Applied State Space Modelling of Non-Gaussian Time Series using Integration-based Kalman-filtering

Frühwirth-Schnatter, Sylvia

DOI:
10.57938/7d02126c-2777-4c02-97d9-04b1a62ba68c

Published: 01/01/1993

Document Version:
Publisher's PDF, also known as Version of record

Applied State Space Modelling of Non-Gaussian Time Series using Integration-based Kalman-filtering

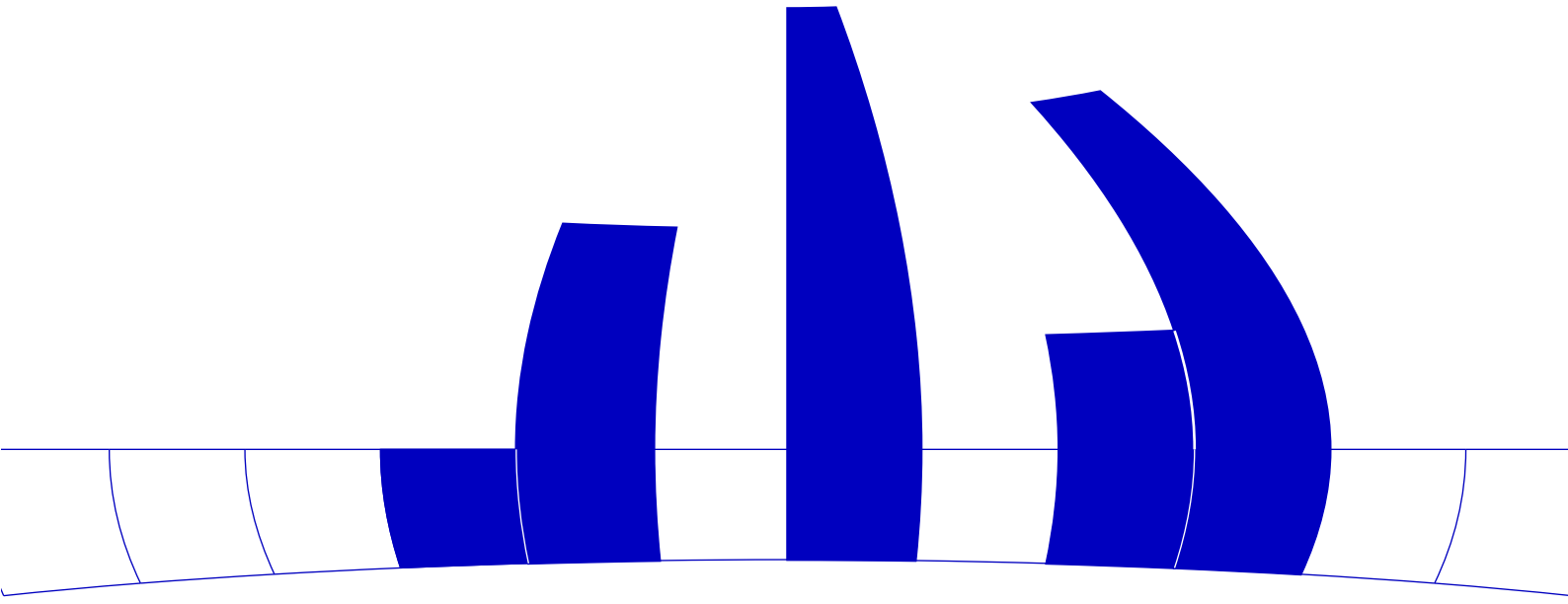
Sylvia Frühwirth-Schnatter

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 35
April 1993

http://statmath.wu-wien.ac.at/



Applied State Space Modelling of Non-Gaussian Time Series using Integration-based Kalman-filtering

Sylvia Frühwirth-Schnatter
Department of Statistics, University of Economics,
Augasse 2-6, A-1090 Vienna, Austria

April 1993

Keywords: Approximate Bayesian Inference; Bayesian Computation; Dynamic Generalized Linear Models; Gauss-Hermite Integration; Kalman-Filtering; Model likelihood; Non-normal State Space Models; Non-Gaussian Time Series; Robust Filtering

Abstract

The main topic of the paper is on-line filtering for non-Gaussian dynamic (state space) models by approximate computation of the first two posterior moments using efficient numerical integration. Based on approximating the prior of the state vector by a normal density, we prove that the posterior moments of the state vector are related to the posterior moments of the linear predictor in a simple way. For the linear predictor Gauss-Hermite integration is carried out with automatic reparametrization based on an approximate posterior mode filter. We illustrate how further topics in applied state space modelling such as estimating hyperparameters, computing model likelihoods and predictive residuals, are managed by integration-based Kalman-filtering. The methodology derived in the paper is applied to on-line monitoring of ecological time series and filtering for small count data.

1 Introduction

The topic of statistical inference within non-Gaussian state space models has received considerable attention during the last two decades. Whereas off-line analysis through Gibbs sampling (Carlin *et al.*, 1992; Fahrmeir *et al.*, 1992) and iterative Gauss-Newton and Fisher-scoring methods (Fahrmeir and Kaufmann, 1991) lead to accurate estimates of the smoothing densities of the state vector and of smoothed posterior modes and curvature, respectively, exact sequential filtering schemes for on-line analysis still seem unfeasible. The method of Kitagawa (1987) who used

crude numerical integration over the state space by piecewise linear approximation is without doubt limited to models with a univariate or at most a bivariate state vector. Approximate sequential filtering schemes have been derived for different classes of non-Gaussian distributions by the use of various approximation methods including the approach of Masreliez (1975), the WHM-Algorithm (West *et al.*, 1985) and approximate posterior mode filtering (Fahrmeir, 1992).

As a gap between these methods and exact numerical integration Schnatter (1992) suggested a filtering scheme based on efficient Gauss-Hermite-integration (Naylor and Smith, 1982; Smith *et al.*, 1985) for the special case of a dynamic generalized linear trend model. This filter has been designed to improve low order approximations to the posterior mode or to the posterior expectation under the assumption that the state prediction density is approximately normal. The scope of the present paper is to illustrate the implementation of integration-based Kalman-filtering for fairly general models. The scheme developped in the paper may be applied to any dynamic generalized linear model (DGLM) with a linear Gaussian transition equation for the state vector x_t,

$$x_t = F_t \cdot x_{t-1} + w_t, \quad w_t \sim \mathrm{N}(0, Q_t), \tag{1}$$

and non-normal observations densities $p(y_t|x_t) = p(y_t|\lambda_t)$ which depend on the state vector x_t only through the linear predictor $\lambda_t = H_t x_t$. The observed process y_t may be multivariate and $p(y_t|x_t)$ need not belong to an exponential family. We will show that integration-based Kalman-filtering may be formulated in such a way that at each time step numerical integration is of the same dimension as the observation vector y_t. Filtering for univariate time series therefore envolves just a one-dimensional integral, even in cases where the dimension of the state vector is high (e.g. equal to 13 for a basic structural model with a form-free monthly seasonal pattern). Numerical integration is carried out by Gauss-Hermite-integration with automatic reparametrization based on an approximate posterior mode filter. The resulting non-linear filter for the first two moments $E(x_t|y^t)$ and $V(x_t|y^t)$ of the posterior density of the state vector x_t given the history y^t is sequential and therefore well-suited for on-line applications.

In Section 2 the filter is derived in detail. Section 3 deals with various additional topics of applied state space modelling such as estimating unknown hyperparameters, approximating the model likelihood, filtering the grand mean within exponential families, and approximating predictive residuals. In Section 4 integration-based Kalman-filtering is applied to on-line monitoring of ecological time series such as daily sulphur dioxide emission or weekly sulphate concentrations in a lake, and Kalman-filtering for time series of small count data.

2 Integration-based Kalman-filtering for non-Gaussian time series

It is well known that for Gaussian time series the first two moments of the posterior density $p(x_t|y^t)$ are easily determined from the first two moments of the density $p(x_{t-1}|y^{t-1})$ by a sequential scheme (Kalman-filtering). If such a sequential scheme for the first two moments is derived for non-Gaussian time series it may be called Kalman-filtering for non-Gaussian time series.

The first two moments of the posterior are given by the following integrals:

$$E(x_t|y^t) = \frac{I^\star(x_t)}{I^\star(1)}$$
$$V(x_t|y^t) = \frac{I^\star(x_t x_t^{\mathrm{T}})}{I^\star(1)} - E(x_t|y^t)E(x_t|y^t)^{\mathrm{T}}$$

with

$$I^\star(h(x_t)) = \int h(x_t)p(y_t|x_t)p(x_t|y^{t-1})\,dx_t. \tag{2}$$

Integration-based Kalman-filtering as proposed in Schnatter (1992) is based on sequential Gauss-Hermite integration of (2) the variable of integration being equal to the state vector x_t. While a straightforward extension of this filter to high dimensional state vectors is unfeasible, it will be shown below, that the filter may be reformulated in such a way that the dimension of integration is equal to the dimension of the observation vector, only.

Although the integration error can be reduced by increasing the number of grid points, integration-based Kalman-filtering is approximate insofar as the prior $p(x_t|y^{t-1})$ which is not completely known for $t > 1$ is approximated by a normal density with the same first and second moment as $p(x_t|y^{t-1})$:

$$E(x_t|y^{t-1}) = F_t \cdot E(x_{t-1}|y^{t-1}),$$
$$V(x_t|y^{t-1}) = F_t \cdot V(x_{t-1}|y^{t-1}) \cdot F_t^{\mathrm{T}} + Q_t.$$

Based on a normal approximation of $p(x_t|y^{t-1})$, however, it is shown in Lemma 1 below that the posterior moments $E(x_t|y^t)$ and $V(x_t|y^t)$ are linear functions of the posterior moments $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ of the predictor $\lambda_t = H_t \cdot x_t$. Numerical integration, therefore, is necessary only for computing the posterior moments $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ of the linear predictor λ_t:

$$E(\lambda_t|y^t) = \frac{I(\lambda_t)}{I(1)} \tag{3}$$
$$V(\lambda_t|y^t) = \frac{I(\lambda_t\lambda_t^{\mathrm{T}})}{I(1)} - E(\lambda_t|y^t)E(\lambda_t|y^t)^{\mathrm{T}} \tag{4}$$

with

$$I(h(\lambda_t)) = \int h(\lambda_t)p(y_t|\lambda_t)p(\lambda_t|y^{t-1})\,d\lambda_t. \tag{5}$$

λ_t has the same dimension p as the observation vector y_t, which typically is smaller and sometimes even much smaller than the dimension of the state vector. Thus for univariate time series and for multivariate time series of moderate dimension the curse of dimensionality may be banned even for very high dimensional state vectors.

Lemma 1. Given the normality of the prior $p(x_t|y^{t-1})$ the first and the second posterior moments $E(x_t|y^t)$ and $V(x_t|y^t)$ are linear functions of the posterior moments $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ of the linear predictor $\lambda_t = H_t \cdot x_t$:

$$E(x_t|y^t) = F_t E(x_{t-1}|y^{t-1}) + A_t(E(\lambda_t|y^t) - \hat{\lambda}_{t|t-1}), \tag{6}$$
$$V(x_t|y^t) = V(x_t|y^{t-1}) + A_t(V(\lambda_t|y^t) - \Lambda_{t|t-1})A_t^{\mathrm{T}}, \tag{7}$$

where

$$\hat{\lambda}_{t|t-1} = H_t F_t E(x_{t-1}|y^{t-1}), \tag{8}$$
$$\Lambda_{t|t-1} = H_t V(x_t|y^{t-1})H_t^{\mathrm{T}}, \tag{9}$$
$$A_t = V(x_t|y^{t-1})H_t^{\mathrm{T}}\Lambda_{t|t-1}^{-1}, \tag{10}$$
$$V(x_t|y^{t-1}) = F_t V(x_{t-1}|y^{t-1})F_t^{\mathrm{T}} + Q_t.$$

A proof of this lemma is given in the appendix. $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ are derived from (3) and (4) using Gauss-Hermite integration of (5). After introducing the necessary notations the detailed results of how this derivation is carried out in practice are summarized in Lemma 2.

Let m_t and S_t^{-1} be approximations to the mode and the Fisher information of the posterior $p(\lambda_t|y^t)$ which are obtained by second order Taylor expansion of the log of the non-normalized posterior around the expected predictor (8) (see Fahrmeir, 1992):

$$m_t = \hat{\lambda}_{t|t-1} + S_t \cdot v_t(y_t, \hat{\lambda}_{t|t-1}), \tag{11}$$
$$S_t = (\Lambda_{t|t-1}^{-1} + V_t(\hat{\lambda}_{t|t-1}))^{-1}, \tag{12}$$

where $v_t(y_t, \lambda_t)$ is the conditional score function and $V_t(\lambda)$ is the expected conditional Fisher information of the observation density:

$$v_t(y_t, \lambda_t) = \frac{\partial \log p(y_t|\lambda_t)}{\partial \lambda_t},$$
$$V_t(\lambda_t) = E(-\frac{\partial^2 \log p(y_t|\lambda_t)}{\partial \lambda_t^2}|y^{t-1}, \lambda_t).$$

We use the expected Fisher information instead of the random Fisher information to ensure positive definiteness of $V_t(\lambda_t)$. Let U_t be such that:

$$2 \cdot S_t = U_t U_t^{\mathrm{T}}, \tag{13}$$

i.e. either the Cholesky decomposition or the symmetric square root decomposition of $2 \cdot S_t$.

Let $(\tau_M^{(i)}, \omega_M^{(i)}), i = 1, \ldots, M$, be the grid points and the weights of univariate Gauss-Hermite integration of order M as tabled e.g. in Abramowitz and Stegun (1970). Let the function $d(\lambda_t)$ denote the ratio between the exact likelihood $p(y_t|\lambda_t)$ and the approximate likelihood obtained from quadratic expansion:

$$d(\lambda_t) = \frac{p(y_t|\lambda_t)}{\exp\{-\frac{1}{2}(\lambda_t - \hat{\eta}_t)^{\mathrm{T}} \cdot V_t(\hat{\lambda}_{t|t-1}) \cdot (\lambda_t - \hat{\eta}_t)\}}, \tag{14}$$

$$\hat{\eta}_t = \hat{\lambda}_{t|t-1} + V_t(\hat{\lambda}_{t|t-1})^{-1} \cdot v_t(y_t, \hat{\lambda}_{t|t-1}) \tag{15}$$

Lemma 2. $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ may be computed in terms of these quantities in the following way:

$$E(\lambda_t|y^t) = m_t + U_t \cdot z_t, \tag{16}$$

$$V(\lambda_t|y^t) = U_t \cdot Z_t \cdot U_t^{\mathrm{T}}, \tag{17}$$

$$c_t = \sum_{(i_1,\ldots,i_p)\in\mathcal{I}} \psi^{(i_1,\ldots,i_p)} \tag{18}$$

$$z_t = \frac{1}{c_t} \sum_{(i_1,\ldots,i_p)\in\mathcal{I}} u^{(i_1,\ldots,i_p)} \psi^{(i_1,\ldots,i_p)}$$

$$Z_t = \frac{1}{c_t} \sum_{(i_1,\ldots,i_p)\in\mathcal{I}} u^{(i_1,\ldots,i_p)} (u^{(i_1,\ldots,i_p)})^{\mathrm{T}} \psi^{(i_1,\ldots,i_p)} - z_t z_t^{\mathrm{T}},$$

$$\psi^{(i_1,\ldots,i_p)} = d(\lambda_t^{(i_1,\ldots,i_p)}) \omega_{M_1}^{(i_1)} \cdots \omega_{M_p}^{(i_p)}, \qquad \lambda_t^{(i_1,\ldots,i_p)} = m_t + U_t \cdot u^{(i_1,\ldots,i_p)},$$

$$u^{(i_1,\ldots,i_p)} = \begin{pmatrix} \tau_{M_1}^{(i_1)} \\ \vdots \\ \tau_{M_p}^{(i_p)} \end{pmatrix}, \qquad \mathcal{I} = \times_{i=1}^{p} \{1, \ldots, M_i\}, \qquad M_1 \geq 2, \ldots, M_p \geq 2.$$

A proof of this lemma is given in the appendix. The formulae given by Lemma 1 and Lemma 2 are a sequential scheme to compute the posterior moments $E(x_t|y^t)$ and $V(x_t|y^t)$ from $E(x_{t-1}|y^{t-1})$ and $V(x_{t-1}|y^{t-1})$. This scheme, which is called integration-based Kalman-filtering, consists of the following steps:

Integration-based Kalman-filtering

1. Extrapolation: compute $\hat{\lambda}_{t|t-1}$ and $\Lambda_{t|t-1}$ from (8) and (9);
2. Approximate posterior mode estimation: compute m_t and S_t from (11) and (12);
3. Gauss-Hermite filtering for the predictor: decompose $2 \cdot S_t$ according to (13) to obtain U_t and compute $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ from (16) and (17);

4. Filtering for the state vector: compute $E(x_t|y^t)$ and $V(x_t|y^t)$ from $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ by means of (6) and (7).

The filter splits into a linear part – filtering for x_t once the posterior moments of λ_t are available – and a non-linear part – filtering for λ_t. The second part is non-linear as for non-Gaussian time series the likelihood ratio $d(\cdot)$ in (14) and therefore $E(\lambda_t|y^t)$ is non-linear in the observation y_t. The linear steps (Step 1 and Step 4) are the same for all distribution families. Only Step 2 and Step 3 do involve assumptions on the distribution family through $p(y_t|\lambda_t)$ and the first and second derivative of its logarithm with respect to λ_t. The quantities which in fact depend on the distribution family are the likelihood ratio $d_t(\lambda_t)$ in (14), and $v_t(y_t, \hat{\lambda}_{t|t-1})$ and $V_t(\hat{\lambda}_{t|t-1})$ in (11) and (12).

For Gaussian time series with $E(y_t|\lambda_t)$ linear in λ_t and $V(y_t|\lambda_t)$ independent of λ_t and y_t, the moments obtained from integration-based Kalman-filtering are identical with those obtained by classical Kalman-filtering. The proof of this fact is a straightforward extension of Example 1 in Schnatter (1992, p.452).

For distributions from exponential families with

$$\mu_t(x_t) = E(y_t|x_t) = g^{-1}(H_t x_t), \quad V(y_t|x_t) = s(\mu_t(x_t)), \tag{19}$$

we use the explicit formula for the posterior mode filter m_t and S_t given by Fahrmeir (1992):

$$\begin{aligned} m_t &= \hat{\lambda}_{t|t-1} + K_t \cdot (y_t - g^{-1}(\hat{\lambda}_{t|t-1})), \\ S_t &= (I - K_t \cdot h_t) \cdot \Lambda_{t|t-1}, \\ K_t &= \Lambda_{t|t-1} \cdot h_t \cdot \left(h_t^{\mathrm{T}} \cdot \Lambda_{t|t-1} \cdot h_t + s(g^{-1}(\hat{\lambda}_{t|t-1})) \right)^{-1}, \\ h_t &= \frac{\partial g^{-1}}{\partial \lambda_t}(\hat{\lambda}_{t|t-1}). \end{aligned} \tag{20}$$

$g^{-1}(\cdot)$ and $s(\cdot)$ are the link and the variance function appearing in (19).

In Lemma 2 the filter is formulated as a corrector of an approximate posterior mode filter which is a low order approximation of $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ – low order in the sense that the posterior mode filter does not depend on the skewness and higher moments of the observation density $p(y_t|\lambda_t)$. Integration-based Kalman-filtering corrects this first low order approximation in reaction to the higher moments of the observation density.

The filter is approximate in two senses: it approximates the prior $p(x_t|y^{t-1})$ which results from a convolution of the normal transition density $p(x_t|x_{t-1})$ with the posterior $p(x_{t-1}|y^{t-1})$ by a normal density, and it approximates an analytical integral by a numerical one. If the posterior $p(x_{t-1}|y^{t-1})$ is close to a normal density, then the first error will be small. Experimental results reported in Schnatter (1992) and Fahrmeir (1992) indicate that with t increasing the posterior tends to be normal even in cases where the observation density is extremely non-normal. If $p(x_{t-1}|y^{t-1})$ is already normal, then we known from Lemma 2 that the first error is 0. In this case

the moments computed by integration-based Kalman-filtering can be made highly accurate by increasing M_i.

The computational burden of the linear part of integration-based Kalman-filtering is comparable with that of classical Kalman-filtering. The computational burden of the non-linear part depends on the non-normality of the the observation density $p(y_t|\lambda_t)$ and on the dimension p of the observation vector y_t. The higher the non-normality of the observation density, the higher M_i has to be. Gauss-Hermite integration of order M_i is exact if the integrand – considered as function of the i-th component of λ_t – is equal to a polynomial of order less equal $2M_i - 1$. As for Gaussian time series $d(\lambda_t)$ is a constant, a lower bound for M_i when computing up to the second moment of the posterior is given by 2. Practical experience suggests that even in cases with very skew observation densities or with observation densities with very heavy tails it is sufficient to choose M_i somewhere between 4 and 7.

With increasing dimension p of the observation vector the computional burden both of the matrix decomposition of $2 \cdot S_t$ and of the $M_1 \cdot \ldots \cdot M_p$ evaluations of the likelihood ratio $d(\lambda_t)$ increases. For univariate time series the computional burden is small: we just need to compute the square root of $2 \cdot S_t$ and to evaluate the ratio of two univariate densities M times. For bivariate time series we may use explicit formulae for Cholesky or symmetric square root decomposition and have to evaluate the ratio of two bivariate densities $M_1 \cdot M_2$, i.e. between 20 and 50, times. This seems to be easy to manage, too. We dare not say where lies the limit for multivariate time series.

3 Some Further Topics of Applied State Space Modelling

The aim of this section is to show that some further interesting results in applied state space modelling are obtained more or less directly from integration-based Kalman-filtering.

3.1 Unknown hyperparameters

In practice some or all hyperparameters θ of the dynamic generalized linear model, e.g. the process variance Q_t, may be unknwon. Data based information about θ is contained in the likelihood function $L(\theta)$ given by the product of the N successive one-step-ahead predictive values $p(y_t|\theta, y^{t-1})$:

$$L(\theta) = \prod_{t=1}^{N} p(y_t|\theta, y^{t-1}).$$

Estimates of the unknown hyperparameters may be obtained by maximizing the likelihood function numerically using algorithms discussed e.g. in Harvey (1989).

In the examples of Section 4 the Bayesian approach of Frühwirth-Schnatter (1992) is used which is a straightforward extension of the multi-process-filtering algorithm (see e.g. Harrison and Stevens, 1976) to non-normal observations. Based on a grid $\theta^{(1)}, \ldots, \theta^{(M_H)}$ and a prior $p(\theta^{(i)}|y^0)$ the posterior is updated in a dynamic manner:

$$p(\theta^{(i)}|y^t) \propto p(y_t|\theta^{(i)}, y^{t-1}) \cdot p(\theta^{(i)}|y^{t-1}), \qquad i = 1, \ldots, M_H.$$

For both the likelihood and the Bayesian approach computational problems occur with non-Gaussian dynamic models, as the predictive densities are not known analytically. However, for a *fixed* value of θ the predictive value $p(y_t|\theta, y^{t-1})$ is identical with the normalising constant $I(1)$ occuring in the filtering step at time t (see (3) and (4)). From formula (31) in the proof of Lemma 2 we find that the Gauss-Hermite integration of $I(1)$ is easily obtained from the value c_t appearing in Eq. (18) of integration-based Kalman-filtering:

$$p(y_t|\theta^{(i)}, y^{t-1}) \approx \frac{\gamma_t^{(i)} c_t^{(i)}}{(\sqrt{\pi})^p}, \tag{21}$$

$$\gamma_t^{(i)} = \frac{|S_t^{(i)}|^{\frac{1}{2}}}{|\Lambda_{t|t-1}^{(i)}|^{\frac{1}{2}}} \exp\{\frac{1}{2} v_t(y_t, \hat{\lambda}_{t|t-1}^{(i)})^{\mathrm{T}} (\Lambda_{t|t-1}^{(i)})^{-1} v_t(y_t, \hat{\lambda}_{t|t-1}^{(i)})\}. \tag{22}$$

Thus integration-based Kalman-filtering *automatically* provides an approximation to the likelihood function $L(\theta)$.

3.2 Model Discrimination

To discriminate among a group of possible dynamic models we could apply Bayesian model discrimination – see e.g. Geisser and Eddy (1979) and Frühwirth-Schnatter (1993) for a recent application of these ideas to state space models. As this procedure is not restricted to nested models we may apply it e.g. to the problem of choosing the appropriate distribution family of the observed process (see Case Study 1 and 2 in Subsection 4.1).

The discrimination procedure requires the computation of the model likelihood from the data. As the model likelihood is a weighted combination of the likelihood function $L(\theta)$, again an approximation is directly available from integration-based Kalman-filtering:

$$\begin{aligned} P(\text{Model}|y^N) &= \left(\prod_{t=1}^{N} p(y_t|y^{t-1})\right) \cdot P(\text{Model}|y^0) \approx \\ &\approx \left(\prod_{t=1}^{N} \sum_{i=1}^{M_H} \frac{\gamma_t^{(i)} c_t^{(i)}}{(\sqrt{\pi})^p}\right) \cdot P(\text{Model}|y^0). \end{aligned}$$

3.3 Filtering the grand mean within exponential families

Within exponential families often not only the state vector, but also the grand mean $\mu_t = g^{-1}(\lambda_t)$ equal to the level of the observation process is a quantity of interest. Harvey and Fernandez (1989) introduced an approximate filter for the grand mean. An alternative approximate filter for the grand mean for uni- and multivariate time series from a very general model is easily derived from integration-based Kalman-filtering.

As μ_t depends on x_t only through λ_t: $\mu_t = g^{-1}(\lambda_t)$, the first two posterior moments of $\mu_t|y^t$ are given by integrals of type (5) which again are approximated in the same way as above, e.g. for the first moment:

$$E(\mu_t|y^t) = \frac{I(g^{-1}(\lambda_t))}{I(1)} \approx \frac{1}{c_t} \sum_{(i_1,\ldots,i_p)\in\mathcal{I}} g^{-1}(\lambda^{(i_1,\ldots,i_p)})\psi^{(i_1,\ldots,i_p)},$$

where the notation is the same as in Lemma 2.

3.4 Predictive Distributions

By predictive distribution we mean the one-step ahead predictive probability $P(Y_t \leq c|y^{t-1})$. One of the possible applications of predictive distributions is model diagnostics of non-normal state-space models (Smith, 1985) which is based on the predictive residuals $u_t = P(Y_t \leq y_t|y^{t-1})$. Another application is the dynamic update of the probability that the process under investigation will exceed a critical threshold in the near future (see Section 4.1). In this subsection we will show how to extend the ideas of integration-based Kalman-filtering to obtain approximate predictive probabilities.

Under suitable conditions on the observation density $p(y_t|\lambda_t)$, the following holds:

$$P(Y_t \leq c|y^{t-1}) = \int_{-\infty}^{c} p(y|y^{t-1})\,dy, = \int_{-\infty}^{\infty} P(c|\lambda_t)p(\lambda_t|y^{t-1})\,d\lambda_t,$$

where $P(y_t|\lambda_t)$ is the distribution function of $p(y_t|\lambda_t)$. The last integral is approximated by Gauss-Hermite-integration after applying the transformation $\tilde{u}_t = \tilde{U}_t^{-1}(\lambda_t - \hat{\lambda}_{t|t-1})$ with $2\Lambda_{t|t-1} = \tilde{U}_t\tilde{U}_t^{\mathrm{T}}$. This leads to the following approximation:

$$P(Y_t \leq c|y^{t-1}) \approx \frac{1}{(\sqrt{\pi})^p} \sum_{(i_1,\ldots,i_p)\in\mathcal{I}} P(c|\tilde{\lambda}_t^{(i_1,\ldots,i_p)})\omega_{M_1}^{(i_1)}\cdots\omega_{M_p}^{(i_p)}, \tag{23}$$

with $\tilde{\lambda}_t^{(i_1,\ldots,i_p)} = \hat{\lambda}_{t|t-1} + \tilde{U}_t u^{(i_1,\ldots,i_p)}$ where $\omega_{M_j}^{(i_j)}$ und $u^{(i_1,\ldots,i_p)}$ are the same as in Lemma 2.

Note that the number $M_1,\ldots,M_p$ of grid points we use for computing the predictive probability from (23) may be different from the one used in Lemma 2, as the choice of these quantities depends on non-linearity of the integrand in λ_t.

4 Applications

In this section we will discuss various practical applications of integration-based Kalman-filtering. These applications differ mainly in the distribution family $p(y_t|\lambda_t)$ which generates the data. The examples cover heavy-tailed error distributions for observations with outliers such as weekly sulphate concentrations (Subsection 4.1), gamma distributions for observations which concentrate near zero and are highly skewed to the right such as daily sulphur dioxide emission data (Subsection 4.1), and Poisson distributions for small count data (Subsection 4.2).

4.1 Monitoring ecological processes

Sequential on-line filtering is a valuable tool for monitoring an ecological process Y_t for which regular observations y_t are available such as noxious emission or waste water. During the last two decades the classical Kalman filter has seen numerous applications to monitoring processes of this kind (see e.g. Chiu, 1978). The classical Kalman filter, however, is known to be optimal only for normally distributed processes and to be sensitive to outlying observations. As ecological processes, especially if they are daily or even hourly observations, are highly skewed and quite often exhibit a couple of outlying observations, the classical Kalman filter is not always the appropriate method of sequential estimation.

Subsequently we will apply models which explicetely assume that the observations are from a skew density or from a density with heavy tails. For practical on-line filtering we apply the method of integration-based Kalman-filtering. As mentioned earlier, integration-based Kalman filtering not only leads to estimates of the state vector, but furthermore to dynamic estimates of unknown hyperparameters, to dynamic estimates of the level $\mu_t|y^t$ of the process and to dynamic updates of the probabilty $P(Y_{t+1} > c|y^t)$ that the process will be exceeded the treshold c in the near future.

We will study two different models. The first model is designed for metric observations which – compared to the variance – concentrate close to 0 and are highly skewed to the right. For an example see the daily sulphur dioxide emission in Brotjachtriegel (FRG) over a period of three months marked by the dots in Figure 1. For such data we assume that y_t follows a gamma distribution with

$$E(y_t|\lambda_t) = \mu_t(\lambda_t) = g^{-1}(\lambda_t)$$
$$V(y_t|\lambda_t) = \mu_t(\lambda_t)^2 \cdot \phi.$$

$\alpha = 1/\phi$ is the shape parameter of the gamma distribution. As a link function we will not use the natural link which may lead to negative value of μ_t, but either the log-link

$$g^{-1}(\lambda_t) = \exp(\lambda_t) \tag{24}$$

or the mixed link

$$g^{-1}(\lambda_t) = \begin{cases} \lambda_t, & \lambda_t \geq 1, \\ \exp(\lambda_t - 1), & \lambda_t \leq 1. \end{cases} \tag{25}$$

For a previous discussion of this model see Frühwirth-Schnatter (1991). The transition equation of the space model depends on the effects that are present. It may range from a simple local level model to a more complex model such as the basic structural model or a model with explanatory variables (for a full account of these models see e.g. Harvey, 1989).

The transition model together with the link function and the gamma observation density completely define the model. For sequential filtering we apply integration-based Kalman-filtering based on the approximate posterior mode filter (20). The unknown scaling parameter ϕ may be regarded as a hyperparameter which is estimated from the data. For an application of this model to monitoring daily sulphur dioxide emissions see Case Study 1 below.

The second model is designed for metric observations which have a fairly symmetric distribution and exhibit some outlying observations. Such time series most often result from avering observations of high frequency, e.g. hourly observations, to obtain observations of a much lower frequency, e.g. weekly or monthly means. For an example see the weekly means of sulphate concentration in Turkey Lakes Watershed in Ontario (Canada) over a period of two years marked by the dots in Figure 4. Filtering for such data is known under the heading of robust Kalman filtering. Following Masreliez and Martin (1977) and West (1981) we assume that y_t follows a heavy tailed distributions:

$$y_t = \lambda_t + \varepsilon_t, \quad \varepsilon_t \sim p(\varepsilon_t; \sigma_R^2),$$

with $p(\varepsilon_t; \sigma_R^2)$ from a heavy-tailed error distribution, e.g. the Student-t-family:

$$p(\varepsilon_t; \sigma_R^2) = \frac{\nu^{\frac{\nu}{2}} \Gamma(\frac{\nu+1}{2})}{\pi^{\frac{1}{2}} \Gamma(\frac{\nu}{2})} \left(\nu + \varepsilon_t^{\mathrm{T}} (\sigma_R^2)^{-1} \varepsilon_t\right)^{-\frac{\nu+1}{2}}.$$

The choice of a heavy-tailed error distribution leads to a filter which is less sensitive to outlying observations than the classical Kalman filter. The filter will not be robust to outliers in the state process, as the transition density still is Gaussian (for a filter which is also robust to outliers in the state process see Meinhold and Singpurwalla, 1989). As above, the transition equation of the space model depends on the effects that are present. The transition model together with the heavy-tailed error distribution completely defines the model.

For sequential filtering we do not apply the approximate filter of West (1981) but integration-based Kalman-filtering based on the approximate posterior mode filter (11) and (12). For the Student-t-family the quantities $v_t(\cdot)$ and $V_t(\cdot)$ may be approximated by:

$$v_t(y_t, \lambda_t) \approx \frac{(\nu+1)(y_t - \lambda_t)}{\sigma_R^2 \nu + (y_t - \lambda_t)^2},$$

$$V_t(\lambda_t) \approx \frac{\nu(\nu+1)(\nu^2-\nu-2)}{\sigma_R^2 \cdot (\nu^2+\nu+2)^2}.$$

The parameter ν which influences the heaviness of the tails has to be prespecified by the user. The parameter σ_R^2 is regarded as a hyperparameter and estimated from the data. For an application of this model to monitoring weekly sulphate concentration in a lake see Case Study 2 below.

Case Study 1. All plots in Figure 1 show the daily sulphur dioxide emission in Brotjachtriegel (FRG) over a period of three months.

Figure 1 about here

As the distribution of the data is very skew with extreme observations on the right, the gamma distribution seems to be an appropriate candidate as observation distribution. The transition model is assumed to be the simple local level model:

$$x_t = x_{t-1} + w_t, \quad w_t \sim N(0, \sigma_\eta^2),$$
$$\lambda_t = x_t.$$

The link function will be either the log-link (24) or the mixed link (25). For illustration we compare these two models which differ in the link function with two models where the observation distribution is assumed to be either normal (classical Kalman-filtering) or a Student-t-distribution with $\nu = 4$.

Table 1 about here

For estimation we applied integration-based Kalman-filtering with $M = 7$ and the prior given in Table 1.

For both gamma filters the hyperparameters are equal to $\boldsymbol{\theta} = (\sigma_\eta^2, \phi)$, for the Kalman and the Student-t filter the hyperparameters are equal to $\boldsymbol{\theta} = (\sigma_\eta^2, \sigma_R^2)$. The hyperparameters were estimated on the grid $(\theta^{(i_1)}, \theta^{(i_2)})$ displayed in Table 1 which furthermore shows the posterior expectation of the hyperparameters given the whole data set.

The four plots in Figure 1 compare the data with the filtered mean $\mu_t|y^t$ for all t for the various models. The Kalman filter is very sensitive to the extreme observations. The Student-t filter, though it has heavy tails, is sensitive, too, because of the skewness of the data. The two gamma filters are less sensitive to observations which under the normal assumption seems to be outliers but under the assumption of a skew density turn out to be quite probable. The log-link is slightly more sensitive than is the mixed link.

To compare the four models we computed the model likelihoods and the posterior probabilities of the models based on uniformly distributed prior weights (see Table 1). There is a high support for the models based on the gamma-distributions with a marked preference for the model with the mixed link.

Figure 2 about here

Figure 2 shows that for both gamma models the empirical distribution function of the predictive residuals u_t (see Subsection 3.4) is very similar to the uniform distribution, as it should be under the assumption of an appropriate model, whereas for the other two models it is not.

Figure 3 about here

Finally, Figure 3 shows, for the preferred gamma model with mixed link, the probability that the process will exceed 30 μg on the next day as a function of time. Such dynamic probabilities are a very helpful tool for monitoring noxious emissions or waste water plants.

Case Study 2. All plots in Figure 4 show the weekly sulphate concentration in Turkey Lakes Watershed in Ontario (Canada) over a period of two years.

Figure 4 about here

There is a single outlying observation in the 31st week and a group of three consecutive outlying observations starting in the 36st week. As – apart from these outlying observations – the data are not markedly skew the Student-t-distribution seems to be an appropriate candidate as observation distribution. Again, we take the simple local level model as transition model. Again, for illustration we will compare this model with the other models used in Case Study 1.

For estimation we applied integration-based Kalman-filtering with $M = 7$ and the prior given in Table 2. The hyperparameters were estimated on the grid $(\theta^{(i_1)}, \theta^{(i_2)})$ displayed in Table 2 which furthermore shows the posterior expectation of the hyperparameters given the whole data set.

Table 2 about here

The four plots in Figure 4 compare the data with the filtered mean $\mu|y^t$ for all t and the various models. The Kalman filter as well as the gamma filters are sensitive both to the single as well as to the group of outlying observations. The Student-t filter is robust to the single outlier, but reacts to the third outlier within the group and indicates too high a shift in the state vector. As mentioned earlier, the Student-t filter is not robust to outliers in the state process. To compare the four models we computed the model likelihoods and the posterior probabilities of the models based on equally distributed prior weights (see Table 2). There is an extremely high preference for the model based on the Student-t-distribution.

Figure 5 about here

Figure 5, however, shows that even for this preferred model, the empirical distribution function of the predictive residuals u_t is far from perfect, though it is better than for the other candidates.

4.2 Filtering for Small Count Data

Filtering for small count data is discussed in great detail in Harvey and Fernandes (1989). For such data the observation density is assumed to be a Poisson distribution with time-varying mean. Analysis of such models through integration-based Kalman-filtering follows much along the same lines as in the previous subsection. For illustration we include an application of integration-based Kalman-filtering to measuring the effect of the seat belt law of January 1983 on deaths in road accidents in Great Britain (Harvey and Durbin, 1986).

Case Study 3. Harvey and Fernandes (1989) as well as Durbin and Koopman (1992) used the number of monthly deaths of drivers of light goods vehicles to illustrate their methods of sequential filtering for non-Gaussian time series. As this data are fairly small counts which exhibit seasonal variation they assume that the observed numbers of deaths follow a Poisson distribution with time-varying mean μ_t. We reanalyze the data with a similar model as Harvey and Fernandes (1989). We extend a basic structural model with a local level a_t, form-free seasons, and no slope, by introducing the interaction parameter δ into the state vector:

$$
\begin{aligned}
a_t &= a_{t-1} + \delta_t + w_{t,1}, \qquad w_{t,1} \sim N(0, \sigma_\eta^2), \\
s_t &= -\sum_{j=1}^{10} s_{t-j} + w_{t,3}, \qquad w_{t,3} \sim N(0, \sigma_\omega^2), \\
\delta_t &= \delta_{t-1}, \\
\lambda_t &= a_t + s_t + \delta_t.
\end{aligned}
$$

The linear predictor λ_t is related to the mean μ_t through the natural link $\lambda_t = \ln \mu_t$. The state space formulation of this model leads to a state vector of dimension 13. The state vector and the unknown hyperparameters σ_η^2 and σ_ω^2 are estimated by the methodology developped in this paper.

The posterior expectations of σ_η^2 and σ_ω^2 are given by $\hat{\sigma}_\eta^2 = 0.00118$ and $\hat{\sigma}_\omega^2 = 0.0000222$. These estimates are based on a multi process filter with 64 different combinations of $(\sigma_\eta^2, \sigma_\omega^2)$ which are obtained by combining the following values of the variances:

$\sigma_\eta^2 \cdot 10^3$:	0.5	1.	1.5	2.	2.5	3.	3.5	4.
$\sigma_\omega^2 \cdot 10^5$:	0.5	1.	1.5	2.	2.5	3.	3.5	4.

To ensure high accuracy of Gauss-Hermite integration for the linear predictor we take $M = 7$. The prior of the state vector is $N(2.5, 1)$ for $a_1|y^0$ and $N(0, 1)$ for the remaining components of the state vector. From integration-based Kalman filtering we find that the posterior expectation and standard deviation of the interaction parameter $\delta|y^N$ are given by $E(\delta|y^N) = -0.2604$ and $V(\delta|y^N) = 0.02778$. As our method of estimating the hyperparameter is Bayesian, the reported variance of δ in contrast to the likelihood approach of Harvey and Fernandes (1989) and Durpin and

Koopman (1992) does take into account the uncertainty due to the unknown hyper-parameter. Although the variance is bigger than the variances reported previously, from Figure 6, which shows the posterior of the proportional decrease $d = \exp(\delta)$ of the mean due to seat belt legislation, we find that there is still a high support for the hypothesis $d < 1$.

Figure 6 about here

Acknowledgements

I want to thank Dr. Paul Michels, Universität Konstanz (FRG) for providing me with the data used in Case Study 1 and Prof. Rainer Schlittgen, Universitt Hamburg (FRG) for providing me with the data used in Case Study 2.

Appendix

Proof of Lemma 1. First we extend the state vector by the linear predictor λ_t. This new state vector will be denoted by β_t. Then we express the posterior moments of x_t as integrals of the appropriate functions $h(x_t)$ of x_t over the non-normalized posterior of the extended state vector β_t:

$$J(h(x_t)) = \int h(x_t)p(y_t|\beta_t)p(\beta_t|y^{t-1})\,d\beta_t. \tag{26}$$

Now the basic point is that for the extended state vector β_t the likelihood $p(y_t|\beta_t)$ depends on x_t only through λ_t, therefore $p(y_t|\beta_t) = p(y_t|\lambda_t)$. Thus the integral in (26) may be written as:

$$J(h(x_t)) = \int \left[\int h(x_t)p(x_t|\lambda_t, y^{t-1})\,dx_t\right] p(y_t|\lambda_t)p(\lambda_t|y^{t-1})\,d\lambda_t. \tag{27}$$

Obviously, under the assumption of a normal prior $p(x_t|y^{t-1})$ the prior of the extended state vector β_t and therefore the conditional prior $p(x_t|\lambda_t, y^{t-1})$ are normal, too. It is easy to verify that the moments of the later are given by:

$$\begin{aligned} E(x_t|\lambda_t, y^t) &= E(x_t|y^{t-1}) + A_t(\lambda_t - \hat{\lambda}_{t|t-1}), \\ V(x_t|\lambda_t, y^t) &= V(x_t|y^{t-1}) - A_t\Lambda_{t|t-1}A_t^{\mathrm{T}}, \end{aligned}$$

with $\hat{\lambda}_{t|t-1}$, $\Lambda_{t|t-1}$ and A_t same as in (8), (9) and (10). As $p(x_t|\lambda_t, y^{t-1})$ is normal, the inner integral in (27) may be carried out analytically. Define

$$I(f(\lambda_t)) = \int f(\lambda_t)p(y_t|\lambda_t)p(\lambda_t|y^{t-1})\,d\lambda_t.$$

Then it is obvious that the following holds:

$$J(h(x_t)) = I(E(h(x_t)|\lambda_t, y^{t-1}))$$

Therefore:

$$
\begin{aligned}
J(1) &= I(1) \\
E(x_t|y^t) &= \frac{J(x_t)}{J(1)} = \frac{I(E(x_t|\lambda_t, y^{t-1}))}{I(1)} = \frac{I(E(x_t|y^{t-1}) + A_t(\lambda_t - \hat{\lambda_{t|t-1}}))}{I(1)} = \\
&= E(x_t|y^{t-1}) + A_t(E(\lambda_t|y^t) - \hat{\lambda_{t|t-1}}) \qquad (28) \\
V(x_t|y^t) &= \frac{J((x_t - E(x_t|y^t))(x_t - E(x_t|y^t))^{\mathrm{T}})}{J(1)} = \\
&= \frac{I(E[(x_t - E(x_t|\lambda_t, y^{t-1}))(x_t - E(x_t|\lambda_t, y^{t-1}))^{\mathrm{T}}|\lambda_t, y^{t-1}]}{I(1)} - \\
&\quad - \frac{I((E(x_t|\lambda_t, y^{t-1}) - E(x_t|y^t))(E(x_t|\lambda_t, y^{t-1}) - E(x_t|y^t))^{\mathrm{T}})}{I(1)} = \\
&= V(x_t|\lambda_t, y^{t-1}) - \frac{I(A_t(\lambda_t - \hat{\lambda}_{t|t-1})(\lambda_t - \hat{\lambda}_{t|t-1})^{\mathrm{T}} A_t^{\mathrm{T}})}{I(1)} = \\
&= V(x_t|y^{t-1}) + A_t(V(\lambda_t|y^t) - \Lambda_{t|t-1})A_t^{\mathrm{T}} \qquad (29)
\end{aligned}
$$

These findings together with

$$E(x_t|y^{t-1}) = F_t E(x_{t-1}|y^{t-1}), \qquad V(x_t|y^{t-1}) = F_t V(x_{t-1}|y^{t-1})F_t^{\mathrm{T}} + Q_t$$

prove Lemma 1.

Proof of Lemma 2. First we rewrite integral (5) in the following way:

$$
\begin{aligned}
I(h(\lambda_t)) &= \int h(\lambda_t) d(\lambda_t) \gamma_t(\lambda_t) p_N(\lambda_t; m_t, S_t)\, d\lambda, \qquad (30) \\
\gamma_t(\lambda_t) &= \frac{\exp\{-\frac{1}{2}(\lambda_t - \hat{\eta}_t)^{\mathrm{T}} \cdot V_t(\hat{\lambda}_{t|t-1}) \cdot (\lambda_t - \hat{\eta}_t)\} p_N(\lambda_t; \hat{\lambda}_{t|t-1}, \Lambda_{t|t-1})}{p_N(\lambda_t; m_t, S_t)},
\end{aligned}
$$

with $m_t, S_t, d(\cdot)$, and $\hat{\eta}_t$ same as in (11) to (15). By completing the square in λ_t it is straightforward to verify that γ_t is independent of λ_t and equal to the following constant:

$$\gamma_t = \frac{|S_t|^{\frac{1}{2}}}{|\Lambda_{t|t-1}|^{\frac{1}{2}}} \exp\{\frac{1}{2} v_t(y_t, \hat{\lambda}_{t|t-1})^{\mathrm{T}} \Lambda_{t|t-1}^{-1} v_t(y_t, \hat{\lambda}_{t|t-1})\}.$$

Gauss-Hermite integration of (30) is carried out after applying the transformation

$$u_t = U_t^{-1}(\lambda_t - m_t), \qquad 2 \cdot S_t = U_t U_t^{\mathrm{T}},$$

which leads to a new scale where the components of the transformed variable are uncorrelated. Therefore it makes sense to use the cartesian product of univariate

Gauss-Hermite grid points on this new scale. Transforming back to the original scale the grid points for λ_t are non-cartesian and given by:

$$\lambda_t^{(i_1,\ldots,i_p)} = m_t + U_t u^{(i_1,\ldots,i_p)}, \qquad u^{(i_1,\ldots,i_p)} = \begin{pmatrix} \tau_{M_1}^{(i_1)} \\ \vdots \\ \tau_{M_p}^{(i_p)} \end{pmatrix}.$$

It is straighforward to verify that the numerical approximation to (30) is given by:

$$I(h(\lambda_t)) \approx \frac{\gamma_t}{(\sqrt{\pi})^p} \sum_{(i_1,\ldots,i_p)\in\mathcal{I}} h(\lambda_t^{(i_1,\ldots,i_p)}) d(\lambda_t^{(i_1,\ldots,i_p)}) \omega_{M_1}^{(i_1)} \cdots \omega_{M_p}^{(i_p)} \tag{31}$$

Finally, the representation of $E(\lambda_t|y^t)$ and $V(\lambda_t|y^t)$ given in Lemma 2 is obtained from (3), (4) and (31).

References

Abramowitz, M. and Stegun, I. (1970) *Handbook of Mathematical Functions.* New York.

Carlin, B., Polson, N. and Stoffer, D. (1992) A Monte Carlo Approach to Nonnormal and Nonlinear State-Space Modeling. *Journal of the American Statistical Association* , **87**, 493-500.

Chiu, Ch. (1978) *Applications of Kalman-Filter to Hydrology, Hydraulics and Water Recources.* Proceedings of AGU Chapman Conference. University of Pittsburgh.

Durbin, J. and Koopman, S.J. (1992) *Kalman Filtering and Smoothing for Non-Gaussian Time Series.* Preprint.

Fahrmeir, L. (1992) Posterior Mode Estimation by Extended Kalman Filtering for Multivariate dynamic Generalized Linear Models. *Journal of the American Statistical Association* , **87**, 501-509.

Fahrmeir, L., Hennevogel, W. and Klemme, K.: Smoothing in Dynamic Generalized Linear Models by Gibbs Sampling. In: Fahrmeir, L., Francis, B., Gilchrist, R. and Tutz, G. (Eds), Advances in GLIM and Statistical Modelling. *Lecture Notes in Statistics*, **78**, 85-90.

Fahrmeir, L. and Kaufmann, H. (1991) On Kalman-Filtering, Posterior Mode Estimation and Fisher-Scoring in Dynamic Exponential Family Regression. *Metrika*, **38**, 37-60.

Frühwirth-Schnatter, S. (1991) Monitoring von ökologischen und biometrischen Prozessen mit statistischen Filtern. In: Seeber, G.U.H. and Minder, CH.E. (Eds.), Multivariate Modelle. Neue Ansätze für biometrische Anwendungen, 89 -122. Springer.

Frühwirth-Schnatter, S. (1992) Approximate Predictive Integrals for Dynamic Generalized Linear Models. In: Fahrmeir, L., Francis, B., Gilchrist, R. and Tutz, G. (Eds), Advances in GLIM and Statistical Modelling. *Lecture Notes in Statistics*, **78**, 101-106.

Frühwirth-Schnatter,S. (1993) Bayesian Model Discrimination and Bayes Factors for State Space Models. Preprint submitted for publication.

Geisser, S. and Eddy, W.F. (1979) A Predictive Approach to Model Selection. *Journal of the American Statistical Association* , **74**, 153-160.

Harrison,P.J. and Stevens,C.F. (1976) Bayesian Forecasting (with discussion). *Jornal of the Royal Statistical Society B, 38*, 205-247.

Harvey,A. (1989) *Forecasting, Structural Time Series Models, and the Kalman Filter.* Cambridge: Cambridge University Press.

Harvey, A. and Durbin, J. (1986) The effect of seat belt legislation on British road casualities: a case study in structural time series modelling (with discussion). *Jornal of the Royal Statistical Society A* , **149**, 187-227.

Harvey, A. and Fernandes, C. (1989) Time Series models for count or qualitative observations (with discussion). *Journal of Business and Economic Statistics*, **7**, 407-422.

Kitagawa, G. (1987) Non-Gaussian State Space Modelling of Nonstationary Time Series (with comments). *Journal of the American Statistical Association*, **82**, 1032-1063.

Naylor,J.C. and Smith, A.F.M. (1982) Application of a Method for the Efficient Computation of Posterior Distributions. *Applied Statistics*, **31**, 214-225.

Masreliez, C.J. (1975) Approximate non-Gaussian filtering with linear state and observation relations. *IEEE Transaction on Automatic Control*, **AC-20**, 107-110.

Masreliez, C.J. and Martin, R.D. (1977) Robust Bayesian Estimation for the linear model and robustifying the Kalman Filter. *IEEE Transaction of Automatic Control*, **AC-22**, 361-371.

Meinhold, R. and Singpurwalla, N. (1989) Robustification of Kalman Filter Models. *Journal of the American Statistical Association*, **84**, 479-486.

Schnatter, S. (1992) Integration-based Kalman-filtering for a dynamic generalized linear trend model. *Computational Statistics & Data Analysis*, **13**, 447-459.

Smith, J.Q. (1985) *Diagnostic Check of Non-standard Time Series Models.* Research Report **61**, Department of Statistics, University of Warwick.

Smith, A.F.M., Skene, A.M. J.E.H. Shaw, J.C. Naylor and Dransfield, M. (1985) The Implementation of the Bayesian Paradigma. *Communications in Statistics – Theory and Methods*, **14**, 1079-1102.

West, M. (1981) Robust Sequential Approximate Bayesian Estimation. *Journal of the Royal Statistical Society B*, **43**, 157-166.

West, M. and Harrison ,P.J. (1989) *Bayesian Forecasting and Dynamic Models.* New York/Heidelberg/Berlin: Springer.

West, M., Harrison, P.J. and Migon, H. (1985) Dynamic Generalized Linear Models and Bayesian Forecasting Models. *Journal of the American Statistical Association*, **80**, 741-750.

Tables

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$E(x_0\|y^0)$	0	0	1	1
$V(x_0\|y^0)$	100	100	100	1
$\theta_1^{(i_1)}$, $i_1 = 1,\ldots,8$	1 2 5 10	.05 .1 .5 1.	.1 .5 1 2	.01 .02 .04 .05
	20 30 40 50	2.5 5. 7.5 10.	3 4 5 10	.06 .07 .08 .1
$\theta_2^{(i_2)}$, $i_2 = 1,\ldots,8$	1 2 5 10	1 2 5 10	.5 .6 .7 .8	.5 .6 .7 .8
	20 30 40 50	20 30 40 50	.9 .95 .99 1.	.9 .95 .99 1.
$E(\theta_1^{(i)}\|y^N)$	41.1	6.86	3.52	0.0391
$E(\theta_2^{(i)}\|y^N)$	38.2	38.9	.945	.935
$P(\mathcal{M}_i\|y^N)$	4E-21	1.7E-20	.729	.271

Table 1: Various assumptions and results for integration-based Kalman-filtering for the different models of Case Study 1 ($\mathcal{M}_1$: normal, $\mathcal{M}_2$: Student-t, $\mathcal{M}_3$: gamma with mixed link, $\mathcal{M}_4$: gamma with log-link)

	$\mathcal{M}_1$	$\mathcal{M}_2$	$\mathcal{M}_3$	$\mathcal{M}_4$
$E(x_0\|y^0)$	65	65	65	4
$V(x_0\|y^0)$	100	100	100	10
$\theta_1^{(i_1)}$, $i_1 = 1,\ldots,8$	1 3 5 10 20 30 40 50	1 3 5 10 20 30 40 50	1 3 5 10 20 30 40 50	.001 .005 .01 .025 .05 .075 .1 .2
$\theta_2^{(i_2)}$, $i_2 = 1,\ldots,8$	5 10 20 40 50 60 80 100	1 2 5 10 20 30 40 50	.001 .005 .01 .015 .02 .025 .05 .1	.001 .005 .01 .025 .05 .075 .1 .2
$E(\theta_1^{(i)}\|y^N)$	25.6	6.46	15.9	0.085
$E(\theta_2^{(i)}\|y^N)$	42.5	9.9	0.01	0.00576
$P(\mathcal{M}_i\|y^N)$	3.2E-14	0.999999	6.8E-9	3.7 E-7

Table 2: Various assumptions and results for integration-based Kalman-filtering for the different models of Case Study 2 ($\mathcal{M}_1$: normal, $\mathcal{M}_2$: Student-t, $\mathcal{M}_3$: gamma with mixed link, $\mathcal{M}_4$: gamma with log-link)

Figures (Originals enclosed)

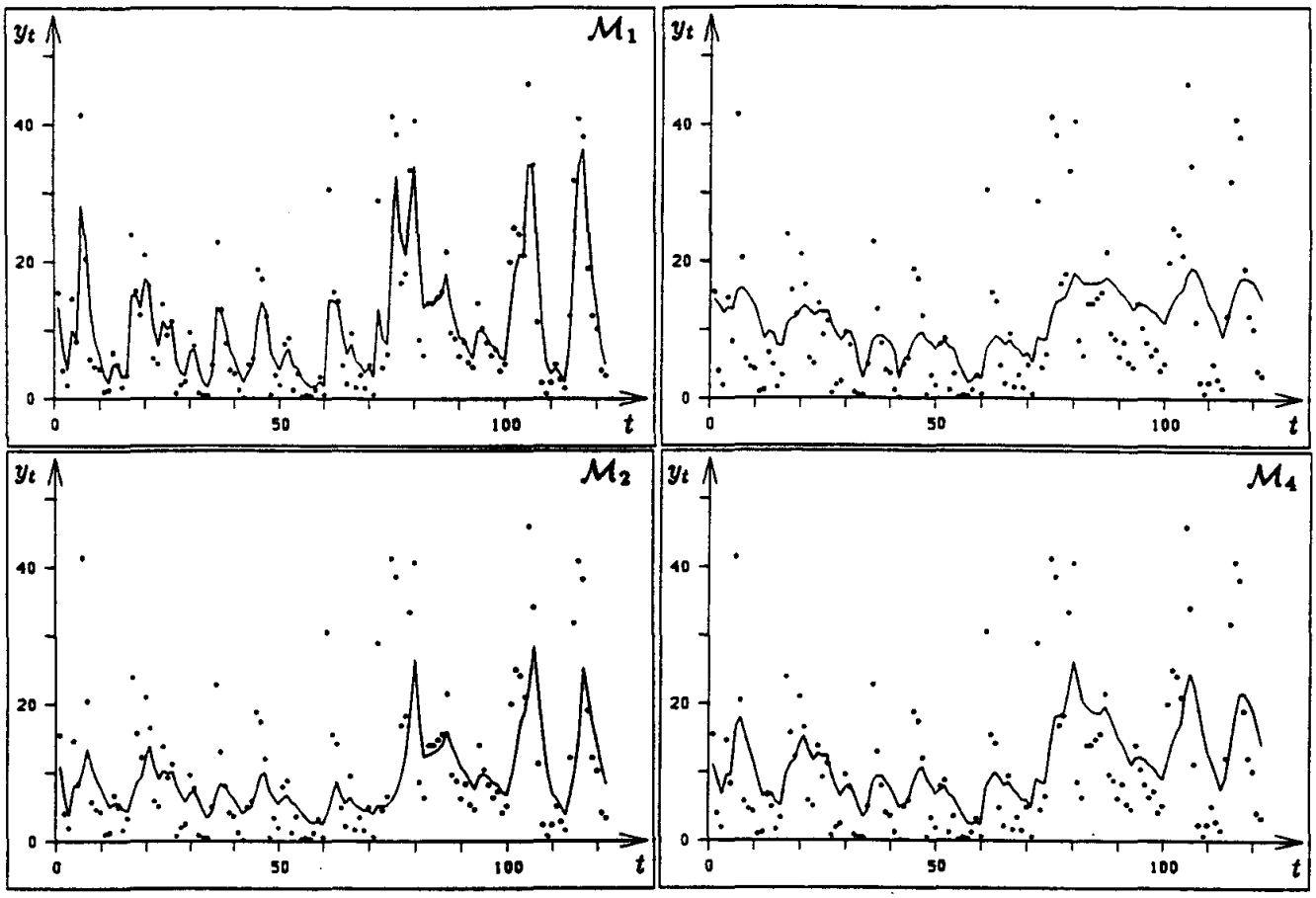


Figure 1: Daily sulphur dioxide emission in Brotjachtriegel (FRG) over a period of three months (marked by the dots) compared with the filtered mean $\mu_t|y^t$ (full line) for the various distribution families ($\mathcal{M}_1$: normal, $\mathcal{M}_2$: Student-t, $\mathcal{M}_3$: gamma with mixed link, $\mathcal{M}_4$: gamma with log-link)

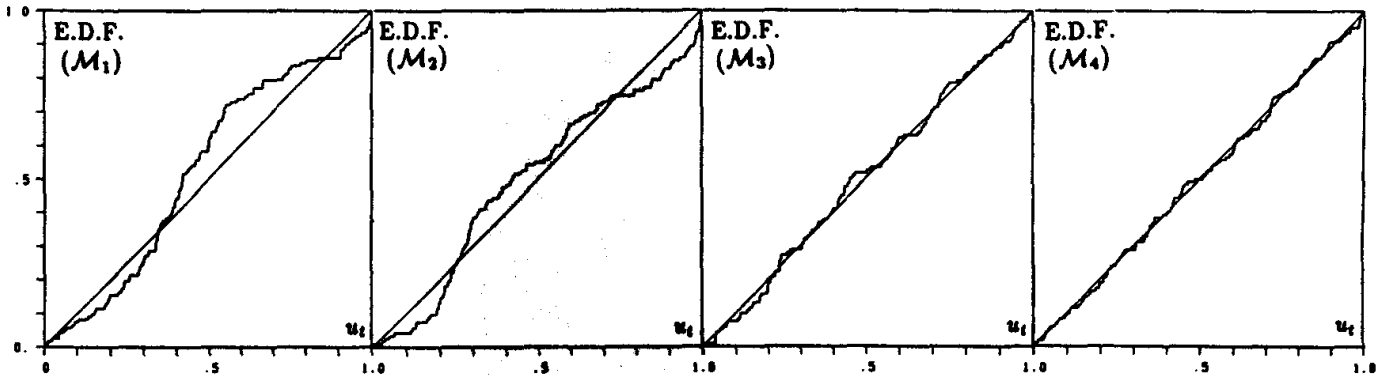


Figure 2: Case Study 1 – empirical distribution function of the predictive residuals u_t for the various distribution families ($\mathcal{M}_1$: normal, $\mathcal{M}_2$: Student-t, $\mathcal{M}_3$: gamma with mixed link, $\mathcal{M}_4$: gamma with log-link)

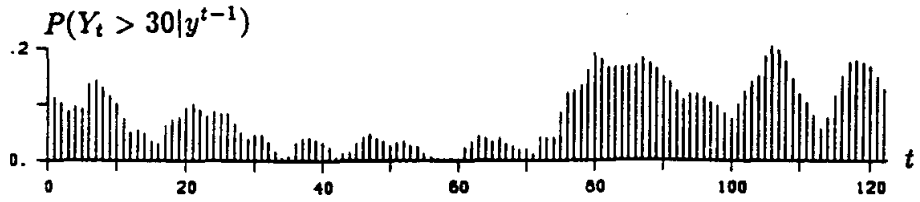


Figure 3: Case Study 1 – dynamic probabilities $P(Y_t > 30|y^{t-1})$

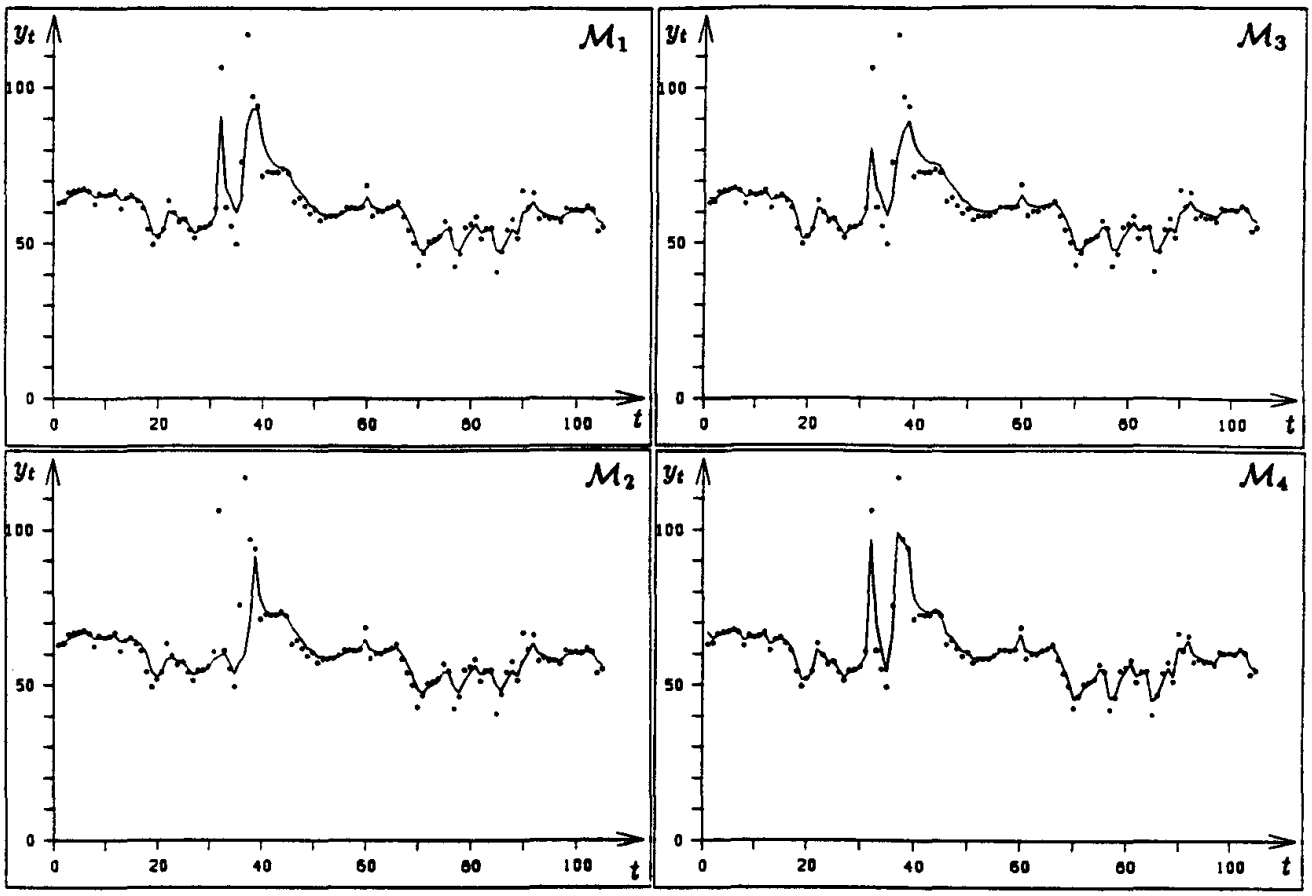


Figure 4: Weekly means of sulphate concentration in Turkey Lakes Watershed in Ontario (Canada) over a period of two years (marked by the dots) compared with the filtered mean $\mu_t|y^t$ (full line) for the various distribution families ($\mathcal{M}_1$: normal, $\mathcal{M}_2$: Student-t, $\mathcal{M}_3$: gamma with mixed link, $\mathcal{M}_4$: gamma with log-link)

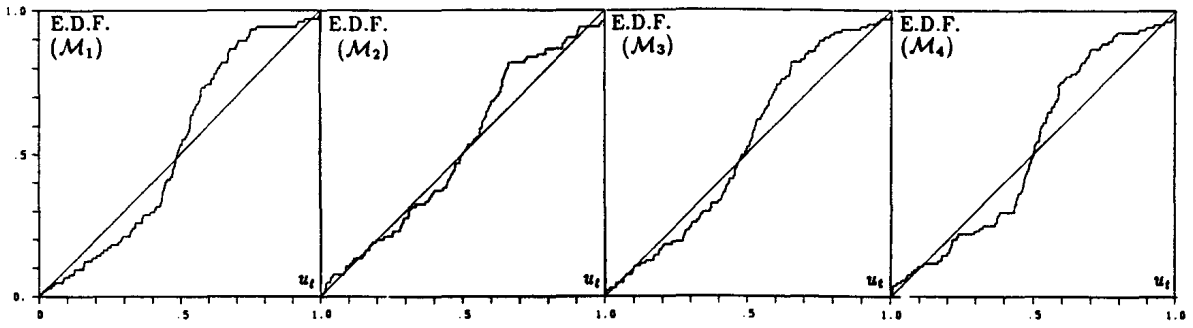


Figure 5: Case Study 2 – empirical distribution function of the predictive residuals u_t for the various distribution families ($\mathcal{M}_1$: normal, $\mathcal{M}_2$: Student-t, $\mathcal{M}_3$: gamma with mixed link, $\mathcal{M}_4$: gamma with log-link)

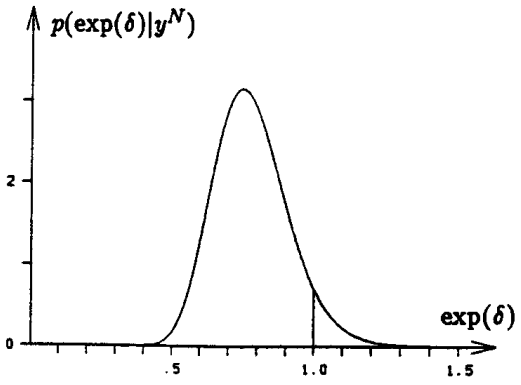


Figure 6: Effect of seat belt legislation on deaths of drivers of light good vehicles posterior density of the proportional decrease $d = \exp(\delta)$ of the mean.