Nonlinear Continuous Time Modeling Approaches in Panel Research

Hermann Singer
FernUniversität in Hagen *

August 22, 2006

Abstract

Stochastic differential equations (SDE) are used as dynamical models for cross sectional discrete time measurements (panel data). Thus causal effects are formulated on a fundamental infinitesimal time scale. Cumulated causal effects over the measurement interval can be expressed in terms of the fundamental effects which are independent of the chosen sampling intervals (e.g. weekly, monthly, annually etc.). The nonlinear continuous-discrete filter is the key tool in deriving a recursive sequence of time and measurement updates. Several approximation methods including the extended Kalman filter (EKF), higher order nonlinear filters (HNF), the unscented Kalman filter (UKF), the Gauss-Hermite filter (GHF) and generalizations (GGHF), as well as simulated filters (functional integral filter FIF) are compared.

Key Words: Itô calculus; Sampling; Continuous-discrete state space models; Nonlinear Kalman filtering; Hermite orthogonal expansion; Numerical integration; Monte Carlo simulation

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

1 Introduction

Continuous time models are natural, since time is a continuously flowing quantity without steps. On the other hand, empirical data in the social sciences and economics are mostly available only at certain time points, e.g. daily, weekly, quarterly etc. or at arbitrary times. Only physical quantities such as voltages, pressures, levels of rivers etc. may be measured on a continuous base. Therefore, there has been a tendency to formulate dynamical models in discrete time (times series and panel analysis). Thus, the causal relations are specified between the arbitrary discrete measurement times. BARTLETT (1946) argues as follows

> It will have been apparent that the discrete nature of our observations in many economic and other time series does not reflect any lack of continuity in the underlying series. Thus theoretically it should often prove more fundamental to eliminate this imposed artificiality. *An unemployment index does not cease to exist between readings, nor does Yule's pendulum cease to swing.* (emphasis H.S.)

Indeed there are many disadvantages of discrete time models. One of the most basic defects is that the dynamics are modeled between the (arbitrarily sampled) measurements and not between the dynamically relevant system states. For example, a physical system like a pendulum (cf. the citation above) fulfils a simple linear relation (Newton's equation) between the state and its velocity change (acceleration), whereas the relation between sampled measurements (e.g. daily) is very complicated and nonlinearly dependent on the parameters (mass, length of the pendulum etc.) and the sampling interval. Moreover, the velocity cannot be measured with discrete time data (latent variable).
Discrete time studies with different sampling intervals cannot be compared, since the causal parameters relate to the chosen interval. Moreover, if the same data set is analyzed with different intervals (select a weekly or monthly data set from daily measurements), one gets estimates corresponding to these intervals which can be in contradiction.
Nevertheless, the *continuous-discrete state space model* is able to combine both points of view:

1. a continuous time dynamical model
2. discrete time (sampled) measurements.

This hybrid model first appeared in engineering (JAZWINSKI, 1970), but is now well known in econometrics, sociology and psychology. One can estimate the parameters of the continuous time model from time series or panel measurements. This is achieved by computing the conditional probability density between the measurement times. In the linear Gaussian case, only the time dependent conditional mean and autocovariance is needed. More generally, in the presence of latent states and errors of measurement, a measurement model can be defined, mapping the continuous time state to observable discrete time data.

2 Nonlinear state space models

Whereas the linear continuous-discrete state space model can be treated completely and efficiently by using the Kalman filter algorithm (Harvey and Stock, 1985, Jones and Ackerson, 1990, Jones and Boadi-Boateng, 1991, Jones, 1993, Singer, 1993, 1995, 1998) or by structural equations models (SEM) with nonlinear parameter restrictions (Oud and Jansen, 1996, 2000, Singer, 2006[40]), 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 methods dominate and also sociology, despite the old tradition of Bergstrom (1966), Coleman (1968) and others.

We define the *nonlinear continuous-discrete state space model* (Jazwinski, 1970, ch. 6.2) for the panel units $n = 1, ..., N$

$$dy_n(t) = f(y_n(t), x_n(t), \psi)dt + g(y_n(t), x_n(t), \psi)dW_n(t) \tag{1}$$

$$z_i = h(y(t_i), x_n(t_i), \psi) + \epsilon_{ni}. \tag{2}$$

with nonlinear drift and diffusion functions f and g. The $x_n(t)$ are deterministic exogenous (control) variables. As usual, stochastic controls are treated by extending the state $y_n(t) \to \{y_n(t), x_n(t)\}$. The dependence on $x_n(t)$ includes the nonautonomous case $x_n(t) = t$. The error terms are mutually independent Gaussian white noise with zero means and covariance $E[(dW_n(t)/dt)(dW_m(s)/ds)] = \delta(t-s)\delta_{nm}$ and $E[\epsilon_{ni}\epsilon_{mj}] = \delta_{ij}\delta_{nm}$.

Since the panel units are independent, the panel index is dropped in the sequel for simplicity of notation. For maximum likelihood estimation, one only has to sum the N likelihood contributions of each panel unit. Alternatively, using Bayesian estimation, the parameter vector is filtered with the other states, and one has to use the extended state vector $\eta(t) = \{y_1(t), ..., y_N(t), \psi(t)\}$

In the nonlinear case it is important to interpret the SDE (1) correctly. We use the Itô interpretation yielding simple moment equations (for a thorough discussion of the system theoretical aspects see Arnold, 1974, ch. 10, van Kampen, 1981, Singer, 1999, ch. 3). 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 density $p(y_{i+1}, t_{i+1}|y_i, t_i)$ must be computed in order to obtain the likelihood function (cf. Aït-Sahalia, 2002, Singer, 2006). 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* (cf. 6), 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; second order nonlinear*

filter SNF; cf. JAZWINSKI, 1970 and section 2.4) 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 of a reference trajectory, the conditional Gaussian schemes are valid only for short measurement intervals $\Delta t_i = t_{i+1} - t_i$. Other linearization methods relate to the diffusion term, but are interpretable in terms of the EKF (NOWMAN, 1997).
A different class of approximations relates to the filter density. In the *unscented Kalman filter (UKF)*, cf. JULIER ET AL. (1997, 2000), the true density is replaced by a singular density with correct first and second moment, whereas the *Gaussian filter (GF)* assumes a normal density. Integrals in the update equations are obtained using Gauss-Hermite quadrature (ITO AND XIONG, 2000).
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 on polynomials ($L = F^\dagger$ is the adjoint of the Fokker-Planck operator (6)). Alternatively, one can use systems of moment differential equations (SINGER, 2006[39]) or numerical integration (CHALLA ET AL., 2000, SINGER, 2006[37, 38]). It seems that this approach is most efficient both in accuracy and computing time (cf. AÏT-SAHALIA, 2002, figure 1, JENSEN AND POULSEN, 2002).
Nonparametric approaches attempt to estimate the drift function f and the diffusion function Ω without assumptions about a certain functional form. They typically involve kernel density estimates of conditional 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).

2.1 Exact Continuous-Discrete Filter

The exact time and measurement updates of the continuous-discrete filter are given by the recursive scheme (JAZWINSKI, 1970) for the conditional density $p(y,t|Z^i)$:[1]

Time update:

$$\begin{aligned} \frac{\partial p(y,t|Z^i)}{\partial t} &= F(y,t)p(y,t|Z^i) \, ; t \in [t_i, t_{i+1}] \\ p(y,t_i|Z^i) &:= p_{i|i} \end{aligned} \tag{3}$$

Measurement update:

$$p(y_{i+1}|Z^{i+1}) = \frac{p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} := p_{i+1|i+1} \tag{4}$$

[1] again dropping panel index n

$$p(z_{i+1}|Z^i) = \int p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)dy_{i+1}, \quad (5)$$

$i = 0, \ldots, T-1$, where

$$\begin{aligned} Fp &= -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t,\psi)p(y,t|x,s)] \\ &+\tfrac{1}{2}\sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t,\psi)p(y,t|x,s)] \end{aligned} \quad (6)$$

is the Fokker-Planck operator, $Z^i = \{z(t)|t \le t_i\}$ are the observations up to time t_i and $p(z_{i+1}|Z^i)$ is the likelihood function of observation z_{i+1}. The first equation describes the time evolution of the conditional density $p(y,t|Z^i)$ given information up to the last measurement and the measurement update is a discontinuous change due to new information using the Bayes formula. The above scheme is exact, but can be solved explicitly only for the linear case where the filter density is Gaussian with conditional moments

$$\mu(t|t_i) = E[y(t)|Z^i] \quad (7)$$

$$\Sigma(t|t_i) = \mathrm{Var}[y(t)|Z^i]. \quad (8)$$

2.2 Exact Moment Equations

Instead of solving the time update equations for the conditional density (3), the moment equations for the first, second and higher order moments are considered. Using the Euler approximation for the SDE (1), one obtains in a short time interval δt ($\delta W(t) := W(t+\delta t) - W(t)$)

$$y(t+\delta t) = y(t) + f(y(t),t)\delta t + g(y(t),t)\delta W(t). \quad (9)$$

Taking the expectation $E[...|Z^i]$ one gets the moment equation

$$\mu(t+\delta t|t_i) = \mu(t|t_i) + E[f(y(t),t)|Z^i]\delta t \quad (10)$$

or in the limit $\delta t \to 0$

$$\dot{\mu}(t|t_i) = E[f(y(t),t)|Z^i]. \quad (11)$$

The higher order central moments

$$m_k(t|t_i) := E[(y(t) - \mu(t|t_i))^k|Z^i] := E[M_k(t|t_i)|Z^i] \quad (12)$$

fulfil (scalar notation, dropping the condition)

$$\begin{aligned} m_k(t+\delta t) &= E[y(t) + f(y(t),t)\delta t - \mu(t+\delta t) + g(y(t),t)\delta W(t)]^k \\ &:= E[a+bc]^k \end{aligned} \quad (13)$$

Using the binomial formula we obtain, utilizing the independence of $y(t)$ and $\delta W(t)$

$$E[a+bc]^k = \sum_{j=0}^{k} \binom{k}{j} E[a^{k-j}b^j] * E[c^j] \quad (14)$$

$$E[c^j] = \begin{cases} (j-1)!!\delta t^{j/2}; j & \text{is even} \\ 0; \quad j & \text{is odd} \end{cases} \quad (15)$$

since odd powers of $E[\delta W(t)]^j$ vanish and $E[\delta W(t)^{2j}] = (2j-1)!!\delta t^j$. For example, the second moment (variance) $m_2 = \sigma^2$ fulfils

$$
\begin{aligned}
E[a+bc]^2 &= E[a^2] + E[b^2]\delta t && (16)\\
&= E[y(t) + f(y(t),t)\delta t - \mu(t+\delta t)]^2 + \\
&+ E[\Omega(y(t),t)^2]\delta t. && (17)
\end{aligned}
$$

Inserting the first moment (10) and setting $a := \alpha + \beta = (y - E(y)) + (f - E(f))\delta t$ one obtains

$$
\begin{aligned}
m_2(t+\delta t) &= m_2(t) + 2E[(y-E(y))(f-E(f))]\delta t \\
&+ E[f-E(f)]^2\delta t^2 + E[\Omega]\delta t && (18)
\end{aligned}
$$

In general, up to $O(\delta t)$ we have ($M_k := (y-\mu)^k$)

$$
\begin{aligned}
m_k(t+\delta t) &= E[a^k] + \tfrac{k(k-1)}{2}E[b^{k-2}]\delta t + O(\delta t^2) \\
&= m_k(t) + kE[(y-E(y))^{k-1}(f-E(f))]\delta t \\
&+ \tfrac{k(k-1)}{2}E[(y-E(y))^{k-2}\Omega]\delta t + O(\delta t^2) \\
&= m_k(t) + kE[f(y,t) * (M_{k-1}(t) - m_{k-1}(t))]\delta t \\
&+ \tfrac{k(k-1)}{2}E[M_{k-2}(t)\Omega(y,t)]\delta t + O(\delta t^2). && (19)
\end{aligned}
$$

The exact moment equations (11, 19) are not differential equations, since they depend on the unknown conditional density $p(y,t|Z^i)$. Using Taylor expansions or approximations of the conditional density one obtains several filter algorithms.

2.3 Continuous-discrete filtering scheme

Using only the first and second moment equation (11,18), and the optimal linear update (normal correlation) one obtains the recursive scheme

Initial condition: $t = t_0$

$$
\begin{aligned}
\mu(t_0|t_0) &= \mu + \mathrm{Cov}(y_0,h_0) \times \\
&\times (\mathrm{Var}(h_0) + R(t_0))^{-}(z_0 - E[h_0]) \\
\Sigma(t_0|t_0) &= \Sigma - \mathrm{Cov}(y_0,h_0) \times \\
&\times (\mathrm{Var}(h_0) + R(t_0))^{-}\mathrm{Cov}(h_0,y_0) \\
L_0 &= \phi(z_0; E[h_0], \mathrm{Var}(h_0) + R(t_0))
\end{aligned}
$$

$i = 0, \ldots, T-1$:
Time update: $t \in [t_i, t_{i+1}]$

$$
\begin{aligned}
\tau_j &= t_i + j\delta t; j = 0, ..., J_i - 1 = (t_{i+1} - t_i)/\delta t - 1 \\
\mu(\tau_{j+1}|t_i) &= \mu(\tau_j|t_i) + E[f(y(\tau_j),\tau_j)|Z^i]\delta t \\
\Sigma(\tau_{j+1}|t_i) &= \Sigma(\tau_j|t_i) +
\end{aligned}
$$

$$
\begin{aligned}
&+ \quad \{\mathrm{Cov}[f(y(\tau_j),\tau_j), y(\tau_j)|Z^i] + \\
&+ \quad \mathrm{Cov}[y(\tau_j), f(y(\tau_j),\tau_j)|Z^i] + \\
&+ \quad E[\Omega(y(\tau_j),\tau_j)|Z^i]\}\delta t
\end{aligned}
$$

Measurement update: $t = t_{i+1}$

$$
\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \; (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^-(z_{i+1} - E[h_{i+1}|Z^i]) \\
\Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, h_{i+1}|Z^i) \times \\
&\times \; (\mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1}))^-\mathrm{Cov}(h_{i+1}, y_{i+1}|Z^i) \\
L_{i+1} &= \phi(z_{i+1}; E[h_{i+1}|Z^i], \mathrm{Var}(h_{i+1}|Z^i) + R(t_{i+1})).
\end{aligned}
$$

Remarks:

1. The time update for the interval $t \in [t_i, t_{i+1}]$ was written using time slices of width δt. They must be chosen small enough to yield a good approximation for the moment equations (11, 18).

2. The measurement update is written using the theorem on normal correlation (Liptser and Shiryayev, 1978, ch. 13, theorem 13.1, lemma 14.1)

$$
\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + \mathrm{Cov}(y_{i+1}, z_{i+1}|Z^i)\mathrm{Var}(z_{i+1}|Z^i)^- \times \\
&\times \; (z_{i+1} - E[z_{i+1}|Z^i]) \qquad (20) \\
\Sigma(t_{i+1}|t_{i+1}) &= \Sigma(t_{i+1}|t_i) - \mathrm{Cov}(y_{i+1}, z_{i+1}|Z^i)\mathrm{Var}(z_{i+1}|Z^i)^- \times \\
&\times \; \mathrm{Cov}(z_{i+1}, y_{i+1}|Z^i). \qquad (21)
\end{aligned}
$$

   Inserting the measurement equation (2) one obtains the measurement update of the filter. The formula is exact for Gaussian variables and the optimal *linear* estimate for $\mu(t_{i+1}|t_{i+1}), \Sigma(t_{i+1}|t_{i+1})$ in the nongaussian case. It is natural to use, if only two moments are considered. Despite the linearity in z_{i+1}, it still contains the measurement nonlinearities in the expectations involving $h(y,t)$. Alternatively, the the Bayes formula (4) can be evaluated directly. This is necessary, if strongly nonlinear measurements are taken (e.g. the threshold mechanism for ordinal data; see sect. 7)

The approximation of the expectation values containing the unknown filter density leads to several well known algorithms:

1. Taylor expansion of f, Ω and h:

   extended Kalman filter EKF, second order nonlinear filter SNF, higher order nonlinear filter HNF$(2, L)$ (Jazwinski, 1970, Singer, 2006[39]). Direct linearization in the SDE (1) using the Itô formula yields the LL approach of Shoji and Ozaki (1997, 1998); cf. Singer (2002).

2. Approximation of the expectations using sigma points:
   Unscented Kalman filter UKF (Julier et al., 1997, 2000).
3. Approximation of the expectations using Gauss-Hermite quadrature:
   Gauss-Hermite filter GHF (Ito and Xiong, 2000).

3 Filter Approximations based on Taylor Expansion

3.1 Extended Kalman Filter EKF

Using Taylor expansions around the conditional mean $\mu(\tau_j|t_i)$ for the nonlinear functions in the filtering scheme, one obtains

$$E[f(y(\tau_j),\tau_j)|Z^i] \approx f(\mu(\tau_j|t_i),\tau_j) \tag{22}$$
$$\mathrm{Cov}[f(y(\tau_j),\tau_j),y(\tau_j)|Z^i] \approx f_y(\mu(\tau_j|t_i),\tau_j)\varSigma(\tau_j|t_i) \tag{23}$$
$$E[\varOmega(y(\tau_j),\tau_j)|Z^i] \approx \varOmega(\mu(\tau_j|t_i),\tau_j). \tag{24}$$

Expanding around $\mu(t_{i+1}|t_i)$ the measurement update is approximately

$$\mathrm{Cov}[y_{i+1},h_{i+1}|Z^i] \approx \varSigma(t_{i+1}|t_i)h_y'(\mu(t_{i+1}|t_i),t_{i+1}) \tag{25}$$
$$\mathrm{Var}[h_{i+1}|Z^i] \approx h_y(\mu(t_{i+1}|t_i),t_{i+1})\varSigma(t_{i+1}|t_i) \times \times\ h_y'(\mu(t_{i+1}|t_i),t_{i+1}) \tag{26}$$
$$E[h_{i+1}|Z^i] \approx h(\mu(t_{i+1}|t_i),t_{i+1}). \tag{27}$$

3.2 Second Order Nonlinear Filter SNF

Expanding up to second order one obtains (using short notation and dropping third moments)

$$E[f(y(\tau_j),\tau_j)|Z^i] \approx f(\mu(\tau_j|t_i),\tau_j) + + \tfrac{1}{2}f_{yy}(\mu(\tau_j|t_i),\tau_j) * \varSigma(\tau_j|t_i) \tag{28}$$
$$\mathrm{Cov}[f(y(\tau_j),\tau_j),y(\tau_j)|Z^i] \approx f_y(\mu(\tau_j|t_i),\tau_j)\varSigma(\tau_j|t_i) \tag{29}$$
$$E[\varOmega(y(\tau_j),\tau_j)|Z^i] \approx \varOmega(\mu(\tau_j|t_i),\tau_j) + + \tfrac{1}{2}\varOmega_{yy}(\mu(\tau_j|t_i)) * \varSigma(\tau_j|t_i). \tag{30}$$

$$\mathrm{Cov}[y_{i+1},h_{i+1}|Z^i] \approx \varSigma(t_{i+1}|t_i)h_y'(\mu(t_{i+1}|t_i),t_{i+1}) \tag{31}$$
$$\mathrm{Var}[h_{i+1}|Z^i] \approx h_y(\mu(t_{i+1}|t_i),t_{i+1})\varSigma(t_{i+1}|t_i) \times \times\ h_y'(\mu(t_{i+1}|t_i),t_{i+1}) \tag{32}$$
$$E[h_{i+1}|Z^i] \approx h(\mu(t_{i+1}|t_i),t_{i+1}) + + \tfrac{1}{2}h_{yy}(\mu(t_{i+1}|t_i),t_{i+1}) * \varSigma(t_{i+1}|t_i), \tag{33}$$

where $(f_{yy} * \Sigma)_i = \sum_{jk} f_{yy,ijk} * \Sigma_{jk}$ etc. Expanding to higher orders in the HNF (higher order nonlinear filter) yields moments of order $k > 2$ on the right hand side, which must be dropped or factorized by the Gaussian assumption $m_k = (k-1)!!m_2^{k/2}, k$ even, $m_k = 0, k$ odd. For details, see JAZWINSKI (1970) or SINGER (2006[39]).

3.3 Local linearization LL

A related algorithm occurs if the drift is expanded directly in SDE (1). Using Itô's lemma one obtains

$$\begin{aligned} & f(y(t),t) - f(y(t_i),t_i) = \\ & = \int_{t_i}^{t} f_y(y(s),s)dy(s) + \\ & + \int_{t_i}^{t} \tfrac{1}{2} f_{yy}(y(s),s) * \Omega(y(s),s)ds + \int_{t_i}^{t} f_t(y(s),s)ds. \end{aligned} \tag{34}$$

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

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

The corresponding moment equations are

$$\begin{aligned} \dot{\mu}(t|t_i) &= f_y(y_i,t_i)(y(t|t_i) - y_i) + f(y_i,t_i) + f_t(y_i,t_i))(t - t_i) + \\ &+ \tfrac{1}{2} f_{yy}(y_i,t_i) * \Omega(t_i)(t - t_i) \quad (35) \\ \dot{\Sigma}(t|t_i) &= f_y(y_i,t_i)\Sigma(t|t_i) + \Sigma(t|t_i) f_y'(y_i,t_i) + \Omega(t_i). \quad (36) \end{aligned}$$

By contrast to the EKF and SNF moment equations 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 (for details, cf. SINGER, 2002).

4 Filter Approximations based on Numerical Integration

The traditional way of nonlinear filtering has been the expansion of the system functions f, Ω and h. Another approach is the approximation of the filtering density $p(y|Z^i)$.

4.1 Unscented Kalman Filtering

The idea of Julier et al. was the definition of so called sigma points with the property that the weighted mean and variance over these points coincides with the true parameters. According to JULIER ET AL. (2000) one can take the $2p+1$ points

$$\begin{aligned} x_0 &= \mu && (37)\\ x_l &= \mu + \sqrt{p+\kappa}\varGamma_{.l}, l = 1, ..., p && (38)\\ x_{-l} &= \mu - \sqrt{p+\kappa}\varGamma_{.l}, l = 1, ..., p && (39) \end{aligned}$$

with weights

$$\begin{aligned} \alpha_0 &= \kappa/(p+\kappa) && (40)\\ \alpha_l &= 1/(2(p+\kappa)) = \alpha_{-l}, && (41) \end{aligned}$$

where $\varGamma_{.l}$ is the lth column of the Cholesky root $\varSigma = \varGamma\varGamma'$, κ is a scaling factor and p is the dimension of the random vector X. For example, in the univariate case $p = 1$ one obtains the three points $\mu, \mu \pm \sqrt{1+\kappa}\,\sigma$.
The UT method may be interpreted in terms of the singular density

$$p_{UT}(x) = \sum_{l=-p}^{p} \alpha_l \delta(x - x_l). \quad (42)$$

Then, however, only nonnegative weights α_l are admissible. Generally, the expectation $E_{UT}[X] = \int x\, p_{UT}(x)dx = \mu$ and

$$\begin{aligned} \mathrm{Var}_{UT}(X) &= \int (x-\mu)(x-\mu)' p_{UT}(x)dx && (43)\\ &= \sum_{l=-p}^{p} (x_l-\mu)(x_l-\mu)\alpha_l && (44)\\ &= \sum_{l=1}^{p} \varGamma_{.l}(\varGamma_{.l})' = \varGamma\varGamma' = \varSigma && (45) \end{aligned}$$

yields the correct first and second moment. Nonlinear expectations are easily evaluated as sums

$$\begin{aligned} E_{UT}[f(X)] &= \int f(x)p_{UT}(x)dx && (46)\\ &= \sum_l f(x_l)\alpha_l && (47)\\ &= \frac{\kappa}{p+\kappa} f(\mu) + \frac{1}{2(p+\kappa)} \sum_{l=-p,p\neq 0}^{p} f(x_l) && (48) \end{aligned}$$

Using large κ, the EKF formula $E_{Taylor}[f(X)] = f(\mu)$ is recovered.
All expectations in the filter are evaluated using the sigma points computed from the conditional moments $\mu(\tau_j|t_i), \varSigma(\tau_j|t_i)$. To display the dependence on the moments, the notation $y_l = y_l(\mu, \varSigma)$ will be used. For example, the terms in the time update are (short notation dropping arguments)

$$\begin{aligned} E[f|Z^i] &\approx \sum_l f(y_l)\alpha_l && (49)\\ \mathrm{Cov}[f, y|Z^i] &= E[f(y)(y-\mu)|Z^i] \approx \sum_l f(y_l)(y_l-\mu)\alpha_l && (50)\\ E[\varOmega|Z^i] &\approx \sum_l \varOmega(y_l)\alpha_l && (51) \end{aligned}$$

with sigma points $y_l = y_l(\mu(\tau_j|t_i), \tau_j), \Sigma(\tau_j|t_i))$. With the new moments $\mu(\tau_{j+1}|t_i)$, $\Sigma(\tau_j|t_i)$ updated sigma points are computed.

4.2 Gauss-Hermite Filtering

For the Gaussian filter, one may assume that the true $p(x)$ is approximated by a Gaussian distribution $\phi(x; \mu, \sigma^2)$ with the same mean μ and variance σ^2. Then, the Gaussian integral

$$E_\phi[f(X)] = \int f(x)\phi(x; \mu, \sigma^2)dx = \int f(\mu + \sigma z)\phi(z; 0, 1)dz \tag{52}$$

$$\approx \sum_{l=1}^{m} f(\mu + \sigma\zeta_l)w_l = \sum_{l=1}^{m} f(\xi_l)w_l \tag{53}$$

may be approximated by Gauss-Hermite quadrature (cf. Ito and Xiong, 2000). The ζ_l, w_l are quadrature points and weights for the standard gaussian distribution $\phi(z; 0, 1)$. If such an approximation is used, one obtains the Gauss-Hermite filter (GHF). Generally, filters using Gaussian densities are called Gaussian filters (GF). The GHF can be interpreted in terms of the singular density $p_{GH}(x) = \sum_{l=1}^{m} w_l\delta(x - \xi_l)$ concentrated at the quadrature points ξ_l. The Gauss-Hermite quadrature rule is exact up to order $O(x^{2m-1})$. Multivariate Gaussian integrals can be computed by transforming to the standard normal distribution and p-fold application of (52).
The Gaussian filter is equivalent to an expansion of f to higher orders L

$$E[f(X)] \approx \sum_{l=0}^{L} \tfrac{1}{l!} f^l(\mu)E[X - \mu]^l = \sum_{l=0}^{L} \tfrac{1}{l!} f^l(\mu)m_l \tag{54}$$

(higher order nonlinear filter HNF$(2, L)$) and factorization of the moments according to the Gaussian assumption $m_l := E[X - \mu]^l = (l - 1)!!\sigma^l$ (l even) and $m_l = 0$ (l odd). This leads to an exact computation of (52) for $L \to \infty$. In this limit, the HNF and GF coincide. In the EKF=HNF(2,1) and SNF=HNF(2,2), the higher order corrections are neglected. Also, third and higher order moments could be used (HNF(K, L); cf. Singer, 2006[39]).
It is interesting that $\kappa = 2, p = 1$ in the UT corresponds to a Gauss-Hermite rule with $m = 3$ sample points (Ito and Xiong, 2000).

4.3 Generalized Gauss-Hermite Filtering

The Gaussian filter assumed a Gauss density $\phi(y; \mu, \sigma^2)$ for the filter distribution $p(y)$. More generally, one can use a Hermite expansion

$$p(y) = \phi(y; \mu, \sigma^2) \sum_{n=0}^{K} c_n H_n((y - \mu)/\sigma) = \phi(y) * H(y, K), \tag{55}$$

with Fourier coefficients $c_0 = 1, c_1 = 0, c_2 = 0$,

$$c_3 := (1/3!)E[Z^3] = (1/3!)\nu_3 \tag{56}$$

$$c_4 := (1/4!)E[Z^4 - 6Z^2 + 3] = (1/24)(\nu_4 - 3), \tag{57}$$

$Z := (Y - \mu)/\sigma$ and orthogonal polynomials $H_0 = 1, H_1 = x, H_2 = y^2 - 1, H_3 = y^3 - 3y, H_4 = y^4 - 6y^2 + 3$ etc. Expectation values occuring in the update equations are again computed by Gauss-Hermite integration, including the nongaussian term

$$H(y; \{\mu, m_2, ..., m_K\}) \quad := \quad \sum_{n=0}^{K} c_n H_n((y - \mu)/\sigma). \tag{58}$$

Since $H(y; \{\mu, m_2, ..., m_K\})$ depends on higher order moments, one must use K moments equations (19). The choice $K = 2$ recovers the usual Gaussian filter, since $c_0 = 1, c_1 = 0, c_2 = 0$. The Hermite expansion can model bimodal, skewed and leptokurtic distributions. For details, see SINGER (2006[39, 38]) and sect. 7.

5 Discussion

1. The density based filters UKF and GHF have the strong advantage, that no derivatives of the system functions must be computed. This is no problem for the EKF and SNF, but for higher orders in the HNF(K, L) complicated tensor expressions arise. Moreover, higher order moments must be dropped or factorized in order to obtain closed equations. In the multivariate case, the formulas for Gaussian moments are involved. The fourth moment is

$$m_{4,ijkl} \quad = \quad \sigma_{ij}\sigma_{kl} + \sigma_{ik}\sigma_{jl} + \sigma_{il}\sigma_{jk}. \tag{59}$$

   For a general formula, see GARDINER, 1996, p. 36.

2. Apart from an implementation point of view (see 1.), the low order EKF and SNF suffer from problems such as filter divergencies, especially when the sampling intervals are large. Simulation studies suggest, that the UKF and GHF are more stable and yield smaller filtering error in the mean (SINGER, 2006[37]).

3. The moment equations and measurement updates as derived in section (2.3) involve expectations with respect to the filter density $p(y)$, but not for the noise processes. Their statistics are already included in these updates (the terms $E[\Omega]dt = E[gdW(gdW)']$ and $R = \mathrm{Var}(\epsilon)$ stem from the noise sequences). Thus no sigma points w.r.t the noises must be computed, as suggested in the literature on the UKF (JULIER ET AL., loc. cit., SITZ ET AL., 2002). This is only necessary if the system is first modeled deterministically and afterwards extended by the noises. This is neither necessary nor efficient.

6 Example: Bifurcation System

The several filtering algorithms can be used to compute the likelihood for each panel unit and the sum of all likelihood contributions is maximized. We study the nonlinear system

$$dy_n \quad = \quad -[\alpha y_n + \beta y_n^3]dt + \sigma dW_n(t) \tag{60}$$

with measurement equation

$$z_{in} = y_{in} + \epsilon_{in}, \tag{61}$$

$n = 1, ..., N = 10$, measured at times $t_i \in \{0, 4, 6, 8, 10, 11, 12, 13.5, 13.7, 15, 15.1, 17, 19, 20\}$. The measurement times could be different for each panel unit. The random initial condition was $y_n(t_0) \sim N(0, 10)$. The nonlinear drift $f(y) = -[\alpha y + \beta y^3]$ is the negative gradient of a potential $\Phi(y) = \frac{1}{2}\alpha y^2 + \frac{1}{4}\beta y^4$ and the motion may be visualized as a Brownian motion in the landscape defined by $\Phi(y)$ (fig. 1). The stationary density is given by $p_{stat} \propto \exp(-(2/\sigma^2)\Phi(y))$ (fig. 2). For $\beta > 0$, the potential can have two minima and one maximum ($\alpha < 0$ or one minimum ($\alpha \geq 0$). Such a qualitative change following a continuous variation of a parameter is called a bifurcation (fig. 3).

The model is interesting from both a theoretical and an application point of view: The density function strongly deviates from Gaussian behavior, at least in the bimodal state $\alpha < 0$. Thus it is a good test for filters relying on only two moments. For applications, it can model systems with two stable states with a sudden transition to only one equilibrum. It has been used for phase transitions (Ginzburg and Landau; cf. Haken, 1977, chs. 6.4, 6.7-8), stability of engineering systems (Frey, 1996) and equilibrium states in economics (Herings, 1996). Fig. 4 shows the true trajectory, the measurements and approximate 67% HPD (highest probability density) confidence intervals (conditional mean ± standard deviation) for all panel units $n = 1, ..., 10$ using the UKF($\kappa = 2$) = GHF($m = 3$) and the true parameter vector $\theta = \{-1, 0.1, 2, 1\}$. Fig. 5 displays the true trajectory, the measurements and 67% HPD confidence intervals for panel unit $n = 10$ using the true parameter vector. It can be shown that some algorithms such as the SNF exhibit divergencies in the first measurement interval [0,4] when the conditional mean approaches zero. Simulation studies demonstrate, that the EKF, SNF and LL are prone to such numerical instabilities (cf. Singer, 2006[37]). Higher order expansion of the drift can avoid such singularities. As noted, the Gaussian filter corresponds to an infinite Taylor expansion with Gaussian factorization of the moments (cf. eqn. 54). Generally, the density based UKF and GHF are numerically more stable than EKF and SNF and lead to smaller filtering error.

The performance of the several filters was compared in a simulation study (see table (1), where the maximum likelihood (ML) estimates were computed for $M = 100$ replications and $N = 10$ panel units. The likelihood was maximized using a quasi Newton algorithm with numerical score function and BFGS secant updates (Dennis and Schnabel, 1983). In terms of RMSE, the several algorithms are comparable. There is a tradeoff between bias and variance of the estimates. For example, the estimates for α in the Taylor based methods EKF, SNF and LL are strongly biased, but the standard deviation is smaller than for the density based algorithms. However, due to the long sampling intervals, EKF, SNF and LL tend to diverge, and large conditional means $|\mu(t|Z^i)| > YMAX = 1000$ were reset to zero. Over all, there is no algorithm with clear advantages, although UKF and GHF are more stable. The UKF furthermore has the problem of choosing the scaling paramter κ. It seems, in this example, that $\kappa = 0$ yields the smallest RMSE, whereas the bias is minimized

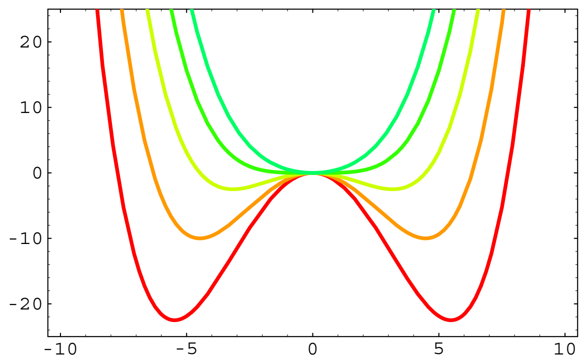

Figure 1: Potential $\Phi(y)$ as a function of y for several parameter values $\alpha = -3, -2, ..., 1$.

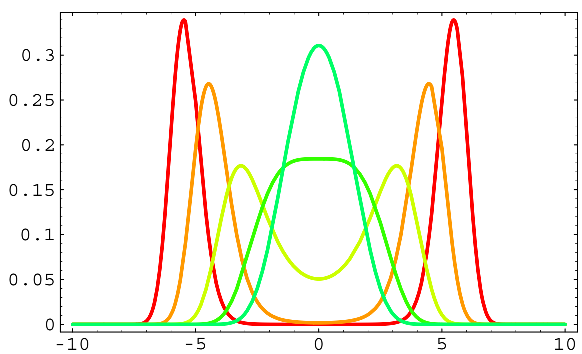

Figure 2: Stationary density $p_{stat} \propto \exp[-(2/\sigma^2)\Phi(y)]$ as a function of y for several parameter values $\alpha = -3, -2, ..., 1$.

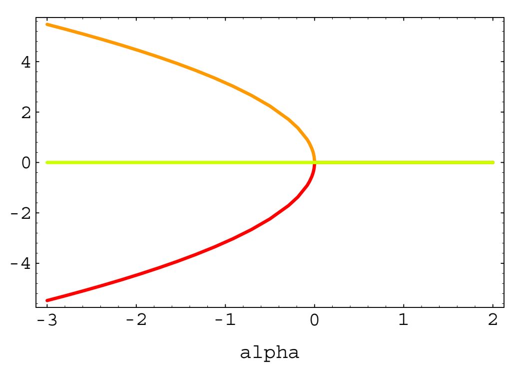

Figure 3: Minima and maxima of $\Phi(y)$ as a function of α (bifurcation diagram).

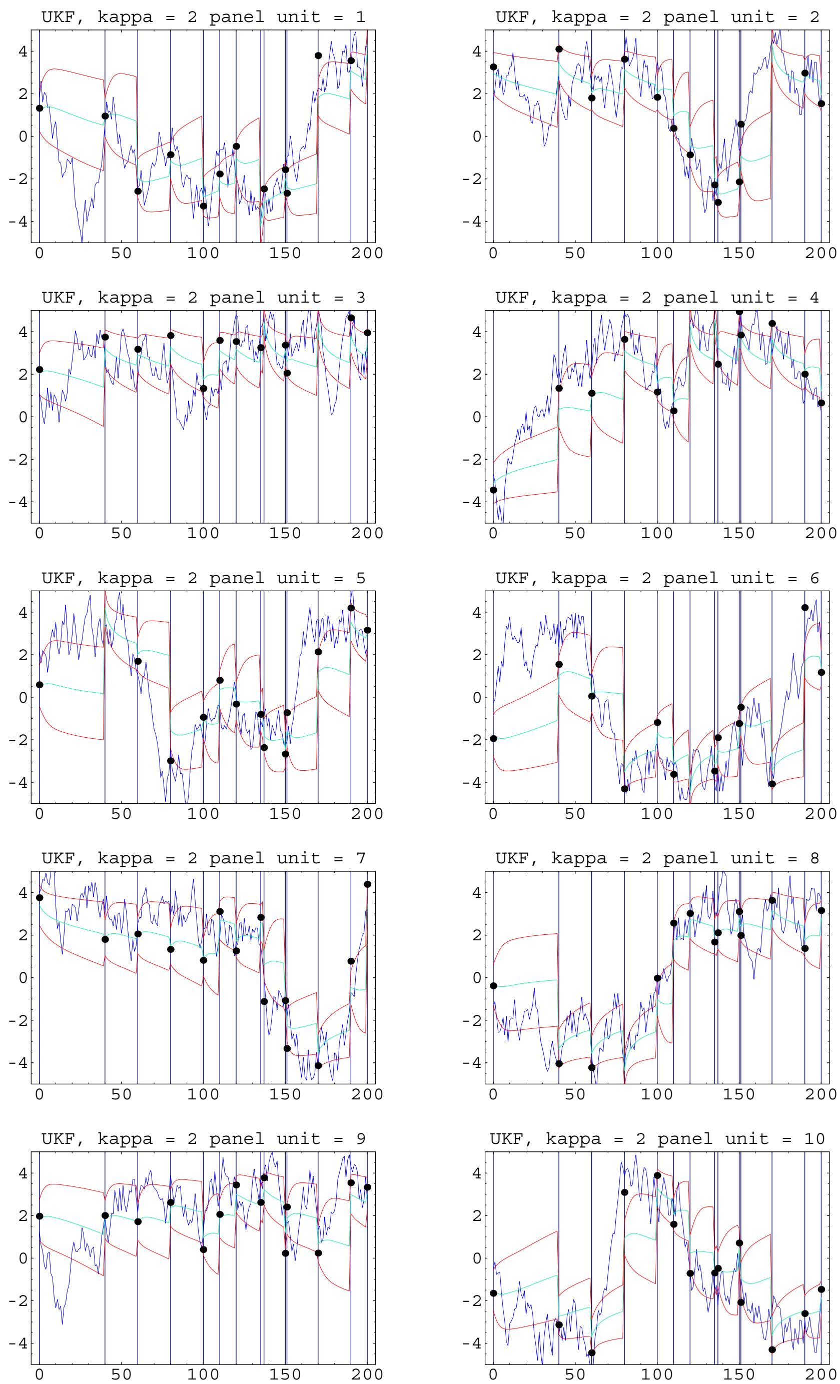

Figure 4: UKF($\kappa = 2$) = GHF($m = 3$) for all $N = 10$ panel units. True trajectory, measurements (dots), and 67% HPD band.

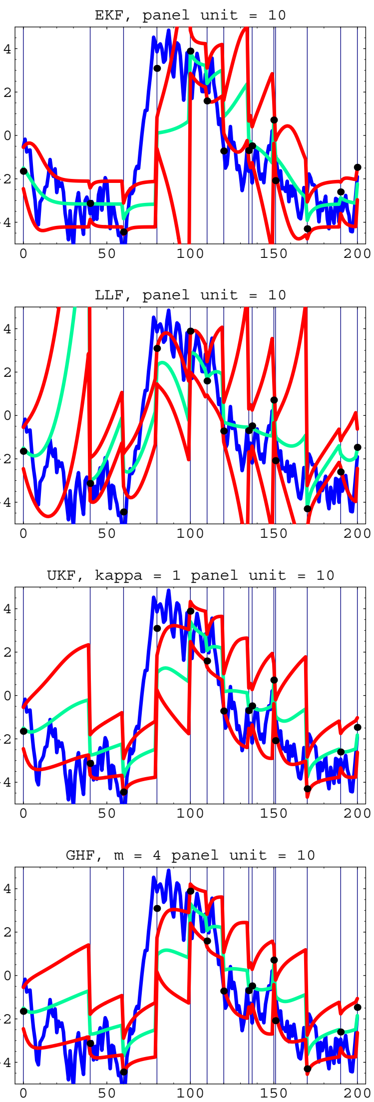

Figure 5: Panel unit $n = 10$: Comparision of several filter algorithms

EKF, $M' = 100$					
parameter	value	mean	std	bias	RMSE
α	-1.	-0.254169	0.413799	0.745831	0.852932
β	0.1	0.0658808	0.0730564	-0.0341192	0.080631
σ	2.	2.01777	0.638036	0.017767	0.638283
R	1.	0.951857	0.444057	-0.0481435	0.446659
SNF, $M' = 84$					
parameter	value	mean	std	bias	RMSE
α	-1.	-0.119926	0.0953236	0.880074	0.885222
β	0.1	0.0231572	0.00795955	-0.0768428	0.0772539
σ	2.	1.46765	0.215205	-0.53235	0.574203
R	1.	1.1505	0.399459	0.150497	0.426869
LL, $M' = 65$					
parameter	value	mean	std	bias	RMSE
α	-1.	-0.0885385	0.117338	0.911461	0.918983
β	0.1	0.0163748	0.00790206	-0.0836252	0.0839977
σ	2.	1.48073	0.246138	-0.519273	0.574655
R	1.	1.25062	0.370842	0.250619	0.447586
UKF, $\kappa = 0, M' = 99$					
parameter	value	mean	std	bias	RMSE
α	-1.	-0.712736	0.95255	0.287264	0.994923
β	0.1	0.076122	0.0853785	-0.023878	0.0886547
σ	2.	1.79875	0.403944	-0.201246	0.451299
R	1.	1.07417	0.442	0.0741727	0.44818
UKF, $\kappa = 1, M' = 96$					
parameter	value	mean	std	bias	RMSE
α	-1.	-1.09344	1.00893	-0.0934443	1.01325
β	0.1	0.102081	0.086685	0.0020806	0.08671
σ	2.	2.11663	0.439907	0.116632	0.455106
R	1.	0.984557	0.450917	-0.0154427	0.451182
UKF, $\kappa = 2, M' = 93$					
parameter	value	mean	std	bias	RMSE
α	-1.	-1.51408	1.19571	-0.514084	1.30154
β	0.1	0.124142	0.0898191	0.024142	0.093007
σ	2.	2.46263	0.489341	0.462625	0.673407
R	1.	0.873614	0.437204	-0.126386	0.455105
GHF, $m = 4, M' = 96$					
parameter	value	mean	std	bias	RMSE
α	-1.	-1.4521	1.14993	-0.4521	1.23561
β	0.1	0.122561	0.0894279	0.0225606	0.0922298
σ	2.	2.39497	0.440323	0.394971	0.591511
R	1.	0.891953	0.433076	-0.108047	0.446351

Table 1: Simulation study for bifurcation model. Distribution of ML estimates in $M = 100$ replications. M'=number of converged samples.

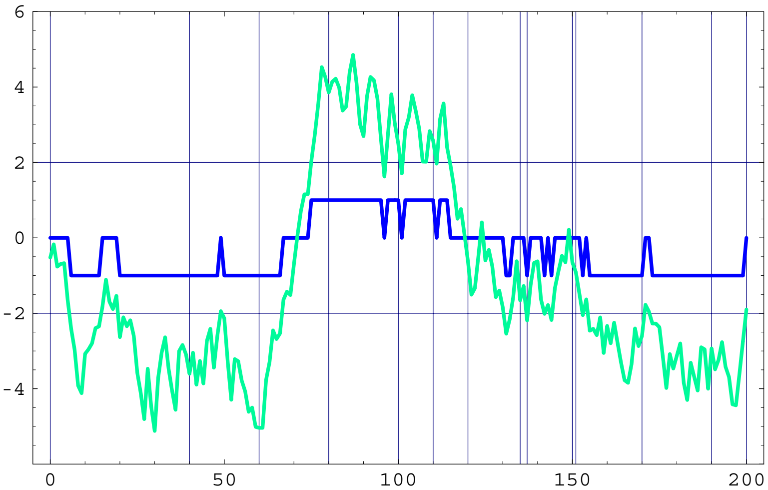

Figure 6: Panel unit $n = 10$: Latent trajectory (green) and ordinal measurements (blue) with thresholds $\{c_1, c_2\} = \{-2, 2\}$. The values of $z(t)$ are -1,0,1.

for $\kappa = 1$.

7 Example: Ordinal measurements

The nonlinear state space model (1–2) is flexible enough to model ordinal measurements via the threshold model

$$z \quad = \quad z_0 + \sum_{l=1}^{L} \theta(y - c_l) + \epsilon := h(y) + \epsilon \tag{62}$$

where θ is the Heaviside step function and c_l are thresholds contained in the parameter vector ψ. The variance of the measurement error $R = \mathrm{Var}(\epsilon)$ is taken small (10^{-6} here), so that the measurement density $p(z|y) = \phi(z; z_0 + \sum_l \theta(y - c_l), R)$ is proportional to the indicator function $\chi_{h^{-1}(z)}(y)$. Now the measurements are strongly nonlinear and the a posteriori density is proportional to the a priori density truncated by the windows $C_l = (c_l, c_{l+1}]$ defined from the thresholds $c = \{-\infty, c_1, ..., c_L, \infty\}$. Fig. 6 shows the trajectory of panel unit $n = 10$ together with the thresholds $c = \{c_1, c_2\} = \{-2, 2\}$ and the ordinal data $z(t) \in \{-1, 0, 1\}$ (setting $z_0 = -1$). The data were filtered using the generalized Gauss-Hermite filter GGHF comparing the normal correlation and the Bayes update. As explained in sect. 4.3, the filter density is represented by the Hermite expansion $\phi(y; \mu, \sigma^2)H(y, k)$ and the measurement update is obtained either by the normal correlation (20) or by the Bayes formula (4). In both cases, Gauss-Hermite integration can be used. Denoting the linear estimates (20) by μ_0 and Σ_0, the normal correlation update is given by the product

$$p(y|z) \quad \approx \quad L_0\, \phi(y; \mu_0, \Sigma_0) H(y; \{\mu, m_2 ..., m_K\})/L \tag{63}$$

$$L \quad = \quad L_0 \int \phi(y; \mu_0, \Sigma_0) H(y; \{\mu, m_2 ..., m_K\}) dy \tag{64}$$

$$L_0 \quad = \quad \phi(z; E[h], \mathrm{Var}(h) + R) \tag{65}$$

In the equations above, the posteriori distribution is nongaussian due to the Hermite part $H(y; \{\mu, m_2..., m_K\})$, where $\{\mu, m_2..., m_K\}$ are the apriori moments. For strongly nonlinear measurements, the Bayes formula yields

$$p(y|z) = \phi(z; h(y), R)\phi(y; \mu, \sigma^2)H(y; \{\mu, m_2..., m_K\})/L \quad (66)$$

$$L = \int \phi(z; h(y), R)\phi(y; \mu, \sigma^2)H(y; \{\mu, m_2..., m_K\})dy \quad (67)$$

The likelihood integral is over the a priori Gauss part $\phi(y; \mu, \sigma^2)$, but the efficiency can be improved by integrating over the normal correlation update $\phi(y; \mu_0, \Sigma_0)$ (analogously to importance sampling). Figures (7–8) compare the updates in the case $K = 2$ (Gaussian filter). The densities are always Gaussian, but the a posteriori moments are either the linear estimates or are computed using the Bayes formula. The latter method yields better updates which more closely approximate the truncated Gaussian a posteriori densities. Note that the measurement density $p(z|y)$ was scaled by 10^{-3} in the graphics.
Using more terms (e.g. $K = 20$) in the Hermite expansion yields a more realistic modeling of the bimodal a priori density (figs. 9–10). Note that the normal correlation update is nongaussian as well, due to the Hermite term $H(y)$. In some cases the Bayes update tends to unrealistic oscillations in the a posteriori density. This is due to locally negative values of the Hermite series.

8 Conclusion

We compared filtering algorithms for nonlinear panel models in continuous time with time discrete measurements. The classical algorithms EKF and SNF are based on Taylor expansions of the moment equations. In contrast, UKF and GHF use numerical integration for the expectation values. The UT transformation is directly applied to the moment equations avoiding the extension of the system state and doubling the dimension. ML estimation of a Ginzburg-Landau model did not yield a uniformly best method, but in terms of bias, the UKF with $\kappa = 1$ was best. Finally, ordinal data were treated using a threshold model using the GHF and the GGHF ($K = 20$). The Bayes update is superior to the linear normal correlation. Since the measurement function is not differentiable, the EKF type algorithms cannot be used here. The Hermite expansion yields a more realistic approximation of the truncated a posteriori density, but already the Gaussian case $K = 2$ leads to sufficient results.

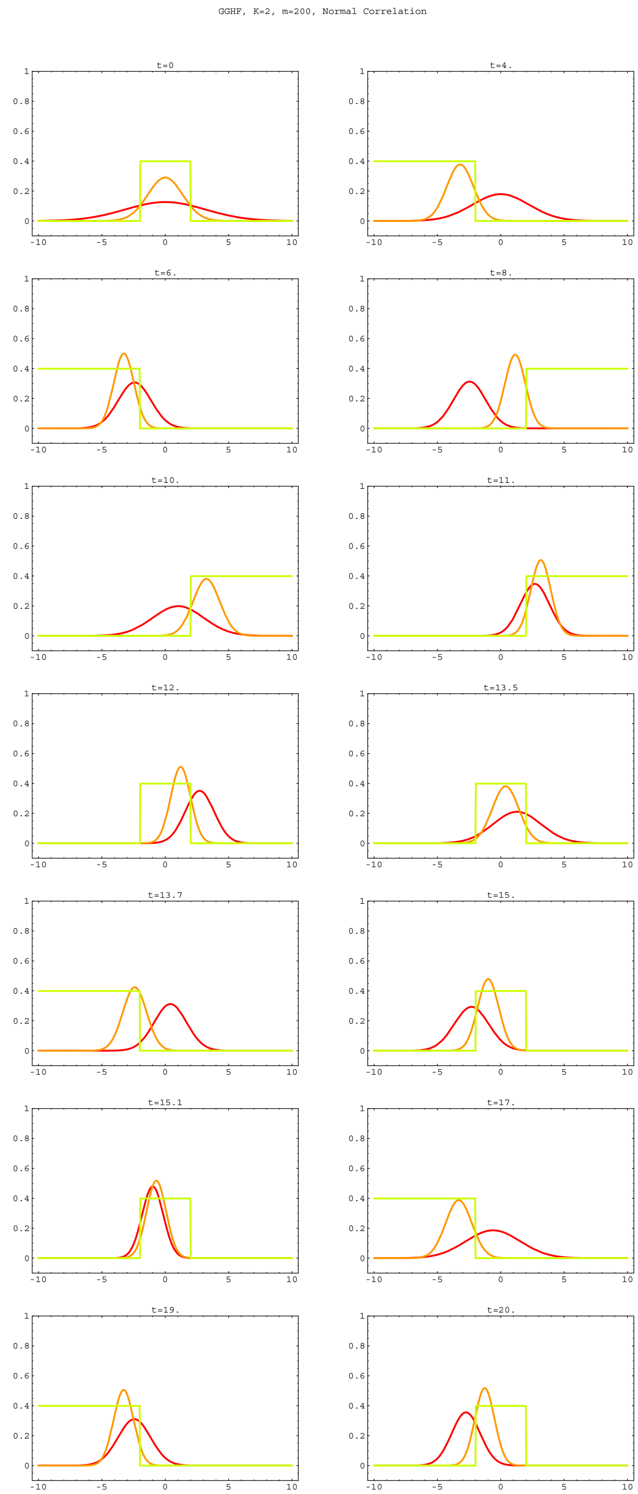

Figure 7: GGHF: Measurement updates of threshold model. Normal correlation, $K = 2$ ($=$ GHF). A priori (red), a posteriori (orange) and measurement density $p(z|y)$ (green).

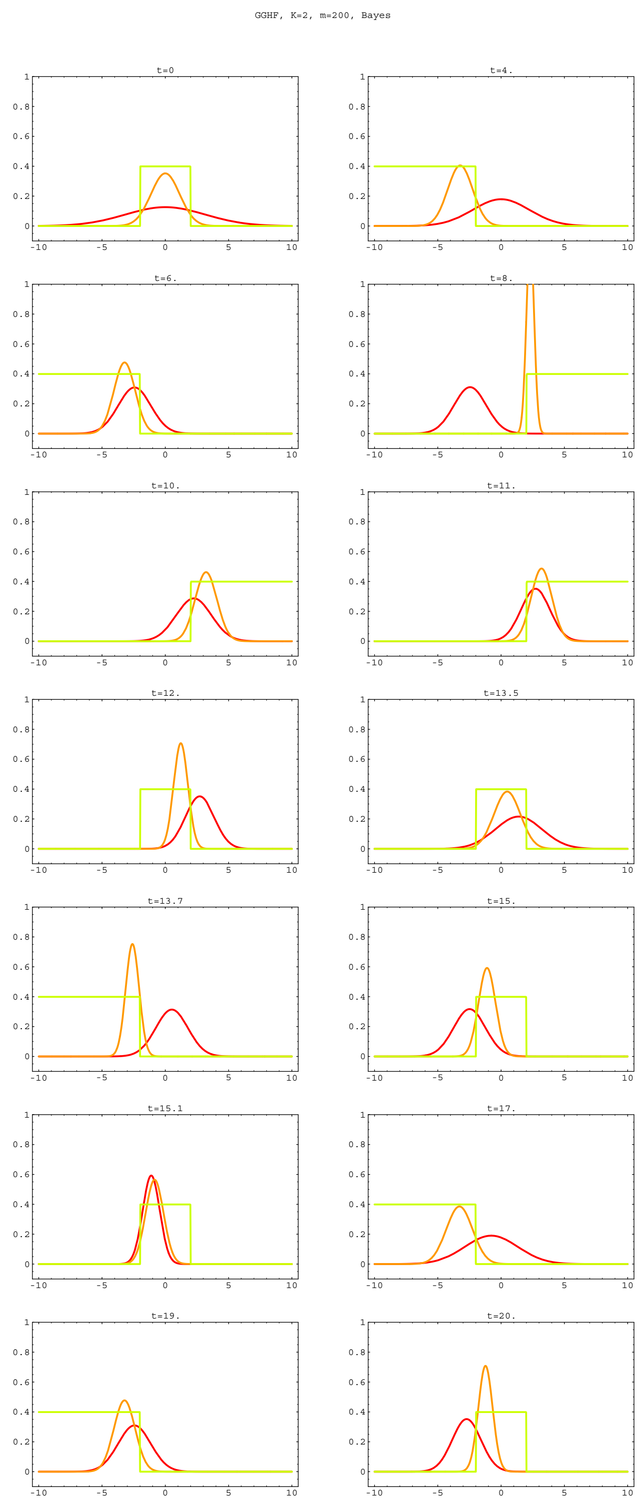

Figure 8: GGHF: Measurement updates of threshold model. Bayes formula, $K = 2$ (= GHF). Note that the update yields a better approximation of $p(y|z)$. A priori (red), a posteriori (orange) and measurement density $p(z|y)$ (green).

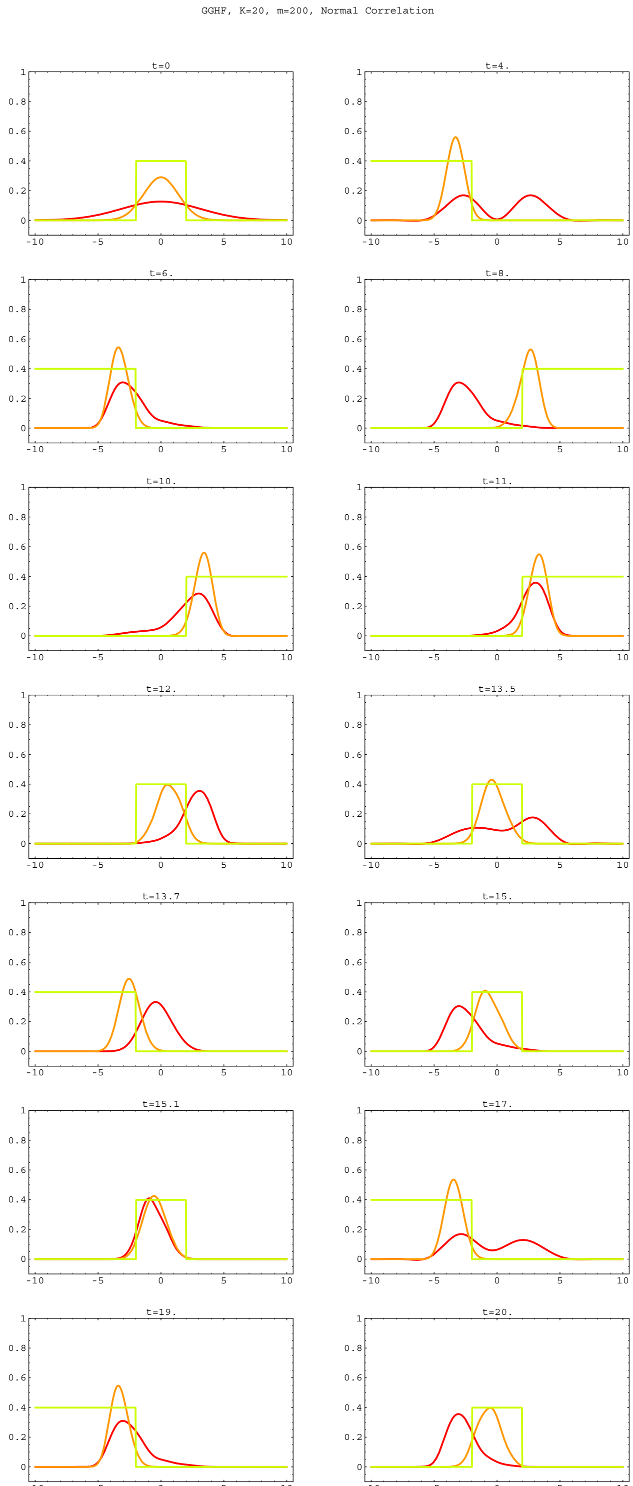

Figure 9: GGHF: Measurement updates of threshold model. Normal correlation, $K = 20$ ($= \text{GHF}$). A priori (red), a posteriori (orange) and measurement density $p(z|y)$ (green).

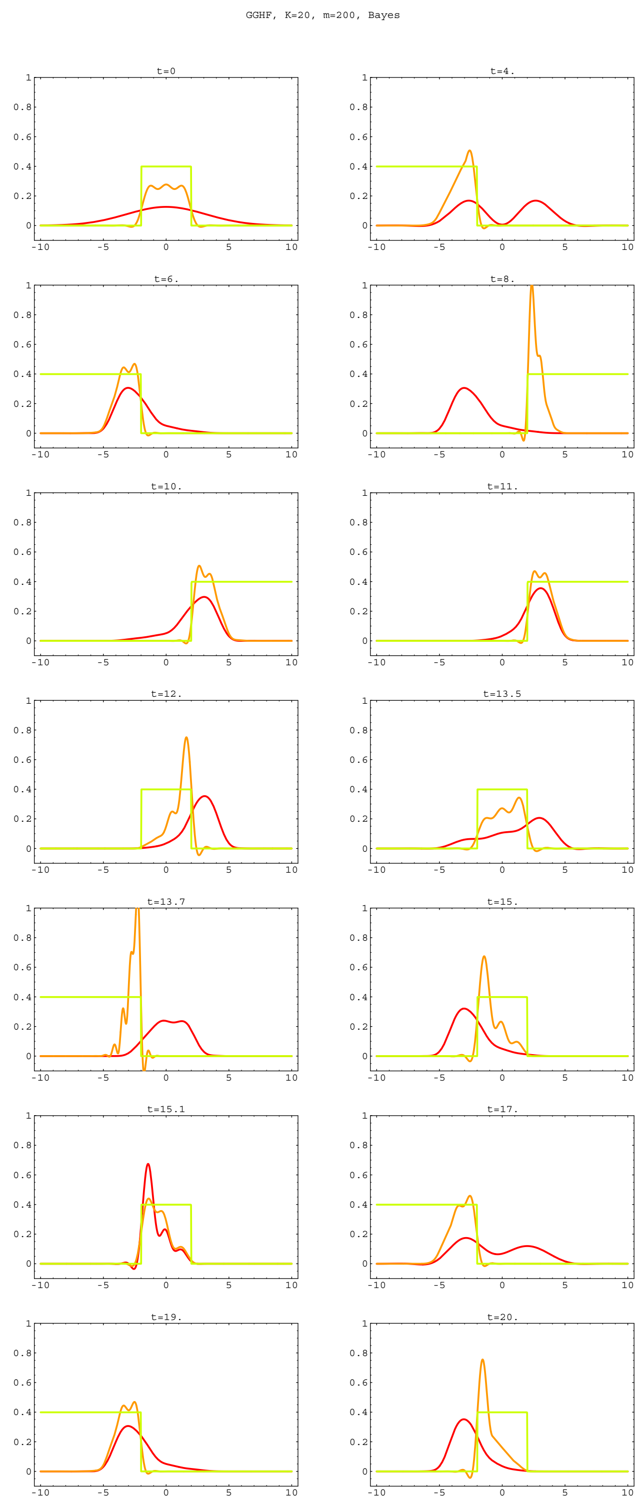

Figure 10: GGHF: Measurement updates of threshold model. Bayes formula, $K = 20$. A priori (red), a posteriori (orange) and measurement density $p(z|y)$ (green). Oscillation at $t = 13.7$ (see text).

References

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

[2] 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.

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

[4] M.S. Bartlett. On the theoretical specification and sampling properties of autocorrelated time-series. *Journal of the Royal Statistical Society (Supplement)*, 7:27–41, 1946.

[5] A.R. Bergstrom. Non Recursive Models as Discrete Approximations to Systems of Stochastic Differential Equations. In A.R. Bergstrom, editor, *Statistical Inference in Continuous Time Models*, pages 15–26. North Holland, Amsterdam, 1976.

[6] S. Challa, Y. Bar-Shalom, and V. Krishnamurthy. Nonlinear Filtering via Generalized Edgeworth Series and Gauss-Hermite Quadrature. *IEEE Transactions On Signal Processing*, 48:1816–1820, 2000.

[7] 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, New York, 1968.

[8] J.E. Dennis Jr. and R.B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice Hall, Englewood Cliffs, 1983.

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

[10] M.R. Frey. A Wiener Filter, State-Space Flux-Optimal Control Against Escape from a Potential Well. *IEEE Transactions on Automatic Control*, 41, 2:216–223, 1996.

[11] C.W. Gardiner. *Handbook of Stochastic Methods*. Springer, Berlin, Heidelberg, New York, second edition, 1996.

[12] H. Haken. *Synergetics*. Springer, Berlin, 1977.

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

[14] J.P. Herings. *Static and Dynamic Aspects of General Disequilibrium Theory*. Kluwer, Boston, London, Dordrecht, 1996.

[15] K. Ito and K. Xiong. Gaussian filters for nonlinear filtering problems. *IEEE Transactions on Automatic Control*, 45, 5:910–927, 2000.

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

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

[18] R.H. Jones. *Longitudinal Data With Serial Correlation: A State Space Approach*. Chapman and Hall, New York, 1993.

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

[20] R.H. Jones and F. Boadi-Boateng. Unequally Spaced Longitudinal Data with AR(1) Serial Correlation. *Biometrics*, 47:161–175, 1991.

[21] S. Julier and J. Uhlmann. A new extension of the Kalman filter to nonlinear systems. In *The 11th International Symposium on Aerospace/Defense Sensing, Simulation and Control*, Orlando, Florida, 1997.

[22] S. Julier, J. Uhlmann, and H. F. Durrant-White. A New Method for the Nonlinear Transformation of Means and Covariances in Filters and Estimators. *IEEE Transactions on Automatic Control*, 45, 3:477–482, 2000.

[23] R.S. Liptser and A.N. Shiryayev. *Statistics of Random Processes, Volumes I and II*. Springer, New York, Heidelberg, Berlin, 1977, 1978.

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

[25] 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.

[26] J.H.L. Oud and R.A.R.G Jansen. Nonstationary Longitudinal LISREL Model Estimation from Incomplete Panel Data Using EM and the Kalman Smoother. In U. Engel and J. Reinecke, editors, *Analysis of Change*, pages 135–159. de Gruyter, Berlin, New York, 1996.

[27] J.H.L. Oud and R.A.R.G Jansen. Continuous Time State Space Modeling of Panel Data by Means of SEM. *Psychometrika*, 65:199–215, 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] I. Shoji. Nonparametric state estimation of diffusion processes. *Biometrika*, 89, 2:451–456, 2002.

[30] 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.

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

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

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

[34] H. Singer. *Finanzmarktökonometrie. Zeitstetige Systeme und ihre Anwendung in Ökonometrie und empirischer Kapitalmarktforschung*. Physica-Verlag, Heidelberg, 1999.

[35] 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.

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

[37] H. Singer. Continuous-Discrete Unscented Kalman Filtering. Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 384, FernUniversität in Hagen, 2006.

[38] H. Singer. Generalized Gauss-Hermite Filtering. Diskussionsbeiträge Fachbereich Wirtschaftswissenschaft 391, FernUniversität in Hagen, 2006.

[39] H. Singer. Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models. *Computational Statistics*, 21, 3, 2006. in print.

[40] H. Singer. Stochastic Differential Equation Models with Sampled Data. In K. van Montfort, H. Oud, and A. Satorra, editors, *Longitudinal Models in the Behavioral and related Sciences*. Erlbaum publishers, 2006. in print.

[41] A. Sitz, U. Schwarz, and J. Kurths. The Unscented Kalman Filter, A Powerful Tool For Data Analysis. *International Journal of Bifurcation and Chaos*, 14, 6:2093–2105, 2002b.

[42] A. Sitz, U. Schwarz, J. Kurths, and H.U. Voss. Estimation of parameters and unobserved components for nonlinear systems from noisy time series. *Physical Review E*, 66:016210–1–016210–9, 2002a.

[43] N.G. Van Kampen. Itô vs. Stratonovich. *Journal of Statistical Physics*, 24:175–187, 1981.