Sylvia Frühwirth-Schnatter

Data Augmentation and Dynamic Linear Models

Working Paper

Data Augmentation and Dynamic Linear Models

Sylvia Frühwirth-Schnatter

Institut für Statistik
Wirtschaftsuniversität Wien

Forschungsberichte

Bericht 28
1992

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

Data Augmentation and Dynamic Linear Models

Sylvia Frühwirth-Schnatter
Department of Statistics, Vienna University of Economics, Vienna, Austria

All correspondence and proofs sould be sent to the author under the following postal address:

Dr. Sylvia Frühwirth-Schnatter
University of Economics
Augasse 2-6
A-1090 Vienna
Austria

Abstract

We define a subclass of dynamic linear models with unknown hyperparameters called d-inverse-gamma models. We then approximate the marginal p.d.f.s of the hyperparameter and the state vector by the data augmentation algorithm of Tanner/Wong. We prove that the regularity conditions for convergence hold. A sampling based scheme for practical implementation is discussed. Finally, we illustrate how to obtain an iterative importance sampling estimate of the model likelihood.

Keywords: Approximate Bayesian Analysis, Data Augmentation, Dynamic Linear Models, Kalman Filtering, Model Likelihood, State Space Models

1 Introduction

In the present paper we define a subclass of dynamic linear models with unknown hyperparameters. This subclass is characterized in Section 2 by the structure of the posterior p.d.f. of the distribution of the hyperparameter conditional on the state and observation process. As this density factorizes into the product of d p.d.f.s of an inverted gamma distribution this class will be called d-inverse-gamma.

In Section 3 we suggest to approximate the marginal p.d.f.s of the hyperparameter and the state vectors by the data augmentation algorithm of Tanner and Wong (1987). We prove that the convergence conditions given by Tanner and Wong (1987) hold for the model class under investigation. For practical implementation a sampling based scheme is presented which is based on sampling trajectories of the state process from the conditional posterior p.d.f. of the state process given the hyperparameter and on sampling the hyperparameter from the conditional posterior p.d.f. given a trajectory of the state process.

In Section 4 we illustrate how an iterative importance sampling estimate of the model likelihood results as a by-product of the data augmentation algorithm.

2 The d-inverse-gamma Class

2.1 Model Definition

Assume that an observed time-series $y_1, \dots, y_N$ is the realization of a dynamic linear model depending on unknown hyperparameters $\boldsymbol{\theta} \in \Theta \subseteq I\!R^d$:

$$\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta} \sim N(\boldsymbol{F}_t\boldsymbol{x}_{t-1}, \boldsymbol{Q}_t(\boldsymbol{\theta})), \quad y_t|\boldsymbol{x}_t, \boldsymbol{\theta} \sim N(\boldsymbol{H}_t\boldsymbol{x}_t, R_t(\boldsymbol{\theta})), \qquad t = 1, 2, \dots \quad (1)$$

$\boldsymbol{x}_t \in I\!R^r$ is the state vector. $R_t(\boldsymbol{\theta})$ is called the observation variance, $\boldsymbol{Q}_t(\boldsymbol{\theta})$ is called the process variance. The state vector $\boldsymbol{x}_0$ is assumed to be normally distributed: $\boldsymbol{x}_0|y^0 \sim N(\hat{\boldsymbol{x}}_{0|0}, \boldsymbol{P}_{0|0})$. $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d)$ is assumed to have a-priori independent, inverted gamma-distributed components: $\theta_j|y^0 \sim IG(\alpha_0^{(j)}, \beta_0^{(j)}), \alpha_0^{(j)} > 0, \beta_0^{(j)} > 0, j = 1, \ldots, d$. $\boldsymbol{x}_0|y^0$ and $\boldsymbol{\theta}|y^0$ are assumed to be independent a-priori. Let $\boldsymbol{x}^N = (\boldsymbol{x}_0, \boldsymbol{x}_1, \ldots, \boldsymbol{x}_N)$ denote the whole process of state vectors from 0 to N (state process) and let $y^N = (y_1, \ldots, y_N)$ denote all observations up to N.

It is well known that the conditional posterior p.d.f. $p(\boldsymbol{x}^N|\boldsymbol{\theta}, y^N)$ of the whole state process $\boldsymbol{x}^N$ given the observations y^N and the hyperparameter $\boldsymbol{\theta}$ is the density of a normal distribution (e.g. Harvey, 1989). The conditional density $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ of $\boldsymbol{\theta}$ given $\boldsymbol{x}^N$ and given all observations y^N has a known analytical form (e.g. Broemeling, 1985):

$$\begin{aligned} p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) &\propto p(y_N|\boldsymbol{x}^N, \boldsymbol{\theta}, y^0) p(\boldsymbol{x}^N|\boldsymbol{\theta}, y^0) = \\ &= \prod_{t=1}^{N} p(y_t|\boldsymbol{x}_t, \boldsymbol{\theta}) \cdot \prod_{t=1}^{N} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) \cdot p(\boldsymbol{x}_0|y^0) p(\boldsymbol{\theta}|y^0). \end{aligned} \tag{2}$$

If $\boldsymbol{Q}_t(\boldsymbol{\theta})$ is regular, then (2) takes the following form:

$$\begin{aligned} & p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) \propto p(\boldsymbol{\theta}|y^0) \cdot \\ & \cdot \prod_{t=1}^{N} |\boldsymbol{Q}_t(\boldsymbol{\theta})|^{-\frac{1}{2}} \cdot \exp\left\{ -\frac{1}{2} \sum_{t=1}^{N} (\boldsymbol{x}_t - \boldsymbol{F}_t \boldsymbol{x}_{t-1})^{\mathrm{T}} \boldsymbol{Q}_t(\boldsymbol{\theta})^{-1} (\boldsymbol{x}_t - \boldsymbol{F}_t \boldsymbol{x}_{t-1}) \right\} \cdot \\ & \cdot \prod_{t=1}^{N} |R_t(\boldsymbol{\theta})|^{-\frac{1}{2}} \cdot \exp\left\{ -\frac{1}{2} \sum_{t=1}^{N} (y_t - \boldsymbol{H}_t \boldsymbol{x}_t)^2 R_t(\boldsymbol{\theta})^{-1} \right\}. \end{aligned} \tag{3}$$

Wether $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ is the density of a standard distribution or not depends on the way how the variances $\boldsymbol{Q}_t(\boldsymbol{\theta})$ and $R_t(\boldsymbol{\theta})$ depend on $\boldsymbol{\theta}$. We now characterize a subclass of dynamic linear models by the structure of this density.

Definition 1. A path $\boldsymbol{x}^N$ of the state process is called realizable from model (1), if $\exists$ a hyperparameter $\boldsymbol{\theta} \in \Theta$ such that

$$p(\boldsymbol{x}_1, \ldots, \boldsymbol{x}_N|\boldsymbol{\theta}, \boldsymbol{x}_0) = \prod_{t=1}^{N} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) > 0. \tag{4}$$

Definition 2. A dynamic linear model with an unknown hyperparameter $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in \Theta = I\!R_+^d$ is called d-**inverse-gamma**, if the conditional posterior p.d.f. $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ of $\boldsymbol{\theta}$ given a realizable path of the state process $\boldsymbol{x}^N$ and given the observation process y^N splits into the product of d densities of inverted gamma distributions:

$$p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) = \prod_{j=1}^{d} p(\theta_j|\boldsymbol{x}^N, y^N), \quad \theta_j|\boldsymbol{x}^N, y^N \sim IG(\alpha_N^{(j)}, \beta_N^{(j)}(\boldsymbol{x}^N, y^N)). \tag{5}$$

2.2 Discussion and Conditional Posterior Analysis

The main property of this model class is the fact that the components of the hyperparameter are conditionally independent given the state and the observation process. A variety of dynamic linear models with unknown hyperparameters, e.g. the steady-state-model, the dynamic trend model or regression models with dynamic regression coefficients, belong to the d-inverse-gamma class. We are going to discuss structures of $\boldsymbol{Q}_t$ and $R_t(\boldsymbol{\theta})$ which ensure that the model is inverse-gamma in this subsection.

Typically, $R_t(\boldsymbol{\theta})$ is of the form

$$R_t(\boldsymbol{\theta}) = v_t(y^{t-1})\theta_d \tag{6}$$

with θ_d unknown and $v_t(\cdot)$ equal to a positive function which is deterministic given y^{t-1}.

As far as $\boldsymbol{Q}_t(\boldsymbol{\theta})$ is concerned we will distinguish between the "fully dynamic" case of $\boldsymbol{Q}_t(\boldsymbol{\theta})$ regular and the "partly dynamic" case of non-regular $\boldsymbol{Q}_t(\boldsymbol{\theta})$. For a fully dynamic model $\boldsymbol{Q}_t(\boldsymbol{\theta})$ typically has one of the following forms:

$$\boldsymbol{Q}_t(\boldsymbol{\theta}) = \theta_1 \boldsymbol{A}_t(y^{t-1}), \tag{7}$$

$$\boldsymbol{Q}_t(\boldsymbol{\theta}) = \boldsymbol{B}_t(y^{t-1})\mathrm{Diag}(\theta_1, \ldots, \theta_r)\boldsymbol{B}_t^{\mathrm{T}}(y^{t-1}), \tag{8}$$

with $\boldsymbol{A}_t(y^{t-1})$ and $\boldsymbol{B}_t(y^{t-1})$ being deterministic, regular matrices, if y^{t-1} is known. $\mathrm{Diag}(\cdot)$ is a diagonal matrix. In simple cases the function $v_t(\cdot)$ in (6) and the matrices $\boldsymbol{A}_t(\cdot)$ in (7) or $\boldsymbol{B}_t(\cdot)$ in (8) will be constant (e.g. steady-state-model, dynamic trend model, regression models with a dynamic regression coefficient). In some cases they are equal to a constant depending on time. More generally, they may also depend on past paths of the observation process (e.g. the conditionally Gaussian model studied in Liptser and Shiryayev, 1977).

It is straightforward to determine the parameters of the conditional posterior p.d.f. $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ in (5) from (3). They are given by:

$$\alpha_N^{(j)} = \alpha_0^{(j)} + \frac{N}{2},$$

$$\beta_N^{(j)}(\boldsymbol{x}^N, y^N) = \beta_0^{(j)} + b_N^{(j)}(\boldsymbol{x}^N, y^N), \quad j = 1, \ldots, d.$$

The parameter $b_N^{(d)}(\cdot)$ which determines the posterior of the scaling factor θ_d of the observation variance is given by:

$$b_N^{(d)}(\boldsymbol{x}^N, y^N) = \frac{1}{2}\sum_{t=1}^{N} \frac{(y_t - \boldsymbol{H}_t\boldsymbol{x}_t)^2}{v_t(y^{t-1})}. \tag{9}$$

The parameters determining the posterior of the other components depend on the form of $\boldsymbol{Q}_t(\boldsymbol{\theta})$. If $\boldsymbol{Q}_t(\boldsymbol{\theta})$ is of the form (7) then the model is 2-inverse-gamma with $b_N^{(1)}(\cdot)$ given by:

$$b_N^{(1)}(\boldsymbol{x}^N, y^N) = \frac{1}{2}\sum_{t=1}^{N}(\boldsymbol{x}_t - \boldsymbol{F}_t\boldsymbol{x}_{t-1})^{\mathrm{T}}\boldsymbol{A}_t^{-1}(y^{t-1})(\boldsymbol{x}_t - \boldsymbol{F}_t\boldsymbol{x}_{t-1}). \tag{10}$$

If $\boldsymbol{Q}_t(\boldsymbol{\theta})$ is of the form (8) then the model is $(r+1)$-inverse-gamma with $b_N^{(j)}(\cdot)$, $j = 1, \ldots, r$, given by:

$$b_N^{(j)}(\boldsymbol{x}^N, y^N) = \frac{1}{2} \sum_{t=1}^{N} \left[\boldsymbol{B}_t^{-1}(y^{t-1})(\boldsymbol{x}_t - \boldsymbol{F}_t \boldsymbol{x}_{t-1})\right]_j^2. \tag{11}$$

$[\cdot]_j$ denotes the j-th component of a vector.

Now let us proceed with the partly dynamic model. A partly dynamic model most often occurs, if in a fully dynamic model with unknown hyperparameter $\tilde{\boldsymbol{\theta}}$ m components of $\tilde{\boldsymbol{\theta}}$ are restricted to 0. Then only $r - m$ components of $\tilde{\boldsymbol{\theta}}$ – denoted by $\boldsymbol{\theta}$ – remain unknown.

Lemma 1. Assume that a fully dynamic model belongs to the d-inverse-gamma class with $\boldsymbol{Q}_t(\tilde{\boldsymbol{\theta}})$ taking the form (7) or (8). Then the partly dynamic model obtained by restricting m components of $\tilde{\boldsymbol{\theta}}$ to 0 is a $(d-m)$-inverse-gamma-model. If $\boldsymbol{x}^N$ is a realizable path from the restricted model, then the following holds: the parameters $b_N^{(j)}(\boldsymbol{x}^N, y^N)$, $j = 1, \ldots, (d-m)$ of the conditional posterior $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ in the restricted model are related to the parameters $\tilde{b}_N^{(i)}(\boldsymbol{x}^N, y^N)$ in the unrestricted model by the following identity:

$$b_N^{(j)}(\boldsymbol{x}^N, y^N) = \tilde{b}_N^{(i_j)}(\boldsymbol{x}^N, y^N), \tag{12}$$

with i_j such that $\theta_j = \tilde{\theta}_{i_j}$.

A proof of this lemma is given in the appendix.

2.3 Examples

Example 1: Steady State Model. The steady state model with the one-dimensional state variable x_t equal to the level of the process (e.g. Harvey, 1989) is characterized by:

$$p(x_t|x_{t-1}, \boldsymbol{\theta}) \sim N(x_{t-1}, \theta_1), \quad p(y_t|x_t, \boldsymbol{\theta}) \sim N(x_t, \theta_2), \qquad t = 1, 2, \ldots$$

This model is 2-inverse-gamma with $v_t(y^{t-1}) = 1$ and $A_t(y^{t-1}) = 1$ in (7), if $\boldsymbol{\theta} = (\theta_1, \theta_2)$ is unknown. Given an inverted gamma prior for $\theta_1|y^0$ and $\theta_2|y^0$ and given a normal prior for $x_0|y^0$ the conditional posterior $p(\boldsymbol{\theta}|x^N, y^N)$ is the product of two independent inverted gamma densities with the parameters in (9) and (10) given by:

$$\alpha_N^{(j)} = \alpha_0^{(j)} + \frac{N}{2}, \qquad j = 1, 2,$$

$$\beta_N^{(1)} = \beta_0^{(1)} + \frac{1}{2} \sum_{t=1}^{N} (x_t - x_{t-1})^2, \quad \beta_N^{(2)} = \beta_0^{(2)} + \frac{1}{2} \sum_{t=1}^{N} (y_t - x_t)^2.$$

Example 2: Dynamic Trend Model. The dynamic trend model introduced by Harrison and Stevens (1976) has a two dimensional state vector $\boldsymbol{x}_t = (\mu_t, a_t)$. μ_t is the mean of the observation density, a_t is the trend component. The model is characterized by (1) with

$$\boldsymbol{F}_t = \boldsymbol{F} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \ \boldsymbol{Q}_t(\boldsymbol{\theta}) = \boldsymbol{F}\text{Diag }(\theta_1, \theta_2)\boldsymbol{F}^{\mathrm{T}}, \ \boldsymbol{H}_t = \begin{pmatrix} 1 & 0 \end{pmatrix}, \ R_t(\boldsymbol{\theta}) = \theta_3.$$

This model is 3-inverse-gamma with $v_t(y^{t-1}) = 1$ and $\boldsymbol{B}_t(y^{t-1}) = \boldsymbol{F}$ in (8), if $\boldsymbol{\theta} = (\theta_1, \theta_2, \theta_3)$ is unknown. Given an inverted gamma prior for each component of $\boldsymbol{\theta}$ and given a normal prior for $x_0|y^0$ the conditional posterior $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ is the product of three inverted gamma densities with the parameters in (9) and (10) given by:

$$\alpha_N^{(j)} = \alpha_0^{(j)} + \frac{N}{2}, \quad j = 1,2,3, \qquad \beta_N^{(1)} = \beta_0^{(1)} + \frac{1}{2}\sum_{t=1}^{N}(\mu_t - (\mu_{t-1} + a_t))^2,$$

$$\beta_N^{(2)} = \beta_0^{(2)} + \frac{1}{2}\sum_{t=1}^{N}(a_t - a_{t-1})^2, \qquad \beta_N^{(3)} = \beta_0^{(3)} + \frac{1}{2}\sum_{t=1}^{N}(y_t - \mu_t)^2.$$

From this model we obtain a partly dynamic model, if the dynamic trend component is restricted to a static one: $\tilde{\theta}_2 = 0$ – note the change of notation from $\boldsymbol{\theta}$ to $\tilde{\boldsymbol{\theta}}$. The new model is 2-inverse-gamma with $\theta_1 = \tilde{\theta}_1$ and $\theta_2 = \tilde{\theta}_3$. The second component of $\boldsymbol{w}_t$ is 0 iff $a_t \equiv a \forall t = 0, \ldots, N$. Thus a path $\boldsymbol{x}^N$ is realizable iff the second component of $\boldsymbol{x}_t$ is the same $\forall t = 0, \ldots, N$. Therefore $\beta_N^{(1)}$ is given by:

$$\beta_N^{(1)} = \beta_0^{(1)} + \frac{1}{2}\sum_{t=1}^{N}(\mu_t - (\mu_{t-1} + a))^2,$$

and $\beta_N^{(2)} = \tilde{\beta}_N^{(3)}$.

3 Approximate Posterior Densities

We now turn to the problem of approximating the marginal density $p(\boldsymbol{\theta}|y^N)$ of the unknown hyperparameter $\boldsymbol{\theta}$ and the smoothing densities $p(\boldsymbol{x}_t|y^N)$ which is a marginal density, too. It should be noted that for $d = 1$ the model belongs to the well-known normal-inverse-gamma-class. This is typically the case when the dynamic model degenerates to a static one with $\boldsymbol{x}_t \equiv \boldsymbol{x}$. Then both the posterior $p(\theta|y^N)$ of the unknown hyperparameter θ and the posterior density $p(\boldsymbol{x}|y^N)$ are known analytically (see e.g. Tanner, 1991):

$$p(\theta|y^N) = IG(\alpha_0^{(1)} + \frac{N-r}{2}, \beta_N^{(1)}(\hat{\boldsymbol{x}}^N))$$

with $\hat{\boldsymbol{x}}^N$ being the posterior mean of $p(\boldsymbol{x}|y^N, \theta = 1)$ with uninformative prior $\hat{\boldsymbol{x}}_{0|0} = 0$ and $\boldsymbol{P}_{0|0}^{-1} = 0$.

For $d > 1$ all these densities have no tractable analytical form and need to be approximated. Dynamic linear models with unknown hyperparameters are discussed e.g. in Harrison and Stevens (1976) and Harvey (1989). A number of authors have approached the problem of unknown hyperparameters from a frequentistic point of view (e.g. Harvey, 1989) by maximizing the likelihood of the hyperparameter either numerically or by the E-M-Algorithm. The Bayesian approach dates back to the early paper of Magill (1965) and was reinvented by Harrison and Stevens (1976). In both papers the posterior of the hyperparameter is approximated by a discretization method called multi-process-filtering. This method has been refined by Pole (1988). One of the main problems with this approach "is the reconstruction of the continous posterior density from the discrete approximation" (see also West and Harrison, 1989, p.533).

In this paper we treat dynamic linear models with unknown hyperparameters as a special case of what has been called an incomplete data problem (Dempster et al., 1985). We then apply the data augmentation algorithm proposed by Tanner and Wong (1987) for approximating marginal densities in incomplete data problems to our problem.

3.1 The Data Augmentation Algorithm

Tanner and Wong (1987) observed that in incomplete data problems the marginal distribution is the fixed point solution of an integral equation. If this result is applied to our problem we obtain that the posterior $p(\boldsymbol{\theta}|y^N)$ of the unknown hyperparameter $\boldsymbol{\theta}$ is a fixed point solution of the following integral equation:

$$p(\boldsymbol{\theta}|y^N) = \int\int p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) p(\boldsymbol{x}^N|\boldsymbol{\theta}', y^N) p(\boldsymbol{\theta}'|y^N) d\boldsymbol{x}^N d\boldsymbol{\theta}'. \tag{13}$$

Tanner and Wong (1987) used the substitution algorithm for finding the solution of this integral equation. Given some approximation $g_{n-1}(\boldsymbol{\theta})$ of the posterior $p(\boldsymbol{\theta}|y^N)$ one may use Eq. (13) to improve it:

$$g_n(\boldsymbol{\theta}) = \int K(\boldsymbol{\theta}, \boldsymbol{\theta}') g_{n-1}(\boldsymbol{\theta}') d\boldsymbol{\theta}', \; K(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) p(\boldsymbol{x}^N|\boldsymbol{\theta}', y^N) d\boldsymbol{x}^N. \tag{14}$$

Iteration (14) converges in L_1-norm to a unique solution which is equal to the marginal posterior $p(\boldsymbol{\theta}|y^N)$ under the following regularity conditions on $K(\boldsymbol{\theta}, \boldsymbol{\theta}')$ (Tanner and Wong, 1987, p.538): $K(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is uniformly bounded (C1); $K(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is equicontinous in $\boldsymbol{\theta}$ (C2); $\forall \boldsymbol{\theta}_0 \in \Theta : \exists$ open neighbourhood $U_\epsilon(\boldsymbol{\theta}_0) : K(\boldsymbol{\theta}, \boldsymbol{\theta}') > 0, \forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in U_\epsilon(\boldsymbol{\theta}_0)$ (C3).

To prove convergence of the substitution algorithm to the posterior of the hyperparameter for the d-inverse-gamma class we only need the following condition on the prior parameters of $\boldsymbol{\theta}$: $\beta_0^{(j)} > 0, j = 1, \ldots, d$ (C4).

Lemma 2. Under Condition (C4) L_1-convergence of $g_n(\boldsymbol{\theta})$ to the posterior $p(\boldsymbol{\theta}|y^N)$ of the hyperparameter holds for d-inverse-gamma dynamic linear models.

A proof of this lemma is given in the appendix.

3.2 Practical Implementation

Integration in (14) cannot be performed analytically for dynamic linear models. For such cases Tanner and Wong (1987) suggested to use Monte-Carlo methods. In this subsection we discuss how such a sampling based scheme may be implemented for dynamic linear models.

First we have to discuss how to sample a path $(\boldsymbol{x}^N)^{(m)}$ from the smoothing density $p(\boldsymbol{x}^N|y^N,\boldsymbol{\theta})$ which is a normal density. We use the following conditional representation of a multivariate normal density:

$$p(\boldsymbol{x}^N|y^N,\boldsymbol{\theta}) = \prod_{t=0}^{N-1} p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1},\ldots,\boldsymbol{x}_N,y^N,\boldsymbol{\theta})p(\boldsymbol{x}_N|y^N,\boldsymbol{\theta}) \tag{15}$$

We then start sampling with the present value $\boldsymbol{x}_N^{(m)}$ of the state vector from the marginal $p(\boldsymbol{x}_N|y^N,\boldsymbol{\theta})$. We then sample back in time by sampling from the conditional densities $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}^{(m)},\ldots,\boldsymbol{x}_N^{(m)},y^N,\boldsymbol{\theta})$, $t = N-1,\ldots,0$. The moments of these conditional densities are given in the following proposition. A proof of this proposition is given in the appendix.

Proposition 1. The conditional densities $p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1},\ldots,\boldsymbol{x}_N,y^N,\boldsymbol{\theta})$, $t = 0, \ldots, N-1$, in (15) are normal densities with moments $\hat{\boldsymbol{x}}_{t|N}(\boldsymbol{x}_{t+1},\boldsymbol{\theta})$ and $\boldsymbol{P}_{t|N}(\boldsymbol{\theta})$ given by:

$$\begin{aligned}
\hat{\boldsymbol{x}}_{t|N}(\boldsymbol{x}_{t+1},\boldsymbol{\theta}) &= (\boldsymbol{I}-\boldsymbol{A}_{t+1}(\boldsymbol{\theta})\boldsymbol{F}_{t+1})\hat{\boldsymbol{x}}_{t|t}(\boldsymbol{\theta}) + \boldsymbol{A}_{t+1}(\boldsymbol{\theta})\boldsymbol{x}_{t+1}\\
\boldsymbol{P}_{t|N}(\boldsymbol{\theta}) &= (\boldsymbol{I}-\boldsymbol{A}_{t+1}(\boldsymbol{\theta})\boldsymbol{F}_{t+1})\boldsymbol{P}_{t|t}(\boldsymbol{\theta})\\
\boldsymbol{A}_{t+1}(\boldsymbol{\theta}) &= \boldsymbol{P}_{t|t}(\boldsymbol{\theta})\boldsymbol{F}_{t+1}^{\mathrm{T}}(\boldsymbol{F}_{t+1}\boldsymbol{P}_{t|t}(\boldsymbol{\theta})\boldsymbol{F}_{t+1}^{\mathrm{T}} + \boldsymbol{Q}_{t+1}(\boldsymbol{\theta}))^{-1}
\end{aligned}$$

$\hat{\boldsymbol{x}}_{t|t}(\boldsymbol{\theta})$ and $\boldsymbol{P}_{t|t}(\boldsymbol{\theta})$ are the moments of the filtering densities $p(\boldsymbol{x}_t|\boldsymbol{\theta},y^t)$ obtained by Kalman-filtering.

Now assume that an approximation $g_{n-1}(\boldsymbol{\theta})$ of the posterior $p(\boldsymbol{\theta}|y^N)$ of $\boldsymbol{\theta}$ is given. We then repeat the following steps for $m = 1,\ldots,M_n$:

1. We sample a hyperparameter $\boldsymbol{\theta}^{(m)}$ from $g_{n-1}(\boldsymbol{\theta})$. We will comment on this step after having explained the approximation step.

2. We sample a path $(\boldsymbol{x}^N)^{(m)}$ of the state process from the conditional smoothing density $p(\boldsymbol{x}^N|\boldsymbol{\theta}^{(m)},y^N)$ by the method suggested above.

3. We determine the moments of the conditional posterior p.d.f. $p(\boldsymbol{\theta}|(\boldsymbol{x}^N)^{(m)},y^N)$ from (5). The simulated path of the state process need not to be stored once these moments are calculated.

From the final Step 3 we obtain the following improved density estimate $g_n(\boldsymbol{\theta})$ of the posterior $p(\boldsymbol{\theta}|y^N)$:

$$p(\boldsymbol{\theta}|y^N) \approx g_n(\boldsymbol{\theta}) = \frac{1}{M_n}\sum_{m=1}^{M_n} p(\boldsymbol{\theta}|(\boldsymbol{x}^N)^{(m)}, y^N). \tag{16}$$

Note that the unknown density is approximated analytically by a density from a finite mixture distribution. The type of the mixture depends on the form of the conditional posterior $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$. For d-inverse-gamma dynamic linear models this density is a mixture of a product of d inverted gamma densities.

Sampling of $\boldsymbol{\theta}^{(m)}$ from $g_{n-1}(\boldsymbol{\theta})$ is clear now. We just sample the number I_m of the member in the mixture which is equally distributed between 1 and M_{n-1} and then sample $\boldsymbol{\theta}^{(m)}$ from the corresponding conditional density $p(\boldsymbol{\theta}|(\boldsymbol{x}^N)^{(I_m)}, y^N)$. For d-inverse-gamma dynamic linear models this density splits into d inverted gamma densities. Thus the components $\theta_i^{(m)}$, $i = 1, \ldots, d$, of $\boldsymbol{\theta}^{(m)}$ are sampled independently.

Finally, let us discuss approximation of the marginal smoothing densities $p(\boldsymbol{x}_t|y^N)$ of state vector which are given by:

$$p(\boldsymbol{x}_t|y^N) = \int p(\boldsymbol{x}_t|\boldsymbol{\theta}, y^N)p(\boldsymbol{\theta}|y^N)d\boldsymbol{\theta}. \tag{17}$$

It is not necessary to compute these densities during iteration for the posterior $p(\boldsymbol{\theta}|y^N)$ of the hyperparameter. Having found an appropriate approximation of $p(\boldsymbol{\theta}|y^N)$ the smoothing density of the state vector $\boldsymbol{x}_t|y^N$ is approximated by a sum of Gaussian densities:

$$p(\boldsymbol{x}_t|y^N) \approx \frac{1}{M_X}\sum_{m=1}^{M_X} p(\boldsymbol{x}_t|\boldsymbol{\theta}^{(m)}, y^N) \tag{18}$$

where $\boldsymbol{\theta}^{(m)}$, $m = 1, \ldots, M_X$, is sampled from the approximation of $p(\boldsymbol{\theta}|y^N)$ as described above.

4 Approximate Model Likelihoods

In this section we will demonstrate how an iterative importance sampling estimate ("IIS-estimate") of the model likelihood $L(y^N|y^0)$,

$$L(y^N|y^0) = \int p(y_1, \ldots, y_N|\boldsymbol{\theta}, y^0)p(\boldsymbol{\theta}|y^0)d\boldsymbol{\theta} \tag{19}$$

results as a by-product of the data augmentation method discussed in section 3.

4.1 An IIS-estimate of the Model Likelihood

A Monte-Carlo estimate of $L(y^N|y^0)$ is given by:

$$\hat{L}(y^N|y^0) = \frac{1}{M}\sum_{m=1}^{M} \frac{p(y_1,\ldots,y_N|\boldsymbol{\theta}^{(m)},y^0)p(\boldsymbol{\theta}^{(m)}|y^0)}{w(\boldsymbol{\theta}^{(m)})} \tag{20}$$

with $\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(M)}$ drawn independently from an importance sampling distribution with density $w(\boldsymbol{\theta})$.

From Bayes'theorem

$$p(\boldsymbol{\theta}|y^N) \propto p(y_1,\ldots,y_N|\boldsymbol{\theta},y^0)p(\boldsymbol{\theta}|y^0) =: \tilde{p}(\boldsymbol{\theta}|y^N)$$

it is obvious that the model likelihood is identical with the normalizing constant of the non-normalized posterior $\tilde{p}(\boldsymbol{\theta}|y^N)$. Thus the best choice for $w(\boldsymbol{\theta})$ is the posterior $p(\boldsymbol{\theta}|y^N)$ of the hyperparameter: we could obtain the exact result with (20), if we drew a single $\boldsymbol{\theta}^{(1)}$. However, the exact posterior is untractable in general.

In Section 3 we discussed in great details how to approximate $p(\boldsymbol{\theta}|y^N)$ iteratively by a density $g_n(\boldsymbol{\theta})$ based on the data augmentation method. The IIS-estimate is therefore defined by sampling iteratively from $g_n(\boldsymbol{\theta})$:

$$\hat{L}_n(y^N|y^0) = \frac{1}{M_n}\sum_{m=1}^{M_n} \frac{p(y_1,\ldots,y_N|\boldsymbol{\theta}^{(m)},y^0)p(\boldsymbol{\theta}^{(m)}|y^0)}{g_{n-1}(\boldsymbol{\theta}^{(m)})}, \quad n = 1,2,\ldots \tag{21}$$

Lemma 4. The IIS-estimate defined in (21) has the following properties:

$$E_{g_{n-1}}(\hat{L}_n) = L, \quad V_{g_{n-1}}(\hat{L}_n) = \frac{L^2}{M_n} E_{p(\boldsymbol{\theta}|y^N)}\left(\frac{p(\boldsymbol{\theta}|y^N)}{g_{n-1}(\boldsymbol{\theta})} - 1\right). \tag{22}$$

The proof is given in the appendix. An application of this IIS-estimate to Bayes tests on variances in the structural time series models defined in Harvey (1989) is discussed in details in Frühwirth-Schnatter (1992).

4.2 Practical Implementation

The implementation of the IIS-estimate causes hardly any additional computational burden when iterating for $p(\boldsymbol{\theta}|y^N)$. The improved estimate $g_n(\boldsymbol{\theta})$ of $p(\boldsymbol{\theta}|y^N)$ is based on a sample $\boldsymbol{\theta}^{(1)},\ldots,\boldsymbol{\theta}^{(M_n)}$ drawn from $g_{n-1}(\boldsymbol{\theta})$ and M_n Kalman-filter run conditional on this sample. The values $p(y_1,\ldots,y_N|\boldsymbol{\theta}^{(m)},y^0)$ are available from the Kalman-filter (see e.g. Harrison and Stevens, 1976). Thus the additional work to obtain $\hat{L}_n(y^N|y^0)$ consists only of evaluating the prior $p(\boldsymbol{\theta}|y^0)$ and the approximation $g_{n-1}(\boldsymbol{\theta})$ at the sample.

5 An Illustrative Case Study

Table 1 contains a time series of the mean yearly ground water level from 1967 - 1988 at a place called "Seewinkel" in Austria.

Table 1: Yearly mean ground water Data from Seewinkel (Austria) [From: Frühwirth-Schnatter, 1991]

1967	1968	1969	1970	1971	1972	1973	1974	1975	1976
124.640	125.748	125.666	125.620	125.676	125.701	125.462	125.601	125.405	124.896
1977	**1978**	**1979**	**1980**	**1981**	**1982**	**1983**	**1984**	**1985**	**1986**
124.822	124.568	124.203	124.541	124.399	124.199	124.270	124.074	123.796	124.019
1987	**1988**								
124.028	124.070								

We will analyse this time series by the dynamic linear trend model discussed in Example 2 of Section 2.3 to estimate the trend component $a_t|y^{1988}$ for this time series. As the variances θ_1, θ_2 and θ_3 are unknown, we approximate the posterior $p(\boldsymbol{\theta}|y^{1988})$ by the algorithm described in Section 3.

The following priors on the state vector and on the variances are assumed:

$$\boldsymbol{x}_{1966|1966}|y^{1966} \sim N(\hat{\boldsymbol{x}}_{1966|1966}, \boldsymbol{P}_{1966|1966}),$$
$$\hat{\boldsymbol{x}}_{1966|1966} = \begin{pmatrix} 125 \\ 0 \end{pmatrix}, \quad \boldsymbol{P}_{1966} = \begin{pmatrix} 10 & 0 \\ 0 & 1 \end{pmatrix},$$
$$\theta_j|y^{1966} \sim IG(0.01, 10^{-5}), \quad j = 1,2,3.$$

The iteration is started with $g_0(\boldsymbol{\theta}) = \prod_{j=1}^{3} p_{IG}(\theta_j; 3, 0.02)$. During the first forty iterations the number M_n of sampled paths and hyperparameters is equal to 100. For the subsequent 20 iterations M_n is increased to 500. The posterior $p(\boldsymbol{\theta}|y^{1988})$ is then approximated by a last iteration step with $M_{61} = 2000$. The bivariate marginals of this approximation together with the contours are shown in Figure 1.

Figure 1 about here

The posterior of the trend component $a_t|y^{1988}$ is then approximated by a mixture of conditionally Gaussian distributions, where the conditioning hyperparameters are sampled from the last posterior $g_{61}(\boldsymbol{\theta})$. 95%-H.P.D.-regions of $a_t|y^{1988}$ are shown in Figure 2.

Figure 2 about here

For further technical details on the choice of M_n and on pooling among the iterations we refer to Tanner and Wong (1987).

6 Concluding Remarks

In this paper a data augmentation algorithm has been suggested for approximating posterior p.d.f. and model likelihoods for dynamic linear models with unknown hyperparameters. We confined ourselves to a particular subclass of such models called d-inverse-gamma class. The relation between data augmentation methods and dynamic models surely holds on a more general class. The algorithm suggested for practical implementation may be applied to *any* dynamic linear model with unknown hyperparameters provided the conditional density $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ is of such a form that we are able to sample from it. The restriction to the d-inverse-gamma class is mainly due to the fact that so far we proved the regularity conditions sufficient for convergence of the algorithm for this class only. However, we are quite confident that the algorithm converges for a wider class of dynamic linear models.

Acknowledgement

The idea of developing the algorithm described in section 3 arose during discussions I had with M. Reidinger who by that time was Ph.D. student at the Department of Statistics and Probability Theory at the Vienna University of Technology and who meanwhile sorrowfully went on a journey with no return.

Part of the work of section 3 was carried out at the Department of Statistics and Probability Theory at the Vienna University of Technology under the support of the Austrian Science Foundation, Project Nr. 7079.

Appendix: Proofs

Proof of Lemma 1. Let us proof the lemma for the case (8). First we assume that the last m components of $\tilde{\boldsymbol{\theta}}$ are reduced to 0: $\theta_j = \tilde{\theta}_j$, $1 \leq j \leq r-m$, $\theta_{r-m+1} = \theta_{d-m} = \tilde{\theta}_d$. $\boldsymbol{x}^N$ is realizable iff the last m components of $\boldsymbol{w}_t = \boldsymbol{B}_t^{-1}(y^{t-1})(\boldsymbol{x}_t - \boldsymbol{F}_t\boldsymbol{x}_{t-1})$ are equal to 0. For such a realizable path the following holds:

$$\prod_{t=1}^{N} p(\boldsymbol{x}_t|\boldsymbol{x}_{t-1}, \boldsymbol{\theta}) \propto \prod_{j=1}^{r-m} \left(\frac{1}{\theta_j}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2}\sum_{t=1}^{N}[\boldsymbol{B}_t^{-1}(y^{t-1})(\boldsymbol{x}_t - \boldsymbol{F}_t\boldsymbol{x}_{t-1})]_j\frac{1}{\theta_j}\right\}.$$

From this result and (2) we see that $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ is the product of $(d-m)$-inverse-gamma densities. Thus (12) is obvious now from (11).

If $\tilde{\boldsymbol{\theta}}$ does not split directly into $(\boldsymbol{\theta}, \mathbf{0})$ we choose a suitable permutation matrix $\boldsymbol{P}$ to obtain a new parametrization $\tilde{\boldsymbol{\theta}}' = \boldsymbol{P}\tilde{\boldsymbol{\theta}} = (\boldsymbol{\theta}, \mathbf{0})$ of the hyperparameter. Then the same proof as above is applied to the new parametrization. The case (7) is proved in a similar manner. □

Proof of Lemma 2. For each $j = 1, \ldots, d$ the p.d.f. $p(\theta_j|\boldsymbol{x}^N, y^N)$ is bounded by the functional value of this density at the mode. This functional value has un upper bound which is independent of $\boldsymbol{x}^N$:

$$p(\theta_j|\boldsymbol{x}^N, y^N) \leq \frac{c(\alpha_N^{(j)})}{\beta_0^{(j)} + b_N^{(j)}(\boldsymbol{x}^N, y^N)} \leq \frac{c(\alpha_N^{(j)})}{\beta_0^{(j)}}, \quad c(\alpha) = \frac{(\alpha+1)^{\alpha+1}}{\Gamma(\alpha)} e^{-(\alpha+1)}. \tag{23}$$

This bound is finite because of Condition (C4) and $c(\alpha) < \infty$ for $\alpha < \infty$. Thus Condition (C1) follows from:

$$K(\boldsymbol{\theta}, \boldsymbol{\theta}') = \int p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) p(\boldsymbol{x}^N|y^N, \boldsymbol{\theta}') d\boldsymbol{x}^N \leq \prod_{j=1}^{d} \frac{c(\alpha_N^{(j)})}{\beta_0^{(j)}} < \infty, \quad \forall \boldsymbol{\theta}, \boldsymbol{\theta}' \in \boldsymbol{\Theta}.$$

As each partial derivative of the conditional posterior $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ of the hyperparameter is continous in $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ the conditional posterior $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ of the hyperparameter is differentiable $\forall \boldsymbol{\theta} \in \boldsymbol{\Theta}$. Now take $\boldsymbol{\theta} \in \boldsymbol{\Theta}$ arbitrarily. Let $M > 0$ be some constant. Then for any $\tilde{\boldsymbol{\theta}}$ with $||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}|| \leq M$ the following representation holds with $\boldsymbol{\xi} \in \overline{\tilde{\boldsymbol{\theta}}\boldsymbol{\theta}}$:

$$p(\tilde{\boldsymbol{\theta}}|\boldsymbol{x}^N, y^N) - p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) = \sum_{j=1}^{d} \frac{\partial p(\boldsymbol{\xi}|\boldsymbol{x}^N, y^N)}{\partial \theta_j} (\tilde{\theta}_j - \theta_j). \tag{24}$$

The first derivative of an inverted gamma density is bounded:

$$\frac{\partial p(\theta_j|\boldsymbol{x}^N, y^N)}{\partial \theta_j} \leq \frac{c'(\alpha_N^{(j)})}{(\beta_0^{(j)} + b_N^{(j)}(\boldsymbol{x}^N, y^N))^2} \leq \frac{c'(\alpha_N^{(j)})}{(\beta_0^{(j)})^2} < \infty,$$
$$c'(\alpha) = \frac{(\alpha+2)^{\frac{\alpha+2}{2}}}{\Gamma(\alpha)} \left(\frac{\alpha+1}{\sqrt{\alpha+2}-1} \right)^{\alpha+3} \exp\left(-\frac{(\alpha+1)\sqrt{\alpha+2}}{\sqrt{\alpha+2}+1} \right). \tag{25}$$

This bound is finite because of Condition (C4) and $c'(\alpha) < \infty$ for $\alpha < \infty$. Combining (23) and (25) we obtain that the first derivatives in formula (24) are bounded:

$$\frac{\partial p(\boldsymbol{\xi}|\boldsymbol{x}^N, y^N)}{\partial \theta_j} \leq \left(\prod_{i=1, i\neq j}^{d} \frac{c(\alpha_N^{(i)})}{\beta_0^{(i)}} \right) \cdot \frac{c'(\alpha_N^{(j)})}{(\beta_0^{(j)})^2} =: c_j''(N, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0).$$

The bound $c_j''(N, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$ depends only on the number N of observations and on the parameters $\boldsymbol{\alpha}_0 = (\alpha_0^{(1)}, \ldots, \alpha_0^{(d)})$ and $\boldsymbol{\beta}_0 = (\beta_0^{(1)}, \ldots, \beta_0^{(d)})$ of the prior of the hyperparameter. Therefore:

$$|K(\boldsymbol{\theta}, \boldsymbol{\theta}') - K(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}')| \leq$$
$$\int |p(\tilde{\boldsymbol{\theta}}|\boldsymbol{x}^N, y^N) - p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)| p(\boldsymbol{x}^N|\boldsymbol{\theta}', y^N)\, d\boldsymbol{x}^N \leq M \sum_{j=1}^{d} c_j''(N, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)$$

Thus $\forall\boldsymbol{\theta} \in \Theta$ and $\forall\epsilon > 0$ the following choice of δ_ϵ :

$$\delta_\epsilon = \frac{\epsilon}{\sum_{j=1}^{d} c_j''(N, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)}$$

leads to: $\forall\tilde{\boldsymbol{\theta}} : ||\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}|| \leq \delta_\epsilon \Rightarrow \quad |K(\boldsymbol{\theta}, \boldsymbol{\theta}') - K(\tilde{\boldsymbol{\theta}}, \boldsymbol{\theta}')| \leq \epsilon \quad \forall\boldsymbol{\theta}' \in \Theta$. This proves Condition (C2).

The p.d.f.s $p(\theta_j|\boldsymbol{x}^N, y^N)$ are strictly positive on $I\!R_+$. Thus $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ is strictly positive on Θ. Now take $\boldsymbol{\theta}_0 \in \Theta$. As Θ is open $\forall\ \epsilon > 0\ \exists$ open neighbourhood $U_\epsilon(\boldsymbol{\theta}_0) \subset \Theta$. As $p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N)$ is strictly positive $\forall\boldsymbol{\theta} \in U_\epsilon(\boldsymbol{\theta}_0)$, we define: $c := \inf_{\boldsymbol{\theta}\in U_\epsilon(\boldsymbol{\theta}_0)} p(\boldsymbol{\theta}|\boldsymbol{x}^N, y^N) > 0$. Then $K(\boldsymbol{\theta}, \boldsymbol{\theta}')$ is positive on $U_\epsilon(\boldsymbol{\theta}_0)$. Thus condition (C3) holds. □

Proof of Proposition 1. From Bayes' theorem we obtain that each of the conditional densities in (15) is proportional to the product of the transition density $p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}, \boldsymbol{x}_t)$ and the filtering density $p(\boldsymbol{x}_t|\boldsymbol{\theta}, y^t)$:

$$\begin{aligned}
&p(\boldsymbol{x}_t|\boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_N, y^N, \boldsymbol{\theta}) \propto p(y_{t+1}, \ldots, y_N, \boldsymbol{x}_t, \boldsymbol{x}_{t+1}, \ldots, \boldsymbol{x}_N|y^t, \boldsymbol{\theta}) = \\
&\quad = \prod_{i=t+1}^{N} p(y_i|\boldsymbol{x}_i, \boldsymbol{\theta}) \prod_{i=t+1}^{N} p(\boldsymbol{x}_i|\boldsymbol{x}_{i-1}, \boldsymbol{\theta}) p(\boldsymbol{x}_t|y^t, \boldsymbol{\theta}) \propto p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}, \boldsymbol{x}_t) p(\boldsymbol{x}_t|y^t, \boldsymbol{\theta}), \\
&p(\boldsymbol{x}_{t+1}|\boldsymbol{\theta}, \boldsymbol{x}_t) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}_{t+1} - \boldsymbol{F}_{t+1}\boldsymbol{x}_t)^{\mathrm{T}} \boldsymbol{Q}_{t+1}(\boldsymbol{\theta})^{-1} (\boldsymbol{x}_{t+1} - \boldsymbol{F}_{t+1}\boldsymbol{x}_t)\right), \\
&p(\boldsymbol{x}_t|\boldsymbol{\theta}, y^t) \propto \exp\left(-\frac{1}{2}(\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t}(\boldsymbol{\theta}))^{\mathrm{T}} \boldsymbol{P}_{t|t}(\boldsymbol{\theta})^{-1} (\boldsymbol{x}_t - \hat{\boldsymbol{x}}_{t|t}(\boldsymbol{\theta}))\right).
\end{aligned}$$

Completing the squares gives the desired result. □

Proof of Lemma 3.

1. $E_{g_{n-1}}(\hat{L}_n) = \dfrac{1}{M_n} \sum_{m=1}^{M_n} \displaystyle\int \tilde{p}(\boldsymbol{\theta}|y^N) d\boldsymbol{\theta} = L.$

2. As $\boldsymbol{\theta}^{(1)}, \ldots, \boldsymbol{\theta}^{(M_n)}$ are independent, the following holds:

$$\begin{aligned}
V_{g_{n-1}}(\hat{L}_n) &= \frac{1}{M_n^2} \sum_{m=1}^{M_n} \int \frac{\tilde{p}(\boldsymbol{\theta}|y^N)^2}{g_{n-1}(\boldsymbol{\theta})} d\boldsymbol{\theta} - \frac{1}{M_n} L^2 = \\
&= \frac{1}{M_n}\left[\int \frac{p(\boldsymbol{\theta}|y^N)}{g_{n-1}(\boldsymbol{\theta})} p(\boldsymbol{\theta}|y^N) d\boldsymbol{\theta} - L^2\right] = \frac{L^2}{M_n} E_{p(\boldsymbol{\theta}|y^N)}\left(\frac{p(\boldsymbol{\theta}|y^N)}{g_{n-1}(\boldsymbol{\theta})} - 1\right).
\end{aligned}$$

References

Broemeling, L. (1985) *Bayesian Analysis of Linear Models.* New York: Marcel Dekker.

Dempster,A.P., Laird,N.M. and Rubin,D.B. (1976) Maximum Likelihood from Incomplete Data Via the EM Algorithm. *J.R.Stat.Soc. B 39*, 1-38.

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., Minder,Ch. (Eds.): *Multivariate Modelle*, 89-122. Springer.

Frühwirth-Schnatter, S. (1992) *Bayes Tests for Dynamic Linear Models.* To be presented at the Workshop "New Direction in Time Series Analysis", Heidelberg, Germany, June 1992.

Harrison,P.J. and Stevens,C.F. (1976): Bayesian Forecasting (with discussion). *J.R.Stat.Soc. B 38*, 205-247.

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

Liptser,R.S. and Shiryayev,A.N. (1977) *Statistics of Random Processes.* New York/Heidelberg/Berlin: Springer.

Magill,D.T. (1965) Optimal Adaptive Estimation of Sampled Stochastic Processes. *IEEE-TAC 10*, 434-439.

Pole,A. (1988) Transfer Response Models: a Numerical Approach. In Bernardo,J.M., DeGroot,M.H., Lindley,D.V. and Smith,A.F.M. (Eds.): *Bayesian Statistics 3*, 733-745. Oxford: Oxford University Press

Tanner, M. (1991) *Tools for Statistical Inference. Observed Data and Data Augmentation Methods.* Lecture Notes in Statistics, **67**. New York: Springer.

Tanner,M. and Wong,W.H. (1987) The Calculation of Posterior Distributions by Data Augmentation. *JASA 83*, 398, 528–550.

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

Figures

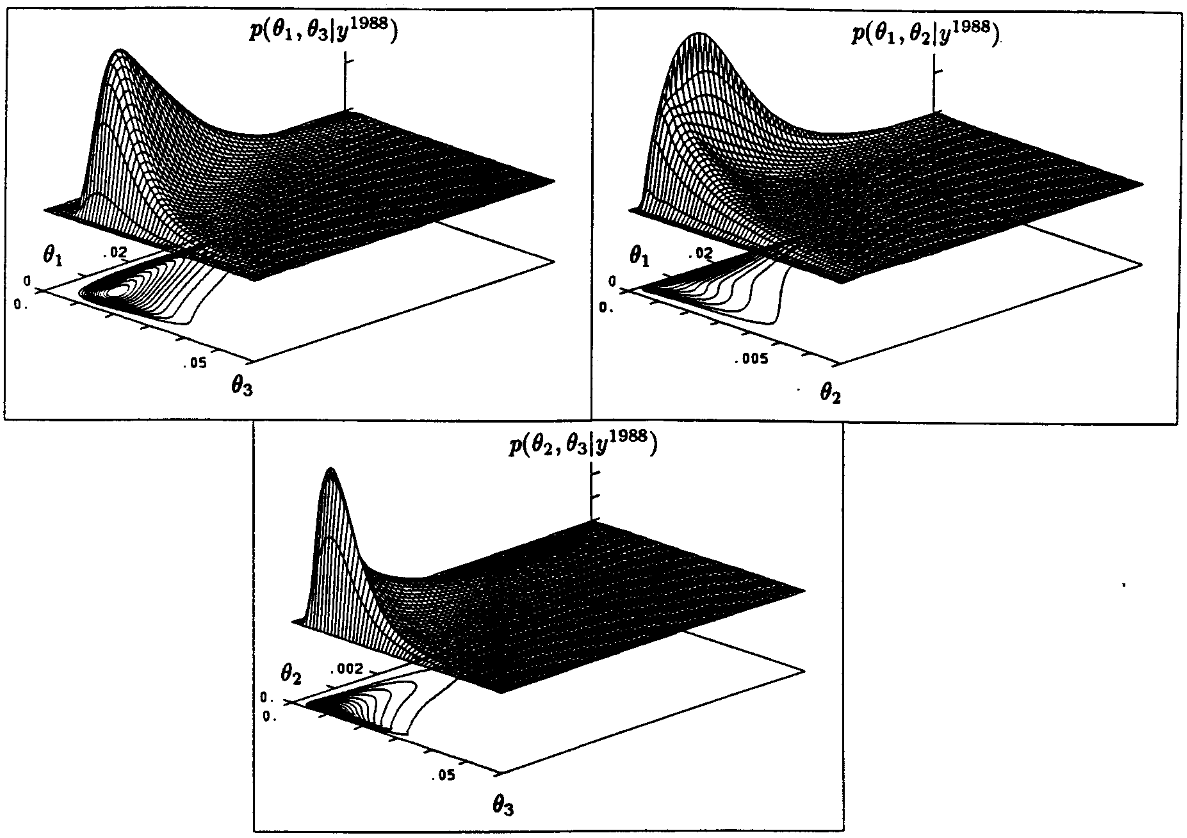

Figure 1: Bivariate marginals of the posterior p.d.f. $p(\boldsymbol{\theta}|y^{1988})$

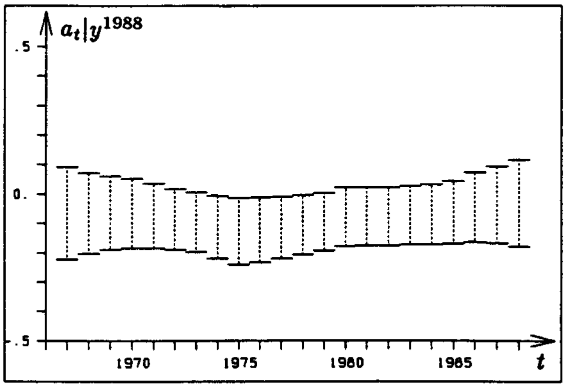

Figure 2: 95%-H.P.D.-Regions of the trend component $a_t|y^{1988}$