Bayesian Estimation of Volatility with Moment-Based Nonlinear Stochastic Filters

Oliver Grothe* and Hermann Singer**

September 2006

Abstract

This article adresses parameter estimation with moment-based stochastic filters that only heed the first two moments of the state densities. This approximation provides good results in numerous cases. However, due to missing linear correlation between diffusion parameters and expected states, Bayesian estimation of diffusion parameters such as volatility is not possible. While other filters overcome this problem by simulations, we present a deterministic algorithm for Bayesian estimation of the diffusion coefficient based on sigma points which can be applied to all moment-based filters. To show the validity of the algorithm we use the continuous-discrete unscented Kalman filter proposed by Singer [18].

Keywords: Bayesian parameter estimation, nonlinear systems, unscented Kalman filter, maximum likelihood estimation, stochastic volatility

*University of Cologne, Research Training Group Risk Management, Meister-Ekkehart-Straße 11, D-50923 Cologne, Germany, tel: +49 (0)221 470 7701, email: grothe@wiso.uni-koeln.de

**University of Hagen, Department of Applied Statistics and Empirical Social Science, Universitätsstr. 41 (ESG), D-58097 Hagen, Germany, tel: +49 (0)2331 987 2615 email: Hermann.Singer@FernUni-Hagen.de

1 Introduction

Nonlinear stochastic filters are powerful tools for simultaneous estimation of parameters and unoberserved states from noisy data. They can be used for a maximum likelihood approach for parameter estimation and provide the possibility of online Bayesian estimation by augmenting the state vector with the parameters. The basic idea of the most nonlinear filters is to apply the Kalman filter to nonlinear systems. This may be extremely difficult as it requires the description of the propagation of the state and parameter probability densities. As these densities may be very complex, a finite number of parameters may not be sufficient to describe them. Therefore, approximations have to be made, dividing the different filters into two sections: the moment-based or Gaussian filters, which reduce the densities to their first two moments, and filters which capture more information about the densities. Examples for the second type are particle filters (see for example Pitt/Shephard et al. [12]) which use simulations for the density approximations.

The first and most widely used moment-based nonlinear filter is the extended Kalman filter (EKF), which uses a Taylor expansion of the nonlinear functions around the estimates up to the first order. Expansions up to the second order lead to the second order nonlinear filter (SNF). Expansions to higher orders lead to the higher order nonlinear filters (HNF). While these classical filters need the explicitly given Jacobians of the nonlinear functions, there exist approaches on numerical derivations. The divided difference filters (DD-i) of Nørgaard [11] use polynomial expansions of the nonlinear functions up to the i-th order which can be solved numerically by evaluations of these functions. Other filters, based on numerical quadrature rules, are the Gauss-Hermite filter (GHF) based on the numerical Gauss-Hermite integration and the central-difference filter (CDF) (see Ito/Xiong [3] for both) based on polynomial interpolation. The second one turns out to be equivalent to the DD-2 Filter (see van der Merwe/Wan [20]).

The cited moment-based filters cited above use approximations of the nonlinear functions.

Whereas, the unscented Kalman filter (UKF) developed by Julier/Uhlmann [6], is based on the *intuition that it is easier to approximate a probability distribution than to approximate an arbitrary nonlinear function or transformation* [7], using the unscented transform (UT) for the approximation of the probability densities that undergo nonlinear transformation. The specification presented by Julier/Uhlman was only formulated for the time discrete case and with respect to the noisy terms in general, treating the noise sequences by including them into the state vector. Recently, Singer [18] formulated a specification for the continuous-discrete case and included the noisy terms directly with no need to extend the state vector.

One important advantage of the moment-based filters over the simulation-based filters is the computing time needed. In general, simulation-based filters need hundreds of times the computing power of moment-based filters due to the simulation of hundreds of trajectories. However, moment-based filters are not able to estimate parameters of the diffusion coefficient of the state equations such as volatility using the Bayesian approach (see for example Sitz et al. [19]). For this reason simulation-based filters have to be used, resulting in more computing power needed (see for example the FIF in Singer [15] or [17]). To tackle this problem, we present a deterministic meta-algorithm in Section 4.2. It may be adapted to all moment-based filters. Using this algorithm allows Bayesian estimation of the diffusion coefficient with a computationally overhead of just factor three compared to the direct use of a moment-based filter. However, if applied directly, the moment-based filter does not estimate the diffusion coefficient. Therefore this overhead has to be compared with the use of simulation-based filters.

The paper is organized as follows: In Section 2 the continuous-discrete state space model is defined. The nonlinear state estimation is discussed deriving the general filter equations for moment-based filters. In Section 3 the unscented transform and the unscented Kalman filter according to Singer [18] are derived. In Section 4 two approaches of parameter estimation with moment-based filters (ML und Bayesian) are discussed. We introduce an abstract notation

and present the Bayesian meta-algorithm. In Section 5 the validity of our approach is shown in simulation studies. We apply the algorithm in order to estimate parameters of an Ornstein-Uhlenbeck model using the Bayesian approach. We also apply the algorithm to a stochastic volatility model. Section 6 concludes.

2 Nonlinear Continuous Discrete State Estimation

2.1 State Space Model

The *continuous-discrete state space* representation (Jazwinski [4]) turns out to be very useful in systems, in which the underlying models are continuous in time and only discrete observations are available. It consists of a continuous state equation for the state $y(t)$ and discrete measurements z_i at times t_i:

$$dy(t) = f(y(t), t, \psi)dt + g(y(t), t, \psi)dW(t) \quad (1)$$

$$z_i = h(y(t_i), t_i) + \epsilon_i. \quad (2)$$

The first equation is a p-dimensional Itô differential equation with an r-dimensional Wiener process $W(t)$. The drift coefficient $f : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \longrightarrow \mathbb{R}^p$ and the diffusion coefficient $g : \mathbb{R}^p \times \mathbb{R} \times \mathbb{R}^u \longrightarrow \mathbb{R}^p \times \mathbb{R}^r$ are functions of the state, the time and a u-dimensional parameter vector ψ.

The measurement equation (2) projects the state vector $y(t)$ onto the time discrete k-dimensional measurements z_i. The measurement may be noisy with the k-dimensional discrete white noise process $\epsilon_i \sim \mathbf{N}(0, R(t_i, \psi))$, ϵ_i, i.d. and independent of $W(t)$.

2.2 Time and Measurement Update for Moment-Based Filters

The idea of filtering is to estimate the actual probability density of the state vector from noisy data. For $t \in]t_i, t_{i+1}[$ (no measurement information) the estimation of $p(y, t|Z^i)$ is based on

prior information and the propagation of the density through the state equation (*time-update, a-priori density*). For $t = t_{i+1}$ with actual measurement information available, the a-posteriori state density $p(y, t|Z^{i+1})$ has to be estimated using the a-priori density and the measurement information (*measurement update*).

Both steps can be made without approximation (Jazwinski [4]). For the time update this means integrating the Fokker-Planck operator resulting in terms that can be solved explicitly only for linear systems and Gaussian densities. There are numerical and Monte-Carlo based methods that lead to approximated solutions for nonlinear systems (see for example Singer [16] and the reviewing introduction there). Furthermore, there are moment-based filters solving the propagation approximately only for the first two moments (expectation μ and variance Σ) of the state densities using the moment equations (**time-update**):

$$\dot{\mu}(t) = E[f(y(t), t)|Z^i] \tag{3}$$

$$\dot{\Sigma}(t) = \mathrm{Cov}[f, y|Z^i] + \mathrm{Cov}[y, f|Z^i] + E[\Omega|Z^i], \tag{4}$$

with $\Omega = gg'$. Since these equations depend on the conditional densities $p(y, t|Z^i)$ they have to be solved approximately. In the EKF this approximation consists in the Taylor expansion of f and g while the filters of Nørgaard [11] use polynomial expansions. These approximations of the nonlinearities lead to approximated differential equations. In the UKF, however, the density $p(y, t|Z^i)$ is approximated by the UT (see Section 3.1) using the full functions f, g.

The exact measurement update is based on Bayesian updating. Given the prior information $p(y|Z^i)$ and the measurement information $p(z_{i+1})$ the a-posteriori state density is given by

$$p(y|Z^{i+1}) = \frac{p(z_{i+1}|y)p(y|Z^i)}{p(z_{i+1})}. \tag{5}$$

Looking solely at the first two moments, implicitly assumes Gaussian densities. In these cases the measurement update $p(y|Z^{i+1})$ can be simplified using the theorem of normal correlation:

Theorem 1 (Normal correlation)

Let X and Y be multivariate normally distributed. Then:

$$E[X|Y] = \mu_X + \Sigma_{XY}\Sigma_{YY}^{-}(Y - \mu_Y) \tag{6}$$

$$\mathrm{Var}[X|Y] = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-}\Sigma_{YX}, \tag{7}$$

with Σ_{YY}^{-}, pseudoinverse of Σ_{YY}, sufficing (Liptser/Shiryayev [8], chap. 13).

Due to the role of the covariance matrices only the states correlated with the measurement are updated. As this does not apply for the parameters of the diffusion coefficient, they are not updated. With $p(y,t|Z^{i+1}) = p(y,t|z_{i+1},Z^i)$ this gives an optimal estimation for the linear case. Using subscripts $_i$ for *at time* t_i and $_{i+1|i}$ for *at time* t_{i+1} *based on information of* t_i etc. it follows:

$$\begin{aligned}\mu_{i+1|i+1} &= \mu_{i+1|i} + \mathrm{Cov}[y_{i+1}, z_{i+1}|Z^i]\mathrm{Var}[z_{i+1}|Z^i]^{-} \times \\ &\quad \times(z_{i+1} - E[z_{i+1}|Z^i]), \end{aligned} \tag{8}$$

$$\begin{aligned}\Sigma_{i+1|i+1} &= \Sigma_{i+1|i} - \mathrm{Cov}[y_{i+1}, z_{i+1}|Z^i]\mathrm{Var}[z_{i+1}|Z^i]^{-} \times \\ &\quad \times\mathrm{Cov}[z_{i+1}, y_{i+1}|Z^i]. \end{aligned} \tag{9}$$

Taking the measurement equation (2) into account leads to (**measurement update**):

$$\begin{aligned}\mu_{i+1|i+1} &= \mu_{i+1|i} + \mathrm{Cov}[y_{i+1}, h_{i+1}|Z^i] \times \\ &\quad \times(\mathrm{Var}[h_{i+1}|Z^i] + R_{i+1})^{-}(z_{i+1} - E[h_{i+1}|Z^i]), \end{aligned} \tag{10}$$

$$\begin{aligned}\Sigma_{i+1|i+1} &= \Sigma_{i+1|i} - \mathrm{Cov}[y_{i+1}, h_{i+1}|Z^i] \times \\ &\quad \times(\mathrm{Var}[h_{i+1}|Z^i] + R_{i+1})^{-}\mathrm{Cov}[h_{i+1}, y_{i+1}|Z^i] \end{aligned} \tag{11}$$

$$\begin{aligned}L_{i+1} &= p(z_{i+1}|Z^i) \\ &= \phi\left(z_{i+1}; E[h_{i+1}|Z^i]\mathrm{Var}[h_{i+1}|Z^i] + R_{i+1}\right), \end{aligned} \tag{12}$$

where L_{i+1} is the likelihood function of the measurement at t_{i+1} with Gaussian density ϕ. It is based on the prediction error $\nu_{i+1} := z_{i+1} - E[z_{i+1}|Z^i]$ and its covariance

$\Gamma_{i+1} := \mathrm{Var}[z_{i+1}|Z^i] = \mathrm{Var}[h_{i+1}|Z^i] + R_{i+1}$. As in the time update step, the expectations and covariances have to be approximated. In the EKF, Taylor expansion of h around $\mu_{i+1|i}$ is done, the polynomial filter of Nørgaard uses numerical expansion. In the UKF, this is done using the unscented transform.

3 Continuous-Discrete Unscented Kalman Filter

The basic ideas and formulas of moment-based filters were proposed in the former section. In this section we present the unscented transform to build the unscented Kalman Filter. In contrast to the classical, more general UKF presented by Julier et al. [6] which uses an augmentation of the state vector to include the noisy terms in the time update and calculate the covariance $\Sigma_{i+1|i}$, we use a version for the continuos-discrete state space model presented by Singer [18]. Therein, the UKF is implemented directly, using the time update equations mentioned above. There is no need for augmenting the state vector with the noisy terms, which results in fewer computing costs. Moreover, this implementation adapts directly to continuous-discrete systems by decoupling of observation width in the measurement equation and sampling width in the time update of the filter.

3.1 Unscented Transform (UT)

The UT is a method for calculating the transformation of the density of a random variable which undergoes a nonlinear transformation (see Julier/Uhlmann [6]). For calculating the moments before and after the transformation, the density $p(y)$ of the random variable $y \in \mathbb{R}^{\tilde{p}}$ is approximated by the sum

$$p_{UT}(y) = \sum_{j=-\tilde{p}}^{\tilde{p}} \omega^{(j)} \delta(y - y^{(j)}), \tag{13}$$

with δ the Dirac delta function, the $n = 2\tilde{p} + 1$ *sigma points* $y^{(j)}$ and the weights $\omega^{(j)}$. Often, $p_{UT}(y)$ is interpreted as a singular probability density. This is not the case in the original

framework of Julier/Uhlmann, where the weights can be negative. Furthermore the sigma points and weights are chosen so that the first two moments of the density of y are replicated:

$$\begin{aligned} E[y] &= \int yp(y)dy \stackrel{!}{=} \int yp_{UT}(y)dy = \sum_{j=1}^{n} \omega^{(j)} y^{(j)} \\ \mathrm{Var}[y] &\stackrel{!}{=} \sum_{j=1}^{n} \omega^{(j)} (y^{(j)} - E[y])(y^{(j)} - E[y])^T. \end{aligned}$$

In contrast to Monte Carlo approaches, this choice is deterministic. Julier/Uhlmann give the following choice:

$$\begin{aligned} y^{(i)} &= E[y] + \left(\sqrt{(p+\kappa)\mathrm{Var}[y]}\right)_i, & \omega^{(i)} &= \frac{1}{2(p+\kappa)}, & i &= 1..p \\ y^{(i+p)} &= E[y] - \left(\sqrt{(p+\kappa)\mathrm{Var}[y]}\right)_i, & \omega^{(i+p)} &= \frac{1}{2(p+\kappa)}, & i &= 1..p \\ y^{(2p+1)} &= E[y], & \omega^{(2p+1)} &= \frac{\kappa}{(p+\kappa)}, \end{aligned}$$

where $(\sqrt{...})_i$ is the i-th row or column of the matrix root. The real parameter κ gives an extra degree of freedom for further *fine tuning*. For Gaussian densities of y Julier/Uhlmann recommend $\tilde{p} + \kappa = 3$, however $\kappa = 0$ works in many cases (see e.g. Julier/Uhlmann [7] or Singer [18]) reducing the number of sigma points by one (for a detailed discussion regarding κ see Julier/Uhlmann [6] or Julier [5]). After undergoing a nonlinear transformation $y \to \tilde{f}(y)$ expectation, covariance of $\tilde{f}(y)$ and cross covariance of $\tilde{f}(y)$ and y can be computed as:

$$\begin{aligned} E[\tilde{f}(y)] &= \sum_{j=1}^{n} \omega^{(j)} \tilde{f}(y^{(j)}) \\ \mathrm{Var}[\tilde{f}(y)] &= \sum_{j=1}^{n} \omega^{(j)} \left(\tilde{f}(y^{(j)}) - E[\tilde{f}(y)]\right) \left(\tilde{f}(y^{(j)}) - E[\tilde{f}(y)]\right)' \\ \mathrm{Cov}[\tilde{f}(y), y] &= \sum_{j=1}^{n} \omega^{(j)} \left(\tilde{f}(y^{(j)}) - E[\tilde{f}(y)]\right) \left(y^{(j)} - E[y]\right)' \end{aligned}$$

with the outer product $(...)(...)'$.

3.2 Filter Algorithm

The unscented transform can be used to evaluate the expectation, variance and covariance terms of the filter equations. The time update is done by Euler integration using equations (3) and (4). With regard to the time continuous nature of the state equations of the system, the Euler scheme uses a finer discretization interval δt than the measurement intervals $t_{i+1} - t_i$ dividing them into $L = (t_{i+1} - t_i)/\delta t$ parts. The time update is done at each point τ_l of the resulting grid. With $\tau_0 = t_i$ and $\tau_L = t_{i+1}$ this is an iteration which calculates $\mu_{(l+1)|i}$ and $\Sigma_{(l+1)|i}$. The expectation and covariance values are evaluated with the UT, building the sigma points using the moments $\mu_{l|i}$ and $\Sigma_{l|i}$ (here the subscripts $_{l|i}$ denote *at time τ_l, based on the information at time t_i*). The measurement update is made as given by the equations (10) and (11). The covariance and expectation values are evaluated with the UT building the sigma points using $\mu_{i+1|i}$ and $\Sigma_{i+1|i}$. The following algorithm summarizes this filter:

Algorithm 1 (Continuous-discrete unscented Kalman filter).

Initialization: $t = t_0$

$$\begin{aligned}
\mu_{0|0} &= \mu + \mathrm{Cov}[y_0, h_0] \times \\
&\times \quad (\mathrm{Var}[h_0] + R_0)^{-}\mathrm{Cov}[h_0, y_0] \\
\Sigma_{0|0} &= \Sigma - \mathrm{Cov}[y_0|h_0] \times \\
&\times \quad (\mathrm{Var}[h_0] + R_0)^{-}\mathrm{Cov}[h_0, y_0] \\
L_0 &= \phi(z_0; E[h_0], \mathrm{Var}[h_0] + R_0) \\
\text{Sigma points} &: \quad y^{(j)} = y^{(j)}(\mu, \Sigma); \mu = E[y_0], \Sigma = \mathrm{Var}[y_0].
\end{aligned}$$

Recursion: $i = 0, ..., T-1$

Time update: $t \in [t_i, t_{i+1}]$

$$\begin{aligned}
\tau_l &= t_i + l \cdot \delta t;\ l = 0, ..., L_i - 1 := (t_{i+1} - t_1)/\delta t - 1 \\
\mu_{l+1|i} &= \mu_{l|i} + E[f(y(\tau_l), \tau_l)|Z^i]\delta t \\
\Sigma_{l+1|i} &= \Sigma_{l|i} + \{\mathrm{Cov}[f(y(\tau_l), \tau_l), y(\tau_l)|Z^i] + \\
&+ \mathrm{Cov}[y(\tau_l), f(y(\tau_l), \tau_l)|Z^i] + E[\Omega(y(\tau_l), \tau_l)|Z^i]\}\delta t \\
\text{Sigma points} &: y^{(j)} = y^{(j)}(\mu_{l|i}, \Sigma_{l|i})
\end{aligned}$$

Measurement Update

$$\begin{aligned}
\mu_{i+1|i+1} &= \mu_{i+1|i} + \mathrm{Cov}[y_{i+1}, h_{i+1}|Z^i] \times \\
&\times (\mathrm{Var}[h_{i+1}|Z^i] + R_{i+1})^{-}(z_{i+1} - E[h_{i+1}|Z^i]), \\
\Sigma_{i+1|i+1} &= \Sigma_{i+1|i} - \mathrm{Cov}[y_{i+1}, h_{i+1}|Z^i] \times \\
&\times (\mathrm{Var}[h_{i+1}|Z^i] + R_{i+1})^{-}\mathrm{Cov}[h_{i+1}, y_{i+1}|Z^i] \\
L_{i+1} &= \phi\left(z_{i+1}; E[h_{i+1}|Z^i]\mathrm{Var}[h_{i+1}|Z^i] + R_{i+1}\right) \\
\text{Sigma points} &: y^{(j)} = y^{(j)}(\mu_{i+1|i}, \Sigma_{i+1|i})
\end{aligned}$$

The subscript $_{i|i}$ denotes *at time t_i based on information at time t_i*, whereas the subscript $_{l|i}$ denotes *at time τ_l based on information at time t_i*.

4 Parameter Estimation

Moment-based filters are powerful tools for the estimation of the state $y(t)$ of the system from of noisy data. In contrast to the classical Kalman filter, the presented UKF is capable to handle nonlinearities in the state space model due to the approximation of the state densities using the unscented transform. Other moment-based filters like the EKF or the DD-i filters of Nørgaard use linearizations of the state space equations for these approximations. All filters can be used to calculate a likelihood function regarding the state space model and the

observations, so that parameter estimation in the sense of maximum likelihood is possible. We use the following filter-independent notation:

$$[\hat{y}, \hat{\Sigma}, L, \text{logL}] \quad = \quad \text{MBF}\,[E[y_0], \text{Var}[y_0], \psi, R, t, z] \tag{14}$$

where MBF stands for moment-based filter. The variables are: the estimated state vector $\hat{y} \in \mathbb{R}^p \times \mathbb{R}^T$ containing the estimated p-dimensional state at T points in time, the estimated covariance $\hat{\Sigma} \in \mathbb{R}^p \times \mathbb{R}^p \times \mathbb{R}^T$ for all T points in time, the likelihood $L = \prod_{i=0}^{T} L_i$ and the log-likelihood $\text{logL} = \sum_{i=0}^{T} \log(L_i)$. On the right side: assumed expectation $E[y_0]$ and covariance matrix $\text{Var}[y_0]$ of the initial state, the parameter vector ψ, a vector $R \in \mathbb{R}^k \times \mathbb{R}^k (\times \mathbb{R}^T)$ containing the covariance matrix of the measurement noise (possibly different for all T measurements) and the time vector $t = t_0, ..., t_{T-1}$ for the measurements and the measurement vector $z = z_0, ..., z_{T-1}$. The not focused variables are suppressed when possible.

The classical maximum (log-)likelihood (ML) approach for parameter estimation with this notation is received as

$$\max_{\psi}\{\text{logL}\} \quad = \quad \text{MBF}\,[E[y_0], \text{Var}[y_0], \psi, R, t, z] \tag{15}$$

and can be performed by numerical maximization.

4.1 Bayesian Approach

Beside the maximum likelihood approach the filters can estimate the unknown parameters by considering them as latent state variables. The state vector is augmented by the parameter vector ($y \to \tilde{y} = \begin{bmatrix} y \\ \psi \end{bmatrix}$) with trivial dynamics ($d\psi = 0$). This leads to the extended state space model:

$$\begin{aligned} dy(t) &= f(y(t), t, \psi)dt + g(y(t), t, \psi)dW(t) \\ d\psi &= 0 \\ z_i &= h(y(t_i), t_i) + \tilde{\epsilon}_i. \end{aligned} \tag{16}$$

By filtering this state space model as

$$\left[\hat{\tilde{y}}, \hat{\tilde{\Sigma}}, L\right] \quad = \quad \text{MBF}\left[\tilde{y}_0, \text{Var}[\tilde{y}_0], R, t, z\right], \tag{17}$$

the estimates of the parameters are updated each time new measurement information comes in. This gives a sequential Bayesian estimator of the unknown parameters: $\hat{\psi} = E[\psi(t)|Z^t]$.

Note, that this approach leads to nonlinear problems even if the state space equations are linear. For this reason, linear filters like the Kalman filter cannot be used for this approach. Nonlinear moment-based filters like the EKF can be used for this approach but do not estimate parameters of the diffusion coefficient $g(y(t), t)$, such as volatility (see for example Sitz et al. [19]). This is due to a missing linear correlation between observations and the diffusion parameters. Equation (10) will not update the diffusion parameters with zero-entries in the covariance. Solutions are given by superior filter designs like the functional integral filter (FIF) (see Singer [18]) or other simulation-based filters (see Pitt/Shephard [12] for a short review of the statistical basics of particle filters), but will be paid with high computing costs due to numerical simulations. In the next section an algorithm for Bayesian estimation of volatility using moment-based filters is presented.

4.2 Meta-Algorithm for Estimating Volatility

Due to the missing correlation between diffusion coefficient and the observation, the parameters of the diffusion coefficient are not estimated by the moment-based filters. This is caused by the approximation of the exact measurement update equation (5) with the theorem of normal correlation. Nevertheless, the moment-based filters provide an accurate likelihood with respect to the value of the diffusion coefficient. Interpreting the exact equation in terms of likelihood, it is

$$p(y|Z^{i+1}) = \text{constant} \cdot \text{likelihood} \cdot \text{prior}, \tag{18}$$

where the constant is a normalizing constant, given by $p(z_{i+1})$, the sum of all likelihoods. The idea behind it is to decouple the parameters of the diffusions coefficients from the rest and to

update their a-priori belief via this Bayesian formula instead of in the filtering algorithms itself. In the resulting meta-algorithm the state vector is extended by the parameter vector ψ as in the usual Bayesian approach. However, it has to be differentiated between the parameters of the drift coefficient (ψ_{drift}) and the parameters of the diffusion coefficient ($\psi_{\text{diff}} \in \mathbb{R}^m$) as given in equation (19). After initialization a recursion begins, in which sigma points of the diffusion parameter vector in the sense of the unscented transform (see Section 3.1) are formed sequentially for each measurement time t_i. The likelihood for all $(2m+1)$ sigma points regarding the next time step is calculated using the MBF as shown in equation (20). The MBF is initialized for only one time step with state covariance zero. Thereafter, relative weights $\alpha^{(\tilde{j})}$ regarding the likelihoods are calculated as given by equation (21). Using these weights an Bayesian estimator as given by (18) is of the sigma points (equation (22)) with a covariance matrix using both the likelihood weights and the sigma point weights as given by equation (23). The MBF is initialized with the new parameter set at the end of the recursion step for one step to estimate the remaining states, parameters and covariances (see equation (24)).

Algorithm 2 (Estimating diffusion coefficient with MBF).

Initialization: $t = t_0$

$$\tilde{y}_0 = [y_0, \psi_{\text{drift};0}, \psi_{\text{diff};0}]' \qquad \tilde{\Sigma}_0 = \text{Var}[\tilde{y}_0] = \begin{pmatrix} \Sigma_{(y,\text{drift});0} & 0 \\ 0 & \Sigma_{\text{diff};0} \end{pmatrix} \tag{19}$$
$$\psi_{\text{diff}} \in \mathbb{R}^m$$

Recursion: $i = 0, ..., T$

Sigma points : $\psi^{(\tilde{j})}_{\text{diff};i} = \psi^{(\tilde{j})}_{\text{diff};i}(\psi_{\text{diff};i}, \Sigma_{\text{diff};i})$; weights: $\omega^{(\tilde{j})}$

$$\tilde{j} = 1, ..., (2m+1) \quad : \quad L^{(\tilde{j})} = \text{MBF}\left[\begin{pmatrix} y_i \\ \psi_{\text{drift};i} \\ \psi^{(\tilde{j})}_{\text{diff};i} \end{pmatrix}, 0, R, [t_i, t_{i+1}], z_{i+1}\right] \tag{20}$$

$$\alpha_{\tilde{j}} = L^{(\tilde{j})} / \left(\sum_{\tilde{j}=1}^{2m+1} L^{(\tilde{j})}\right) \tag{21}$$

$$\hat{\psi}_{\text{diff}} = \sum_{\tilde{j}=1}^{2m+1} \alpha_{\tilde{j}} \psi^{(\tilde{j})}_{\text{diff};i} \tag{22}$$

$$\hat{\Sigma}_{\text{diff}} = (2m+1) \sum_{\tilde{j}=1}^{2m+1} \alpha_{\tilde{j}} \omega_{\tilde{j}} (\hat{\psi}_{\text{diff}} - \psi^{(\tilde{j})}_{\text{diff};i})^2 \tag{23}$$

$$\tilde{\Sigma}_{\text{tmp}} := \begin{pmatrix} \Sigma_{(y,\text{drift});i} & 0 \\ 0 & \hat{\Sigma}_{\text{diff}} \end{pmatrix}$$

$$\left[\tilde{y}_{i+1}, \tilde{\Sigma}_{i+1}, L_{i+1}\right] = \text{MBF}\left[\begin{pmatrix} y_i \\ \psi_{\text{drift};i} \\ \hat{\psi}_{\text{diff}} \end{pmatrix}, \tilde{\Sigma}_{\text{tmp}}, R, \begin{bmatrix} t_i \\ t_{i+1} \end{bmatrix}', z_{i+1}\right] \tag{24}$$

$$\tilde{y}_{i+1} =: [y_{i+1}, \psi_{\text{drift};(i+1)}, \psi_{\text{diff};(i+1)}]'$$

$$\tilde{\Sigma}_{i+1} =: \begin{pmatrix} \Sigma_{(y,\text{drift});(i+1)} & 0 \\ 0 & \Sigma_{\text{diff};(i+1)} \end{pmatrix}$$

5 Simulation Studies

5.1 Ornstein-Uhlenbeck Process

Mean reversion processes of the Ornstein-Uhlenbeck (O-U) type are used to model the price behavior of different commodities which underly a long price equilibrium like oil and gas (see

Schwartz [14]), or electricity prices (see Lucia/Schwartz [9]). Apart from commodities, interest rates may be modelled using O-U type models as for example in the Cox/Ingersoll/Ross [1] models. The following continuous-discrete state space model describes such a mean reversion process with a continuous system equation and a discrete measurement equation, meaning the prices or interest rates at certain points in time:

$$\begin{aligned} dy(t) &= \psi_1 \left[\psi_2 - y(t)\right] dt + \psi_3 dW(t) \\ z_i &= y(t_i) \end{aligned} \tag{25}$$

with the parameter set $\psi = [\psi_1, \psi_2, \psi_3]$. For simplicity, we suppress all units, the time unit is one, which is $(t_{i+1} - t_i)$. By augmenting the state vector $y(t)$ as $\tilde{y}(t) = [y(t)', \psi]'$, the extended state space model for the Bayesian estimation is received:

$$\begin{aligned} d\tilde{y}_1 = dy &= \tilde{y}_2 \left[\tilde{y}_3 - \tilde{y}_1(t)\right] dt + \tilde{y}_4 dW(t) \\ d\tilde{y}_2 = d\psi_1 &= 0 \\ d\tilde{y}_3 = d\psi_2 &= 0 \\ d\tilde{y}_4 = d\psi_3 &= 0 \\ z_i &= \tilde{y}_1(t_i). \end{aligned} \tag{26}$$

The data for 1000 time units is simulated using an Euler scheme on a grid of one tenth of the measurement interval with the parameter set $\psi = [\psi_1, \psi_2, \psi_3] = [0.5, 3, 2]$. The data is then filtered using the extended state with initial parameters $\psi_0 = [1, 4, 10]$ and a diagonal initial covariance matrix with variances of 1. Figure 1 shows the estimation results for the parameters ψ (2. to 4. component of $\tilde{y}$) using the continuous-discrete UKF ($L = 10$). The parameters of the drift coefficient are estimated correctly, the confidence intervals (3 times the standard deviation on each side) shrink with time. As expected, the diffusion coefficient i.e. the volatility is not estimated.

Figure 2 shows the results using the meta-algorithm adapted to the continuous-discrete UKF. Obviously the diffusion coefficient is estimated. The filter needs about 50 recursions to decrease

ψ_3 from 10 down to approximately 2. The diffusion coefficient with a higher lieklihood leads to better results of the other parameters with faster shrinking confidence intervals than in the UKF case.

5.2 Stochastic Volatility

We extend the former model by a time dependent diffusion coefficient in the sense of a generalized Vasicek or a Hull-White [2] model. The time dependent diffusion coefficient (volatility) is of O-U type as well. Similar approaches for extending the commodity models by stochastic volatility of O-U type were recently used by Nielson/Schwartz [10] and Ribeiro/Hodges [13]. We assume the following state space model:

$$\begin{aligned} dy_1(t) &= \psi_1 \left[\psi_2 - y_1(t)\right] dt + y_1(t) y_2(t) dW_1(t) \\ dy_2(t) &= \psi_3 \left[\psi_4 - y_2(t)\right] dt + \psi_5 dW_2(t) \\ z_i &= y_1(t_i), \end{aligned} \tag{27}$$

with independent Wiener processes $W_1(t), W_2(t)$. As above we suppress all units, the time unit is one, which is $(t_{i+1} - t_i)$.

The data is simulated for 365 time units using an Euler scheme on a grid of one tenth of the measurement interval with the parameter set $\psi = [0.5, 3, 0.5, 0.2, 0.1]$.

We use the following extended state space model to estimate the first two parameters (ψ_1, ψ_2) and the stochastic volatility y_2:

$$\begin{aligned} d\tilde{y}_1 = dy_1 &= \tilde{y}_2 \left[\tilde{y}_3(t) - \tilde{y}_1(t)\right] dt + \tilde{y}_1(t) \tilde{y}_4(t) dW(t) \\ d\tilde{y}_2 = d\psi_1 &= 0 \\ d\tilde{y}_3 = d\psi_2 &= 0 \\ d\tilde{y}_4 = dy_2 &= 0 \\ z_i &= \tilde{y}_1(t_i). \end{aligned} \tag{28}$$

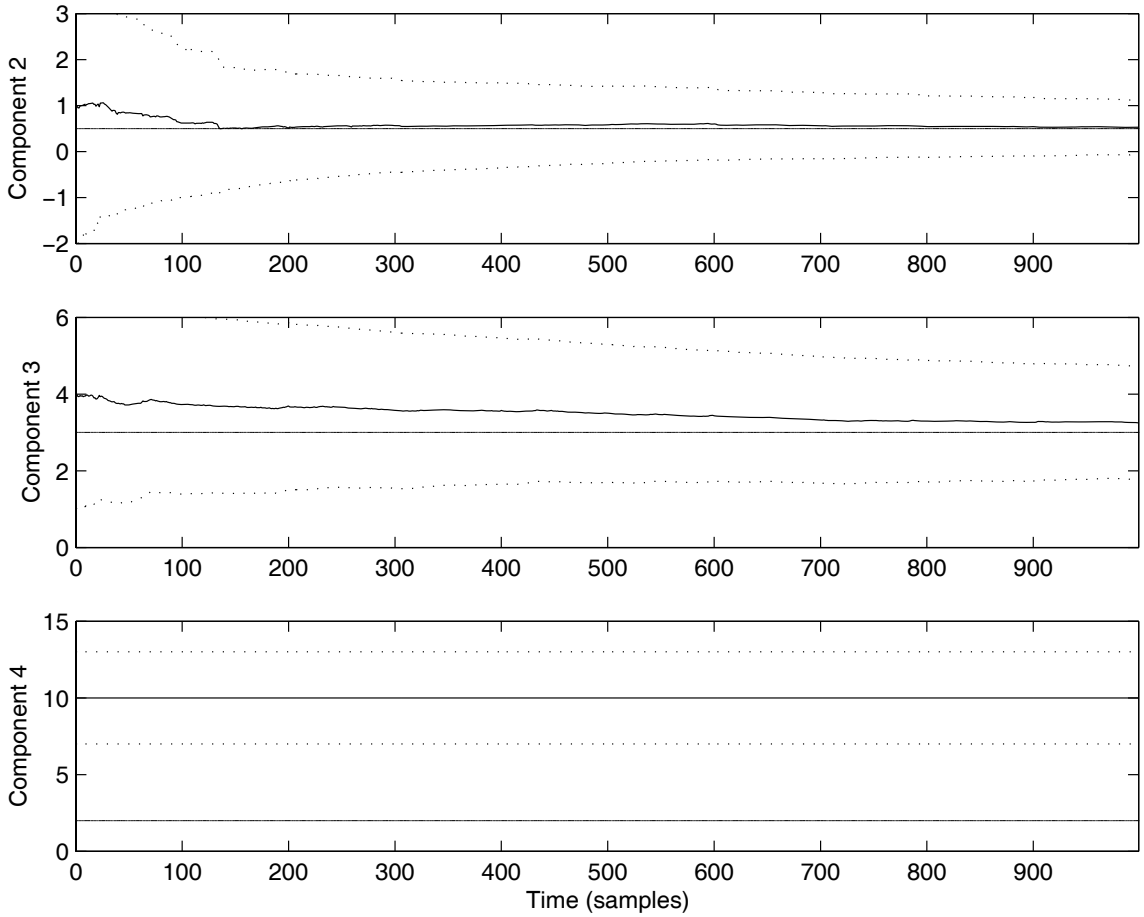

Figure 1: Estimation of the parameter set $[\tilde{y}_2, \tilde{y}_3, \tilde{y}_4]$ using the continuous-discrete unscented Kalman filter initialized with [1,4,10]. The true parameter values are marked. Confidence intervals are three standard deviations to each side. The diffusion parameter (component 4) is not estimated by the UKF.

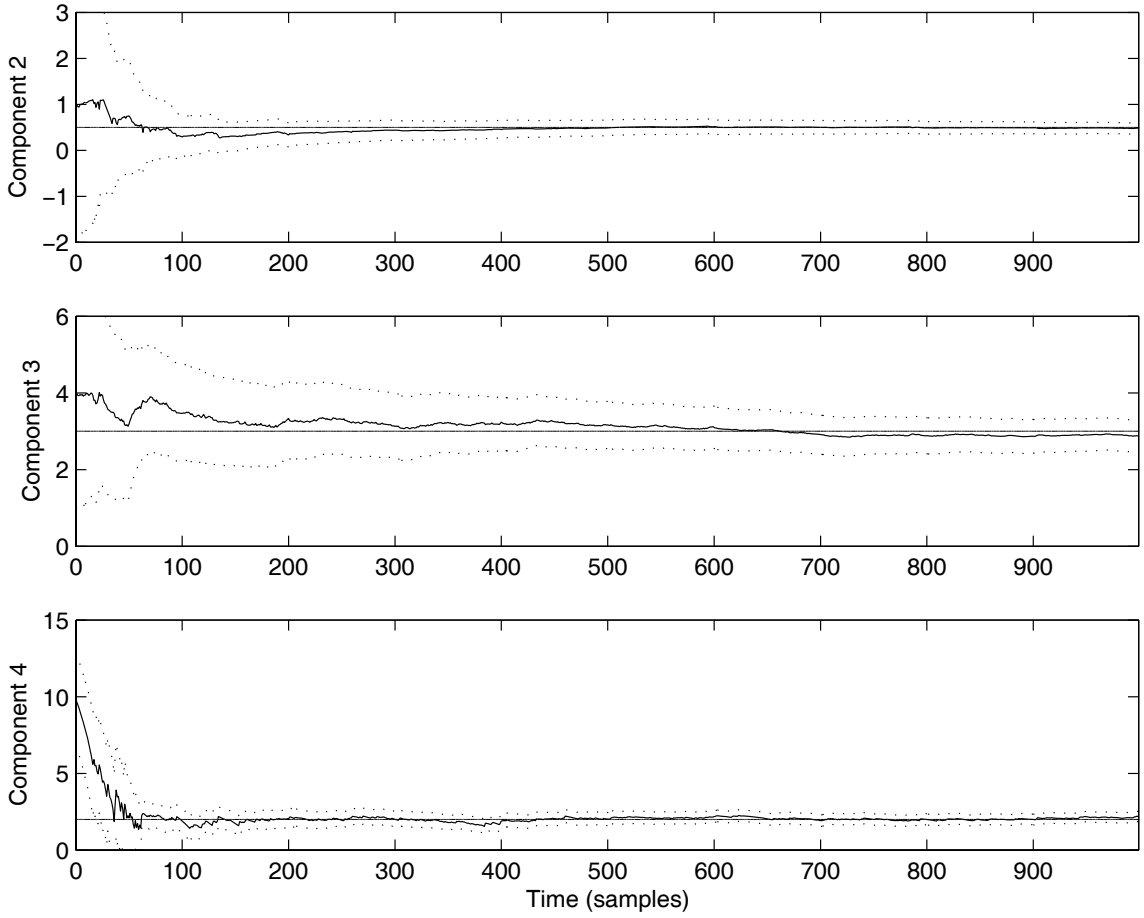

Figure 2: Estimation of the parameter set $[\tilde{y}_2, \tilde{y}_3, \tilde{y}_4]$ using the meta-algorithm adapted to the continuous-discrete unscented Kalman filter and initialized with [1,4,10]. The true parameter values are marked. Confidence intervals are three standard deviations to each side.

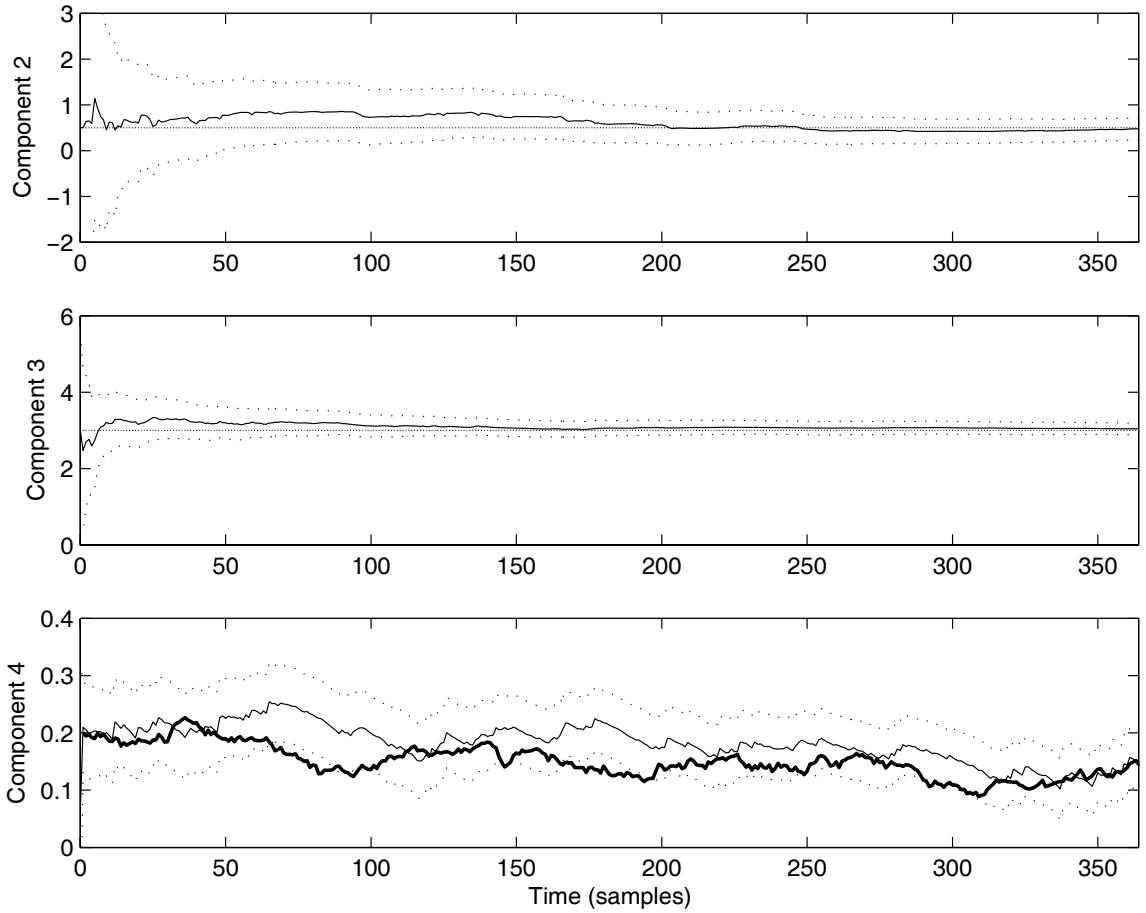

Figure 3: Estimation of the parameters $\tilde{y}_2, \tilde{y}_3$ and $\tilde{y}_4$ (=stochastic volatility) using the meta-algorithm adapted to the continuous-discrete unscented Kalman filter and initialized with the true parameter values. The true parameters are marked. Confidence intervals are three standard deviations to each side.

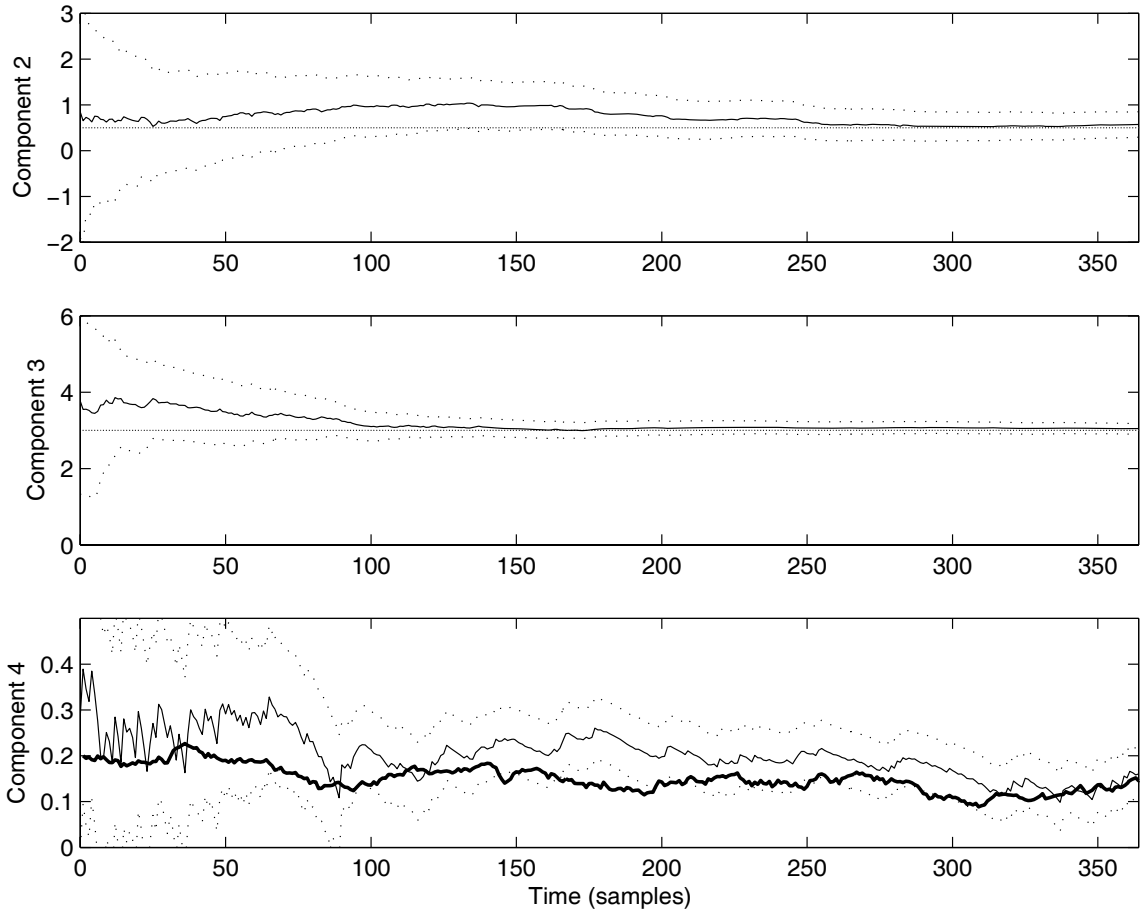

Figure 4: Estimation of the parameters $\tilde{y}_2, \tilde{y}_3$ and $\tilde{y}_4$ (=stochastic volatility) using the meta-algorithm adapted to the continuous-discrete unscented Kalman filter and initialized with [1,4,0.2]. The true parameters are marked. Confidence intervals are three standard deviations to each side.

The estimation results using the meta-algorithm are shown in Figure 3. The algorithm is initialized using the true parameter set and a diagonal covariance with variances of 1 for ψ_1 and ψ_2 and of 0.01 for the volatility y_2. As before, the first two parameters are estimated correctly. Furthermore, the algorithm tracks the volatility fluctuations. In Figure 4, the algorithm is initialized with the same covariance but with the parameter set $[\tilde{y}_2, \tilde{y}_3, \tilde{y}_4] = [1, 4, 0.2]$. This time, the algorithm needs more time to estimate the parameters due to the necessary correction time of the first two parameters. In both cases, the confidence intervals of all parameters do not shrink as fast as in the case with constant diffusion coefficient as the fluctuations cannot be captured perfectly.

6 Conclusion

In this paper Bayesian estimation of parameters of diffusion coefficients such as volatility in nonlinear state space models using moment-based filters is conducted. As this cannot be done by applying the moment-based filters directly, we present a meta-algorithm that can be adapted to the moment-based filters, so that Bayesian estimation of diffusion coefficients becomes possible. Our approach shows large advantages with respect to the computational costs over simulation-based filtering methods. We present the algorithm in the context of a continuous-discrete state space model using the recently proposed continuous-discrete unscented Kalman filter (Singer [18]). However, the algorithm can be used in a much broader context, as with all moment-based filters or in purely discrete state space models.

References

[1] Cox, J. C. ; Ingersoll, J. E. ; Ross, S. A.: A Theory of the Term Structure of Interest Rates. In: *Econometrica* 53 (1985), S. 385–407

[2] Hull, J. ; White, A.: The Pricing of Options with Stochastic Volatilities. In: *Journal of Finance* 42 (1987), Nr. 2, S. 281–300

[3] Ito, K. ; Xiong, K.: Gaussian Filters for Nonlinear Filtering Problems. In: *IEEE Transaction on Automatic Control* 45 (2000), Nr. 5, S. 910–927

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

[5] Julier, S.: The scaled unscented transformation. In: *Proceedings of the American Control Conference* (2002)

[6] Julier, S. ; Uhlmann, J. K.: A New Extension of the Kalman Filter to Nonlinear Systems. In: *Proc. of AeroSense: The 11th Int. Symp. A.D.S.S.C.* (1997)

[7] Julier, S. ; Uhlmann, J. K.: Unscented Filtering and Nonlinear Estimation. In: *Proc. of the IEEE* 92 (2004), Nr. 3

[8] Liptser, R.S. ; Shiryayev, A.N.: *Statistics of Random Processes, Volumes I and II.* Springer, 1978

[9] Lucía, J. J. ; Schwartz, E. S.: Electricity process and power derivates: Evidence from the nordic power exchange. In: *Review of Derivates Research* 5 (2002), S. 5–50

[10] Nielsen, M. J. ; Schwartz, E. S.: Theory of storage and the pricing of commodity claims. In: *Review of Derivatives Research* 7 (2004), Nr. 1, S. 5–24

[11] Nørgaard, M. ; Poulsen, N. K. ; Ravn, O.: Advances in Derivate-Free State Estimation for Nonlinear Systems. In: *Technical Report IMM-REP-1998-15, Department of Mathematical Modelling, DTU* (2000)

[12] Pitt, M. K. ; Shephard, N.: Filtering via Simulation: Auxiliary Particle Filters. In: *Journal of the American Statistical Association* 94 (1999), Nr. 446

[13] Ribeiro, D. R. ; Hodges, S. D.: A Two-Factor Model for Commodity Prices and Futures Valuation. In: *EFMA 2004 Basel Meetings Paper* (2004)

[14] Schwartz, E. S.: The stochastic behavior of commodity prices: Implications of valuation and hedging. In: *The Journal of Finance* 52 (1997), Nr. 3, S. 923–973

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

[16] Singer, H.: Nonlinear Continuous-Discrete Filtering using Kernel Density Estimates and Functional Integrals. In: *Journal of Mathematical Sociology* 27 (2003), S. 1–28

[17] Singer, H.: Simulated Maximum Likelihood in Nonlinear Continuous-Discrete State Space Models: Importance Sampling by Approximate Smoothing. In: *Computational Statistics* 18 (2003)

[18] Singer, H.: Continuous-Discrete Unscented Kalman Filtering. In: *Discussion paper 384, Fernuniversität Hagen* (2005)

[19] Sitz, A. ; Schwartz, U. ; Kurths, J. ; Voss, H.U.: Estimation of parameters and unobserverd components for nonlinear systems from noisy time series. In: *Phys. Rev. E* 66 (2002)

[20] van der Merwe, R. ; Wan, E. A.: Efficient Derivative-Free Kalman Filters for Online Learning. In: *ESANN'2001 proceedings* (2001), S. 205–210. ISBN 2–930307–01–3