu(t|t_i) &= f(\mu(t|t_i),t) && (34)\\ (d/dt)\Sigma(t|t_i) &= A(\mu(t|t_i),t)\Sigma(t|t_i) + \Sigma(t|t_i)A'(\mu(t|t_i),t) + \Omega(\mu(t|t_i),t). && (35) \end{aligned}$$

with Jacobian $A(\mu(t|t_i),t) = (\partial f/\partial y)(\mu(t|t_i),t) : p\times p$. In contrast to (32) the EKF equations are a closed system of differential equations to be solved with standard numerical techniques (Runge-Kutta etc). Second order derivatives lead to the so called second order nonlinear filter (SNF).
At measurement time t_{i+1}, the output vector $h()$ may be expanded around the approximate conditional mean $\mu(t_{i+1}|t_i)$ to yield a locally linear measurement equation

$$Z_{i+1} = h(\mu(t_{i+1}|t_i),t_{i+1}) + H(\mu(t_{i+1}|t_i),t_{i+1}) * (Y_{i+1} - \mu(t_{i+1}|t_i)) + \epsilon_{i+1}. \quad (36)$$

with Jacobian $H(\mu(t_{i+1}|t_i),t_{i+1}) = (\partial h/\partial y)(\mu(t_{i+1}|t_i),t_{i+1}) : k \times p$. This permits the usage of the linear measurement update equations (13).
For large measurement intervals or strongly nonlinear systems the conditionally Gaussian approach (EKF, SNF or LL) is not sufficient and other approximation methods can be applied. The computation of the a priori density $p(y_{i+1},t_{i+1}|Z^i)$ requires the solution of the Fokker-Planck equation and the measurement update is the Bayes formula leading to the general nonlinear filtering scheme

time update:

$$\begin{aligned} \frac{\partial p(y,t|Z^i)}{\partial t} &= F(y,t)p(y,t|Z^i)\ ; t \in [t_i,t_{i+1}] \quad (37)\\ p(y,t_i|Z^i) &:= p(y_i|Z^i)\\ p(y,t_{i+1}|Z^i) &:= p(y_{i+1}|Z^i) \end{aligned}$$

measurement update:

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &= \frac{p(z_{i+1}|y_{i+1},Z^i)p(y_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} \quad (38)\\ &:= p_{i+1|i+1}\\ p(z_{i+1}|Z^i) &= \int p(z_{i+1}|y_{i+1},Z^i)p(y_{i+1}|Z^i)dy_{i+1}, \quad (39) \end{aligned}$$

$i = 0,\ldots,T-1$, where F is the Fokker-Planck operator

$$\begin{aligned} F(y,t)p(y,t|x,s) &= -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t)p(y,t|x,s)]\\ &\quad +\tfrac{1}{2}\sum_{ij}\frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t)p(y,t|x,s)]. \quad (40) \end{aligned}$$

The filtering scheme yields a recursive computation of the likelihood function

$$l(\psi;Z) = \sum_{i=0}^{T-1} \log p(z_{i+1}|Z^i;\psi) + \log p(z_0;\psi). \quad (41)$$

Simulation based filtering methods in discrete time have been used such as Markov chain Monte Carlo (MCMC; Carlin et al., 1992, Kim et al., 1998), rejection sampling using density estimators (Tanizaki, 1996, Tanizaki and Mariano, 1995, Hürzeler and Künsch,

1998), importance sampling and antithetic variables (Durbin and Koopman, 1997, 2000) and recursive bootstrap resampling (Gordon et al., 1993, Kitagawa, 1996).
In the continuous-discrete case, the time update can be accomplished by using the Chapman-Kolmogorov equation (Pedersen, 1995, Singer, 2003). Importance sampling is implemented by approximate smoothing in order to reduce the variance of the simulated likelihood function. For this purpose the **Gaussian sum filter** of Alspach and Sorenson (1972) is used. This is an extension of the EKF where several filters run simultaneously and the nongaussian densities are represented by Gaussian mixture distributions. Moreover, **numerical integration** procedures have been used (Kitagawa, 1987).

4 CONCLUSION

Whereas the linear continuous/discrete state space model (including conditionally gaussian models) can be efficiently estimated by the Kalman filter algorithm, the nonlinear case (with and without measurement model) is an active field of research with competing methods (both analytical and numerical). Especially in the multivariate case further research is needed.

References

[1] Y. Aït-Sahalia. Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, 70,1:223–262, 2002.

[2] D.L. Alspach and H.W. Sorenson. Nonlinear Bayesian estimation using Gaussian sum approximations. *IEEE Transactions on Automatic Control*, 17:439–448, 1972.

[3] T.G. Andersen and J. Lund. Estimating continuous-time stochastic volatility models of the short-term interest rate. *Journal of Econometrics*, 77:343–377, 1997.

[4] L. Arnold. *Stochastic Differential Equations*. John Wiley, New York, 1974.

[5] F.M. Bandi and P.C.B Philips. Fully Nonparametric Estimation of Scalar Diffusion Models. Discussion paper 1332, Cowles Foundation, Yale University, 2001.

[6] A.R. Bergstrom, editor. *Statistical Inference in Continuous Time Models*. North Holland, 1976.

[7] A.R. Bergstrom. Continuous time stochastic models and issues of aggregation over time. In Griliches u. Intriligator, editor, *Handbook of Econometrics*. North Holland, 1984.

[8] A.R. Bergstrom. The history of continuous-time econometric models. *Econometric Theory*, 4:365–383, 1988.

[9] B.P. Carlin, N.G. Polson, and D.S. Stoffer. A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association*, 87:493–500, 1992.

[10] J.S. Coleman. The mathematical study of change. In H.M. Blalock and Blalock A.B., editors, *Methodology in Social Research*, pages 428–478. McGraw-Hill, 1968.

[11] J. Durbin and S.J. Koopman. Monte Carlo maximum likelihood estimation for non-Gaussian state space models. *Biometrika*, 84,3:669–684, 1997.

[12] J. Durbin and S.J. Koopman. Time series analysis of non-Gaussian observations based on state space models from both classical and Bayesian perspectives. *Journal of the Royal Statistical Association B*, 62, 1:3–56, 2000.

[13] O. Elerian, S. Chib, and N. Shephard. Likelihood Inference for Discretely Observed Nonlinear Diffusions. *Econometrica*, 69, 4:959–993, 2001.

[14] N.J. Gordon, D.J. Salmond, and A.F.M. Smith. Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEEE Transactions on Radar and Signal Procesing*, 140, 2:107–113, 1993.

[15] A. Hamerle, W. Nagl, and H. Singer. Problems with the estimation of stochastic differential equations using structural equations models. *Journal of Mathematical Sociology*, 16, 3:201–220, 1991.

[16] A. Hamerle, W. Nagl, and H. Singer. Identification and estimation of continuous time dynamic systems with exogenous variables using panel data. *Econometric Theory*, 9:283–295, 1993.

[17] A.C. Harvey and J. Stock. The estimation of higher order continuous time autoregressive models. *Econometric Theory*, 1:97–112, 1985.

[18] A.H. Jazwinski. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.

[19] B. Jensen and R. Poulsen. Transition Densities of Diffusion Processes: Numerical Comparision of Approximation Techniques. *Institutional Investor*, Summer 2002:18–32, 2002.

[20] R.H. Jones and L.M. Ackerson. Serial correlation in unequally spaced longitudinal data. *Biometrika*, 77:721–731, 1990.

[21] S. Kim, N. Shephard, and S. Chib. Stochastic Volatility: Likelihood Inference and Comparision with ARCH Models. *Review of Economic Studies*, 45:361–393, 1998.

[22] G. Kitagawa. Non-Gaussian state space modeling of nonstationary time series. *Journal of the American Statistical Association*, 82:1032 – 1063, 1987.

[23] G. Kitagawa. Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models. *Journal of Computational and Graphical Statistics*, 5,1:1–25, 1996.

[24] R.S. Liptser and A.N. Shiryayev. *Statistics of Random Processes, Volumes I and II*. Springer, New York, Heidelberg, Berlin, second edition, 2001.

[25] R.C. Merton. *Continuous-Time Finance*. Blackwell, Cambridge MA, Oxford UK, 1990.

[26] K.B. Nowman. Gaussian estimation of single-factor continuous time models of the term structure of interest rates. *Journal of Finance*, 52:1695–703, 1997.

[27] J.H.L. Oud and R.A.R.G Jansen. Continuous Time State Space Modeling of Panel Data by Means of SEM. *Psychometrika*, 65, 2000.

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

[29] P.C.B. Phillips. The estimation of linear stochastic differential equations with exogenous variables. In A.R. Bergstrom, editor, *Statistical Inference in Continuous Time Models*, pages 135–173. North Holland, 1976.

[30] P.C.B. Phillips. The structural estimation of a stochastic differential equation system. In A.R. Bergstrom, editor, *Statistical Inference in Continuous Time Models*, pages 97–122. North Holland, 1976b.

[31] I. Shoji. Nonparametric state estimation of diffusion processes. *Biometrika*, 89, 2:451–456, 2002.

[32] I. Shoji and T. Ozaki. Comparative Study of Estimation Methods for Continuous Time Stochastic Processes. *Journal of Time Series Analysis*, 18, 5:485–506, 1997.

[33] I. Shoji and T. Ozaki. A statistical method of estimation and simulation for systems of stochastic differential equations. *Biometrika*, 85, 1:240–243, 1998.

[34] H. Singer. Continuous-time dynamical systems with sampled data, errors of measurement and unobserved components. *Journal of Time Series Analysis*, 14, 5:527–545, 1993b.

[35] H. Singer. Analytical score function for irregularly sampled continuous time stochastic processes with control variables and missing values. *Econometric Theory*, 11:721–735, 1995.

[36] H. Singer. Continuous Panel Models with Time Dependent Parameters. *Journal of Mathematical Sociology*, 23:77–98, 1998.

[37] H. Singer. Parameter Estimation of Nonlinear Stochastic Differential Equations: Simulated Maximum Likelihood vs. Extended Kalman Filter and Itô-Taylor Expansion. *Journal of Computational and Graphical Statistics*, 11,4:972–995, 2002.

[38] H. Singer. Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models: Importance Sampling by Approximate Smoothing. *Computational Statistics*, 18,1:79–106, 2003.

[39] H. Singer. Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models. preprint, FernUniversität in Hagen, 2004.

[40] H. Tanizaki. *Nonlinear filters: estimation and applications*. Springer, Berlin, second edition, 1996.

[41] H. Tanizaki and R.S. Mariano. Prediction, Filtering and Smoothing in Nonlinear and Non-normal Cases using Monte-Carlo Integration. In H.K. Van Dijk, A. Monfort, and B.W. Brown, editors, *Econometric Inference using Simulation Techniques*, pages 245–261. John Wiley, 1995.