Integration-based Kalman-filtering for a Dynamic Generalized Linear Trend Model

Schnatter, Sylvia

Integration-based Kalman-filtering for a Dynamic Generalized Linear Trend Model

Sylvia Schnatter

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 9
1991

Integration-based Kalman-filtering for a Dynamic Generalized Linear Trend Model

Sylvia SCHNATTER
Department of Statistics,
Vienna University of Economics and Business Administration, Austria

Keywords: Approximate Bayesian Inference; Bayesian Computation; Dynamic Generalized Linear Models; Gauss-Hermite Integration; Kalman-Filtering for Non-Gaussian Data; Non-Gaussian State Space Models; Trend Modelling.

Abstract

The topic of the paper is filtering for non-Gaussian dynamic (state space) models by approximate computation of posterior moments using numerical integration. A Gauss-Hermite procedure is implemented based on the approximate posterior mode estimator and curvature recently proposed in [2].

This integration-based filtering method will be illustrated by a dynamic trend model for non-Gaussian time series. Comparision of the proposed method with other approximations ([15], [2]) is carried out by simulation experiments for time series from Poisson, exponential and Gamma distributions.

1 Introduction

The paper deals with the subject of filtering for non-Gaussian dynamic (state space) models. For illustration, we concentrate on a special model, namely an extension of the dynamic linear trend model ([6]) to non-Gaussian time series. This model is described in Section 2.

Complete inference for non-Gaussian dynamic models requires sequential procedures for determining the posterior densities of the state variable from the observed time series $y_1, \ldots, y_N$. There are no exact procedures for non-Gaussian dynamic models and approximations are inevitable. In Section 3 we suggest approximate sequential computation of posterior moments by numerical integration. A Gauss-Hermite procedure is implemented based on the approximate posterior mode estimator and curvature recently proposed in [2].

Comparision of this method with other approximations ([15], [2]) is carried out by simulation experiments for time series from Poisson, exponential and Gamma distributions. Comparative results are summarized in Section 4.

2 The Dynamic Generalized Linear Trend Model

For Gaussian time series $\{y_1, \ldots, y_N\}$ [6] introduced a dynamic linear trend model with the state variable $\boldsymbol{\beta}_t$ defined in the following way: the first element $\boldsymbol{\beta}_{t,1}$ of the state variable is equal to the level of the process, whereas the second element $\boldsymbol{\beta}_{t,2}$ is a parameter measuring systematic increase or decrease per time unit (trend component):

$$\boldsymbol{\beta}_t = \boldsymbol{F} \cdot \boldsymbol{\beta}_{t-1} + \boldsymbol{w}_t, \qquad \boldsymbol{F} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \tag{1}$$

$$\boldsymbol{w}_t \sim \mathrm{N}\left(\boldsymbol{0}, \boldsymbol{Q}(\theta_1, \theta_2)\right), \qquad \boldsymbol{Q}(\theta_1, \theta_2) = \begin{pmatrix} \theta_1 + \theta_2 & \theta_2 \\ \theta_2 & \theta_2 \end{pmatrix}. \tag{2}$$

$\boldsymbol{\beta}_0$ is a Gaussian random variable. The state variable $\boldsymbol{\beta}_t$ is related to the mean μ_t of the Gaussian distribution of the observation y_t by:

$$\mu_t = \boldsymbol{H} \cdot \boldsymbol{\beta}_t, \qquad \boldsymbol{H} = \begin{pmatrix} 1 & 0 \end{pmatrix}. \tag{3}$$

What if the observed time series is non-Gaussian? Non-Gaussian time series often occur in ecological applications e.g. when observing small count data such as daily rainfall occurence or very small, non-negative quantities such as hourly sulphur dioxide concentrations. The following extension of the dynamic linear trend model to a generalized one follows directly along the lines of [15].

We assume that the stochastic behaviour of the observed time series y_t may be described by a distribution from the exponential family ([9]):

$$y_t \sim \exp\left\{\left(y_t \eta_t - b(\eta_t)\right)/a(\phi) + c(y_t, \phi)\right\}, \tag{4}$$

with canonical parameter η_t, scaling parameter ϕ and

$$E(y_t) = b'(\eta_t) = \mu_t, \qquad V(y_t) = b''(\eta_t) \cdot a(\phi) = v(\mu_t) \cdot a(\phi). \tag{5}$$

Extending the Gaussian to the non-Gaussian case, the relationship between the mean μ_t and the state variable $\boldsymbol{\beta}_t$ is modelled by

$$g(\mu_t) = \boldsymbol{H} \cdot \boldsymbol{\beta}_t, \qquad \boldsymbol{H} = \begin{pmatrix} 1 & 0 \end{pmatrix}, \tag{6}$$

with a monotone, differentiable transformation $g(\cdot)$. (6) in combination with (2) defines the dynamic generalized linear trend model. Note that the model is a state space model with a Gaussian transition density $p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})$ and a non-Gaussian observation density $p(y_t|\boldsymbol{\beta}_t)$.

3 Kalman-Filtering for Non-Gaussian Data

3.1 Introduction

An entire solution of the problem of statistical inference requires the determination of the posterior density $p(\boldsymbol{\beta}_t|y^t)$ of the state variable $\boldsymbol{\beta}_t$ given observations up to t for each $t = 1, \ldots, N$. For a Gaussian observation density this density is Gaussian, too, and thus characterized by the first and second moment. These moments are easily determined from the moments of the density $p(\boldsymbol{\beta}_{t-1}|y^{t-1})$ by a sequential scheme (Kalman-filtering).

For the non-Gaussian case the posterior densities are not of well-known type and can only be approximated. Global approximations ([8]) still envolve time consuming computations. Therefore practical inference is reduced to approximate some characteristics of the posterior density $p(\boldsymbol{\beta}_t|y^t)$ given approximations of the same characteristics of the density $p(\boldsymbol{\beta}_{t-1}|y^{t-1})$. Such a sequential scheme may be called Kalman-filtering for non-Gaussian data.

Conjugate Bayesian analysis is not possible for the state variable $\boldsymbol{\beta}_t$ as a whole. [15] introduced a sequential scheme based on conjugate Bayesian analysis only for the canonical parameter η_t of the observation density (4) (WHM-Algorithm). Using the non-linear relationship

$$\eta_t(\boldsymbol{\beta}_t) = (b')^{-1}(g^{-1}(\boldsymbol{H}\boldsymbol{\beta}_t)) \tag{7}$$

between the canonical parameter η_t and the state variable $\boldsymbol{\beta}_t$ the first and the second (approximated) moments of the posterior density of $\boldsymbol{\beta}_{t-1}$ are matched to certain characteristics of a conjugate prior of η_t. Using linear Bayes' arguments the characteristics of the posterior of η_t are matched back to obtain approximate first and second moments of the posterior density of $\boldsymbol{\beta}_t$.

Such a reduction of Bayesian analysis to the canonical parameter is not necessary. Sequential Bayesian analysis is possible directly for the state variable $\boldsymbol{\beta}_t$ (see [14]):

$$p(\boldsymbol{\beta}_t|y^t) \propto p(y_t|\boldsymbol{\beta}_t)p(\boldsymbol{\beta}_t|y^{t-1}). \tag{8}$$

However, the problem of approximation still remains. [2] suggested characterizing the posterior density of the state variable $\boldsymbol{\beta}_t$ by the posterior mode and curvature. He developed a sequential estimation scheme for approximating these quantities. A method of approximating the mode of the posterior $p(\boldsymbol{\beta}_0, \ldots, \boldsymbol{\beta}_N|y^N)$ of the whole state vector is presented in [5].

In Subsection 3.2 we will approximate the expectation and variance of the posterior density of the state variable $\boldsymbol{\beta}_t$ using numerical integration. This integration-based Kalman-filtering scheme is designed to improve the above-mentioned approximate posterior mode estimator of [2].

3.2 Approximating Posterior Moments by Integration-based Kalman-filtering

From Bayes' theorem (8) the expectation and the variance of the posterior density of the state variable $\boldsymbol{\beta}_t$ are determined by the following integrals:

$$\boldsymbol{E}(\boldsymbol{\beta}_t|y^t) = \frac{I(\boldsymbol{\beta}_t)}{I(1)} \tag{9}$$

$$\boldsymbol{V}(\boldsymbol{\beta}_t|y^t) = \frac{I(\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{\mathrm{T}})}{I(1)} - \boldsymbol{E}(\boldsymbol{\beta}_t|y^t)\boldsymbol{E}(\boldsymbol{\beta}_t|y^t)^{\mathrm{T}} \tag{10}$$

with

$$I(h(\boldsymbol{\beta}_t)) = \int h(\boldsymbol{\beta}_t)p(y_t|\boldsymbol{\beta}_t)p(\boldsymbol{\beta}_t|y^{t-1})\,d\boldsymbol{\beta}_t. \tag{11}$$

As no conjugate analysis is possible, these integrals have to be approximated. What is important to realize is that numerical integration of even high accuracy will not leed to the exact moments, as the prior $p(\boldsymbol{\beta}_t|y^{t-1})$ is not completely known for $t > 1$. It is given by the integral

$$p(\boldsymbol{\beta}_t|y^{t-1}) = \int p(\boldsymbol{\beta}_t|\boldsymbol{\beta}_{t-1})p(\boldsymbol{\beta}_{t-1}|y^{t-1})\,d\boldsymbol{\beta}_{t-1}. \tag{12}$$

The densitiy $p(\boldsymbol{\beta}_{t-1}|y^{t-1})$ is characterized only by the first and second moments $\boldsymbol{E}(\boldsymbol{\beta}_{t-1}|y^{t-1})$ and $\boldsymbol{V}(\boldsymbol{\beta}_{t-1}|y^{t-1})$. If we approximate this density by a Gaussian density with the same first and second moments, then extrapolation by (12) leads to an approximate Gaussian prior $p(\boldsymbol{\beta}_t|y^{t-1})$ of $\boldsymbol{\beta}_t$ with the following moments:

$$\boldsymbol{E}(\boldsymbol{\beta}_t|y^{t-1}) = \boldsymbol{F}\cdot\boldsymbol{E}(\boldsymbol{\beta}_{t-1}|y^{t-1}), \tag{13}$$

$$\boldsymbol{V}(\boldsymbol{\beta}_t|y^{t-1}) = \boldsymbol{F}\cdot\boldsymbol{V}(\boldsymbol{\beta}_{t-1}|y^{t-1})\cdot\boldsymbol{F}^{\mathrm{T}} + \boldsymbol{Q}(\theta_1,\theta_2). \tag{14}$$

Because of the non-linear relationship (7) between the state variable $\boldsymbol{\beta}_t$ and the canonical parameter η_t a Gaussian prior for the state variable $\boldsymbol{\beta}_t$ corresponds to a non-Gaussian prior for the canonical parameter η_t. For the examples discussed below (Poisson and Gamma time series with a log-transformation) the prior of the natural parameter is log-normal.

The filtering scheme we propose in this paper is based on sequential numerical integration of (11) with the prior $p(\boldsymbol{\beta}_t|y^{t-1})$ of the state variable being substituted by the approximate Gaussian prior. The computation of posterior moments via numerical (Gauss-Hermite) integration has been successfully implemented in Bayesian Statistics (eg. [10], [12]). Especially for multivariate integrals the practical performance of Gauss-Hermite integration depends on a sensible choice of parameter transformation. If the posterior of $\boldsymbol{\beta}_t$ is approximately normal with mean $\boldsymbol{m}_t$ and covariance matrix $\boldsymbol{S}_t$, then the transformation

$$\boldsymbol{\tau} = \boldsymbol{U}_t^{-1}(\boldsymbol{\beta}_t - \boldsymbol{m}_t), \qquad 2\boldsymbol{S}_t = \boldsymbol{U}_t\boldsymbol{U}_t^{\mathrm{T}}, \tag{15}$$

If we write (11) as

$$I(h(\boldsymbol{\beta}_t)) = \int h(\boldsymbol{\beta}_t) d(\boldsymbol{\beta}_t) p_N(\boldsymbol{\beta}_t; \boldsymbol{m}_t, \boldsymbol{S}_t)\, d\boldsymbol{\beta}_t \tag{26}$$

with

$$d(\boldsymbol{\beta}_t) = \frac{p(y_t|\boldsymbol{\beta}_t) \cdot p(\boldsymbol{\beta}_t|y^{t-1})}{p_N(\boldsymbol{\beta}_t; \boldsymbol{m}_t, \boldsymbol{S}_t)}. \tag{27}$$

then $I(1)$, $I(\boldsymbol{\beta}_t)$ and $I(\boldsymbol{\beta}_t\boldsymbol{\beta}_t^{\mathrm{T}})$ are integrals of the form (16). Thus their approximation is straightforward now. After a few calculations integration-based Kalman-filtering may be expressed in terms of the posterior mode $\boldsymbol{m}_t$ and curvature $\boldsymbol{S}_t = 0.5\, \boldsymbol{U}_t\boldsymbol{U}_t^{\mathrm{T}}$:

$$\boldsymbol{E}(\boldsymbol{\beta}_t|y^t) = \boldsymbol{m}_t + \boldsymbol{U}_t \cdot \boldsymbol{z}_t \tag{28}$$

$$\boldsymbol{V}(\boldsymbol{\beta}_t|y^t) = \boldsymbol{U}_t\boldsymbol{Z}_t\boldsymbol{U}_t^{\mathrm{T}} \tag{29}$$

$$z_t = \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \psi^{(i,j)} \tag{30}$$

$$\boldsymbol{z}_t = \frac{1}{z_t}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2} \boldsymbol{\tau}^{(i,j)}\psi^{(i,j)} \tag{31}$$

$$\boldsymbol{Z}_t = \frac{1}{z_t}\sum_{i_1=1}^{M_1}\sum_{j=1}^{M_2} \boldsymbol{\tau}^{(i,j)}(\boldsymbol{\tau}^{(i,j)})^{\mathrm{T}}\psi^{(i,j)} - \boldsymbol{z}_t\boldsymbol{z}_t^{\mathrm{T}}, \tag{32}$$

$$\psi^{(i,j)} = d(\boldsymbol{\beta}_t^{(i,j)})w^{(i,j)}. \tag{33}$$

This scheme may be viewed as a scheme for correcting the difference between posterior expectation and posterior mode. Note that z_t is an approximate value of the normalising constant $I(1)$.

3.3 Example 1: Gaussian Time Series with Identity Transformation

For Gaussian time series with identity transformation $g(\mu_t) = \mu_t$ both the WHM-Algorithm and the posterior mode approximation are identical with Kalman-filtering. The same holds for the scheme introduced in the previous section.

It follows from Bayes' theorem that

$$p_N(\boldsymbol{\beta}_t; \boldsymbol{m}_t, \boldsymbol{S}_t) = \frac{1}{c} p(y_t|\boldsymbol{\beta}_t)p(\boldsymbol{\beta}_t|y^{t-1})$$

with a constant c. Therefore:

$$d(\boldsymbol{\beta}_t) \equiv c, \quad z_t = \sum_{i=1}^{M_1}\sum_{j=1}^{M_2} cw^{(i,j)} = c.$$

For given M_1 and M_2 integration of polynomials

$$f(\boldsymbol{\beta}_t) = \beta_{t,1}^{m_1}\beta_{t,2}^{m_2}$$

of order $m_1 \leq 2M_1 - 1$ and $m_2 \leq 2M_2 - 1$ is exact ([13]). Therefore for $M_1 \geq 2, M_2 \geq 2$ we obtain:

$$\boldsymbol{z}_t = \frac{1}{c}\sum_{i=1}^{M_1}\sum_{j=1}^{M_2}\boldsymbol{\tau}^{(i,j)} c\, w^{(i,j)} = \int \boldsymbol{\tau} \cdot p_N(\boldsymbol{\tau}; \mathbf{0}, 0.5\boldsymbol{I})\, d\boldsymbol{\tau} = \mathbf{0},$$

$$\boldsymbol{Z}_t = \frac{1}{c}\sum_{i_1=1}^{M_1}\sum_{j=1}^{M_2}\boldsymbol{\tau}^{(i,j)}(\boldsymbol{\tau}^{(i,j)})^{\mathrm{T}} c\, w^{(i,j)} = \int \boldsymbol{\tau}\boldsymbol{\tau}^{\mathrm{T}} p_N(\boldsymbol{\tau}; \mathbf{0}, 0.5\boldsymbol{I}) = 0.5\boldsymbol{I}$$

Therefore actually the classical Kalman-filtering routine results.

3.4 Example 2: Poisson Time Series with Log Transformation

For Poisson time series with log-transformation $g(\mu_t) = \ln \mu_t$ the WHM-Algorithm is discussed in [15]. The posterior mode approximation is given by (20) - (23) with

$$E(y_t|y^{t-1}) = \exp(\boldsymbol{H}\boldsymbol{E}(\boldsymbol{\beta}_t|y^{t-1})), \quad V(y_t|y^{t-1}) = E(y_t|y^{t-1}), \quad h_t = E(y_t|y^{t-1}).$$

For integration-based Kalman-filtering we need the likelihood $p(y_t|\boldsymbol{\beta}_t)$:

$$p(y_t|\boldsymbol{\beta}_t) = \frac{\mu_t^{y_t}}{\Gamma(y_t+1)}\exp(-\mu_t), \qquad \mu_t = \exp(\boldsymbol{H}\boldsymbol{\beta}_t).$$

3.5 Example 3: Gamma Time Series with Log Transformation

As far as we know, the WHM-Algorithm has not been discussed so far for Gamma time series with log-transformation $g(\mu_t) = \ln \mu_t$. We will apply their ideas to Gamma data and perform approximate conjugate analysis for the mean μ_t. From

$$\lambda_t = \boldsymbol{H}\boldsymbol{\beta}_t \sim \mathrm{N}(l_t, L_t)$$

prior moments of $\mu_t = \exp(\lambda_t)$ are given by:

$$E(\mu_t|y^{t-1}) = \exp(l_t + 0.5L_t), \qquad V(\mu_t|y^{t-1}) = \exp(2l_t + L_t)(\exp(L_t) - 1). \quad (34)$$

A conjugate prior of μ_t for Gamma distributed data is given by an inverted Gamma $IG(\nu_{t-1}, \tau_{t-1})$. Matching (34) to an inverted Gamma p.d.f. with the same first and second moment leads to

$$\nu_{t-1} = 2 + \frac{E(\mu_t|y^{t-1})^2}{V(\mu_t|y^{t-1})}, \qquad \tau_{t-1} = \nu_{t-1}E(\mu_t|y^{t-1}).$$

Conjugate analysis results in the following posterior moments of μ_t:

$$E(\mu_t|y^t) = \frac{\tau_{t-1} + y_t/\phi}{\nu_{t-1} + 1/\phi - 1}, \qquad V(\mu_t|y^t) = \frac{(E(\mu_t|y^t))^2}{\nu_{t-1} + 1/\phi - 2}$$

Matching these moments back to λ_t through the log-transformation leads to

$$V(\lambda_t|y^t) = \ln\left(1 + \frac{E(\mu_t|y^t)}{V(\mu_t|y^t)}\right), \qquad E(\lambda_t|y^t) = \ln E(\mu_t|y^t) - 0.5V(\lambda_t|y^t).$$

The posterior of $\boldsymbol{\beta}_t$ is determined from these moments with the usual linear Bayes' arguments ([15]).

The posterior mode approximation for Gamma time series is given by (20) - (23) with

$$E(y_t|y^{t-1}) = \exp(\boldsymbol{H}\boldsymbol{E}(\boldsymbol{\beta}_t|y^{t-1})), \quad V(y_t|y^{t-1}) = E(y_t|y^{t-1})^2 \cdot \phi, \quad h_t = E(y_t|y^{t-1}).$$

The likelihood $p(y_t|\boldsymbol{\beta}_t)$ necessary for integration-based Kalman-filtering is given by:

$$p(y_t|\boldsymbol{\beta}_t) = \frac{1}{\Gamma(\nu)}(y_t \cdot \eta_t)^{\nu-1}\exp(-y_t\eta_t), \quad \nu = 1/\phi, \quad \eta_t = \exp(\boldsymbol{H}\boldsymbol{\beta}_t)/\phi.$$

4 Experimental Studies for Simulated Time Series

In order to gain insight into characteristic properties of the different approximation methods we performed some simulation experiments.

4.1 Comments on the Practical Implementation

For practical implementation we have to fix the grid size $M_1 \times M_2$. The minimum grid size is 2×2 leading to the *exact* posterior moments for *Gaussian* observation densities (see Subsection 3.3). However, for very skew observation densities such a grid will be too small. Comparative results for the examples given below using different grid sizes suggested using use a 5×5 grid.

For each time step just one integration based on the approximate posterior mode estimator (20) - (23) is carried out. In principle, the iterative method suggest by [10] may well be applied to our problem by using the approximate posterior moments $\boldsymbol{E}(\boldsymbol{\beta}_t|y^t)$ and $\boldsymbol{V}(\boldsymbol{\beta}_t|y^t)$ as modified parameters $\boldsymbol{m}_t$ and $\boldsymbol{S}_t$ of the transformation (15) until convergence occurs.

4.2 Time Series Simulated from Poisson Distributions

We simulated three paths of the state variable $\boldsymbol{\beta}_t$ from the dynamic linear trend model (2) with identical process $\boldsymbol{w}_t$, identical starting value for the trend component $\boldsymbol{\beta}_{0,2}$:

$$\boldsymbol{\beta}_{0,2} = 0.01, \quad \theta_1 = 0.005, \quad \theta_2 = 10^{-6},$$

but different starting values for the level $\boldsymbol{\beta}_{0,1}$:

$$\boldsymbol{\beta}_{0,1} = -1, \quad \boldsymbol{\beta}_{0,1} = 1, \quad \boldsymbol{\beta}_{0,1} = 5.$$

Using the log-transformation $g(\mu_t) = \ln(\mu_t)$ three paths of Poisson distributed time series of length $T = 100$ were simulated from the three processes $\boldsymbol{\beta}_t$ (see Figure 1). For the first process μ_0 is equal to 0.37 and the likelihood is very skew, for the second process μ_0 is equal to 2.72 with a moderate skewed likelihood and for the third process with $\mu_0 = 148.4$ the likelihood practically is Gaussian.

For all approximations we take the exact value $\boldsymbol{\beta}_0$ as mean of the prior of $p(\boldsymbol{\beta}_0|y^0)$. The covariance of the prior is chosen as:

$$\boldsymbol{V}(\boldsymbol{\beta}_0|y^0) = \begin{pmatrix} 1 & 0 \\ 0 & 0.01 \end{pmatrix}.$$

Figure 2 shows the sequentially estimated naive confidence bands $E(\boldsymbol{\beta}_{t,i}|y^t) \pm 2\sqrt{V(\boldsymbol{\beta}_{t,i}|y^t)}$ of both components $\boldsymbol{\beta}_{t,i}$ of the state variable $\boldsymbol{\beta}_t$ for the three methods under investigation. The differences between the methods are discussed in Section 4.4.

4.3 Time Series Simulated from Gamma Distributions

We simulated a path of the state variable $\boldsymbol{\beta}_t$ from the dynamic linear trend model (2) with

$$\boldsymbol{\beta}_0 = \begin{pmatrix} 1. \\ 0.01 \end{pmatrix}, \quad \theta_1 = 0.005, \quad \theta_2 = 10^{-6}.$$

Using the log-transformation $g(\mu_t) = \ln(\mu_t)$ three paths of Gamma distributed time series of length $T = 100$ were simulated from the process $\boldsymbol{\beta}_t$ with different scaling factors ϕ. For the first time series with $\phi = 1$ the simulated data have an exponential distribution with a very skew likelihood, for the second time series with $\phi = 0.2$, the simulated data have a Gamma distribution with modest skewness and for the third time series with $\phi = 0.02$ the simulated data are nearly normal with mean dependent variance (see Figure 3).

Figure 1: Time Series Simulated from a Poisson distribution

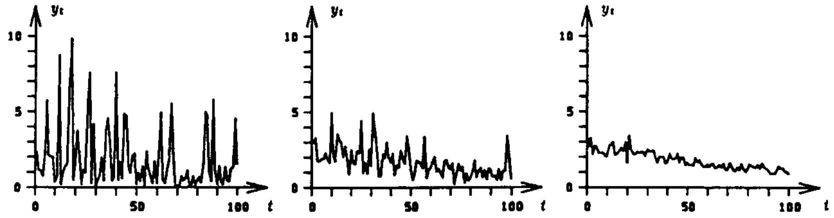


Figure 2: Naive confidence bands of the state variable

(dotted:WHM-Algorithm, broken: posterior mode estimation, full: integration-based Kalman-filtering, full line in the middle: simulated path of the state variable)

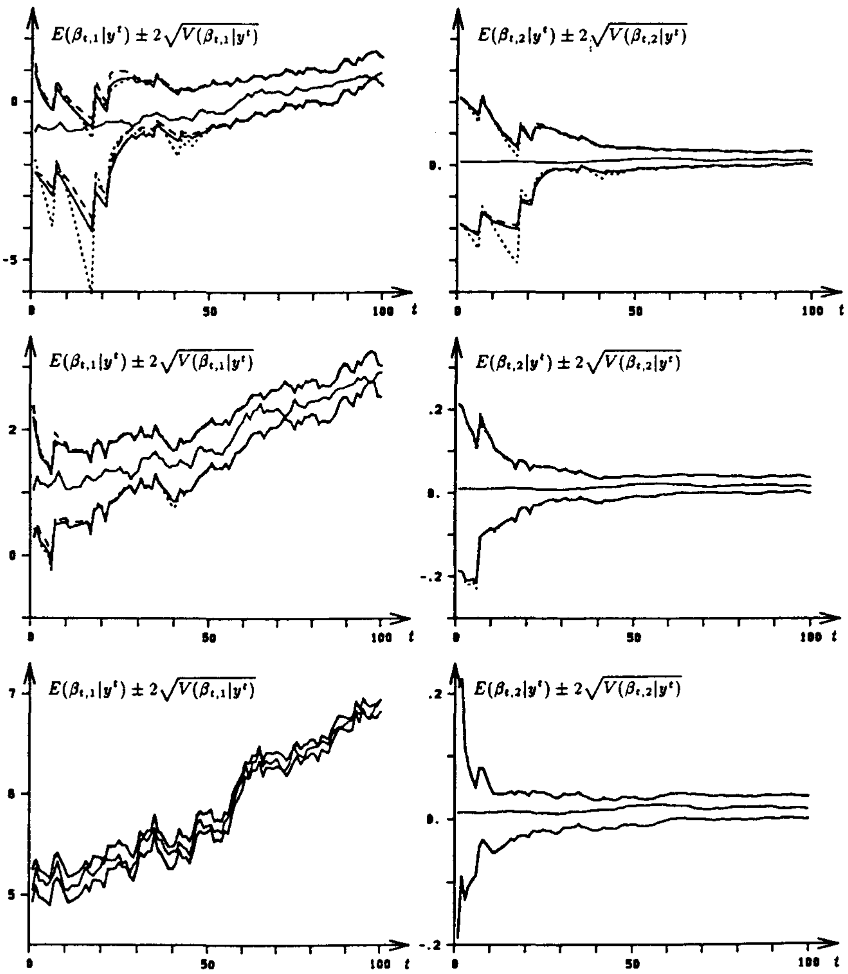


Figure 3: Time Series Simulated from a Gamma distribution

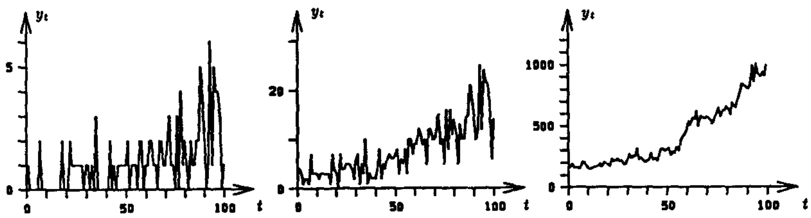
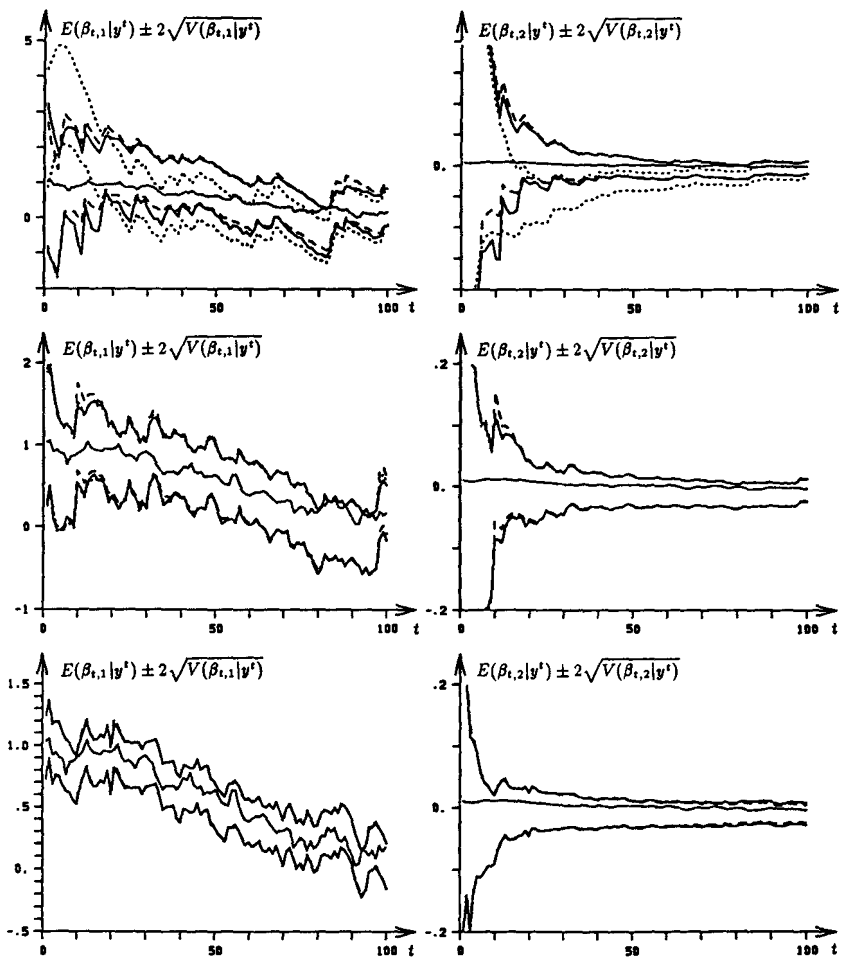


Figure 4: Naive confidence bands of the state variable

(dotted:WHM-Algorithm, broken: posterior mode, full: integration-based Kalman-filtering, full line in the middle: simulated path of the state variable)



For all approximations we take the exact value $\boldsymbol{\beta}_0$ as mean of the prior of $p(\boldsymbol{\beta}_0|y^0)$. The covariance of the prior is chosen as:

$$V(\boldsymbol{\beta}_0|y^0) = \begin{pmatrix} 4\phi & 0 \\ 0 & \phi \end{pmatrix}.$$

Figure 4 shows the sequentially estimated naive confidence bands $E(\boldsymbol{\beta}_{t,i}|y^t) \pm 2\sqrt{V(\boldsymbol{\beta}_{t,i}|y^t)}$ of both components $\boldsymbol{\beta}_{t,i}$ of the state variable $\boldsymbol{\beta}_t$ for the three methods.

4.4 Discussion of Experimental Results

Figure 2 and Figure 4 show that the differences between the naive confidence bands depend on the skewness of the observation density and are the smaller the nearer the observation density is to a Gaussian density. However, even for the case of very skew likelihoods (first time series in both examples) these differences decay as the number of filtering steps increases. Such a result might indicate asymptotic normality of the posterior density. It would be an interesting task of asymptotic statistics to confirm these findings from experimental statistics, e.g. by extending the results of [4] to dynamic generalized linear models.

For the exponential time series the WHM-algorithm shows considerable deviations from posterior mode estimation and integration-based Kalman-filtering. These deviations decrease rather slowly with increasing number of filtering steps – a phenomenon possibly indicating bias. However, this bias vanishes for time series with nearly Gaussian observation densities.

5 Concluding Remarks

In this paper sequential estimation was solved for the special case of a dynamic generalized linear trend model using integration-based Kalman-filtering. Evidently, this method is not restricted to this special non-Gaussian dynamic model. It may be applied to other non-Gaussian dynamic models with state variables of moderate dimension, say up to 5 or 6.

If the dimension of the state variable is high, Monte-Carlo integration may be more appropriate than Gauss-Hermite integration. The integration-based Kalman-filtering scheme is easily adapted to sampling-based integration method. We just have to sample M grid points $\boldsymbol{\beta}_t^{(i)}$ from the approximate posterior $p_N(\boldsymbol{\beta}_t; \boldsymbol{m}_t, \boldsymbol{S}_t)$ with the parameters being equal to the posterior mode scheme of [2] instead of using transformed Gauss-Hermite grid points.

A last remark may be added on the problem of unknown hyperparameters $\boldsymbol{\theta}$. In the case of the dynamic generalised linear trend model, for example, the process variances θ_1 and θ_2 will be unknown in practice. Data based information about $\boldsymbol{\theta}$ is contained in the likelihood function $L(\boldsymbol{\theta})$ given by the product of the N successive one-step-ahead predictive values $p(y_t|\boldsymbol{\theta}, y^{t-1})$:

$$L(\boldsymbol{\theta}) = \prod_{t=1}^{N} p(y_t|\boldsymbol{\theta}, y^{t-1}). \tag{35}$$

For non-Gaussian dynamic models the predictive densities are not known analytically. However, for a *fixed* value of $\boldsymbol{\theta}$ the predictive value $p(y_t|\boldsymbol{\theta}, y^{t-1})$ is identical with the normalising constant $I(1)$ occuring in the filtering step at time t (compare (9) and (10)). As stated above, an approximate value z_t of $I(1)$ directly resulted from numerical integration (see (30)). Thus the proposed integration-based Kalman-filtering scheme *automatically* provides an approximation of the likelihood function $L(\boldsymbol{\theta})$. Estimates of the unknown hyperparameters may be obtained by maximizing this approximate likelihood function numerically using algorithms discussed e.g. in [7]. A Bayesian approach may be implemented by a straightforward extension of the multi-process-filtering algorithm ([6]).

Acknowledgements

This work is part of a research project at the Department of Statistics and Probability Theory at Vienna University of Technology under the supervision of Prof. R. Viertl. It has been supported by the Austrian Science Foundation, Project Nr. 7079.

References

[1] M. Abramowitz, I. Stegun: *Handbook of Mathematical Functions.* New York, 1970.

[2] L. Fahrmeir: *Posterior Mode Estimation by Extended Kalman-Filtering for Multivariate Dynamic Generalized Linear Models.* Regensburger Beiträge zur Statistik und Ökonometrie, **10**. Regensburg, 1988.

[3] L. Fahrmeir: Extended Kalman-Filtering for Non-normal Longitudinal Data. In: A. Decarli, B.J. Francis, R. Gilchrist, G.U.H. Seeber (Eds.): *Statistical Modelling*, Lecture Notes in Statistics, **57**, 151-156. Springer, Berlin/ Heidelberg, 1989.

[4] L. Fahrmeir, H. Kaufmann: Consistency and Asymptotic Normality of the Maximum Likelihood Estimator in Generalized Linear Models. *Ann. Statist.*, 1985, **13**, 342-368.

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

[6] P.J. Harrison, C.F. Stevens: Bayesian Forecasting. *J.R.Stat.Soc.,* 1976, **38**, 205 - 247.

[7] A. Harvey: *Forecasting, Structural Time Series Models and the Kalman Filter.* Cambridge University Press, Cambridge, 1989.

[8] G. Kitagawa: Non-Gaussian State Space Modelling of Nonstationary Time Series (with comments). *JASA*, 1987, **82**, 1032-1063.

[9] P. McCullagh, J.A.Nelder: *Generalized Linear Models.* 2^{nd} ed. Chapman and Hall, London/New York, 1989.

[10] J.C. Naylor, A.F.M. Smith: Application of a Method for the Efficient Computation of Posterior Distributions. *Applied Statistics*, 1982, **31**, 214-225.

[11] S. Schnatter: *Dynamic Bayesian Models and their Application to Hydrological Short-term Forecasting* (in German). Ph.D. Dissertation, University of Technology, Vienna, Austria, 1988 (unpublished).

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

[13] A. Stroud: *Approximate Calculation of Multiple Integrals.* Prentice Hall, New Jersey, 1971.

[14] M. West, P.J. Harrison: *Bayesian Forecasting and Dynamic Models.* Springer Series in Statistics. Springer, 1989.

[15] M. West, P.J. Harrison, H.S. Migon: Dynamic Generalized Linear Models and Bayesian Forecasting. *JASA*, 1985, **80**, 389, 73-97.