



FernUniversität in Hagen

**Expected A Posteriori Estimation
in Financial Applications**

Thomas Mazzoni

Diskussionsbeitrag Nr. 424
April 2008

Diskussionsbeiträge der Fakultät für Wirtschaftswissenschaft
der FernUniversität in Hagen

Herausgegeben von der Dekanin der Fakultät

Alle Rechte liegen bei den Autoren

Expected A Posteriori Estimation in Financial Applications

Thomas Mazzoni

04-09-2008

Abstract

This paper introduces a new method for online estimation and prediction of states and parameters of nonlinear stochastic differential equations. In this setup parameters are considered as random variables in a *Bayesian* sense, which requires integration over parameter distributions. This is accomplished by well suited quadratures.

The suggested procedure is incorporated into a state space architecture, which allows for sequential calculation of likelihood functions. This is done by normal correlation updates and prediction error decomposition. The resulting EAP-Filter can process a variety of nonlinear problems, including latent states. Additionally, estimates and predictions for system states and parameters can be calculated online, without iterative loops.

Keywords: Expected a posteriori estimation; State space model; Maximum likelihood; *Gaußian* Sum-Filter.

1. Introduction

Stochastic differential equations (SDEs) are an inherent part of recent financial models. Especially diffusion problems, which are equivalent to stochastic differential equations of the *Itô*-type, have become an extensive field of scientific research. There are two key aspects related to this topic. Firstly, for nonlinear problems the transition density has to be approximated in an adequate way, because only in rare special cases there exists a closed form solution. Secondly, the parameters of the diffusion model have to be estimated. Additionally, recent research indicates that there may exist some micro structure noise, possibly resulting from discretized market prices of assets and other trade based artifacts. Hence, a measurement model should be taken into consideration, which leads to a (nonlinear) filtering problem.

Since the first extension of the exact linear Filter (Kalman 1960) to the extended *Kalman*-Filter (Schmidt 1966) extensive progress has been accomplished. The difference between particular procedures consists in the way, how they approximate the transition densities (or some properties of the transition densities, like moments etc.) and how they deal with nonlinearities of the measurement model. In financial applications nonlinear measurement models are of minor interest so the approximation of transition densities should be focused. It is always

possible, but time consuming, to simulate these densities via Monte Carlo (see for example Jäckel 2002). Therefore it is often more appropriate to approximate the unknown transition density, for example by a *Gaußian* density. Methods pursuing this strategy are Extended *Kalman*-Filter, Unscented *Kalman*-Filter (Julier and Uhlmann 1997, 2004, Sitz et al. 2002, 2004), *Gauß-Hermite*-Filter (Ito and Xiong 2000), DD1- & DD2-Filter (Nørgaard et al. 2000) and many hybrids. Basically they approximate expectation values of the nonlinear drift function in different ways, for example by *Taylor*-expansion, quadrature or by applying finite differences. More sophisticated ideas involving non *Gaußian* aspects of the transition density are for example *Gaußian* Sum-Filters (Sorenson and Alspach 1971, Ito and Xiong 2000), collecting nonlinear information from parallelization of a whole set of *Gaußian* filters. Another strategy is calculating a finite dimensional matrix representation of the diffusion operator. Therefore a complete set of orthogonal functions, in this case the *Hermite*-polynomials, is required. Several Methods have been proposed in this context. Ait-Sahalia (2002) calculates a closed-form solution involving higher moments for short time steps. Singer (2006) derived ordinary differential equations for higher order moments. Zhang et al. (1997) used distributed approximating functionals (DAFs) to obtain an orthogonal series representation of *Diracs* delta function. All methods based on orthogonal series expansion suffer from the same problems; it is exceedingly difficult to apply them on multivariate problems and an inherent feature of the finite orthogonal series approximation is the possible occurrence of negative values for the resulting density function. Remarkably Daum (1986) was able to derive an exact nonlinear filter for a small class of diffusion problems extending the work of Beneš (1981, 1985).

Another important question in financial applications is that of parameter estimation. The natural way to accomplish this task inside a *Gaußian* filter framework is using the prediction error decomposition (Schweppe 1965), which is a byproduct of the sequential filter update, for maximum likelihood estimation. Unfortunately the likelihood is a complex nonlinear function of the data and therefore has to be maximized numerically, that is, the whole observation series has to be filtered in every step of the optimization procedure. For a large number of parameters this is a computationally demanding task, not well suited for online estimation and prediction. Alternatively, drift parameters can be estimated in a *Bayesian* framework by extension of the state space. This approach does not require iterative constructions but it works only for drift parameters because of structural properties of the *Kalman*-update (see Mazzoni 2007, chap. 5.5). More generally, the diffusion parameters of the SDE are squared inside the *Fokker-Planck*-equation and therefore do not provide linear correlation (Grothe and Singer 2006). Non *Gaußian* filters can overcome this drawback (Singer 2008) but suffer from the above mentioned handicaps and require a more general measurement update, based directly on the *Bayes*-formula.

In this paper a new method for parameter and state estimation is suggested that does not require iterative loops, combining the advantage of sequential filter designs with asymptotic properties of maximum likelihood estimation. The paper is organized as follows: Section 2 introduces the theoretical basics and shows which strategies are applied. In section 3 the performance of the concept is

surveyed with simulation studies in different scenarios. In section 4 the proposed method is tested with real data for the Dow Jones Euro Stoxx 50 index. Section 5 concludes the results with a brief discussion.

2. Formal Derivation

The point of origin for any further step is the stochastic (*Itô*-) differential equation

$$dx_t = f(x_t, t, \psi)dt + g(x_t, t, \psi)dW_t, \quad (1)$$

with drift vector f , diffusion matrix g and a vector of independent *Brownian* motions W_t . In the following we assume that all conditions guaranteeing existence and uniqueness of a solution are satisfied as needed. The time dependency is written as subscript to clarify notation. Further we presume the *Itô*-process (1) autonomous, because an explicit time dependency of drift or diffusion terms can always be eliminated by extension of the state space. The parameter set ψ is usually treated as a fixed but unknown (constant) vector that has to be estimated. *Bayesian* theory suggests that this point of view may be too restrictive and consequently allows for a distribution $P(\psi)$ of the parameter vector. This distribution embodies the uncertainty about the parameter values due to incomplete knowledge. If we follow this idea consequently, we have to accept that the *Brownian* motion is not the only random source of the increment dx_t , but that it is blurred in some way by the distribution of ψ . Hence, we can calculate an expectation, which means integrating (1) with respect to the unknown distribution of ψ

$$E_\psi[dx_t] = \int_{\Psi} f(x_t, \psi)dP(\psi)dt + \int_{\Psi} g(x_t, \psi)dP(\psi)dW_t. \quad (2)$$

The syntax $E_\psi[...]$ indicates that the expectation is calculated regarding the parameter distribution. Now the basic idea is to approximate the integrals in (2) by quadrature in order to make the problem available for a *Gaußian* Sum-Filter architecture. This can be seen easily if we write the m -point quadrature for (2)

$$E_\psi[dx_t] \approx \sum_{j=1}^m w^j \left[f(x_t, \psi^j)dt + g(x_t, \psi^j)dW_t \right] = \sum_{j=1}^m w^j dx_t^j. \quad (3)$$

The quadrature points and weights are indicated by contravariant index in order to avoid confusion with notation of time dependency. Equation (3) suggests that the estimation problem can be treated as weighted sum over m single SDEs/diffusions, located at different points in the parameter space Ψ . But which quadrature should be used? Because we do not know anything about the distribution of ψ we have to make some reasonable assumptions. In most cases we are interested at least in estimating the first two moments, because with that information we have access to the most important characteristics of the unknown distribution.

2.1. Unscented Transform

Julier and Uhlmann (1997) introduced a quadrature called unscented transform (UT), operating with the smallest possible set of quadrature points (they call this set sigma points) that reproduces the first and second moment correctly. For an n -dimensional vector this quadrature takes $m = 2n + 1$ points

$$\psi^j = \mu_\psi + \text{sgn}[j]\sqrt{n + \kappa} S_{|j|} \quad \text{for } j = -n, \dots, n \quad (4a)$$

with S_j indicating the j -th column of the *Cholesky*-root resulting from the decomposition $\Sigma_\psi = SS'$. The corresponding weights are

$$w^0 = \frac{\kappa}{n + \kappa} \quad \text{and} \quad w^j = \frac{1}{2(n + \kappa)} \quad \text{for } j \neq 0. \quad (4b)$$

κ is a free scaling parameter and the quadrature sum runs from $-n$ to n . Julier and Uhlmann (1997) showed that the first and second moment is reconstructed correctly if calculated in the usual way from the weighted sum of quadrature points.

The unscented transform is attractive in two ways. Firstly, it requires only a very small set of quadrature points, the procedure is of order $\mathcal{O}(n)$, and therefore guarantees efficient calculations in high dimensional parameter spaces. Secondly, very few assumptions about the distribution of ψ are necessary, merely existence and finiteness of the first two moments are required.

2.2. Gauß-Hermite Quadrature

Applying *Gauß-Hermite* quadrature requires more assumptions, but its theoretical content is stronger than that of the unscented transform. If we are given the first two moments of an unknown distribution, then the *Gaußian* distribution fulfills the maximum entropy criterion. In other words, given two moments, the least manipulative distribution, regarding the information available, is the *Gaußian*. Making this assumption we can write (2) as expectation integral with a normal density $\phi(\psi) = N(\psi; \mu_\psi, \Sigma_\psi)$ inside

$$E_\psi[dx_t] = \int_{\Psi} f(x_t, \psi) \phi(\psi) d\psi dt + \int_{\Psi} g(x_t, \psi) \phi(\psi) d\psi dW_t. \quad (5)$$

Such integrals can be solved by *Gauß-Hermite* quadrature of the general form

$$\int_{\Psi} h(\psi) \phi(\psi) d\psi \approx \sum_{j=1}^m w^j h(S\psi^j + \mu_\psi), \quad (6)$$

with a continuous function h and S resulting from the decomposition $\Sigma_\psi = SS'$. The quadrature points ψ^j are n -tuples of the roots of univariate *Hermite* polynomials and the weights w^j are products of the one-dimensional weight functions. Both quantities can be calculated from the eigensystem of a particular tridiagonal matrix (see Golub 1973). The quadrature (6) is exact, if the integration domain is $\Psi = \mathbb{R}^n$ and h is a polynomial of degree $k \leq 2m - 1$.

The *Gauß-Hermite* quadrature has a strong theoretical rationale, but it requires more assumptions about the unknown distribution of ψ . Furthermore it is of order $\mathcal{O}(m^n)$, which means, if we consider a five-dimensional parameter space and three quadrature points per dimension for example, we are dealing with 243 quadrature points in a *Gauß-Hermiteian* framework. On the other hand, the unscented transform manages the quadrature with 11 points.

2.3. Calculating Transition Moments

At this point it might not be quite obvious, why one should treat parameters as random variables without dynamics and spend efforts in approximating expectations by discretizing the parameter space. The answer is related to the usual procedure of parameter estimation in the *Kalman-Filter* framework. Schweppe (1965) showed that the measurement update provides a prediction-error decomposition, from which the likelihood can be computed. Thus, the *Kalman-Filter* may not provide update information about extended state variables without linear correlation, but it provides the correct likelihood. Equation (3) indicates that the original SDE (and its corresponding diffusion) can be treated approximately as expectation regarding the parameter distribution, and therefore, as weighted sum over a number of different SDEs located at particular points in the parameter space. Each of these SDEs provides a different likelihood in every update step and delivers information about the parameters. Hence, we can use a *Gaussian* filter design, which means propagating only the first two moments in time and applying the *Kalman*-update, which is stable and easy to implement. The possibly nonlinear information about the parameters is gained from the weighted ensemble likelihood.

Propagating moments in time is easy for linear systems, but an exceedingly difficult task in nonlinear cases. In this paper a method is suggested unifying the stable and economic properties of the Extended *Kalman-Filter* and the flexibility of modern particle filters (UKF, GHF etc.). The starting point is the *Fokker-Planck*-equation, which governs the diffusion process corresponding to the SDE (1)

$$\frac{\partial}{\partial t} p(x, t) = -\frac{\partial}{\partial x_i} f_i(x, t) p(x, t) + \frac{1}{2} \frac{\partial^2}{\partial x_i \partial x_j} \Omega_{ij}(x, t) p(x, t), \quad (7)$$

or each single SDE of the quadrature ensemble, respectively. $p(x, t)$ is the probability density of the random process x at time t with $\Omega = gg^T$ on the right hand side of (7). The dependency of the parameter vector ψ is suppressed for simplicity. Further, Einsteins sum-convention has been used, indicating a sum over all elements in terms containing identical indices. To simplify further manipulations (7) is rewritten in a compacter form

$$\frac{\partial}{\partial t} p(x, t) = L_{FP} p(x, t), \quad (8)$$

with the *Fokker-Planck*-operator L_{FP} according to (7). For small times $\Delta t \ll 1$ the formal solution of the partial differential equation (8) can be approximated

$$p(x, t + \Delta t) = e^{L_{FP} \Delta t} p(x, t). \quad (9)$$

The exact solution of (8) includes integration over L_{FP} in the exponent of (9). This integral is approximated by a rectangle. In order to evaluate the solution, the exponential has to be expanded into a power series which unfolds multi derivative effects of the *Fokker-Planck*-operator. But we are mainly interested in propagating the moments of $p(x, t)$ and therefore, we can take advantage of (9) to obtain an equation for the expectation at $t + \Delta t$

$$\begin{aligned} \mu(t + \Delta t) &= \int x e^{L_{FP}\Delta t} p(x, t) dx = E[e^{L_{FP}^\dagger \Delta t} x] \quad \text{with} \\ L_{FP}^\dagger &= f_i(x, t) \frac{\partial}{\partial x_i} + \frac{1}{2} \Omega_{ij}(x, t) \frac{\partial^2}{\partial x_i \partial x_j}. \end{aligned} \quad (10)$$

The adjoint *Fokker-Planck*-operator L_{FP}^\dagger is the *Kolmogoroff*-backward-operator, which is the infinitesimal generator of the *Markovian* semigroup (cf. Aït-Sahalia 2002). Equation (10) can be easily verified by applying partial integration. The expectation is evaluated regarding $p(x, t)$, which is *Gaußian* by definition inside a *Gaußian* filter. Now we can expand the exponential into a *Taylor*-series. This is done up to second order to enhance accuracy and to obtain nonlinear correction terms

$$\begin{aligned} \mu(t + \Delta t) &= \mu(t) + E[f(x, t)]\Delta t + E\left[f_i(x, t) \frac{\partial}{\partial x_i} f(x, t)\right] \frac{\Delta t^2}{2} \\ &+ E\left[\Omega_{ij}(x, t) \frac{\partial^2}{\partial x_i \partial x_j} f(x, t)\right] \frac{\Delta t^2}{4}. \end{aligned} \quad (11)$$

Again, Einsteins sum-convention is used in (11). The first two terms are identical to the EKF *Euler*-step because they result from linear expansion of the backward-operator, but there are additional correction terms of order $\mathcal{O}(\Delta t^2)$, promising a better adjustment in nonlinear cases.

We are still left with expectation integrals regarding $p(x, t)$. Because we treated the unknown density *Gaußian*, one possible strategy is to approximate the integrals by *Gauß-Hermite*-Quadrature again. A different course of action is suggested here for the following reason. The *Gauß-Hermite*-quadrature is best suited for polynomials but we cannot control the character of nonlinearities that may occur. Possibly we are dealing with an exponential, which can be expanded into an infinite series generating polynomials of any degree. Therefore we would have to apply an adequate number of quadrature points, which can be time-consuming if the state space is high-dimensional. Hence, the expectations in (11) are expanded into a linear *Taylor*-series around $\mu(t)$, corresponding to the idea of the Extended *Kalman*-Filter. If we further restrict tensor dimensions, resulting from differentiating vectors and matrices, to three, we obtain an analytical formula for the *Euler*-step of the state expectation

$$\begin{aligned} \mu_{t+\Delta t} &= \mu_t + f(\mu_t)\Delta t + A(\mu_t)f(\mu_t) \frac{\Delta t^2}{2} \\ &+ \nabla \otimes A(\mu_t) \text{vec} \left[A(\mu_t)\Sigma_t + \frac{1}{2}\Omega(\mu_t) \right] \frac{\Delta t^2}{2}, \end{aligned} \quad (12)$$

with the *Jacobian* $A = \frac{\partial}{\partial x} f(x)$, the gradient vector $\nabla = \left(\frac{\partial}{\partial x_1}, \dots, \frac{\partial}{\partial x_n} \right)$ and ' \otimes ' representing the ordinary (right handed) *Kronecker*-product. The *vec*-operator

partitions an arbitrary matrix into a single vector column by column (see Magnus and Neudecker 2007, chap. 2.4). Equation (12) can still be easily computed by vector and matrix products and therefore is very efficient for implementation purposes. Nevertheless, the term $A(\mu_t)\Sigma_t$ may be neglected because it is potentially hazardous. The reason will become clear immediately, if we derive the appropriate formula for short time propagation of the state expectation error.

The state expectation error is a covariance matrix which means, it is positive definite by definition. This property is not necessarily preserved if we derive the approximation formula as sketched above. In order to guarantee positive definiteness only second order terms arising from outer products should be kept. After algebraic manipulations analog to (10) and (11) we obtain the short time propagation formula for the state expectation error

$$\begin{aligned} \Sigma_{t+\Delta t} &= (I + A(\mu_t)\Delta t)\Sigma_t(I + A(\mu_t)\Delta t)^T \\ &\quad + \Omega(\mu_t)\Delta t + f(\mu_t)f^T(\mu_t)\Delta t^2. \end{aligned} \quad (13)$$

This formula obviously generates a positive definite update for Σ_t but a few second order correction terms are neglected. These ghost-corrections are terms potentially reducing $\Sigma_{t+\Delta t}$ and therefore approximation (13) may be burdened with significant errors for large times. Thus, the Σ_t -term in (12) may have a hazardous influence particularly in long range scenarios.

2.4. Calculating Observation Updates

With the time update equations (12) and (13) and any quadrature (3), we have a weighted ensemble of priori moments for different parameter combinations. This is a *Gaußian* Sum-Filter setup (see Sorenson and Alspach 1971, Tanizaki 1996, chap. 4.2). Following Scheppe (1965), the likelihood of the linear observation $y_i = H_i x_i^j + \epsilon_i$ at t_i can be expressed as prediction error decomposition

$$L_i^j = \frac{1}{\sqrt{\det[2\pi\Gamma_i^j]}} e^{-\frac{1}{2}\nu^T\Gamma_i^{-1}\nu_i^j}, \quad (14a)$$

with

$$\nu_i^j = y_i - H_i \mu_{i|1:i-1}^j \quad \text{and} \quad \Gamma_i^j = H_i \Sigma_{i|1:i-1}^j H_i^T + R_i, \quad (14b)$$

where R_i is the covariance of the measurement error $\epsilon_i \sim N(0, R_i)$. The a priori moments have to be considered conditioned on the information from observations available up to the present time. This is indicated by the conditional subscript. Now by updating the quadrature weights along the particular likelihoods, the contribution of each parameter combination can be calculated. Sorenson and Alspach (1971) proved that the weights are updated along

$$w_i^j = \frac{w^j L_i^j}{w^T L_i}, \quad (15)$$

with $L_i = (L_i^1, \dots, L_i^m)^T$. For each single parameter vector (quadrature vector), the observation update is performed with the well known *Kalman*-Filter formula

$$K_i^j = \Sigma_{i|1:i-1}^j H_i^T [\Gamma_i^j]^{-1} \quad (16a)$$

$$\mu_{i|1:i}^j = \mu_{i|1:i-1}^j + K_i^j \nu_i^j \quad (16b)$$

$$\Sigma_{i|1:i}^j = (I - K_i^j H_i) \Sigma_{i|1:i-1}^j \quad (16c)$$

with the identity matrix I . The *Kalman*-update is simple and restricted to *Gaussian* priors and observation densities (normal correlation update), but it is also stable and efficient in computation.

2.5. Resampling

At this point two possible choices for the further course of action are available. On the one hand, one can calculate the new weights for the parameter quadrature along (15) and proceed with this new weights and the particular posterior moments calculated from (16a) to (16c). On the other hand a resampling step can be conducted which may turn out beneficial. The principal reason for resampling is the problem of weight degeneration, known from Monte Carlo simulation methods. Additionally, if no resampling is applied, the locations of the quadrature points for the parameter expectation integral are not altered, which is crucial, if the initial parameter density is non informative. In this case the variance and hence the spread of the quadrature points is very large, which may affect the state estimation.

Thus, the posterior moments after resampling can be calculated as weighted arithmetic mean

$$\mu_{i|1:i} = \sum_{j=1}^m w_i^j \mu_{i|1:i}^j \quad (17a)$$

$$\Sigma_{i|1:i} = \sum_{j=1}^m w_i^j \Sigma_{i|1:i}^j. \quad (17b)$$

For the posterior moments of the parameter distribution one obtains after resampling

$$E[\psi_i] = \sum_{j=1}^m w_i^j \psi_{i-1}^j \quad (18a)$$

$$\text{Cov}[\psi_i] = \sum_{j=1}^m w_i^j (\psi_{i-1}^j - E[\psi_i]) (\psi_{i-1}^j - E[\psi_i])^T. \quad (18b)$$

There exists a significant difference between state and parameter estimation, indicated in the notation of the appropriate moment equations. Because the driving process is formulated as stochastic differential equation, state estimates are available at any time, especially inside sampling intervals. Parameters are treated in a *Bayesian* way as random variables without dynamics. Therefore information can be obtained solely associated with new sampling information

(expected a posteriori estimation). The state expectation and the expectation error are solutions of moment differential equations in a continuous time domain, whereas moments of the parameter vector occur as time series. The whole algorithm is summarized in figure 1 for convenience.

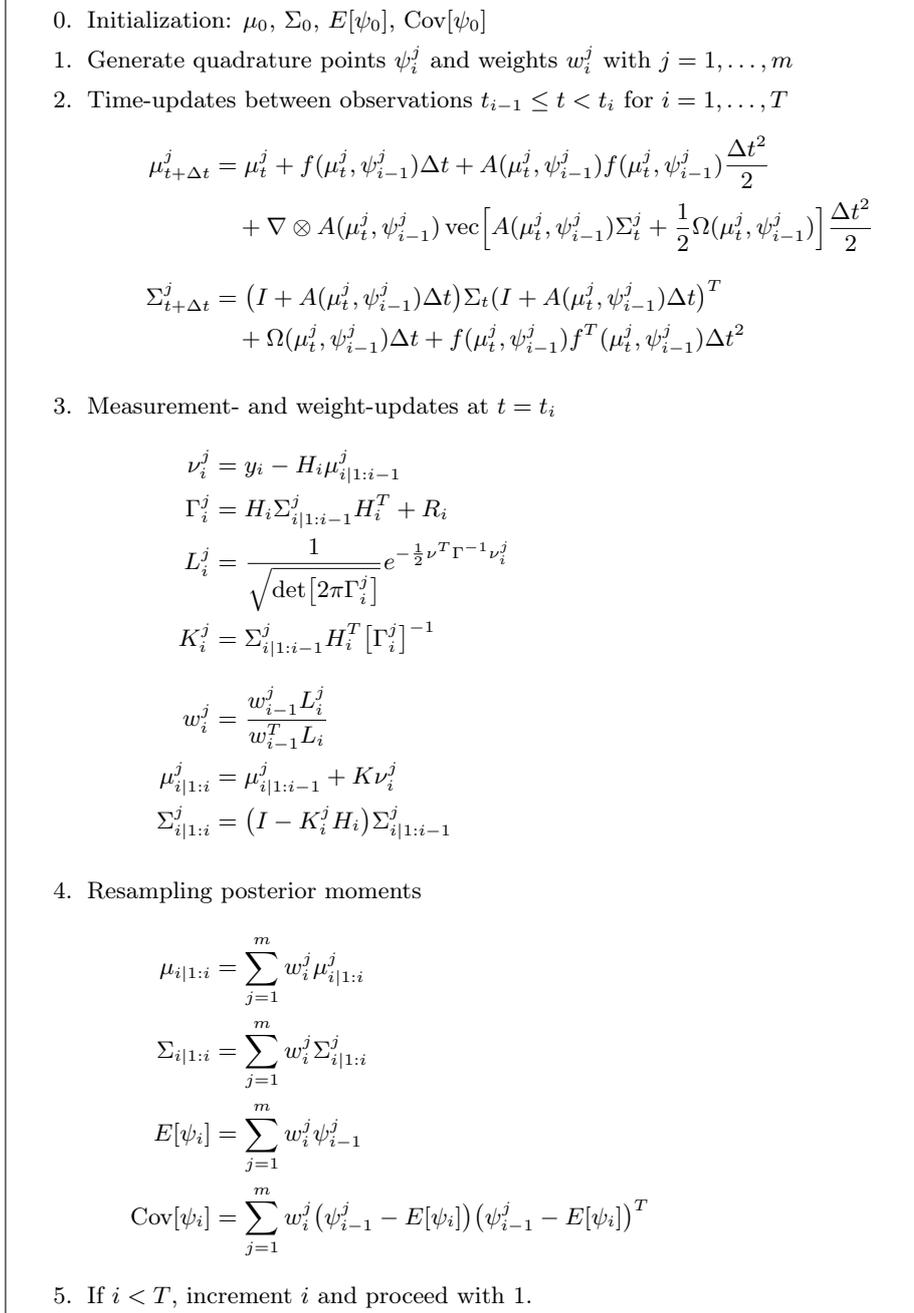


Figure 1: EAP-Filter

3. Performance and Properties of the EAP-Filter

The first example scenario for the EAP-Filter is the geometric *Brownian* motion (GBB)

$$dx_t = \mu x_t dt + \sigma x_t dW_t. \quad (19a)$$

The GBB is important for two reasons. At first, it is a fundamental model for stock price changes, used for example in the *Black-Scholes*-model for option pricing (Black and Scholes 1973). At second, it is a very simple model, which allows to pinpoint the properties of the suggested procedure in a very transparent way.

Because the real stock price is not observed exactly, it is rounded up to the second decimal place, a measurement model needs to be involved

$$y_i = x_{t_i} + \epsilon_i. \quad (19b)$$

Properly, the observation error ϵ_i is uniformly distributed, but in order to avoid complications it is assumed *Gaussian* with variance $R_i = 1.2 \times 10^{-5}$, minimizing the squared integrated distance between the distribution functions.

3.1. Filtering Parameters of the GBB

Both parameters in (19a) can be estimated inside the traditional *Kalman*-Filter with maximum likelihood. The price to pay is numerical maximization of the likelihood-function which is a computationally demanding task in case of an extensive number of parameters. Thus, numerical maximization procedures are to be avoided here. Alternatively, parameters can be estimated in a *Bayesian* way by extension of the state space

$$d \begin{pmatrix} x_t \\ \mu \\ \sigma \end{pmatrix} = \begin{pmatrix} \mu x_t \\ 0 \\ 0 \end{pmatrix} dt + \begin{pmatrix} \sigma x_t \\ 0 \\ 0 \end{pmatrix} dW_t. \quad (20)$$

Obviously, we are now dealing with a nonlinear drift and diffusion function because state variables are multiplied with each other. This is not a vital problem because the original *Kalman*-Filter can be extended by *Taylor*-expansion to the Extended *Kalman*-Filter (Schmidt 1966), which is fully sufficient for this task.

Another problem occurs, which cannot be corrected in the traditional *Gaussian* filter framework. To illustrate the nature of this problem a stock price process is simulated with annual drift $\mu = 5\%$ and volatility $\sigma = 20\%$. The simulated process is observed daily for $T = 2500$ days. Figure 2 shows the filter-solution for the extended state variables according to (20). The Filter was initialized with $E[\mu, \sigma] = (0.1, 0.1)^T$ and $\text{Cov}[\mu, \sigma] = I$. It is easily seen that the drift is estimated well but no information about the volatility is gained from observation. Figure 2 additionally provides the true parameter value (green) and HPD-bands with ± 2 standard deviations (red, yellow) around the mean (blue).

The same estimation problem can be treated with the outlined EAP-Filter. Here, no artificial extension of the state space is necessary in order to estimate

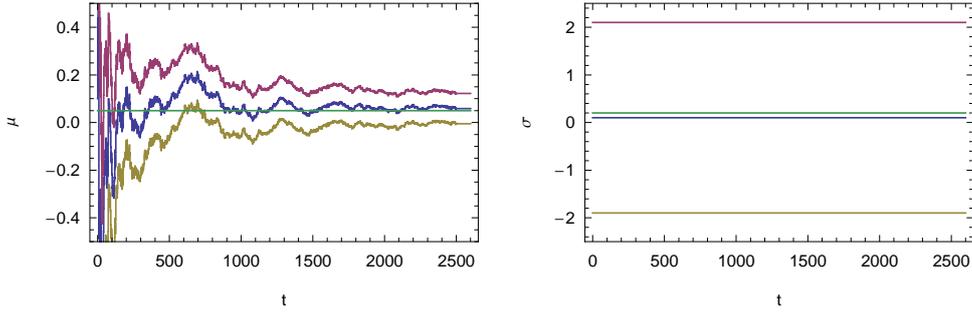


Figure 2: Filtered Parameters of the GBB (Extended *Kalman*-Filter)

the parameters without iterative procedures. A minor modification of the GBB is suggested to ease quadrature operations. Because volatility is always positive, σ is transformed into $\exp[h]$, where h is the parameter to be estimated. Thus, the GBB reads

$$dx_t = \mu x_t dt + e^h x_t dW_t \quad \text{with} \quad e^h = \sigma. \quad (21)$$

With this modification the admissible range for the parameters is $-\infty$ to ∞ and the quadrature rules discussed above can be applied. Alternatively, a *Gauß-Laguerre* quadrature could have been used for the original volatility, ranging from 0 to ∞ .

Figure 3 illustrates the parameter estimation for the geometric *Brownian* motion (21), where a *Gauß-Hermite* quadrature with $m = 3$ was used. The initial parameter distribution was assumed *Gaußian* with $E[\mu, h] = (0.1, \log[0.1])^T$ and $\text{Cov}[\mu, h] = I$. According to Figure 2, the true parameter is given by a green line. Because σ is distributed asymmetrically (log-normal), the 2.5%, 50% and 97.5% quantiles are given (yellow, blue and red). Obviously, the mean estimation shows the same characteristics as in the conventional case (Figure 2), but now an online estimation for the volatility is available. The true parameters are covered by the 95% HPD-bands (High Probability Density) of the filtered parameter densities.

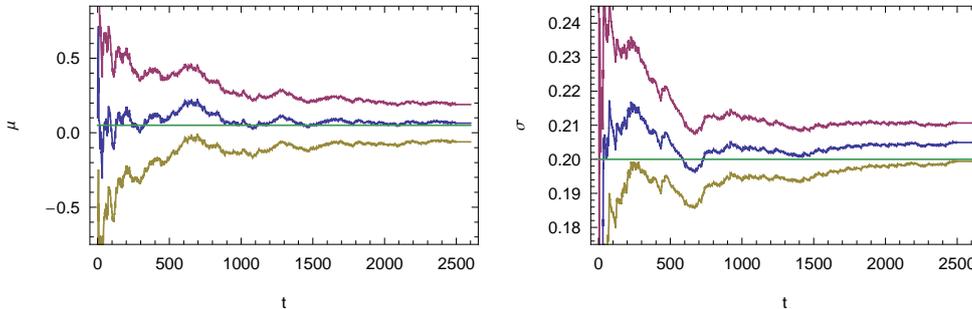


Figure 3: Filtered Parameters of the GBB (EAP-Filter)

3.2. Filtering Latent States of Stochastic Volatility Models

Another drawback of conventional *Gaußian* filters is the inability to track and predict latent states, if they occur only in the diffusion function of the original model. An example for such models are those with stochastic volatility (cf. Taylor 1994)

$$dx_t = \mu x_t dt + e^{h_t} x_t dW_t \quad (22a)$$

$$dh_t = \lambda(\bar{h} - h_t) + \nu dV_t. \quad (22b)$$

Equation (22b) is an ordinary mean reversion process with mean reversion speed λ , mean reversion level \bar{h} and a *Brownian* motion V_t , independent of W_t . This premise can be relaxed in order to deal with a leverage effect (e.g. Heston 1993). The volatility $\sigma_t = \exp[h_t]$ now is a state variable with its own stochastic dynamics. Because it is not observed, information is only gathered from (simulated) movement of the observable stock price. This information is sparse, so the mean reversion speed and diffusion coefficient are given with $(\lambda, \nu) = (0.05, 0.05)$. The latent state, drift coefficient μ and mean reversion level \bar{h} are estimated; the UT-method with scaling parameter $\kappa = 1$ was used as quadrature rule. The results are shown in figure 4. Initial conditions and organization of the illustrations are analog to the previous example.

Obviously the latent state (upper) is estimated adequately. The major characteristics of the simulated trajectory are tracked by the state expectation and the 95% HPD-band covers the trajectory almost everywhere. Nevertheless, the

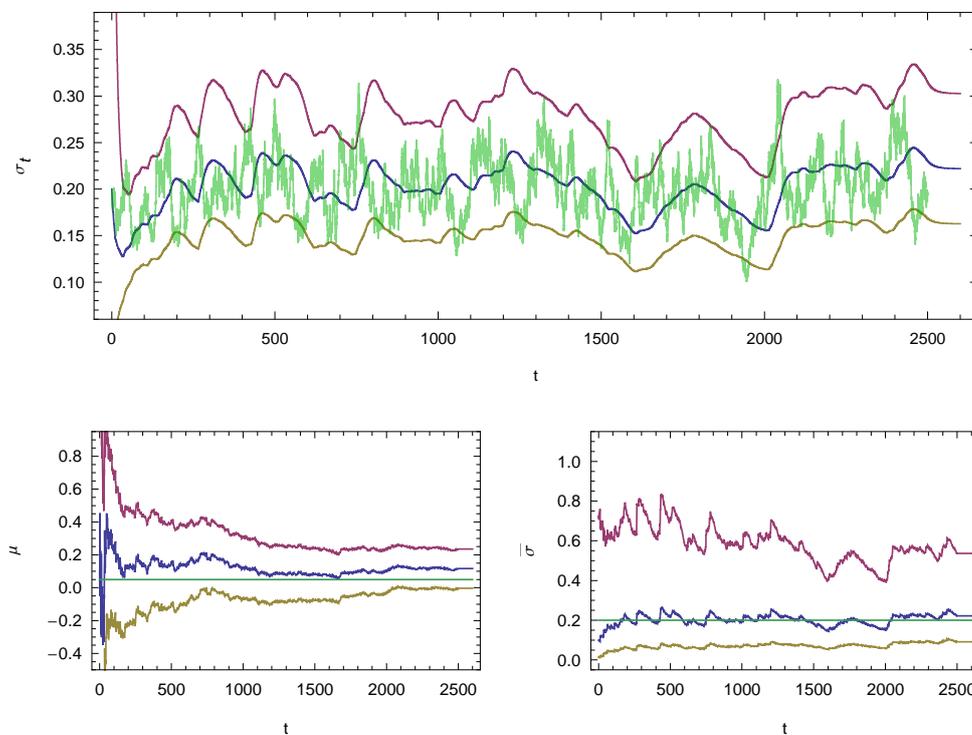


Figure 4: Filtered Latent State and Parameters of the SVola Model

lack of information due to non observability of the volatility process results in large quantile spans for the estimated mean volatility level $\bar{\sigma} = \exp[\bar{h}]$ (lower right). The parameter estimation method suggested here is far more sensitive to sparse information than maximum likelihood procedures. This topic is worth a brief discussion.

3.3. Properties of Parameter Estimation

The advantage of the method suggested above, is the dispensability of iterative loops. The price to pay is on the one hand, the parallelization of a whole diffusion ensemble and on the other hand, the dependence on prior distributions of the parameter vector. The latter is crucial, if the prior distribution is non informative. This property is worth a closer look.

Because the parameter vector is calculated as weighted average of likelihood contributions and the likelihood itself at time t_i is conditioned on the available information provided by the data $y_{1:i}$, the estimator $\hat{\psi}_i$ is an expected a posteriori estimator, $\hat{\psi}_i = E[\psi|y_{1:i}]$. The maximum likelihood estimator is composed in a completely different way. Neglecting the prior distribution, it is calculated as

$$\hat{\psi}_{ML} = \arg \max_{\psi} \sum_{i=1}^T l(x_i, \psi|y_{1:i}). \quad (23)$$

Here, $l(\dots)$ indicates the log-likelihood function. From (23) one can easily see that the ML-estimator is engineered to find a parameter set, which is best suited for the whole sequence of ascending segments of the trajectory. Because the specific segments are weighted equally, it is robust against outliers, no matter where they appear. The expected a posteriori estimator (EAP) is based on the present observation sequence $y_{1:i}$ only, and hence uses less information than the ML estimator. Heuristically the behavior of the EAP-estimator can be understood as follows: the process is filtered with changing parameter values, which are adjusted after every observation. As the observation count increases, adjustments become smaller, because more information of the past is stored in the present value of ψ , resulting in sharper localized prior densities.

This strategy avoids iterative constructions but it can also cause problems, if initial distributions are non informative. In such cases outliers can cause serious problems, because rapid adjustments of the EAP-estimator may be the consequence. This can vitally threaten the algorithm due to numerical destabilization. A very cheap and effective trick to avoid this kind of problem is to calculate an exponentially smoothed update of the parameter estimator

$$\tilde{\psi}_i = (1 - \lambda_i)\psi_i + \lambda_i\tilde{\psi}_{i-1} \quad \text{with} \quad 0 \leq \lambda_i < 1. \quad (24)$$

The weight coefficient λ_i is indicated as function of time t_i , because with increasing observations, the available information grows and the adjustments can be trusted to be more moderate. So λ_i can be defined as monotonic function with $\lim_{i \rightarrow \infty} \lambda_i = 0$. The resulting stabilized parameter estimates can be used either directly or as priors for another unpatched estimation cycle.

4. Application to Financial Market Data

To provide a real world application to the outlined estimation algorithm, financial time series of the Dow Jones Euro Stoxx 50 index and the corresponding volatility, estimated from implied volatility of available index options (figure 5) are analyzed¹. The example data is observed daily and ranges from 03/18/2003 to 03/17/2008. The model under consideration is the stochastic volatility model (22a) and (22b), discussed in the previous section, with the extension of correlated noise processes

$$dW_t dV_t = \rho dt. \quad (25)$$

Because correlation can take values only in the interval $[-1, 1]$, the transformation $\rho = \tanh[\varrho]$ is applied, where ϱ again ranges from $-\infty$ to ∞ . Alternatively ρ can be treated with *Gauß-Legendre* quadrature in the original integration domain. Figure 6 shows the parameter density estimates for the Dow Jones Euro Stoxx 50 data. A *Gauß-Hermite* scheme with four quadrature points was used. The initial distribution was assumed *Gaussian* with $\mu_\psi = (0.1, 0.05, 0.2, 0.05, -0.5)^T$ and $\Sigma_\psi = 0.1 \times I$.

Obviously parameters and states are estimated well. State predictions can be seen in Figure 5 including the usual quantiles of 2.5%, 50% and 97.5%.

5. Conclusions

A new method, based on recursive *Gaussian* Sum-Filter architecture, was introduced, allowing online estimation of states and parameters of a stochastic differential equation model with discrete noisy observations. The method contains two innovations; firstly the idea of treating parameters as random variables and calculating an expected a posteriori estimate by quadrature of the parameter space, secondly a sophisticated derivation of analytical nonlinear time update equations for the state expectation and its error covariance. This update was calculated by adjoining the *Fokker-Planck* operator and *Taylor* expansion up to second order terms, guaranteeing a positive semidefinite update for the expectation error covariance.

¹The data is available at <http://www.stoxx.com>.

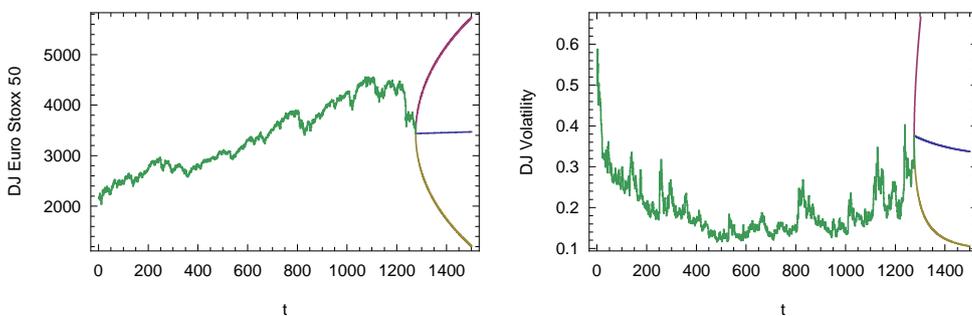


Figure 5: Dow Jones Euro Stoxx 50 Index and Volatility

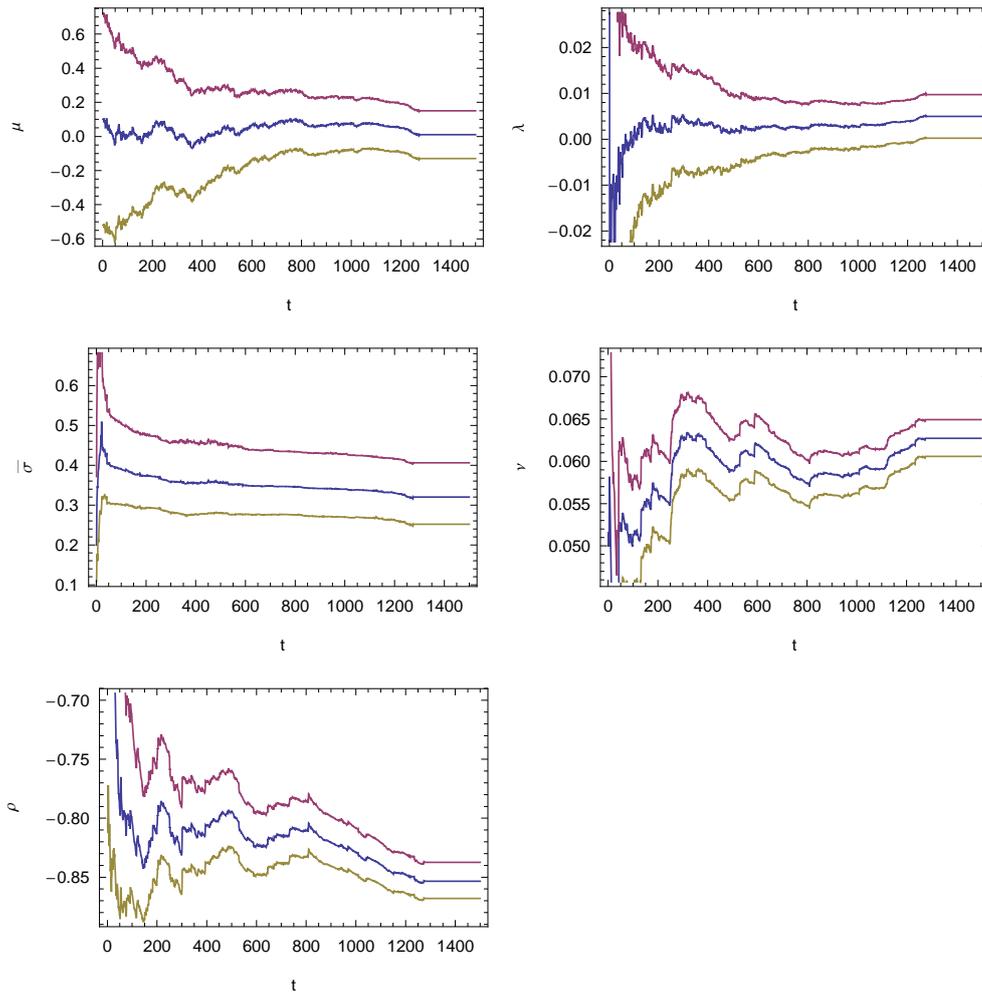


Figure 6: Estimated Parameter Densities for Dow Jones Euro Stoxx 50

The EAP-Filter has proven well suited both in simulation and in application to real financial data. It is theoretically attractive, because unlike other strategies for online estimation, like artificial extension of the state space, it separates the role of system states and parameters very clearly. Furthermore, EAP-Filtering is very economical, because the state space remains unextended and no iteration has to be conducted in order to maximize some kind of optimization function. Therefore it is best suited for applications where estimates and predictions have to be calculated very fast (online) and prior information is available.

References

- Aït-Sahalia, Y. (2002): Maximum Likelihood Estimation of Discretely Sampled Diffusions: A Closed-Form Approximation Approach. *Econometrica*, 70(1):223–262.
- Beneš, A.R. (1981): Exact finite-dimensional for certain diffusions with nonlinear drift. *Stochastics*, 5(1/2):65–92.

- Beneš, A.R. (1985): New exact nonlinear filters with large Lie algebras. *Systems & Control Letters*, 5:217–221.
- Black, F. and M. Scholes (1973): The Pricing of Options and Corporate Liabilities. *Journal of Political Economy*, 81:637–654.
- Daum, F.E. (1986): Exact finite dimensional nonlinear filters. *IEEE Transactions on Automatic Control*, 31(7):616–622.
- Golub, G.H. (1973): Some Modified Matrix Eigenvalue Problems. *SIAM Review*, 15(2):318–334.
- Grothe, O. and H. Singer (2006): *Bayesian Estimation of Volatility with Moment-Based Nonlinear Stochastic Filters*. Tech. Rep. 401, FernUniversität in Hagen.
- Heston, S.L. (1993): A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bonds and Currency Options. *The Review of Financial Studies*, 6(2):327–343.
- Ito, K. and K. Xiong (2000): Gaussian Filters for Nonlinear Filtering Problems. *IEEE Transactions on Automatic Control*, 45(5):910–927.
- Jäckel, P. (2002): *Monte Carlo Methods in Finance*. John Wiley & Sons, New York, Brisbane, Toronto.
- Julier, S. and J. Uhlmann (1997): A New Extension of the Kalman Filter to Nonlinear Systems. Paper presented at the 11th International Symposium on Aerospace/Defense Sensing, Simulation and Control, Orlando, Florida.
- Julier, S. and J. Uhlmann (2004): Unscented Kalman Filtering and Nonlinear Estimation. *Proceedings of the IEEE*, 92(3):401–422.
- Kalman, R.E. (1960): A New Approach to Linear Filtering and Prediction Problems. *Transactions of the ASME—Journal of Basic Engineering*, 82 (Series D):35–45.
- Magnus, J.R. and H. Neudecker (2007): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley & Sons, New York, Brisbane, Toronto, 3rd edn.
- Mazzoni, T. (2007): *Stetig/diskrete Zustandsraummodelle dynamischer Wirtschaftsprozesse*. Shaker, Aachen.
- Nørgaard, M.; N.K. Poulsen; and O. Ravn (2000): New Developments in State Estimation for Nonlinear Systems. *Automatica*, 36:1627–1638.
- Schmidt, S.F. (1966): Application of State-Space Methods to Navigation Problems. In: *Advances in Control Systems. Theory and Applications*, ed. C.T. Leondes, Academic Press, New York, San Francisco, London, vol. 3, pp. 293–340.
- Schweppe, F. (1965): Evaluation of Likelihood Functions for Gaussian Signals. *IEEE Transactions on Information Theory*, 11:61–70.
- Singer, H. (2006): Moment Equations and Hermite Expansion for Nonlinear Stochastic Differential Equations with Application to Stock Price Models. *Computational Statistics*, 21(3):385–397.
- Singer, H. (2008): Generalized Gauss-Hermite Filtering. *ASTA Advances in Statistical Analysis*. DOI: 10.1007/s10182-008-0068-z.

- Sitz, A.; U. Schwarz; and J. Kurths (2004): The Unscented Kalman Filter, a powerful Tool for Data Analysis. *International Journal of Bifurcation and Chaos*, 14(6):2093–2105.
- Sitz, A.; U. Schwarz; J. Kurths; and H.U. Voss (2002): Estimation of Parameters and Unobserved Components for Nonlinear Systems from Noisy Time Series. *Physical Review E*, 66(016210):1–9.
- Sorenson, H.W. and D.L. Alspach (1971): Recursive Bayesian estimation using Gaussian sums. *Automatica*, 7:465–479.
- Tanizaki, H. (1996): *Nonlinear Filters. Estimation and Applications*. Springer, Berlin, Heidelberg, New York, 2nd edn.
- Taylor, S.J. (1994): Modelling Stochastic Volatility. *Mathematical Finance*, 4:183–204.
- Zhang, D.S.; G.W. Wei; D.J. Kouri; and D.K. Hoffman (1997): Numerical method for the nonlinear Fokker-Planck equation. *Physical Review E*, 56(1):1197–1206.