Stochastic Differential Equation Models with Sampled Data

Hermann Singer
FernUniversität in Hagen *

Abstract

Stochastic differential equations (SDE) are used as dynamical models for discrete time measurements (time series and panel data). Thus causal effects are formulated on a fundamental infinitesimal time scale. 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.).

In the linear case, the Kalman filter algorithm yields a recursive representation of likelihood and filtered states. Alternatively, structural equations models (SEM) with nonlinear parameter restrictions permit a nonrecursive computation of the likelihood function. The pro's and contras of both approaches are discussed. The Kalman approach is applied to a clinical data set with irregular sampling and missing data. Finally, the estimation of nonlinear models is discussed, using the extended Kalman filter (EKF) and Hermite expansions of the transition density.

Key Words: Itô calculus; Sampling; Exact discrete model; Continuous-discrete state space models; Kalman filtering; SEM modeling; Conditionally Gaussian models; Hermite orthogonal expansion; Nonlinear filtering; Ordinal measurements

Invited contribution to:

Kees van Montfort, Han Oud and Albert Satorra (eds.),
Longitudinal models in the behavioral and related sciences,
The European Association of Methodology (EAM)
Methodology and Statistics series, vol. II,
Erlbaum publishers
second draft; under review

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

1 Introduction

1.1 Continuous time models

Continuous time models coincide with the feeling, that time is a continuously flowing quantity without steps. On the other hand, data are mostly available at certain time points, e.g. daily, weekly, quarterly etc. or at arbitrary times. Therefore, there has been a tendency to formulate dynamical models in discrete time (times series analysis). Thus, the causal relations are specified between the arbitrary 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 system states. For example, a physical system like a pendulum (cf. the citation above) fulfils a simple linear relation 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.
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. In the sequel it is shown, that the strength and even sign of causal relations depends on the chosen sampling interval, even if the underlying process is the same.
Nevertheless there is a theory which can combine both points of view:

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

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

1.2 Differential equations

Mathematically speaking, **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 implicitly given in terms of time derivatives. For example, the simple **growth model**

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

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 ($a < 0$). The solution of (1) with initial condition $Y(t_0)$ is given by

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

which can be verified by computing the derivative $dY(t)/dt$. In a social science context, the simple deterministic equation must be extended by a random initial condition $Y(t_0, \omega)$ and 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)

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

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

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

Since the continuous time white noise process is a generalized random function (cf. Arnold, 1974, ch. 3), the solution is usually rewritten by the replacement $\zeta(s)ds = dW(s)$, where $W(s)$ is the Wiener process, a continuous, but not differentiable random walk process (cf. Arnold, ch. 4). Thus, in order to avoid derivatives of nondifferentiable processes, one writes symbolically

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

and interprets 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, 1974). 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 discrete time series (**exact discrete model; EDM**; Bergstrom 1976, 1988)

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

$Y_i := Y(t_i)$, or more concisely as

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

where Φ is the fundamental matrix (scalar case in (8)), but it should be noted that the parameters of the EDM 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 nonlinear restrictions**, especially in the multivariate case where (time ordered) matrix exponentials are involved (some references are Phillips, 1976, Jones, 1984, Hamerle et al., 1991, 1993, Singer, 1998). The term **time ordered** means, that the matrices $A(t)$ (see equation 9), which contain the causal effects, do not commute in general, i.e. $A(t)A(s) \neq A(s)A(t)$. This is a general property of matrices which must be considered in computing the fundamental matrix and the EDM (14).
Models with time-varying matrices are of empirical interest, since fluctuating exogenous variables may influence a time invariant system or the system itself changes its dynamics over time. For example, in the context of development psychology, the children get older in the course of a longitudinal study, so the causal effects are time dependent (cf. Singer, 1998). Similarly, the factor structure of a depression questionaire may be time dependent due to the psychological state of the subjects.

1.3 Advantages of differential equation models

Although the estimation of differential equations is more difficult than for discrete time models, there are many **advantages** of the approach (cf. Möbus and Nagl, 1983) :

1. The model specification of the system dynamics is independent of the measurement scheme and is given at the process level of the phenomenon (micro causality in the infinitesimal time interval dt).
2. The design of the study is specified by a measurement model, independently of the systems dynamics.
3. Changes of the variables can occur at any time, at and between the measurements. The state is defined for any time point, even if it can't be measured.
4. Extrapolation and interpolation of the data points can be obtained for arbitrary times and is not constrained to the sampling interval of the panel waves.
5. Studies with different or irregular sampling intervals can be compared since the continuous time structural parameters of the system do not depend on the measurement intervals.
6. Data sets with different sampling intervals can be analyzed together as one vector series.
7. Irregular sampling and missing data are treated in a unified framework. The parametrization is parsimonious since only the fundamental continuous time parameters must be estimated.

8. Cumulated or integrated data (flow data) can be represented explicitly.
9. Nonlinear transformations of data and variables can be handled by a differential calculus (Itô calculus).

2 Linear continuous-discrete state space models

2.1 Exact discrete model

In order to incorporate higher order derivatives (ARMA models), latent factors and errors of measurement, the **linear continuous/discrete state space model** (Jazwinski, 1970, ch. 7.2)

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

with random initial condition $Y(t_0) \sim N(\mu, \Sigma)$ is introduced. In the state equation (9), $W(t)$ denotes an 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 matrix $A : p \times p$ is called drift and $G : p \times r; \Omega = GG'$ is the square root of the diffusion matrix Ω. Furthermore, the vector b models deterministic control variables (stochastic control variables are already included in the model by extending the state vector $Y(t)$).
In the measurement equation (10), the measurement error $\epsilon_i \sim N(0, R(t_i,\psi)), R : k \times k$ is discrete time white noise and $H : k \times p$ contains factor loadings.
It should be noted that model (9–10) is very general and includes continuous time autoregressive integrated moving average (CARIMA) specifications, models with coloured noise, regression models with CARMA errors and dynamic factor analysis.
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). In this case, the system matrices may depend on deterministic regressors $x_n(t)$, i.e. $A(t,\psi) \to A(t, x_n(t), \psi) := A_n(t)$, etc.
Random effects may be added by specifying $d\pi_n = 0$ in the state equation and by extending the state $Y_n \to \{Y_n, \pi_n\}$. In the panel case, the parameters of the random initial condition $Y_n(t_0) \sim N(\mu(x_n(t_0), \psi), \Sigma(x_n(t_0), \psi))$ can be estimated as well.

Example 1: CAR(2) model

As mentioned above, 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 the sampling times t_i. It should be noted that no approximation whatsoever is involved. For example, the CAR(2) process (pendulum, linear oscillator; γ = friction,

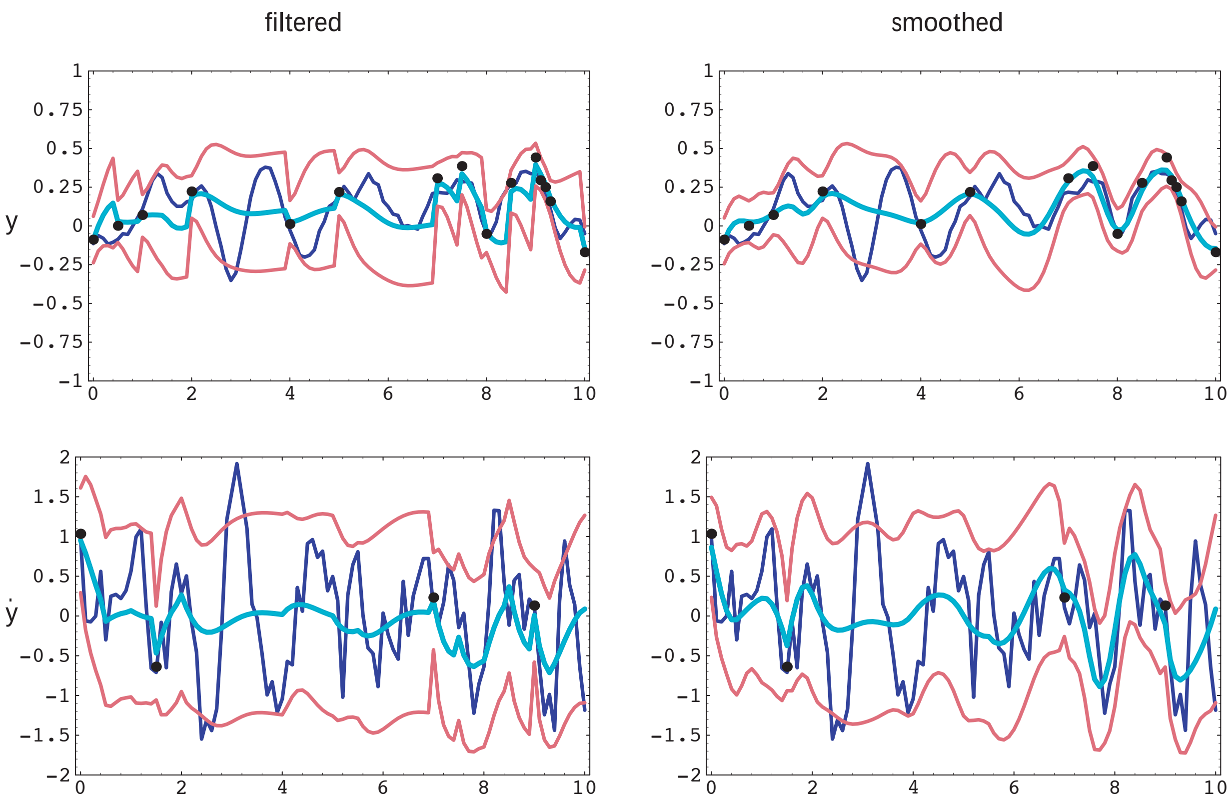

Figure 1: Linear oscillator with irregularly measured states (dots): Filtered state (left), smoothed state (right) with 95%-HPD confidence intervals. Measurements at $\tau_1 = \{0, .5, 1, 2, 4, 5, 7, 7.5, 8, 8.5, 9, 9.1, 9.2, 9.3, 10\}$ (first component; 1 st line), $\tau_2 = \{0, 1.5, 7, 9\}$ (2 nd component, 2 nd line). Discretization interval $\delta t = 0.1$. The controls $x(t)$ were measured at $\tau_3 = \{0, 1.5, 5.5, 9, 10\}$ (see main text).

$\omega_0 = 2\pi/T_o$ = angular frequency, T_o = period of oscillation)

$$\ddot{y} + \gamma\dot{y} + \omega_0^2 y = bx(t) + g\zeta(t) \tag{11}$$

with exogenous controls $x(t)$ has state space representation

$$d\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} := \begin{bmatrix} 0 & 1 \\ -\omega_0^2 & -\gamma \end{bmatrix}\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix}dt + \begin{bmatrix} 0 \\ b \end{bmatrix}x(t)dt + \begin{bmatrix} 0 & 0 \\ 0 & g \end{bmatrix}d\begin{bmatrix} W_1(t) \\ W_2(t) \end{bmatrix} \tag{12}$$

$$z_i := \begin{bmatrix} 1 & 0 \end{bmatrix}\begin{bmatrix} y_1(t_i) \\ y_2(t_i) \end{bmatrix} + \epsilon_i \tag{13}$$

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; cf. Singer, 1993). Approximating the derivatives by finite differences introduces unnecessary specification

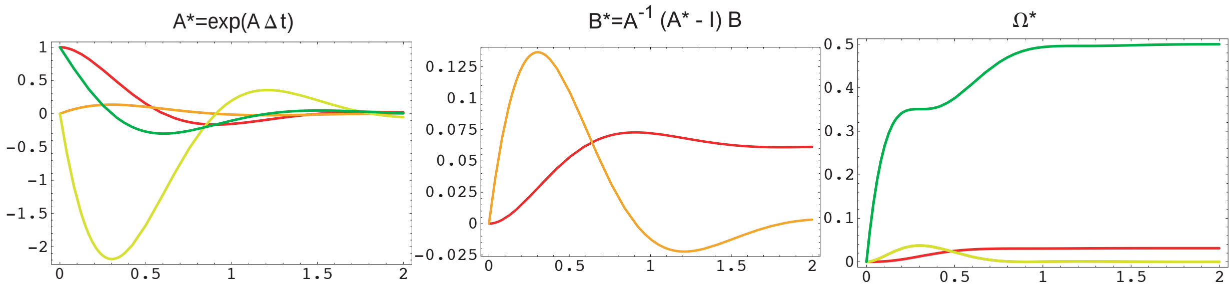

Figure 2: Linear oscillator: Exact discrete matrices $A^* = \exp(A\Delta t)$, $B^* = A^{-1}(A^* - I)B$, $\Omega^* = \int_0^{\Delta t} \exp(As)\Omega\exp(A's)ds$ as a function of measurement interval Δt. Note that the discrete time coefficients change their strength and even sign.

error and biased estimates. Figure 1 shows the true, filtered and smoothed trajectories of a irregularly sampled oscillator with measurement error $R = \mathrm{diag}(.01, .01)$ (cf. Singer, 1995) and 95% highest probability density (HPD) confidence intervals. Here the second component is measured, but at other time points than the first one. The ML estimates were inserted in the filter. The EDM in this example was formulated on a grid with spacing $\delta t = 0.1$ and all time points are expressed as multiples of this (arbitrary) discretization interval (cf. section 2.4). Figure 2 demonstrates, that the strength and even sign of the discrete time matrices of the EDM may change over time. **Thus, the interpretation of data should rely on the fundamental structural parameters of the SDE and not on the arbitrarily sampled discrete time models.**
The AR(2) model has been used for the sunspot activity (Bartlett, 1946; Singer, 1993), and more recently, for modeling the dynamics of married couples (Boker, 2004). In this case, two oscillators (pendulums) are connected by a spring with coupling term $d(y_1 - y_2)$ representing the force ∎

The state equation (9) can be solved for the times of measurement $t_i, i = 0, ..., T$ (**exact discrete model; EDM**)

$$\begin{aligned} Y(t_{i+1}) &= \Phi(t_{i+1}, t_i)Y(t_i) + \\ &+ \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s)b(s)ds + \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s)G(s)dW(s) \end{aligned} \tag{14}$$

with the parameter functionals (for a proof, see Arnold, 1974)

$$A_i^* := \Phi(t_{i+1}, t_i) = \overleftarrow{T} \exp[\int_{t_i}^{t_{i+1}} A(s)ds] \tag{15}$$

$$b_i^* := \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s)b(s)ds \tag{16}$$

where $\overleftarrow{T}\ A(t)A(s) = A(s)A(t); t < s$ is the Wick time ordering operator (cf. Abrikosov et al., 1963) and $\Phi(t_{i+1}, t_i)$ is the state transition matrix solving

$$\begin{aligned} \tfrac{d}{dt}\Phi(t, t_i) &= A(t)\Phi(t, t_i) && (17) \\ \Phi(t_i, t_i) &= I. && (18) \end{aligned}$$

In shorthand, one obtains the vector autoregression VAR(1) scheme

$$\begin{aligned} Y_{i+1} &= A_i^* Y_i + b_i^* + u_i && (19) \\ Z_i &= H_i Y_i + d_i + \epsilon_i. && (20) \end{aligned}$$

with covariance matrix

$$\mathrm{Var}(u_i) \;:=\; \Omega_i^* = \int_{t_i}^{t_{i+1}} \Phi(t_{i+1}, s)G(s)G'(s)\Phi'(t_{i+1}, s)ds. \qquad (21)$$

Example 2: Growth model with time dependent rates

The time dependent model of growth rates with constant part

$$A_0 = \begin{bmatrix} \lambda_{11} & \lambda_{12} \\ \lambda_{21} & \lambda_{22} \end{bmatrix} \qquad (22)$$

and a linearly changing part (due to development processes)

$$A_1(t) = \begin{bmatrix} \alpha\, t & 0 \\ 0 & \beta\, t \end{bmatrix}. \qquad (23)$$

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 is not simply given by the matrix exponential

$$\Phi(t|t_i) \;=\; \exp[\int_{t_i}^{t} A(s)ds] \qquad (24)$$

but by the time ordered expression

$$\Phi(t|t_i) \;=\; \overleftarrow{T}\exp[\int_{t_i}^{t} A(s)ds] \qquad (25)$$

Using the moment equations (26,27) with a respective numerical approximation (Euler- or higher order Runge-Kutta scheme) does automatically produce the time ordered expressions in the Kalman filter or the exact discrete model (14). This is the crucial step a software package must take in order to implement the correct parameter functionals (for details see Singer, 1998) ■

2.2 Filtering and likelihood function

The likelihood function can be computed recursively by means of the **Kalman filter algorithm** (Jazwinski, 1970, Liptser and Shiryayev, 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 $\Sigma(t|t_i) = \text{Var}[Y(t)|Z^i]$, where $Z^i = \{Z_i, ..., Z_0\}$ are the measurements up to time t_i. The **time updates** fulfil ($t \in [t_i, t_{i+1}]$)

$$
\begin{aligned}
(d/dt)\mu(t|t_i) &= A(t,\psi)\mu(t|t_i) + b(t,\psi) && (26)\\
(d/dt)\Sigma(t|t_i) &= A(t,\psi)\Sigma(t|t_i) + \Sigma(t|t_i)A'(t,\psi) + \Omega(t,\psi) && (27)
\end{aligned}
$$

where $\Omega(t,\psi) = GG'(t,\psi)$ is the diffusion matrix. Initial conditions are $\mu(t_i|t_i)$ and $\Sigma(t_i|t_i)$ (a posteriori moments at time t_i). The moment equations (26–27) may be solved explicitly. Using the fundamental matrix one obtains (dropping ψ)

$$
\begin{aligned}
\mu(t_{i+1}|t_i) &= \Phi(t_{i+1},t_i)\mu(t_i|t_i) + \\
&+ \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)b(s)ds && (28)\\
\Sigma(t_{i+1}|t_i) &= \Phi(t_{i+1},t_i)\Sigma(t_i|t_i)\Phi'(t_{i+1},t_i) + \\
&+ \int_{t_i}^{t_{i+1}} \Phi(t_{i+1},s)\Omega(s)\Phi'(t_{i+1},s)ds. && (29)
\end{aligned}
$$

These results may be obtained as well by taking the conditional (on Z^i) mean and variance of the EDM (14).

At the measurement times, the time update (optimal prediction) is corrected by the measurement Z_{i+1} using the Bayes formula leading to the **measurement update**

$$
\begin{aligned}
\mu(t_{i+1}|t_{i+1}) &= \mu(t_{i+1}|t_i) + K(t_{i+1}|t_i)\nu(t_{i+1}|t_i) && (30)\\
\Sigma(t_{i+1}|t_{i+1}) &= [I - K(t_{i+1}|t_i)H(t_{i+1})]\Sigma(t_{i+1}|t_i) && (31)\\
\nu(t_{i+1}|t_i) &= Z_{i+1} - Z(t_{i+1}|t_i) && (32)\\
Z(t_{i+1}|t_i) &= H(t_{i+1})\mu(t_{i+1}|t_i) + d(t_{i+1}) && (33)\\
\Gamma(t_{i+1}|t_i) &= H(t_{i+1})\Sigma(t_{i+1}|t_i)H'(t_{i+1}) + R(t_{i+1}) && (34)\\
K(t_{i+1}|t_i) &:= \Sigma(t_{i+1}|t_i)H'(t_{i+1})\Gamma(t_{i+1}|t_i)^{-1} && (35)
\end{aligned}
$$

where $K(t_{i+1}|t_i)$ is the **Kalman gain**, $Z(t_{i+1}|t_i)$ 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

$$
l(\psi; Z) = \log p(Z_T, \ldots, Z_0; \psi) = \sum_{i=0}^{T-1} \log p(Z_{i+1}|Z^i;\psi)p(Z_0), \tag{36}
$$

where the transition densities are given in terms of the Gaussian distribution

$$
p(Z_{i+1}|Z^i;\psi) = \phi(\nu(t_{i+1}|t_i); 0, \Gamma(t_{i+1}|t_i)). \tag{37}
$$

Thus the Kalman filter computes the likelihood recursively in terms of predictions, prediction errors and their conditional variance (cf. Jazwinski, 1970). The time update is the dynamical prediction starting from the information at time t_i, whereas the measurement update incorporates the new measurement information at t_{i+1}. It has the form of a linear regression model (30), where the Kalman gain is the regression parameter matrix and the time update is the intercept. This replicates the well known fact, that in a Gaussian system information is optimally incorporated by a linear regression model. If there are deviations from normality, it is still the best linear estimate (cf. Liptser and Shiryayev, 2001, vol. 2, lemma 14.1).
It may be stated that the **parameter estimation task for the linear state space model can be carried out 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), Jones and Boadi-Boateng (1991), Jones (1993), Singer (1995, 1998).

2.3 Conditionally Gaussian models

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. 13). Thus, although the latent states $Y(t_i)$ may be nongaussian due to the Z-dependence, the conditionally Gaussian filtering scheme is still valid. For example, the diffusion matrix $\Omega(t)$ may contain delayed prediction errors $\Omega(t, \nu(t_i|t_{i-1}), ...)$ and thus autoregressive conditional heteroskedasticity (ARCH) effects in a continuous time model.

2.4 Computational aspects and simplifications

The moment equations (28) and the EDM (14) contain integrals over the fundamental matrix which must be solved numerically. Since the solution

$$\Phi(t_{i+1}, t_i) = \overleftarrow{T} \exp[\int_{t_i}^{t_{i+1}} A(s)ds] \tag{38}$$

for the fundamental matrix is only formal, one seeks a numerical solution of

$$\frac{d}{dt}\Phi(t, t_i) = A(t)\Phi(t, t_i) \tag{39}$$

$$\Phi(t_i, t_i) = I, \tag{40}$$

e.g. the Euler approximation

$$\Phi(\tau_{j+1}, t_i) \approx [I + A(\tau_j)\delta t]\Phi(\tau_j, t_i) \tag{41}$$

$$\tau_j = t_i + j\delta t; j = 0, ..., J-1; J = \Delta t_i/\delta t \tag{42}$$

on a fine grid with (arbitrary) spacing $\delta t \to 0$. This expression is analogous to the formula $e^x = \lim_{n\to\infty}(1 + x/n)^n$ and automatically incorporates the time ordering of $A(s)$. More generally, one solves the moment equations (26–27) by appropriate discrete schemes in the intervals $[t_i, t_{i+1}]$, e.g.

$$\mu(\tau_{j+1}|t_i) \quad \approx \quad [I + A(\tau_j)\delta t]\mu(\tau_j|t_i) + b(\tau_j)\delta t \tag{43}$$

etc. or higher order Runge Kutta schemes. The EDM may be treated analogously. In many cases, simplified model are sufficient, for example the panel model with constant coefficients (Singer, 1991, 1993)

$$dY_n(t) \quad = \quad [A(\psi)Y_n(t) + B(\psi)x_n(t)]dt + G(\psi)dW_n(t) \tag{44}$$
$$Z_{ni} \quad = \quad H(\psi)Y_n(t_i) + D(\psi)x_n(t_i) + \epsilon_{ni}. \tag{45}$$

Here, the influence of exogenous (control) variables is parametrized by setting $b(t) = Bx_n(t)$. Since $x_n(t)$ is in general only known at times t_i, the parameter functional (16) can be computed only approximately if $x_n(t_i)$ is interpolated. For example, the EDM is given by the explicit matrices

$$A_i^* \quad = \quad \exp(A\Delta t_i) \tag{46}$$
$$B_i^* \quad = \quad A^{-1}(A_i^* - I)B \tag{47}$$
$$\Omega_i^* \quad = \quad \int_0^{\Delta t_i} \exp(As)\Omega\exp(A's)ds \tag{48}$$

if the control variables are (approximated by) piecewise constant regressors (step functions). Similar, but more complicated formulas may be obtained if the controls are polygonal lines or other approximations between measurements (cf. Phillips, 1976, Hamerle, Nagl and Singer, 1993).
The numerical solution of the moment equations (26–27) is more convenient, however, since arbitrary interpolations of $x_n(t_i)$ can be used without explicit computation of the EDM (cf. Singer, 1995, 1998). The exogenous variables may even be measured at other times than the endogenous state.
A variant of the EDM is formulated on a fine uniform grid with spacing $\delta t, \tau_j = t_0 + j\delta t; t_i = t_0 + j_i\delta t, i = 0, ..., T$, such that all sampling times can be expressed as multiples of this unit. Then, one obtains the EDM scheme

$$Y_{j+1} \quad = \quad A_j^* Y_j + b_j^* + u_j \tag{49}$$
$$Z_j \quad = \quad H_j Y_j + d_j + \epsilon_j. \tag{50}$$

$j = 0, ..., j_T - 1; j_T = (t_T - t_0)/\delta t$, with matrices

$$A_j^* \quad := \quad \Phi(\tau_{j+1}, \tau_j) = \overleftarrow{T}\exp[\int_{\tau_j}^{\tau_{j+1}} A(s)ds] \approx \exp(A_j\delta t) \tag{51}$$

$$b_j^* \quad := \quad \int_{\tau_j}^{\tau_{j+1}} \Phi(\tau_{j+1}, s)b(s)ds \approx b_j\delta t \tag{52}$$

$$\mathrm{Var}(u_j) \quad := \quad \Omega_j^* = \int_{\tau_j}^{\tau_{j+1}} \Phi(\tau_{j+1}, s)\Omega(s)\Phi'(\tau_{j+1}, s)ds \approx \Omega_j\delta t. \tag{53}$$

If the discretization interval δt is short, one can approximate (piecewise constant) $A_j^* \approx \exp(A_j\delta t); A_j = A(\tau_j)$, $b_j^* \approx A_j^{-1}(A_j^* - I)b_j \approx b_j\delta t$, $\mathrm{row}(\Omega_j^*) \approx L_j^{-1}(A_j^* \otimes A_j^* -$

$I)\mathrm{row}(\Omega_j) \approx \mathrm{row}(\Omega_j)\delta t$, $L_j = A_j \otimes I + I \otimes A_j$ (cf. Singer, 1995, 1998).[1] Then, the irregular sampling intervals $\Delta t_i = (j_{i+1} - j_i)\delta t$ can be bridged by a missing data treatment if Z_j is not observed (cf. 1). In the limit $\delta t \to 0$ one obtains an alternative numerical scheme for the parameter functionals of the irregular EDM (15) and/or the conditional moments (28). In the case of constant parameter matrices, only one set of EDM matrices must be computed and the exogenous variables $x(t)$ can vary arbitrarily between sampling times.

2.5 Estimation with structural equations models (SEM)

The EDM

$$\begin{aligned} Y_{i+1} &= A_i^* Y_i + b_i^* + u_i && (54) \\ Z_i &= H_i Y_i + d_i + \epsilon_i. && (55) \end{aligned}$$

$i = 0, ..., T-1$ may be represented by the matrix equation (cf. Oud et al., 1993, Oud and Jansen, 2000)

$$\begin{aligned} \eta &= B\eta + \Gamma + \zeta && (56) \\ Y &= \Lambda\eta + \tau + \epsilon && (57) \end{aligned}$$

where $\eta' = [Y_0', ..., Y_T']$ is the sampled trajectory, Γ is a **deterministic** intercept term, $\zeta' = [\zeta_0', u_0', ..., u_{T-1}']$ is the vector of process errors,

$$B = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ A_0^* & 0 & 0 & \dots & 0 \\ 0 & A_1^* & 0 & \dots & 0 \\ \vdots & 0 & \ddots & 0 & 0 \\ 0 & 0 & \dots & A_{T-1}^* & 0 \end{bmatrix} \quad (58)$$

$$\Gamma = \begin{bmatrix} \mu \\ b_0^* \\ b_1^* \\ \vdots \\ b_{T-1}^* \end{bmatrix} \quad (59)$$

$$\mathrm{Var}(\zeta) = \begin{bmatrix} \Sigma & 0 & 0 & \dots & 0 \\ 0 & \Omega_0^* & 0 & \dots & 0 \\ 0 & 0 & \Omega_1^* & \dots & 0 \\ \vdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & \Omega_{T-1}^* \end{bmatrix} \quad (60)$$

are the structural matrices, and

$$\Lambda = \begin{bmatrix} H_0 & 0 & 0 & \dots & 0 \\ 0 & H_1 & 0 & \dots & 0 \\ 0 & 0 & H_2 & \dots & 0 \\ \vdots & 0 & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & H_T \end{bmatrix} \quad (61)$$

[1] row is the row-wise vector operator and $\otimes$ is the Kronecker product.

$$\tau = \begin{bmatrix} d_0 \\ d_1 \\ d_2 \\ \vdots \\ d_T \end{bmatrix} \tag{62}$$

$$\mathrm{Var}(\epsilon) = \begin{bmatrix} R_0 & 0 & 0 & \dots & 0 \\ 0 & R_1 & 0 & \dots & 0 \\ 0 & 0 & R_2 & \dots & 0 \\ \vdots & \vdots & 0 & \ddots & 0 \\ 0 & 0 & \dots & 0 & R_T \end{bmatrix} \tag{63}$$

are the factor loading, **deterministic** intercept and error matrices of the measurement model. Solving for η one obtains the solution of the SDE for the time points t_i

$$\eta = (I - B)^{-1}(\mathit{\Gamma} + \zeta). \tag{64}$$

In this equation, the initial condition is represented by $\eta_0 = y(t_0) = \mu + \zeta_0 \sim N(\mu, \mathit{\Sigma})$. If the matrices do not depend on measurements Z_t, the system is multivariate Gaussian[2] and the log likelihood function is given by (omitting constants)

$$l = -\tfrac{1}{2}(\log|\mathit{\Sigma}_y| + \mathrm{tr}[\mathit{\Sigma}_y^{-1}(Y - \mu_y)(Y - \mu_y)']), \tag{65}$$

where

$$\mu_y = E[Y] = \mathit{\Lambda}(I - B)^{-1}\mathit{\Gamma} + \tau \tag{66}$$
$$\mathit{\Sigma}_y = \mathrm{Var}(Y) = \mathit{\Lambda}(I - B)^{-1}\mathit{\Sigma}_\zeta(I - B)^{-T}\mathit{\Lambda}' + \mathit{\Sigma}_\epsilon. \tag{67}$$

In the panel case, one has N trajectories $Y_n, n = 1, ...N$ and the likelihood is given by (assuming $b(t) = b_n(t), d(t) = d_n(t)$)

$$l = -\tfrac{N}{2}(\log|\mathit{\Sigma}_y| + \mathrm{tr}[\mathit{\Sigma}_y^{-1}\tfrac{1}{N}\sum(Y_n - \mu_{yn})(Y_n - \mu_{yn})']). \tag{68}$$

where $\mu_{yn} = E[Y_n] = \mathit{\Lambda}(I - B)^{-1}\mathit{\Gamma}_n + \tau_n$, and $\mathit{\Gamma}'_n = [\mu'_n, (b^*_{0n})', ..., (b^*_{T-1,n})']$, $\tau'_n = [d'_{0n}, ..., d'_{T-1,n}]$.[3]
These expressions may again be simplified by assuming

$$b_n(t) = B(t, \psi)x_n(t) \tag{69}$$
$$d_n(t) = D(t, \psi)x_n(t) \tag{70}$$

and stepwise constant controls yielding

$$b^*_{ni} = [\int_{t_i}^{t_{i+1}} \mathit{\Phi}(t_{i+1}, s)B(s, \psi)ds]x_{ni} \tag{71}$$
$$:= B^*_i x_{ni}. \tag{72}$$

[2]but cf. section 2.3
[3]Assuming that the matrices A, G etc. do not depend on n, but see section 2.1.

In this case, one can factorize $\Gamma_n = \Gamma X_n$, $\tau_n = \tau X_n$, where

$$X_n = \begin{bmatrix} 1 \\ x_{n0} \\ x_{n1} \\ \vdots \\ x_{nT} \end{bmatrix} : (T+1)q + 1 \times 1 \tag{73}$$

$$\Gamma = \begin{bmatrix} \mu & 0 & 0 & \dots & 0 & 0 \\ 0 & B_0^* & 0 & \dots & 0 & 0 \\ 0 & 0 & B_1^* & \dots & 0 & 0 \\ \vdots & 0 & \ddots & 0 & 0 & 0 \\ 0 & 0 & \dots & 0 & B_{T-1}^* & 0 \end{bmatrix} : (T+1)p \times (T+1)q + 1 \tag{74}$$

$$\tau = \begin{bmatrix} 0 & D_0 & 0 & \dots & 0 & 0 \\ 0 & 0 & D_1 & 0 & \dots & 0 \\ 0 & 0 & 0 & D_2 & \dots & 0 \\ 0 & \vdots & 0 & \ddots & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & D_T \end{bmatrix} : (T+1)k \times (T+1)q + 1. \tag{75}$$

The SEM now reads (X_n are **deterministic** controls) [4]

$$\eta_n = B\eta_n + \Gamma X_n + \zeta_n \tag{76}$$

$$Y_n = \Lambda\eta_n + \tau X_n + \epsilon_n \tag{77}$$

and one can rewrite the likelihood for the N independent panel units in terms of data matrices $Y' = [Y_1, ...Y_N] : (T+1)k \times N$, $X' = [X_1, ...X_N] : (T+1)q + 1 \times N$ as

$$l = -\tfrac{N}{2}(\log|\Sigma_y| + \mathrm{tr}[\Sigma_y^{-1}(M_y + CM_xC' - M_{yx}C' - CM_{xy})]), \tag{78}$$

where

$$E[Y_n] = [\Lambda(I-B)^{-1}\Gamma + \tau]X_n := CX_n \tag{79}$$

and the moment matrices are $M_y = Y'Y : (T+1)k \times (T+1)k$, $M_x = X'X : (T+1)q + 1 \times (T+1)q + 1$, $M_{yx} = Y'X : (T+1)k \times (T+1)q + 1$.
In the case of constant parameter matrices, further simplifications are possible.

Discussion

SEM software must be able to incorporate the nonlinear restrictions as created by the EDM or by the moment equations (26,27). It is very easy to implement the likelihood function (78) in a matrix language (e.g. Mathematica or SAS/IML) and to write all matrices as nonlinear functions of a parameter vector ψ.
The Kalman filter (KF) and the SEM approach are different in some respects:

1. The KF computes the likelihood recursively for the data $Z = \{Z_0, ...Z_T\}$, i.e. the conditional distributions $p(Z_{t+1}|Z^t)$ are updated step by step, whereas the SEM representation utilizes the joint distribution of the vector $\{Z_0, ...Z_T\}$.

[4] This model was implemented as a Mathematica program (SEM, Singer, 2004b) and is available from the author.

2. Therefore, the KF can work online, since new data update the conditional moments and the likelihood, whereas the SEM uses the batch of data $Z = \{Z_0, ...Z_T\}$ with dimension $(T+1)k$. The KF only involves the data point $Z_t : k \times 1$ and one has to invert matrices of order $k \times k$ (prediction error covariance). The SEM must invert the matrices $\Sigma_y : (T+1)k \times (T+1)k$ and $B : (T+1)p \times (T+1)p$ in each likelihood computation.

   This will be a serious problem if long data sets $T > 100$ are analyzed, but not for short panels.

3. The KF also works in the conditionally Gaussian case, since $p(Z_{t+1}|Z^t)$ is still Gaussian, whereas the joint distribution of $Z = \{Z_0, ...Z_T\}$ is not Gaussian any more.

4. As a consequence, the KF approach can be easily generalized to nonlinear systems (extended Kalman filter EKF etc.), since the transition probabilites are still approximately conditionally Gaussian (see ch. 3).

5. The SEM approach is more familiar to many scientists used to work with LISREL and other programs. In the early days of SEM modeling, only linear restrictions could be implemented, but now the system (76–77) and its likelihood (78) can be easily programmed and maximized using matrix software like Mathematica, SAS/IML etc.

6. Filtered estimates of the latent states are computed recursively by the KF (the conditional moments), and smoothed trajectories can be computed by a (fixed interval) smoother algorithm. On the other hand, in the SEM approach, one can compute the conditional expectations $E[\eta|Y]$ and $\mathrm{Var}[\eta|Y]$ yielding the smoothed estimates, but again matrices of order $(T+1)k \times (T+1)k$ are involved.

7. Missing data may be treated in both cases by modifying the measurement model. The Kalman filter processes the data $z_n(t_i) : k \times 1$ for each time point and panel unit. Thus, the missing data treatment can be automatically included in the measurement update by dropping missing entries in the matrices. In the SEM approach, the so called individual likelihood approach may be utilized.

In my opinion, software with a 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.

2.6 Other issues

2.6.1 Flow data

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.

2.6.2 Missing data

Often, **missing data** occur in panel designs or data are measured at **arbitrary frequencies**. Sometimes, on may want 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; see clinical example section 2.7). Missing data can be treated easily 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.

2.6.3 Random effects

Individual specific random effects π_n can be treated in the state space model by writing $(n = 1, \ldots, N)$

$$
\begin{aligned}
d\begin{bmatrix} y_n \\ \pi_n \end{bmatrix} &= \begin{bmatrix} A & I \\ 0 & 0 \end{bmatrix}\begin{bmatrix} y_n \\ \pi_n \end{bmatrix}dt + \begin{bmatrix} G \\ 0 \end{bmatrix}dW_n(t) \qquad (80)\\
z_{ni} &= \begin{bmatrix} H & 0 \end{bmatrix}\begin{bmatrix} y_{ni} \\ \pi_{ni} \end{bmatrix} + \epsilon_{ni},
\end{aligned}
$$

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

2.7 Clinical application

The differential equation method was applied to clinical data which were collected in a psychiatric hospital at arbitrary time points. [5] The data set contains many behavioral, socioeconomic and medical variables. Here we choose the three variables $y(t) = \{weight, neuroleptica\ dose, clinical\ impression\}$, where 'clinical impression': 2,...,8, is a rating scale with lower values corresponding to a better health state.
The data set is highly irregular, since the number $T+1$ of measurement times ranges between persons from 2 to 21 time points, and the sampling intervals Δt_i vary from 1 to 1574 days corresponding to 4.31 years (cf. figures 4–6). Moreover, missing values are present in some components of the state vector. Nevertheless, the proposed methodology is able to model and estimate the dynamical connection between the state components $y = \{y_1, y_2, y_3\}$. In a first step I used a linear SDE of the form (vector CAR(1) model; the panel index n is dropped)

$$dy(t) = A[y(t) - m]dt + GdW(t) \qquad (81)$$

[5] I would like to thank Dr. Matthias Dobmeier, Arbeitsgruppe Psychopharmakologie, Harald Binder, Dipl. Psych., Arbeitsgruppe Versorgungsforschung, Klinik und Poliklinik für Psychiatrie und Psychotherapie der Universität Regensburg am Bezirksklinikum, Direktor: Prof. Dr. H.E. Klein, for the sampling, preparation and permission to analyze the data set.

or explicitly

$$dy_1 = [a_{11}(y_1 - m_1) + a_{12}(y_2 - m_2) + a_{13}(y_3 - m_3)]dt + g_{11}dW_1(t) \quad (82)$$
$$dy_2 = [a_{21}(y_1 - m_1) + a_{22}(y_2 - m_2) + a_{23}(y_3 - m_3)]dt + g_{22}dW_2(t) \quad (83)$$
$$dy_3 = [a_{31}(y_1 - m_1) + a_{32}(y_2 - m_2) + a_{33}(y_3 - m_3)]dt + g_{33}dW_3(t) \quad (84)$$

where y_1 = weight (kg), y_2 = neuroleptica dose (mg), y_3 = clinical impression (2 (better),...,8 (worse)). The parameter vector m can be interpreted as the asymptotic mean value ($t \to \infty$) of $y(t)$. Using only subjects with at least 2 measurement times a number of $N = 384$ persons was selected and the parameters in $A, m, G, \mu = E[y_{n0}], \Sigma = \mathrm{Var}(y_{n0})$ were estimated with the exact discrete model (EDM; $A^* = \exp(A\delta t), etc.$) and a linearized EDM ($A^* \approx I + A\delta t$) by using the Kalman filter. Since all sampling times can be expressed as multiples of $\delta t = 1$ day, the exact discrete model is formulated for this discretization interval (cf. section 2.4).
The model specification is motivated by the fact that the panel is nonstationary, since neuroleptica are known to increase the weight of the subjects. Thus we have an initial level expressed by the parameters μ and Σ and a mean level expressed by m. Indeed, comparing the estimates $\hat{\mu}_1$ and $\hat{m}_1$ we see a mean increase of about 5 kg (over all persons and times). The coefficients a_{ij} are rate constants of the action of variables $i \leftarrow j$, i.e. the relative change of variable $dy_i/((y_j - m_j)dt)$ in a small time interval dt, by keeping the other variables at their mean level m. The action over the interval Δt is given by the regression parameter $A^*(\Delta t) := \exp(A\Delta t)$. Thus the results of the SDE estimation are independent of any sampling intervals and different persons can be analyzed jointly in a panel. The parameter matrices relate to the fundamental infinitesimal change in the interval dt. In contrast, a discrete time model would have to use many autoregressive and error parameter matrices, each one for the different sampling intervals within and between subjects. Thus, a SDE model is parsimonious since all measured quantities are expressed in terms of only one set of matrices A, B, Ω.

Results

Table (5) shows the ML estimates, asymptotic standard deviations and t values. Comparing the initial level μ and the asymptotic saturation mean level m, the weight changes from 75.42 to 80.01 kilograms. A Wald test

$$H_0 : R\theta = b \quad (85)$$
$$W = (R\hat{\theta} - b)'(RCR)^{-1}(R\hat{\theta} - b) \sim \chi^2(\mathrm{rank}R), \quad (86)$$

$C = F^{-1}$ (asymptotic covariance of $\hat{\theta}$; F= Fisher information matrix), shows that this difference is significant ($W = 6.62881$, $\chi^2(0.95, 1)$-quantile $= 3.84146$). Here, the hypothesis matrix is $R = [0, 0, ...1, 0.... - 1, 0, ...0]$, where the 1 entries are at columns 10 and 16 (cf. the parameter list in Table 1) and $b = 0$.
The neuroleptica dose slightly increases (14.89 to 16.01; $W = 4.48754, \chi^2(0.95, 1) = 3.84146$) and the clinical impression is about the same (4.39 to 4.51; $W = 3.13101$, $\chi^2(0.95, 1) = 3.84146$). There is a nonsignificant negative initial covariance $\hat{\sigma}_{12} = -3.16446$ between weight and dose, but the action weight$\leftarrow$ neuroleptica is positive ($\hat{a}_{12}$

$= 0.00299021$ with t-value of 0.990259). The only significant ($\alpha = 5\%$) interaction parameter is $\hat{a}_{32}$=0.00167256 (clinical impression← neuroleptica), which means that higher dose leads to a higher clinical impression score (worse) later. The reverse effect is negative, but nonsignificant (this effect would mean an influence of clinical impression on the medication). The model shows that an initial level μ changes asymptotically to a saturation mean level m with higher weight and slightly higher dose (figure 3), but this effect cannot be significantly explained by the causal action of dose on weight. Figures (4–8) display the filtered and smoothed trajectories of 3 selected persons. At the points of measurement, the extrapolated state $E[y(t)|Z^i], t > t_i$ is corrected by newly incoming information. Between measurements, it tends to the asymptotic mean level m. Fig. 7 shows the case, when a nonmeasured state component (weight at time $t = 63$ days) is corrected by the other measured states. Finally, the smoothed trajectory $E[y(t)|Z^T]$ uses information from the past and future and yields a smooth interpolation of the data points.
Further analysis of the data would impose restrictions on the parameter matrices followed by a comparision of the models with information criteria like AIC (Akaike's information criterion) etc. Also, higher order derivatives could be added, e.g. a CAR(2) vector autoregression. Furthermore, using the EKF (cf. section 3.1), nonlinear specifications could be estimated. However, in contrast to the linear case, an infinity of possible drift and diffusion specifications is conceivable.

Computational aspects

The data set was analyzed by using a Mathematica/C-implementation of the SAS/IML package LSDE (Singer, 1991) [6], where each panel unit n can have different sample size T_n. The Mathematica code for the linear drift function $f(y, x, \psi) = Ay+b = A(y-m)$ is similar to (82–84)

```
f[{y1_,y2_,y3_},x_,
  {a11_,a12_,a13_,a21_,a22_,a23_,a31_,a32_,a33_,
    m1_,m2_,m3_,g11_,g22_,g33_,
   mu1_,mu2_,mu3_,s11_,s12_,s13_,s22_,s23_,s33_}]:=

  {a11 (y1-m1) + a12 (y2-m2) + a13 (y3-m3),
   a21 (y1-m1) + a22 (y2-m2) + a23 (y3-m3),
   a31 (y1-m1) + a32 (y2-m2) + a33 (y3-m3)}
```

In the above code, x is the q-vector $x(t)$ of exogenous variables (not used here). One only has to specify the drift $f(y, x, \psi)$, diffusion function $G(x, \psi)$ and the measurement model given by $h(y, x, \psi) = Hy + d$ and $R(x, \psi)$. Also nonlinear specifications are possible (extended Kalman filter EKF; see section 3.1). The software implements the EDM (49–50). The likelihood for unit n is obtained by the command

```
cEKFEDM[{Z, X}, psi, {{f, A, G, h, H, R, mue, sigma},
```

[6] A new version called SDE (Stochastic Differential Equations) running on Mathematica/C will appear soon. It covers the linear models as well as nonlinear algorithms such as EKF, SNF etc.

θ	$\hat{\theta}$	Std	t-value	interpretation
a_{11}	-0.00529433	0.000598578	-8.84483	weight$\leftarrow$ weight
a_{12}	0.00299021	0.00301962	0.990259	weight$\leftarrow$ neuroleptica
a_{13}	0.0234382	0.0437436	0.535809	weight$\leftarrow$ clinical impression
a_{21}	0.00198898	0.00143129	1.38964	neuroleptica$\leftarrow$ weight
a_{22}	-0.0505835	0.00470394	-10.7534	neuroleptica$\leftarrow$ neuroleptica
a_{23}	-0.0090749	0.0614409	-0.147701	neuroleptica$\leftarrow$ clinical imp.
a_{31}	0.0000800229	0.000453001	0.176651	clinical impression$\leftarrow$ weight
a_{32}	0.00167256	0.000670036	2.49623	clinical imp.$\leftarrow$ neuroleptica
a_{33}	-0.0933535	0.0127247	-7.33641	clinical imp.$\leftarrow$ clinical imp.
m_1	80.0107	1.50742	53.0779	mean weight
m_2	16.0973	0.36628	43.9481	mean dose
m_3	4.50905	0.0439121	102.684	mean impression
g_{11}	1.72074	0.0497599	34.5808	error weight
g_{22}	2.54972	0.0979298	26.0362	error dose
g_{33}	0.38952	0.0237472	16.4028	error impression
μ_1	75.4199	0.922667	81.7412	initial weight
μ_2	14.8932	0.43461	34.2681	initial dose
μ_3	4.39445	0.0472563	92.9919	initial impression
σ_{11}	283.029	21.9171	12.9136	initial weight variance
σ_{12}	-3.16446	13.6215	-0.232314	initial weight-dose covar.
σ_{13}	0.5679	0.872826	0.650645	initial weight-imp. covar.
σ_{22}	72.533	5.23379	13.8586	initial dose variance
σ_{23}	0.667828	0.452169	1.47694	initial dose-imp. covar.
σ_{33}	0.794195	0.0601353	13.2068	initial dose variance

Table 1: ML estimates of the continuous time panel model based on EDM with variables weight, neuroleptica dose and clinical impression.

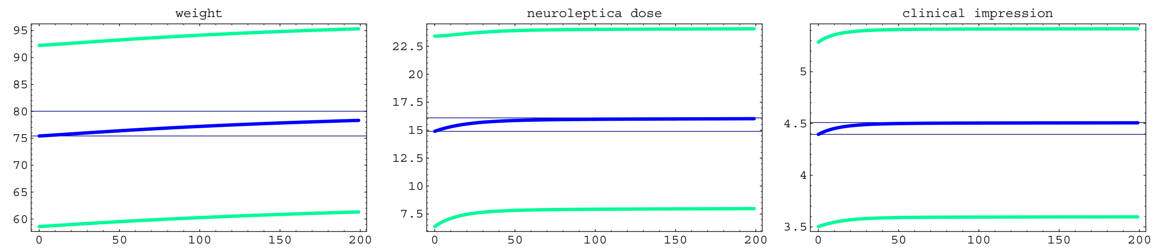

Figure 3: Mean level change from initial condition $\hat{\mu}(t_0) = \hat{\mu}$ to $\hat{\mu}(t \to \infty) = \hat{m}$ (interval $t = 0, ..., 200$ days). Also shown is the standard deviation $\mathrm{std} = \sqrt{\widehat{\mathrm{Var}}(y(t))}$.

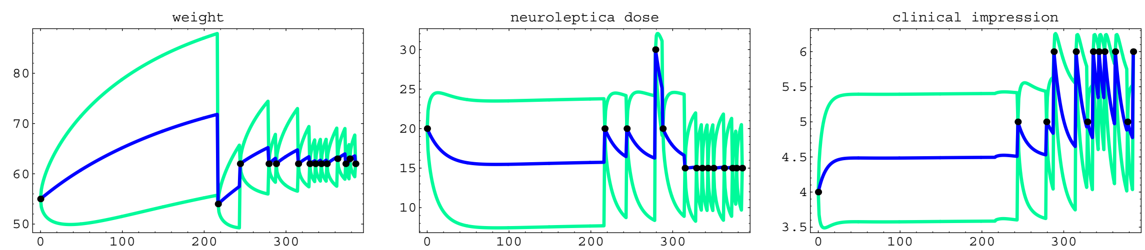

Figure 4: Filtered estimates with data points and 67%-HPD confidence intervals. Female with age 49 and ICD 10 diagnosis F20.

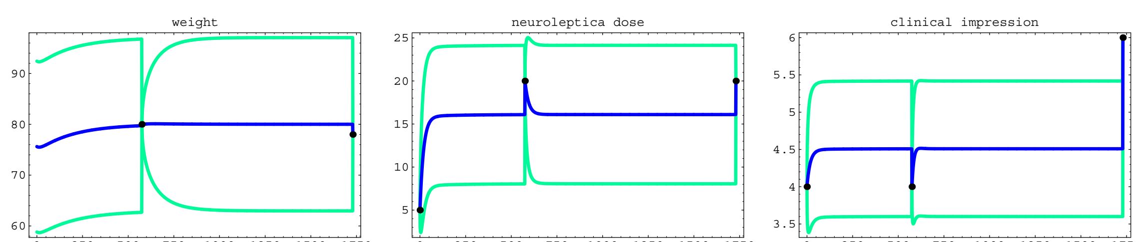

Figure 5: Filtered estimates with data points and 67%-HPD confidence intervals. Male, age 28, ICD diagnosis F20.

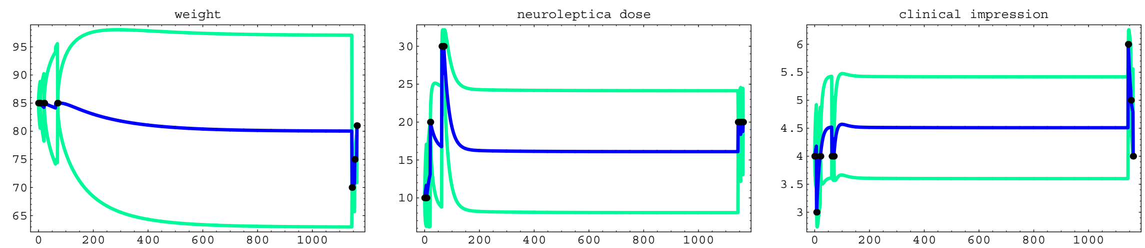

Figure 6: Filtered estimates with data points and 67%-HPD confidence intervals. Female, age 48, diagnosis F20.

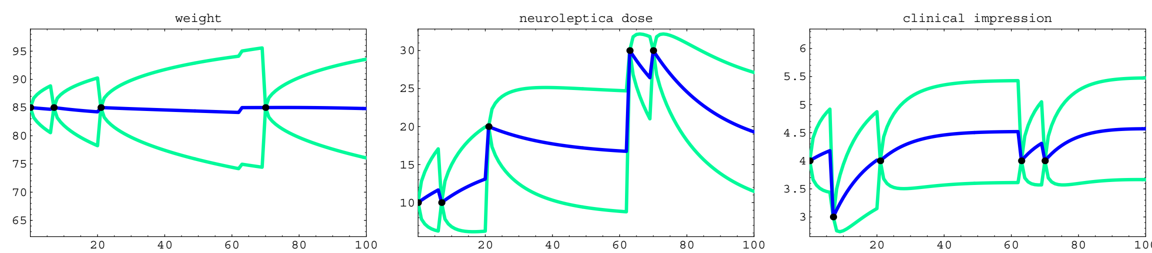

Figure 7: Same person, enlarged graph in the interval [0,100]. The weight is missing at time point $t = 63$, but corrected due to the measurements of dose and impression at the same time.

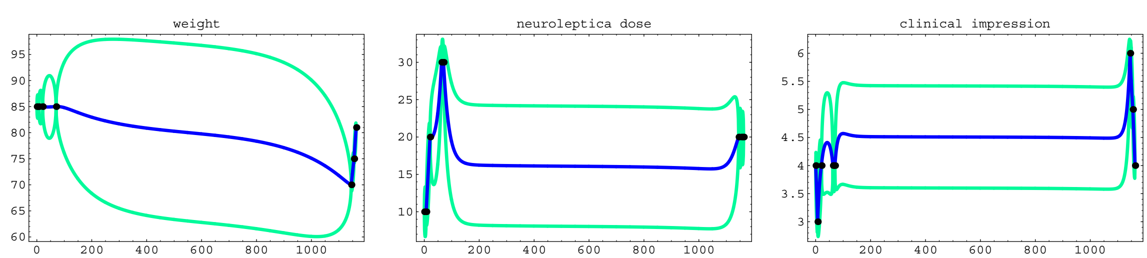

Figure 8: Same person. Smoothed estimates with data points and 67%-HPD confidence intervals.

```
{dt=1, k=3, p=3, q=1, r=3, u=24, T}},
{miss, option, nonlin, constant}]
```

where `dt` $= \delta t$ and `mue` $= \mu(x, \psi)$, `sigma` $= \Sigma(x, \psi)$ is the initial condition of the Kalman filter. Maximization of the likelihood is obtained by using a quasi-Newton algorithm with numerical derivatives and BFGS secant updates. The asymptotic Fisher information matrix is computed from a numerical Hessian with double sided differences.

3 Nonlinear differential equations and state space models

Whereas the linear case can be treated completely and efficiently, 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 Coleman (1968) and others. Here we have the **nonlinear continuous-discrete state space model** (Jazwinski, 1970, ch. 6.2)

$$dY(t) = f(Y(t), t, \psi)dt + g(Y(t), t, \psi)dW(t) \quad (87)$$

$$Z_i = h(Y(t_i), t_i, \psi) + \epsilon_i. \quad (88)$$

with nonlinear drift and diffusion functions f and g. In the nonlinear case it is important to interpret the SDE 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. 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. 96), 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 3.1) 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).

Another class of approximations relates to the filter density. In the **unscented Kalman filter (UKF)**, cf. Julier et al. (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). More generally, the density may be represented by Gaussian sums (Alspach and Sorenson, 1972).
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 [7]. Alternatively, one can use systems of moment differential equations (sect. 3.2, or Singer, 2004a). 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).

3.1 Extended Kalman filter EKF

The continuous-discrete state space model (87–88) may be treated approximately by linearized moment equations, if one computes the exact evolution equations (the dependence on ψ is dropped; $Z^i = \{Z_i, ...Z_0\}$)

$$\begin{aligned}(d/dt)\mu(t|t_i) &= E[f(Y,t)|Z^i] \quad (89)\\ (d/dt)\Sigma(t|t_i) &= E[f(Y,t)(Y(t)-\mu(t|t_i))'|Z^i] + \\ &\quad E[(Y(t)-\mu(t|t_i))f(Y,t)'|Z^i] + E[\Omega(Y,t)|Z^i]. \quad (90)\end{aligned}$$

These are not differential equations, however, since they contain the conditional density $p(y,t|Z^i)$ which already is 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**

$$\begin{aligned}(d/dt)\mu(t|t_i) &= f(\mu(t|t_i),t) \quad (91)\\ (d/dt)\Sigma(t|t_i) &= A(t)\Sigma(t|t_i) + \Sigma(t|t_i)A'(t) + \Omega(\mu(t|t_i),t). \quad (92)\end{aligned}$$

with Jacobian $A(t) = (\partial f/\partial y)(\mu(t|t_i),t) : p \times p$. In contrast to (89-90) the EKF equations are a closed system of differential equations to be solved with standard

[7] $L = F^\dagger$ is the adjoint of the Fokker-Planck operator (99)

numerical techniques (Runge-Kutta etc). Second order derivatives lead to the 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

$$\begin{aligned} Z_{i+1} &= h(\mu(t_{i+1}|t_i), t_{i+1}) + H(t_{i+1}) * (Y_{i+1} - \mu(t_{i+1}|t_i)) + \epsilon_{i+1} && (93) \\ &:= H(t_{i+1})Y_{i+1} + d(t_{i+1}) + \epsilon_{i+1} && (94) \end{aligned}$$

with Jacobian $H(t_{i+1}) = (\partial h/\partial y)(\mu(t_{i+1}|t_i), t_{i+1}) : k \times p$. This permits the usage of the linear **measurement update equations** (30-35). Since we approximate the probability density $p(Z_{i+1}|Z^i;\psi)$ by a Gaussian, we obtain the approximate likelihood function of observation Z_{i+1}

$$p(Z_{i+1}|Z^i;\psi) = \phi(Z_{i+1}; Z(t_{i+1}|t_i), \Gamma(t_{i+1}|t_i)). \quad (95)$$

In the linear case $f(Y,t) = A(t)Y + b(t)$, $g(Y,t) = G(t)$, $h(Y,t) = H(t)Y + d(t)$, the formulas of the KF are recovered. Panel data can be treated by summing the N likelihood contributions and filtered estimates may be computed for each panel unit.

3.2 General nonlinear filtering scheme

For large measurement intervals Δt_i or strongly nonlinear systems the conditionally Gaussian approach (EKF, SNF or LL) is not sufficient and other approximation methods must 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** (Jazwinski, 1970, ch. 6.3)

time update:

$$\begin{aligned} \frac{\partial p(y,t|Z^i)}{\partial t} &= F(y,t)p(y,t|Z^i)\ ; t \in [t_i, t_{i+1}] && (96) \\ p(y, t_i|Z^i) &:= p(y_i|Z^i) \\ p(y, t_{i+1}|Z^i) &:= p(y_{i+1}|Z^i) \end{aligned}$$

measurement update:

$$\begin{aligned} p(y_{i+1}|Z^{i+1}) &= \frac{p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)}{p(z_{i+1}|Z^i)} && (97) \\ &:= p_{i+1|i+1} \\ p(z_{i+1}|Z^i) &= \int p(z_{i+1}|y_{i+1}, Z^i)p(y_{i+1}|Z^i)dy_{i+1}, && (98) \end{aligned}$$

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

$$\begin{aligned} F(y,t)p(y,t|x,s) &= -\sum_i \frac{\partial}{\partial y_i}[f_i(y,t)p(y,t|x,s)] \\ &\quad + \tfrac{1}{2}\sum_{ij} \frac{\partial^2}{\partial y_i \partial y_j}[\Omega_{ij}(y,t)p(y,t|x,s)]. && (99) \end{aligned}$$

The filtering scheme yields a recursive computation of the likelihood function

$$l(\psi; Z) = \sum_{i=0}^{T-1} \log p(Z_{i+1}|Z^i; \psi) + \log p(Z_0; \psi). \tag{100}$$

Although the problem is completely described by the filtering scheme, the computation of the updates is difficult. Some numerical approximations use Monte Carlo integration (Pedersen, 1995, Singer, 2003). Other **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). Moreover, **numerical integration** procedures have been utilized (Kitagawa, 1987).

3.3 Filtering using Hermite expansions

In the EKF approach, the nongaussian transition density $p(y, t|Z^i)$ was approximated by a Gaussian $p(y, t|Z^i) \approx \phi(y; \mu(t|t_i), \Sigma(t|t_i))$ and the conditional moments were obtained as solutions of approximate moment equations. Higher order approximations can be derived if the density is expanded into a Hermite orthogonal series with leading Gaussian term. The first terms in the series read (scalar case)

$$p(y, t|Z^i) := (1/\sigma)\phi(z)[1 + (1/6)\nu_3 H_3(z) + (1/24)(\nu_4 - 3)H_4(z) + ...] \tag{101}$$

where $z = (y - \mu)/\sigma$, and the standardized moments $\nu_k = E[(Y - \mu)^k]/\sigma^k := m_k/\sigma^k$ can be expressed in terms of the *centered moments*

$$m_k := E[M_k] := E[(Y - \mu)^k];\ \sigma^2 := m_2;\ \mu := E[Y]. \tag{102}$$

In the above expressions, the condition on the measurements Z^i was dropped and $H_n(x)$ are the Hermite polynomials, an orthogonal function system w.r.t the standard normal distribution $\phi(x)$, i.e.

$$\int_{-\infty}^{\infty} H_n(x)H_m(x)w(x)dx = \delta_{nm} n! \tag{103}$$

The Hermite polynomials $H_n(x)$ are defined by (cf. Courant and Hilbert, 1968, ch. II, 9, Abramowitz and Stegun, 1965, ch. 22)

$$\phi^{(n)}(x) := (d/dx)^n \phi(x) = (-1)^n \phi(x) H_n(x). \tag{104}$$

and are given explicitly by $H_0 = 1, H_1 = x, H_2 = x^2 - 1, H_3 = x^3 - 3x, H_4 = x^4 - 6x^2 + 3$ etc.

time update:

Between measurements $t_i \le t < t_{i+1}$, the centered moments fulfil

$$\dot{\mu}(t|t_i) = E[f(Y(t), t)|Z^i] \tag{105}$$

$$\begin{aligned} \dot{m}_2(t|t_i) &= 2E[f(Y(t), t) * (Y(t) - \mu(t|t_i))|Z^i] + \\ &+ E[\Omega(Y(t), t)|Z^i] \end{aligned} \tag{106}$$

$$\dot{m}_k = kE[f(Y,t) * (M_{k-1} - m_{k-1})|Z^i] + \tfrac{1}{2}k(k-1)E[\Omega(Y,t) * M_{k-2}|Z^i] \quad (107)$$

with initial condition $m_k(t_i|t_i) = E[(Y(t_i) - \mu(t_i|t_i))^k|Z^i]$. These moment equations may be obtained by inserting the Fokker-Planck equation (96) in the time derivatives of the moments m_k. These are not differential equations, however, and Taylor expansion of f and Ω around μ yields

$$\begin{aligned} \dot{\mu} &:= \sum_{l=0}^{\infty} f^{(l)}(\mu,t)\frac{m_l}{l!} \quad (108) \\ &= f(\mu,t) + \tfrac{1}{2}f''(\mu,t)m_2 + \tfrac{1}{6}f'''(\mu,t)m_3 + ... \end{aligned}$$

and $(k \geq 2)$

$$\begin{aligned} \dot{m}_k &= k\sum_{l=1}^{\infty} \frac{f^{(l)}(\mu,t)}{l!}(m_{l+k-1} - m_l m_{k-1}) + \\ &\quad + \tfrac{1}{2}k(k-1)\sum_{l=0}^{\infty} \frac{\Omega^{(l)}(\mu,t)}{l!}m_{l+k-2}. \quad (109) \end{aligned}$$

measurement update:

At the next measurement Z_{i+1}, the a priori density $p(y, t_{i+1}|Z^i)$ must be updated according to (97). Since the density is represented in terms of the Hermite series, the likelihood and the a posteriori moments were computed using numerical integration. After the determination of $m_k(t_{i+1}|t_{i+1})$, the next time update is performed.

Example: CEV model with ordinal measurements

The algorithm was implemented in Mathematica (Wolfram, 1999) and tested with a heteroskedastic stock price model. The system model reads

$$dY(t) = rY(t)dt + \sigma Y(t)^{\alpha/2}dW(t), \quad (110)$$

the constant elasticity of variance (CEV) process. It is a generalization of the geometric Brownian motion (GBB) which models random rates of change, i.e.

$$dY(t)/dt = [r + \sigma dW(t)/dt]Y(t). \quad (111)$$

This equation can be interpreted as equation (1) with a stochastic rate $a(t) = r + \sigma dW(t)/dt$. The CEV model yields this as a special case ($\alpha = 2$). The measurement model

$$\begin{aligned} Z_i &= h(Y(t_i)) + \epsilon_i \quad (112) \\ h(y) &= \theta(y - c_1) + \theta(y - c_2) \quad (113) \\ \mathrm{Var}(\epsilon_i) &= R \quad (114) \end{aligned}$$

gives ordinal measurements ($Z \in \{0, 1, 2\}$) at sampling times t_i (θ = unit step function). Examples are rating classes for firms or dynamic threshold models in psychometrics. The true parameter values were $\psi = \{r = .1, \alpha = 1, \sigma = .2, R =$

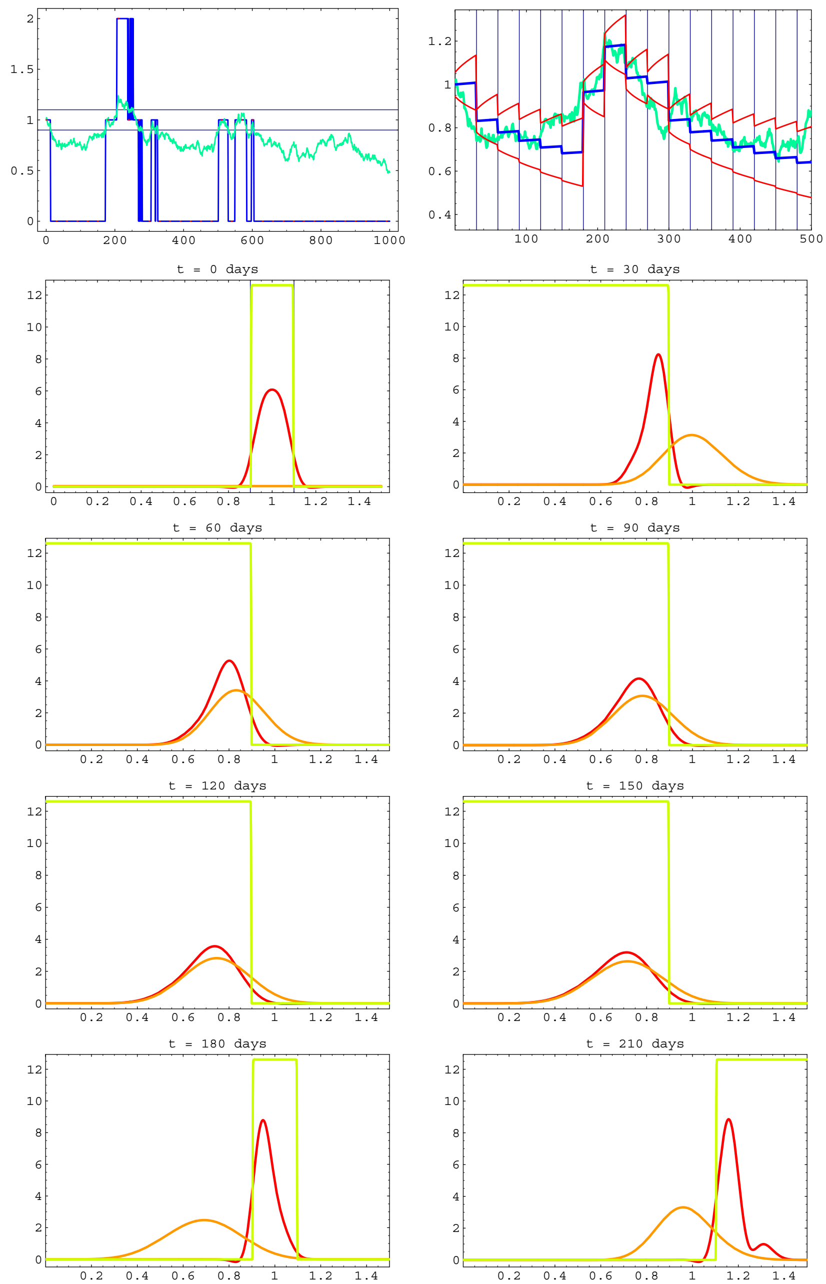

Figure 9: CEV model: trajectory $Y(t)$ and ordinal measurements $h(Y(t))$ with threshold values (left above). Filtered estimates with data points and 67%-HPD confidence intervals (right above). Film of a priori (light grey) and a posteriori densities (dark grey) at measurement points. Also shown is the measurement density ($\propto$ indicator function).

$.001, c_1 = .9, c_2 = 1.1\}$ (square root model). Note that the measurement density $p(z|y) = \phi(z; h(y), R)$ is proportional to the indicator function of the interval $h^{-1}(z)$, i.e. $\chi_{\{y|z=h(y)\}}(y)$. Fig. (9) illustrates the sequence of measurement updates of a nongaussian density using the Hermite expansion (101). The moment equations were solved with $K = 4$ moments and the Taylor expansion was up to order $L = 3$. The unknown moments on the right hand side of 108 – 109) were factorized by the Gaussian assumption

$$m_k = \begin{cases} (k-1)!!m_2^{k/2}; k > K & \text{is even} \\ 0; \quad\quad\quad\quad\quad k > K & \text{is odd} \end{cases} \tag{115}$$

(for a discussion see Singer, 2004a).

4 Conclusion

We have shown that the continuous/discrete state space model is a very flexible dynamical specification with many theoretical as well as practical advantages. Completely irregularly sampled data can be analyzed in a parsimonious way by separating the dynamical model from the measurement process. Moreover, the system state is defined for any continuous time point. Models with higher order derivatives (CAR(p)) can be estimated efficiently without approximating the derivatives by finite differences.

Clinical survey data with irregular sampling intervals and different sample size for each person including missing values could be easily specified and estimated within the proposed framework. Kalman filter software like LSDE or SEM software with nonlinear parameter restrictions (Mx, SEM etc.) can be used to estimate the exact discrete model.

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 presently an active field of research with competing methods (both analytical and numerical). The methods of extended Kalman filtering (EKF) and Hermite expansions were shortly discussed. Especially in the multivariate case further research is needed.

Acknowledgement

I would like to thank two anonymous referees for their detailed and useful comments which helped to improve the presentation of the paper.

References

[1] M. Abramowitz and I. Stegun. *Handbook of Mathematical Functions*. Dover, New York, 1965.

[2] A.A Abrikosov, L.P. Gorkov, and I.E. Dzyaloshinsky. *Methods of Quantum Field Theory in Statistical Physics*. Dover, New York, 1963.

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

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

[5] 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.

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

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

[8] 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.

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

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

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

[12] S. M. Boker and S. Poponak. Multivariate Multilevel Models of Dynamical Systems using Latent Differential Equations. In C. van Dijkum et al., editor, *Sixth International Conference on Social Science Methodology*. SISWO, CD ROM, ISBN 90-6706-176-x, 2004.

[13] 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.

[14] 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.

[15] R. Courant and D. Hilbert. *Methoden der Mathematischen Physik*. Springer, Berlin, Heidelberg, New York, 1968.

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

[17] 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.

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

[19] 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.

[20] 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.

[21] 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.

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

[23] M. Hürzeler and H. Künsch. Monte Carlo Approximations for General State-Space Models. *Journal of Computational and Graphical Statistics*, 7,2:175–193, 1998.

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

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

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

[27] R.H. Jones. Fitting multivariate models to unequally spaced data. In E. Parzen, editor, *Time Series Analysis of Irregularly Observed Data*, pages 158–188. Springer, New York, 1984.

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

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

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

[31] 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.

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

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

[34] 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.

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

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

[37] C. Möbus and W. Nagl. Messung, Analyse und Prognose von Veränderungen (Measurement, analysis and prediction of change; in german). In *Hypothesenprüfung, Band 5 der Serie Forschungsmethoden der Psychologie der Enzyklopädie der Psychologie*. Hogrefe, 1983.

[38] 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.

[39] 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.

[40] J.H.L. Oud, J.F.J van Leeuwe, and R.A.R.G Jansen. Kalman Filtering in discrete and continuous time based on longitudinal LISREL models. In J.H.L. Oud and R.A.W. van Blokland-Vogelesang, editors, *Advances in longitudinal and multivariate analysis in the behavioral sciences*, pages 3–26, Nijmegen, Netherlands, 1993. ITS.

[41] 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.

[42] 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, Amsterdam, 1976.

[43] 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, Amsterdam, 1976b.

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

[45] 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.

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

[47] H. Singer. *LSDE - A program package for the simulation, graphical display, optimal filtering and maximum likelihood estimation of Linear Stochastic Differential Equations, User's guide*. Meersburg, 1991.

[48] 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.

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

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

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

[52] 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.

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

[54] H. Singer. Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models. In C. van Dijkum, J. Blasius, H. Kleijer, and B. Van Hilten, editors, *Sixth International Conference on Social Science Methodology.* Netherlands Institute for the Social Sciences (SISWO), CD ROM, ISBN 90-6706-176-x, 2004b. Paper presented at the Sixth International Conference on Social Science Methodology.

[55] H. Singer. SEM – Linear Structural Equations Models with Arbitrary Nonlinear Parameter Restrictions, Version 0.1. Mathematica Program, http://www.fernuni-hagen.de/STATISTIK/forschung/SEMArchive.exe, Fernuniversität in Hagen, Hagen, Germany, 2004d.

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

[57] 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, New York, 1995.

[58] C. van Dijkum, J. Blasius, H. Kleijer, and B. Van Hilten, editors. *Sixth International Conference on Social Science Methodology.* Netherlands Institute for the Social Sciences (SISWO), CD ROM, ISBN 90-6706-176-x, 2004.

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

[60] S. Wolfram. *The Mathematica Book.* Cambridge University Press, Cambridge, U.K., 4th edition, 1999.