Ann. Inst. Statist. Math.
Vol. 53, No. 1, 31–49 (2001)

FULLY BAYESIAN ANALYSIS OF SWITCHING GAUSSIAN STATE SPACE MODELS

SYLVIA FRÜHWIRTH-SCHNATTER

Department of Statistics, University of Business Administration and Economics, Augasse 2-6, A-1090 Vienna, Austria, e-mail: sfruehwi@isis.wu-wien.ac.at

(Received April 12, 2000; revised July 27, 2000)

Abstract. In the present paper we study switching state space models from a Bayesian point of view. We discuss various MCMC methods for Bayesian estimation, among them unconstrained Gibbs sampling, constrained sampling and permutation sampling. We address in detail the problem of unidentifiability, and discuss potential information available from an unidentified model. Furthermore the paper discusses issues in model selection such as selecting the number of states or testing for the presence of Markov switching heterogeneity. The model likelihoods of all possible hypotheses are estimated by using the method of bridge sampling. We conclude the paper with applications to simulated data as well as to modelling the U.S./U.K. real exchange rate.

Key words and phrases: Bayesian analysis, bridge sampling, Markov switching models, MCMC methods, model selection, state space models.

1. Introduction

State space models are a well-studied tool to analyse time series $y^N = (y_1, \dots, y_N)$, where the distribution of y_t depends on a latent continuous state process $x^N = (x_0, \dots, x_N)$. A switching state space model is obtained if we assume that in addition to the latent, continuous process x^N a discrete latent switching variable I_t taking values in $\{1, \dots, K\}$ influences the distribution of y_t. Such models have been studied by various authors (e.g. Harrison and Stevens (1976); Shumway and Stoffer (1991); Kim (1993, 1994); Carter and Kohn (1994, 1996); Shephard (1994)). Kim and Nelson (1999) present an excellent review of the current state of art.

A typical way of including a switching mechanism into a Gaussian state space model is to assume that one of the variances, e.g. the variance appearing in the observation equation, is heteroskedastic and switches between various values $\theta_1^I, \dots, \theta_K^I$ depending on the state of the latent process I_t (Peña and Guttman (1988)):

$$x_t = F_t x_{t-1} + u_t + G_t w_t, \qquad w_t \sim \mathrm{N}(0, \mathrm{Diag}(\sigma_1^2, \dots, \sigma_d^2)),$$
$$y_t = H_t x_t + v_t, \qquad v_t \sim \mathrm{N}(0, \theta_{I_t}^I).$$

Alternatively, one or all of the variances $\sigma_1^2, \dots, \sigma_d^2$ of the transition equation may be Markov switching heteroskedastic (Kim (1993); Engle and Kim (1999)). Another way of including a switching mechanism is to assume that a drift term is present in the transition equation which switches between various values $\theta_1^I, \dots, \theta_K^I$ (Kim (1994)). For an application of the related switching dynamic factor model to modelling business cycles see Kim and Nelson (1998) and Kaufmann (2000); an interesting application of a state space model with two switching mechanisms to audio signal processing appears in Godsill (1997).

Estimation of a switching state space model is far from trivial. The classical maximum likelihood approach cannot be applied directly as the marginal likelihood, where both latent processes x^N and I^N are integrated out, is not available in a closed form. Approximate filters which lead to an unknown approximation error have been used e.g. in Kim (1993, 1994). Alternatively, a Bayesian approach can be applied by means of Markov chain Monte Carlo (MCMC) methods (see e.g. Smith and Roberts (1993) for a general introduction to MCMC methods). The design of suitable MCMC methods to estimate a Gaussian state space model with switching has been studied in various papers. A general discussion may be found in Carter and Kohn (1994, 1996), Shephard (1994), and Kim and Nelson (1999). MCMC sampling for specific Gaussian state space models with switching appear e.g. in Godsill (1997), Kim and Nelson (1998), Engle and Kim (1999), and Kaufmann (2000). In Section 2 of the present paper we address once more the issue of Bayesian estimation of switching Gaussian state space models via MCMC methods. Emphasis will lie on the unidentifiability of switching state space models and its implication for MCMC estimation. We will discuss the problem of label switching, the importance of finding a sensible identifiability constraint, and the fact that important information such as smoothed estimates of the latent state process x^N or estimates of time varying parameters will be available without the need to identify the model.

In Section 3 we head for a fully Bayesian analysis of switching Gaussian state space models discussing issues in model selection such as selecting the number of states and testing for the presence of a hidden Markov switching process. We will apply the method of bridge sampling to compute the model likelihood for a switching Gaussian state space model.

Applications to simulated data appear in Section 4. The paper is concluded with an application of a switching state space model to the analysis of the U.S./U.K. real exchange rate in Section 5.

2. Bayesian estimation of a Gaussian state space model with switching

Unknown parameters which have to be estimated from the data are the fixed parameters θ^C, which are common to all states, the state specific parameters $\theta_1^I, \ldots, \theta_K^I$, and the parameter η appearing in the definition of the distribution of $I^N = (I_0, \ldots, I_N)$. These parameters will be summarized by ϕ: $\phi = (\theta^C, \theta_1^I, \ldots, \theta_K^I, \eta)$. Within the Bayesian approach both the discrete, latent process I^N as well as the continuous, latent process x^N are viewed as missing data and estimated along with the model parameter ϕ. In what follows we are going to estimate the augmented parameter vector $\psi = (\phi, I^N, x^N)$ by sampling from the posterior density $\pi(\psi \mid y^N)$ by means of MCMC methods.

2.1 *Model structure and choice of the priors*

Bayesian estimation of the model is based on the hierarchical structure of the model:

1. Conditionally on known realisations x^N and I^N of the continuous state process and the switching process, respectively, and on a known model parameter ϕ the observations are independent with the distribution of y_t depending on x_t and I_t,

only. The "complete data likelihood" $f(y^N \mid x^N, I^N, \phi)$ given x^N, I^N and ϕ reads

$$f(y^N \mid x^N, I^N, \phi) = \prod_{t=1}^{N} f_N(y_t \mid x_t, \theta^I_{I_t}, \theta^C),$$

where $f_N(y_t \mid \cdot)$ is the density of a normal distribution.

2. Conditionally on a known realisation I^N of the switching process and on a known model parameter ϕ, the density of the (prior) distribution of the continuous latent process $x^N = (x_0, x_1, \ldots, x_N)$ is given by:

$$\pi(x^N \mid I^N, \phi) = \prod_{t=1}^{N} f_N(x_t \mid x_{t-1}, \theta^I_{I_t}, \theta^C)\pi(x_0 \mid \phi).$$

3. Conditionally on a known model parameter ϕ, the density of the (prior) distribution of the latent switching process I^N depends on η, only: $\pi(I^N \mid \phi) = \pi(I^N \mid \eta)$.
4. Finally, ϕ has a prior distribution $\pi(\phi)$.

Note that the "prior" on I^N and x^N appearing in the second and the third level are not subjective priors, but part of the model. One possible prior structure on I^N is the *exchangeable prior*, where I_t is assumed to be an iid process with $\Pr\{I_t = j\} = \eta_j$, $j = 1, \ldots, K$. This prior has been used in combination with Gaussian state space models e.g. in Shumway and Stoffer (1991). An alternative choice is the *Markovian switching prior* where I_t is assumed to be a stationary Markov process with discrete state space $\{1, \ldots, K\}$ and $\Pr\{I_t = j \mid I_{t-1} = i\} = \eta_{ij}$. This structure has been introduced by Hamilton (1989). It is combined with Gaussian state space models e.g. in Kim (1994), Engle and Kim (1999) and Kim and Nelson (1998, 1999).

Only the prior on ϕ appearing on the fourth level has a subjective flavour. We assume that $\eta = (\eta_{1\cdot}, \ldots, \eta_{K\cdot})$, where $\eta_{i\cdot} = (\eta_{i1}, \ldots, \eta_{iK})$, is independent from the remaining parameters of ϕ and that all conditional transition distributions $\eta_{i\cdot}$, $i = 1, \ldots, K$, are independent *a priori* from each other. A "natural" prior distribution $\pi(\eta_{i\cdot})$ for $\eta_{i\cdot}$ is the Dirichlet prior which is the conjugate prior in the complete data setting, where I^N is assumed to be known. Concerning the state specific parameters $\theta^I_1, \ldots, \theta^I_K$, we assume that they are independent *a priori* and that each θ^I_j has the same prior distribution depending on hyperparameters which are not state specific. This allows different parameters for the various states, however with a slight restriction expressed by the prior. Furthermore this prior is invariant to relabeling the states.

2.2 *Bayesian posterior analysis and the problem of label switching*

From the hierarchical structure of the model we obtain that the (unconstrained) posterior density $\pi(\psi \mid y^N)$ is given by:

$$\pi(\psi \mid y^N) \propto f(y^N \mid x^N, I^N, \phi)\pi(x^N \mid I^N, \phi)\pi(I^N \mid \phi)\pi(\phi).$$

For models including a latent, discrete structure such as I^N the unconstrained posterior has some characteristic properties (see Stephens (1997); Celeux (1998); Frühwirth-Schnatter (2001)). The unconstrained parameter space contains $K!$ subspaces, each one corresponding to a different way of labeling the states. The "complete data likelihood" $f(y^N \mid x^N, I^N, \phi)$, and the "priors" $\pi(x^N \mid I^N, \phi)$ and $\pi(I^N \mid \phi)$ are invariant to relabeling the states. Therefore, if the prior $\pi(\phi)$, is invariant, too, the unconstrained posterior typically is multimodal and invariant to relabeling the states. This special structure of the posterior has a lot of consequences for estimation. The unconstrained model is not identified in a strict sense. When sampling from the unconstrained posterior via MCMC methods, we do not know to which of the labeling subspaces the sampled value belongs. Therefore, we do not know to which state a sampled parameter belongs as label switching (jumping between the various labeling subspaces) might have occured. Thus we are not allowed to estimate functionals $f(\psi)$ of ψ which are not invariant to relabeling the states from MCMC simulations from the unconstrained posterior.

Note, that important information is available from MCMC simulations from the unconstrained posterior *without* introducing unique labeling, as we are allowed to estimate functionals $f(\psi)$ of ψ which are invariant to relabeling the states of I_t. This is important especially for applied time series modelling, as "smoothing" in terms of estimating the latent continuous process x^N from the posterior $\pi(x^N \mid y^N)$ is possible without caring about identifiability. Smoothed estimates $\hat{x}_{t|N}$ of x_t, for instance, are simply given by $\hat{x}_{t|N} = \frac{1}{M}\sum_{m=1}^{M}(x^N)^{(m)}$, where $(x^N)^{(1)}, \ldots, (x^N)^{(M)}$ are MCMC simulations from the unconstrained posterior. This estimate is unaffected by label switching, because the marginal posterior $\pi(x^N \mid y^N)$ is invariant to relabeling the states of I^N.

Another important application is estimation of time varying parameters. In practice, the discrete values $\theta_1^I, \ldots, \theta_K^I$ often will be just an approximation to modelling time varying model parameters ξ_t of the state space model (Harrison and Stevens (1976)) $\xi_t = \theta_j^I$, if $I_t = j$. From rewriting this functional in the form $\xi_t = \sum_{j=1}^{K}\theta_j^I S_t^{(j)}$, where for each j, $S_t^{(j)} \equiv 1$, iff $I_t = j$ and zero otherwise, invariance to relabeling the states of I_t is obvious. Therefore it is possible to obtain individual estimates of ξ_t for each t from the MCMC output of an unconstrained model by averaging $(\theta_s^I)^{(m)}$, where $s = I_t^{(m)}$, over all $m = 1, \ldots, M$. Finally, estimation of state independent parameters θ^C is possible without caring about identifiability.

In order to estimate functionals $f(\psi)$ of ψ which are not invariant to relabeling the states such as the state specific parameters $\theta_1^I, \ldots, \theta_K^I$, η or the probability $\Pr(I_t = j \mid y^N)$ of being in a certain state j at time t we have to identify the model in the sense that we allow for MCMC simulations from a unique labeling subspace, only. A common way to do this is to include an identifiability constraint. However, the problem with identifiability constraints is that they don't necessarily induce a unique labeling if the geometry of the unconstrained posterior density is ignored (see Frühwirth-Schnatter (2001)). Only a carefully selected constraint will separate the labeling subspaces and induce unique labeling. We will demonstrate in our case studies how suitable identifiability constraints may be found by exploring MCMC simulations from unconstrained posterior densities.

2.3 *MCMC Methods*

The design of suitable MCMC methods to generate a (dependent) sample $\psi^{(1)}$, $\psi^{(2)}, \ldots$ from the posterior $\pi(\psi \mid y^N)$ of a dynamic linear model with switching has been studied in various papers (Carter and Kohn (1994, 1996); Shephard (1994); Godsill (1997); Kim and Nelson (1998); Engle and Kim (1999)). MCMC techniques for sampling from a complicated posterior density split the joint unknown parameter into blocks and sample then from the conditional posterior densities of each block given the fixed values for the other blocks. Sampling ψ from the posterior of a switching Gaussian state space model is possible within the following four blocks:

(i) Sample I^N from $\pi(I^N \mid x^N, \eta, \theta^C, \theta^I, y^N)$
(ii) Sample η from $\pi(\eta \mid I^N)$
(iii) Sample $(\theta_1^I, \ldots, \theta_K^I, \theta^C)$ from $\pi(\theta_1^I, \ldots, \theta_K^I, \theta^C \mid x^N, I^N, y^N)$.
(iv) Sample x^N from $\pi(x^N \mid I^N, \theta_1^I, \ldots, \theta_K^I, y^N)$

Step (i) and (ii) are standard steps occuring in MCMC estimation of any model including a latent Markov switching variable. Step (i) may be carried out in a multimove manner as in Carter and Kohn (1994, 1996), Shephard (1994), or Chib (1996). Sampling η is completely standard, as the conditional posterior $\pi(\eta \mid \phi, I^N, x^N, y^N)$ depends only on I^N and each conditional distribution $\eta_{1\cdot}, \ldots, \eta_{K\cdot}$ follows a Dirichlet distribution. The precise procedure applied within step (iii) depends on the specific Gaussian state space model under consideration. For many important examples of state space models such as the basic structural model, the dynamic trend model (Harvey (1989)) or for regression models with random coefficients, H_t, F_t, G_t and u_t are predetermined, and the variances are the only model parameters. For such a conditionally heteroscedastic Gaussian state space model the variances are conditionally independent inverted gamma random variables and easy to sample. Further blocking is necessary if F_t or H_t depend on unknown model parameters. Finally, step (iv) may be carried out by one of the multimove methods discussed in Carter and Kohn (1994), Frühwirth-Schnatter (1994) and DeJong and Shephard (1995).

There exist various ways to run through this scheme and the reader is referred to Frühwirth-Schnatter (2001) for a detailed discussion of this issue. An unconstrained model may be estimated by *unconstrained Gibbs sampling* running through step (i)–(iv) without any constraint on the state specific parameters. Unconstrained Gibbs sampling, however, does not explore the whole unrestricted parameter space, but tends to stick at the current labeling subspace with occasionally switching to other labeling subspaces, while others will never be visited. An alternative method of estimating an unconstrained model is *random permutation sampling* (Frühwirth-Schnatter (2001)). This method is simply an unconstrained Gibbs sampler concluded by a randomly selected permutation $\rho(1), \ldots, \rho(K)$ of the current labeling $1, \ldots, K$. After sampling ψ by an unconstrained Gibbs sampler, state dependent parameters are permuted in the following way:

$$
\begin{aligned}
(\theta_1^I, \ldots, \theta_K^I) &:= (\theta_{\rho(1)}^I, \ldots, \theta_{\rho(K)}^I), \\
(\eta_{i1}, \ldots, \eta_{iK}) &:= (\eta_{\rho(i),\rho(1)}, \ldots, \eta_{\rho(i),\rho(K)}), \qquad i = 1, \ldots, K, \\
(I_1, \ldots, I_N) &:= (\rho(I_1), \ldots, \rho(I_N)),
\end{aligned}
\tag{2.1}
$$

whereas the state independent parameters θ^C and x^N remain unchanged. The permuted parameters are the starting point for the next Gibbs step. This sampler delivers a sample from the unconstrained posterior where balanced label switching occurs as all labeling subspaces are visited with the same probability.

To estimate a constrained model, an identifiability constraint may be introduced into the sampling scheme. One way of forcing the identifiability constraint is to introduce some truncation or rejection method into step (iii) in order to obtain simulations which fulfill the constraint. This *constrained Gibbs sampling* is the standard method applied so far (see e.g. Engle and Kim (1999)). As mentioned before, identifiability constraints do not necessarily induce a unique labeling, and therefore, constrained sampling may introduce a bias towards a poor constraint. The poorness of the constraint may go undetected if we use constrained Gibbs sampling and the sampler sticks at the current labeling subspace.

An alternative method for constrained sampling is *permutation sampling under an identifiability constraint* (Frühwirth-Schnatter (2001)). Unconstrained Gibbs sampling is concluded by a permutation as in (2.1), but this time the permutation is selected in such a way that the identifiability constraint is fulfilled. This method delivers a sample from the constrained posterior. If the constraint is poor in the sense that it does not induce a unique labeling, the permutation sampler will indicate this fact and exhibit label switching. In this case more suitable identifiability constraint may be found by exploring the MCMC output of random permutation sampling.

3. Issues in model selection

3.1 *The Bayesian approach to model selection*

In practical time series analysis we may end up with various dynamic linear models $\mathcal{M}_1, \ldots, \mathcal{M}_L$ —with or without switching— as possible explanation of our observations, and we have to address some issues in model selection such as selecting the number K of states or testing for the presence of Markov switching heterogeneity. Both testing problems are not easily solved within the classical framework of maximum likelihood. Although a non-switching state space model could be viewed as that special case of a switching state space model where $K = 1$, the regularity conditions for justifying the χ^2-approximation to the likelihood ratio statistic do not hold, as the state dependent parameters are unidentified under the hypothesis that there is really one state.

Bayesian approaches to selecting the number of states are the jump diffusion approach (Richardson and Green (1997)) which to our knowledge has not been applied to switching state space models and the variable selection approach of Carlin and Chib (1995) which has been applied to dynamic factor models with switching by Kim and Nelson (2001). In the present paper we pursue the classical Bayesian approach of comparing all possible models under consideration through their model likelihoods $L(y^N \mid \mathcal{M}_l)$ which are obtained by integrating the complete data likelihood $f(y^N \mid \psi)$ with respect to the prior of all unknown quantities including the latent processes:

$$L(y^N \mid \mathcal{M}_l) = \int f(y^N \mid \psi)\pi(\psi)\nu(d\psi). \tag{3.1}$$

Within state space modelling this approach has been applied to non-switching state space models in Frühwirth-Schnatter (1995), to stochastic volatility models in Kim *et al.* (1998) and to dynamic factor models with switching in Kaufmann (2000).

The computation of the model likelihood, which by definition is the normalizing constant of the non-normalized posterior $f(y^N \mid \psi)\pi(\psi)$ has, however, proven to be extremely challenging. Model likelihoods have been estimated from the MCMC output using methods such as the candidate's formula (Chib (1995)), importance sampling based on mixture approximations (Frühwirth-Schnatter (1995)), combining MCMC simulations and asymptotic approximations (Gelfand and Dey (1994); DiCiccio *et al.* (1997)) and bridge sampling (Meng and Wong (1996)). Computing the model likelihood from the MCMC output of a switching model without a latent continuous state process x^N has been discussed in detail in Frühwirth-Schnatter (1999) with the following main results. First, estimation of the model likelihood turns out to be sensitive to the problem of label switching. Especially the candidate's formula (Chib (1995)) should not be applied if label switching is present. Second, the best result with the lowest standard error is obtained by using the method of bridge sampling (Meng and Wong (1996)). In the present paper we extend this method to Gaussian state space models with switching. We will give details in the next subsection.

3.2 *Computing the model likelihood from the MCMC output*

For switching state space models it is possible to reduce the dimension of the integration in (3.1) by integrating out the switching process I^N:

$$L(y^N) = \int L(y^N \mid x^N, \phi)\pi(x^N \mid \phi)\pi(\phi)d(x^N, \phi), \tag{3.2}$$

where closed formulae for the marginal likelihood $L(y^N \mid x^N, \phi)$ and the marginal prior $\pi(x^N \mid \phi)$ are available. Subsequently, we will discuss evaluation of integral (3.2) using the bridge sampling technique (Meng and Wong (1996)). Bridge sampling is a method for computing ratios of normalising constants from MCMC simulations of the posterior and has been applied to the problem of computing the model likelihood by DiCiccio *et al.* (1997). It is obvious from (3.2) that the model likelihood is equal to the normalising constant of the posterior density $\pi(x^N, \phi \mid y^N)$ given by:

$$\pi(x^N, \phi \mid y^N) \propto \pi^\star(x^N, \phi \mid y^N) = L(y^N \mid x^N, \phi)\pi(x^N \mid \phi)\pi(\phi). \tag{3.3}$$

Let $q(x^N, \phi)$ be a density with known normalising constant, which is some simple approximation to the posterior $\pi(x^N, \phi \mid y^N)$. Let $\alpha(x^N, \phi)$ be an arbitrary function such that $\int \alpha(x^N,\phi)\pi(x^N,\phi \mid y^N)q(x^N,\phi)d(x^N,\phi) > 0$. Bridge sampling is based on the following result:

$$\begin{aligned} 1 &= \frac{\int \alpha(x^N,\phi)\pi(x^N,\phi \mid y^N)q(x^N,\phi)d(x^N,\phi)}{\int \alpha(x^N,\phi)q(x^N,\phi)\pi(x^N,\phi \mid y^N)d(x^N,\phi)} \\ &= \frac{\int \alpha(x^N,\phi)\pi^\star(x^N,\phi \mid y^N)q(x^N,\phi)d(x^N,\phi)}{L(y^N)\int \alpha(x^N,\phi)q(x^N,\phi)\pi(x^N,\phi \mid y^N)d(x^N,\phi)}, \end{aligned}$$

which yields the key identity:

$$L(y^N) = \frac{E_q(\alpha(x^N,\phi)\pi^\star(x^N,\phi \mid y^N))}{E_\pi(\alpha(x^N,\phi)q(x^N,\phi))}, \tag{3.4}$$

where E_f is the expectation with respect to a density $f(\cdot)$. If —dependent or independent— samples $(x^N,\phi)^{(m)}$, $m = 1,\ldots,M$, and $(\tilde{x}^N,\tilde{\phi})^{(l)}$, $l = 1,\ldots,L$ from the posterior $\pi(x^N,\phi \mid y^N)$ and the approximate density $q(x^N,\phi)$, respectively, are available, then both expectations are substituted by the appropriate averages and we obtain the bridge sampling estimator $\hat{L}_{BS}(y^N)$:

$$\hat{L}_{BS}(y^N) = \frac{\hat{E}_q}{\hat{E}_\pi} = \frac{L^{-1}\sum_{l=1}^{L}\alpha((\tilde{x}^N,\tilde{\phi})^{(l)})\pi^\star((\tilde{x}^N,\tilde{\phi})^{(l)} \mid y^N)}{M^{-1}\sum_{m=1}^{M}\alpha((x^N,\phi)^{(m)})q((x^N,\phi)^{(m)})}. \tag{3.5}$$

There are two functions for tuning: the function $\alpha(x^N,\phi)$ and the importance density $q(x^N,\phi)$. If one uses $\alpha(x^N,\phi) = 1/q(x^N,\phi)$, then we obtain the following importance sampling estimator of the model likelihood:

$$\hat{L}_{IS}(y^N) = \frac{1}{L}\sum_{l=1}^{L}\frac{\pi^\star((\tilde{x}^N,\tilde{\phi})^{(l)} \mid y^N)}{q((\tilde{x}^N,\tilde{\phi})^{(l)})}. \tag{3.6}$$

This estimator is an extension of the importance sampling estimator of Frühwirth-Schnatter (1995) to Gaussian state space models with switching. If one uses $\alpha(x^N, \phi) = 1/\pi^\star(x^N, \phi \mid y^N)$, the basic identity for reciprocal importance sampling (Gelfand and Dey (1994)) results leading to the following reciprocal importance sampling estimator:

$$\hat{L}_{RI}(y^N) = \left\{ M^{-1} \sum_{m=1}^{M} \frac{q((x^N, \phi)^{(m)})}{\pi^\star((x^N, \phi)^{(m)} \mid y^N)} \right\}^{-1}. \tag{3.7}$$

Meng and Wong (1996) discuss an asymptotically optimal choice of $\alpha(x^N, \phi)$, which minimizes the expected relative error of the estimator $\hat{L}_{BS}(y^N)$ for iid draws from $\pi(x^N, \phi \mid y^N)$ and $q(x^N, \phi)$:

$$\alpha(x^N, \phi) \propto \frac{1}{L \cdot q(x^N, \phi) + M \cdot \pi(x^N, \phi \mid y^N)}. \tag{3.8}$$

We will refer to the corresponding bridge sampling estimator as the "optimal" bridge sampling estimator. As the optimal choice depends on the *normalized* posterior, we apply the following iterative procedure: based on a previous estimate $\hat{L}_{BS}^{(t-1)}(y^N)$ of the normalizing constant, the posterior is normalized, $\hat{\pi}(x^N, \phi \mid y^N) = \pi^\star(x^N, \phi \mid y^N)/\hat{L}_{BS}^{(t-1)}(y^N)$, and a new estimate $\hat{L}_{BS}^{(t)}(y^N)$ is computed by (3.5). This leads to the following recursion:

$$\hat{L}_{BS}^{(t)}(y^N) = \hat{L}_{BS}^{(t-1)}(y^N) \frac{L^{-1} \sum_{l=1}^{L} \frac{\hat{\pi}((\tilde{x}^N, \tilde{\phi})^{(l)} \mid y^N)}{L \cdot q((\tilde{x}^N, \tilde{\phi})^{(l)}) + M \cdot \hat{\pi}((\tilde{x}^N, \tilde{\phi})^{(l)} \mid y^N)}}{M^{-1} \sum_{m=1}^{M} \frac{q((x^N, \phi)^{(m)})}{L \cdot q((x^N, \phi)^{(m)}) + M \cdot \hat{\pi}((x^N, \phi)^{(m)} \mid y^N)}}. \tag{3.9}$$

Either the importance sampling estimator or the reciprocal importance sampling estimator may serve as starting value $\hat{L}_{BS}^{(0)}(y^N)$. Whereas importance sampling as well as reciprocal importance sampling are known to be sensitive to the tail behaviour of the importance density $q(\cdot)$, it has been shown in Frühwirth-Schnatter (1999) that the "optimal" bridge sampling estimator is much more robust in this respect. This is of particular importance for switching state space models, where an importance density in the space of the unobserved, latent process x^N has to be constructed. The choice of an importance density which has only fat or only thin tails in *all* directions of the parameter space appears unattainable and robustness to the tail behaviour will be of great importance.

Now we turn to the choice of $q(x^N, \phi)$ for switching state space models. For much simpler models, DiCiccio *et al.* (1997) suggested to construct a normal importance density from the MCMC output. For an unconstrained switching state space model, however, the posterior usually is multimodal and a normal importance density might be an extremely bad choice. Therefore we construct the importance density in an unsupervised manner from MCMC simulations $\psi^{(1)}, \ldots, \psi^{(M_L)}$ from the unconstrained posterior $\pi(\psi \mid y^N)$ in a similar way as in Frühwirth-Schnatter (1995, 1999):

$$\begin{aligned} &q(x^N, \phi) \\ &= \frac{1}{M_L} \sum_{n=1}^{M_L} \pi(\eta \mid (I^N)^{(n)}) \pi(x^N \mid \theta^{(n)}, (I^N)^{(n)}, y^N) K_\theta(\theta \mid (I^N, x^N, \theta)^{(n)}, y^N), \end{aligned} \tag{3.10}$$

where $K_\theta(\theta \mid I^N, x^N, \theta', y^N)$ is the density of the transition kernel appearing in the unconstrained Gibbs sampler. The mixture importance density (3.10) is based on averaging over the conditional densities, where the argument (x^N, ϕ) is fixed and the conditioning indicator process I^N is sampled from the unconstrained posterior switching between the different ways of labeling. Therefore the mixture importance density (3.10) will be multimodal, too. In order to reproduce all modes of the posterior, however, it is essential to base the mixture approximation (3.10) on a MCMC method which forces *balanced* label switching such as the random permutation sampler.

Construction of $q(\cdot)$ and sampling from $q(\cdot)$ is possible during MCMC sampling. Typically, M_L is smaller than M as there is no need to use the whole MCMC sample to construct the importance density. We select the M_L mixture indices randomly from the possible indices $1, \dots, M$ prior to MCMC sampling. By sampling L times randomly from the selected indices we decide how many simulations are necessary from each conditional density in (3.10). Sampling from a conditional density takes place whenever we reach the corresponding mixture index during MCMC sampling. As these conditional densities are the same as those used for MCMC sampling, this is especially easy to implement. In order to evaluate $q(\cdot)$ later on, we store the moments of these conditional densities.

To compute the model likelihood from (3.9) it is necessary to evaluate the functions $\pi^*(x^N, \phi \mid y^N)$ given by (3.3) and $q(\cdot)$ given by (3.10) with the argument taking all values from the MCMC sample as well as from the $q(\cdot)$-sample. The first function could be evaluted during sampling as $L(y^N \mid x^N, \phi)\pi(x^N \mid \phi)$ in (3.3) results as a by-product when sampling from $\pi(I^N \mid x^N, \phi, y^N)$. The evaluation of $q(\cdot)$ is possible only after having finished sampling. Finally, we compute the importance sampling estimator (3.6) and the reciprocal importance sampling estimator (3.7) as starting value for the recursion (3.9) and compare the final estimators for convergence diagnostics. During recursion the non-normalized posterior is only divided by the last estimate of the normalising constant, and no additional function evalutions are necessary.

4. Application to simulated data

4.1 *Data simulated from a model with switching*

First, we apply the methods of the previous sections to a time series of length $N = 400$ simulated from a local level model with switching observation variance:

$$\begin{aligned} x_t &= x_{t-1} + w_t, \qquad w_t \sim \mathrm{N}(0, Q_t), \\ y_t &= x_t + v_t, \qquad v_t \sim \mathrm{N}(0, R_t), \end{aligned}$$

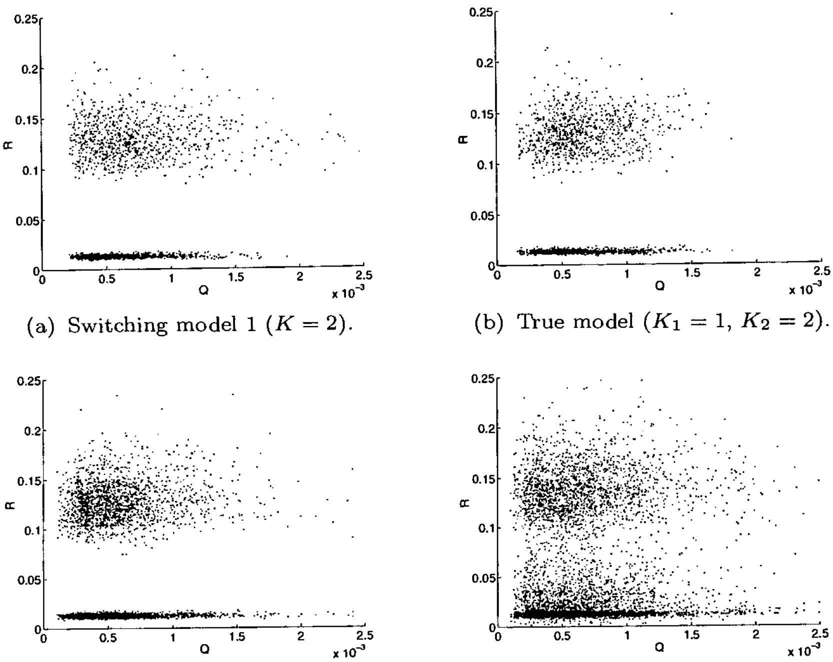
where Q_t is constant ($Q_t = 0.001$), R_t is a switching variance, and I_t is simulated as a 2-state Markov chain with transition matrix η:

$$R_t = \begin{cases} 0.01, & I_t = 1, \\ 0.1, & I_t = 2, \end{cases} \qquad \eta = \begin{pmatrix} 0.95 & 0.05 \\ 0.05 & 0.95 \end{pmatrix}.$$

We test the "true" model against the following alternatives:

- a local level model with jointly Markov switching variances ("switching model 1"):

$$(Q_t, R_t) = \begin{cases} (Q^{[1]}, R^{[1]}), & I_t = 1, \\ (Q^{[2]}, R^{[2]}), & I_t = 2; \end{cases}$$



(a) Switching model 1 ($K = 2$). (b) True model ($K_1 = 1$, $K_2 = 2$).

(c) Switching model 2 ($K_1 = K_2 = 2$). (d) Switching model 2 ($K_1 = 2$, $K_2 = 3$).

Fig. 1. Explorative Bayesian analysis for simulated data set 1.

Table 1. Formal model selection for simulated data set 1 (standard errors given in parenthesis).

Model	$\log L(y^N \mid \text{Model})$
True Model ($K_1 = 1$, $K_2 = 2$)	57.3052 (0.0819)
Switching Model 1 ($K = 2$)	49.6716 (0.0665)
Switching Model 2 ($K_1 = 2$, $K_2 = 2$)	45.8414 (0.0845)
Switching Model 2 ($K_1 = 2$, $K_2 = 3$)	33.9585 (0.0928)
No switching	−6.2878 (0.0427)

- a local level model with independently Markov switching variances ("switching model 2"):

$$Q_t = \begin{cases} Q^{[1]}, & I_t^1 = 1, \\ Q^{[2]}, & I_t^1 = 2, \end{cases} \qquad R_t = \begin{cases} R^{[1]}, & I_t^2 = 1, \\ R^{[2]}, & I_t^2 = 2; \end{cases}$$

  an extension of this model, where R_t switches between three states;
- and a local level model without switching variances.

In Fig. 1 we perform explorative Bayesian analysis of the MCMC output obtained from random permutation sampling of these models. We have used $M = 1000$ simulations after a burn-in of 1500 iterations. For a model with a single switching variable all simulations $(Q^{[j]}, R^{[j]})^{(l)}$, $l = 1, \ldots, L$, are projected onto the (Q, R)-plane for all states $j = 1, \ldots, K$. Similarly, for all models with two switching variables all simulations $(Q^{[j]}, R^{[j']})^{(l)}$, $l = 1, \ldots, L$, are projected onto the (Q, R)-plane for all combinations of states $j \in \{1, \ldots, K_1\}$ of the first and states $j' \in \{1, \ldots, K_2\}$ of the second switching variable.

There are a lot of interesting hints in these figures concerning the model selection issues discussed earlier. For "switching model 1" there are two groups, as expected, and although the model is the wrong one, the simulations show that there is practically no difference between Q in the first and the second state. This finding would lead to hypothesize the correct model. If we carry out formal model selection between "switching model 1" and the true model (Table 1), we find that the true model is clearly preferred. For the model with two switching variables "switching model 2" again there are just two groups both for two and three states in the observation variance R. This is not what would be expected, if the selected number of states were correct. If we had two states in both variances, we would expect four groups in this figure and similarly, six groups, if one variance has two and the other variance has three different states. The fact that there are just two groups is a hint that the number of states is too large for both models. The MCMC output of both models clearly indicates, that there are two states in R and just one state in Q. Again we would hypothesize the true model from analyzing the MCMC output of a model which is wrong. If we test all wrong models against the true one (Table 1), we find that the true model has the largest model likelihood. All model likelihoods were computed with $L = M = 1000$ and $M_L = 50 \cdot K!$. For completeness, we also report the model likelihood for a model without switching which is much smaller than the model likelihood for any of the switching models.

After having selected the model, we need an identifiability constraint, if we are interested in state specific estimation. From Fig. 1 it is clear that the constraint $R^{[1]} < R^{[2]}$ separates the groups. Including this constraint into the constrained permutation sampler leads to an identified model without label switching.

4.2 *Data simulated from a model without switching*

Next, we have simulated a time series of length $N = 400$ from a local level model without switching:

$$\begin{aligned} x_t &= x_{t-1} + w_t, \qquad w_t \sim \mathrm{N}(0, Q_t) \\ y_t &= x_t + v_t, \qquad v_t \sim \mathrm{N}(0, R_t), \end{aligned}$$

where Q_t and R_t are constant: $Q_t \equiv 0.001$, $R_t \equiv 0.01$. We test this model against the following alternatives:

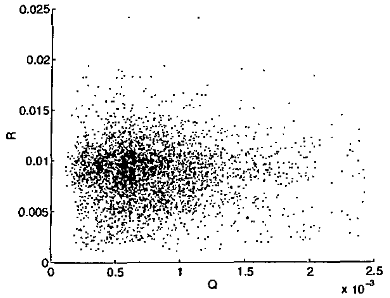
- a local level model with jointly switching variances ("switching model 1"):

$$(Q_t, R_t) = \begin{cases} (Q^{[1]}, R^{[1]}), \ I_t = 1, \\ (Q^{[2]}, R^{[2]}), \ I_t = 2; \end{cases}$$

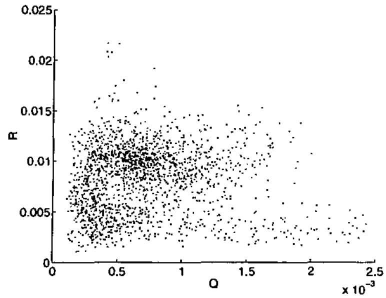
- and a local level model with independently switching variances ("switching model 2"):

$$Q_t = \begin{cases} Q^{[1]}, \ I_t^1 = 1, \\ Q^{[2]}, \ I_t^1 = 2, \end{cases} \qquad R_t = \begin{cases} R^{[1]}, \ I_t^2 = 1, \\ R^{[2]}, \ I_t^2 = 2. \end{cases}$$

In Fig. 2 we perform explorative Bayesian analysis of the MCMC output obtained from random permutation sampling of these models, where the simulations are projected



(a) Switching model 2 ($K_1 = 2$, $K_2 = 2$).



(b) Switching model 1 ($K = 2$).

Fig. 2. Explorative Bayesian analysis for simulated data set 2.

Table 2. Formal model selection for simulated data set 2 (standard errors given in parenthesis).

Model	$\log L(y^N \mid \text{Model})$
True Model (No switching)	299.0033 (0.0251)
Switching Model 1 ($K = 2$)	282.9084 (0.0879)
Switching Model 2 ($K_1 = 2$, $K_2 = 2$)	280.8277 (0.0817)

in the same way as in Fig. 1 (again $M = 1000$ simulations after a burn-in of 1500 iterations). The simulations from "switching model 2" should show four groups, if both variances were switching independently, however the only group we see is a clear hint that this model should be compared with a model without switching which would be the true one. Similarily, the simulations from "switching model 1" seem like coming from one group. If we compute model likelihoods for all models (Table 2) we find that the true model, the one without switching, has the highest model likelihood. Again, all model likelihoods were computed with $L = M = 1000$ and $M_L = 50 \cdot K!$.

5. Application to modelling exchange rate data

For further illustration we reanalyze the U.S./U.K. real exchange rate from January 1885 to November 1995, originally published in Grilli and Kaminsky (1991) and reanalyzed by Engle and Kim (1999). The real exchange rate is defined as the relative price of U.K. to U.S. producer goods, i.e. U.S./U.K. nominal exchange rate times the U.K. producer price index divided by the U.S. producer price index. Engle and Kim (1999) suggested to decompose the log of the real exchange rate y_t into a permanent component p_t and a transitory component c_t : $\log y_t = p_t + c_t$, where p_t follows a random walk process,

$$p_t = p_{t-1} + w_{2,t}, \qquad w_{2,t} \sim \mathrm{N}(0, \sigma^2_{2,t}),$$

whereas c_t is assumed to follow an AR(r)-process;

$$c_t = \phi_1 c_{t-1} + \cdots + \phi_r c_{t-r} + w_{t,2}, \qquad w_{t,2} \sim \mathrm{N}(0, \sigma^2_{1,t}).$$

The variance $\sigma^2_{1,t}$ of the transitory component c_t is assumed to switch between K_1 states $\theta^{1,I}_1, \ldots, \theta^{1,I}_{K_1}$ according to a Markov switching process I^1_t with transition matrix η^1, whereas the variance $\sigma^2_{2,t}$ of the permanent component is assumed to switch between K_2 states $\theta^{2,I}_1, \dots, \theta^{2,I}_{K_2}$ according to a Markov switching process I^2_t with transition matrix η^2:

$$\sigma^2_{1,t} = \theta^{1,I}_{I^1_t}, \qquad \sigma^2_{2,t} = \theta^{2,I}_{I^2_t}. \tag{5.1}$$

As the model can be reformulated as a dynamic linear model with state vector

$$x_t = (p_t, c_t, \dots, c_{t-r+1}),$$

it can be viewed as a special case of the switching Gaussian state space model discussed in the present paper. Various issues in model selection will be discussed in the context of this case study:

- Is the (conditional) variance of the permanent and/or the transitory component really Markov switching?
- How many states do the switching variances exhibit?
- How should we select the order r of the AR(r)-process?

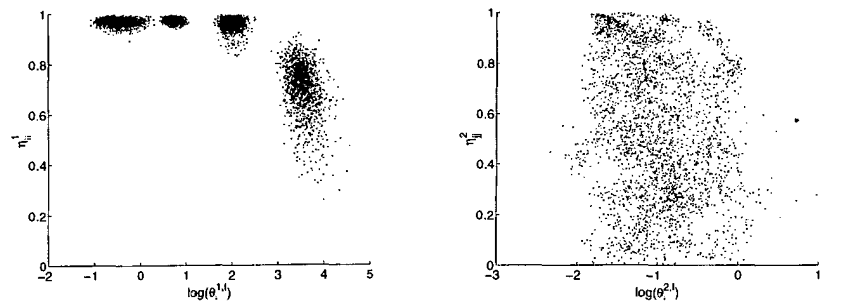
Engle and Kim (1999) selected a model where the (conditional) variance of the transitory component is determined from a 3-state Markov switching process, the (conditional) variance of permanent component is constant, and the order of the AR-process is equal to two (i.e. $K_1 = 3, K_2 = 1$, $r = 2$). They adopt this specification by exploring the posterior distributions without formal Bayesian model selection. In the present paper the best model will be one where the (conditional) variance of the transitory component is determined from a four-state Markov switching process, the (conditional) variance of permanent component is constant, and the order of the AR-process is equal to two (i.e. $K_1 = 4, K_2 = 1$, $r = 2$). Our findings will be confirmed by formal Bayesian model selection.

Our estimation method differs from the one adopted in Engle and Kim (1999) in various respects. First, we start with estimating unidentified models of various orders and try to find the "best" one through formal Bayesian model selection. Estimation is based on random permutation sampling. Within permutation sampling we may sample all variances $\theta^{1,I}_i$, $i = 1, \dots, K_1$, and $\theta^{2,I}_j$, $j = 1, \dots, K_2$, at the same time, as they are conditionally independent, inverted Gamma distributed for an unconstrained model given $(I^{1,N}, I^{2,N}, x^N, y^N)$. This is different from Engle and Kim (1999) who impose a priori an identifiability constraint on the variances and sample the variances in a single move manner from the constrained posterior.

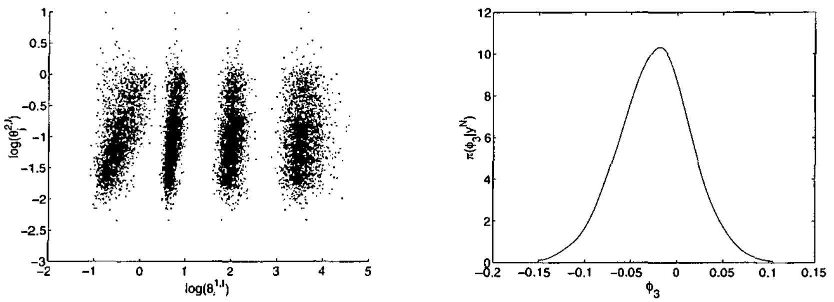
Second, we do not condition on the first values of the state process, but sample the whole processes $c_{1-r}, \dots, c_0, \dots, c_N$ and $p_0, \dots, p_N$ including the starting values by applying the multi-move sampler of Frühwirth-Schnatter (1994), and initializing the filtering step with the prior $x_0 \sim \mathrm{N}(\hat{x}_{0|0}, P_{0|0})$, where $\hat{x}_{0|0} = (\log y_1 \; 0 \cdots 0)'$ and $P_{0|0} = \mathrm{Diag}(1000\, M)$ where

$$\mathrm{vec}(M) = (I_{r\times r} - F^S_{r\times r} \otimes F^S_{r\times r})^{-1} \begin{pmatrix} \sigma^2_{1,0} \\ 0_{(r^2-1)\times 1} \end{pmatrix}, \quad F^S_{r\times r} = \begin{pmatrix} \phi_1 \cdots \phi_{r-1} & \phi_r \\ I_{(r-1)\times(r-1)} & 0_{(r-1)\times 1} \end{pmatrix},$$

and $\otimes$ is the Kronecker product of two matrices. This choice is based on the suggestion of DeJong and Chu-Chun-Lin (1994) for combining a vague prior with a stationary prior for state vectors containing both non-stationary and stationary components. Sampling



(a) $\log(\theta_i^{1,I})$ versus η_{ii}^1 for all possible i. (b) $\log(\theta_j^{2,I})$ versus η_{jj}^2 for all possible j.



(c) $\log(\theta_i^{1,I})$ versus $\log(\theta_j^{2,I})$ for all possible (i,j). (d) Posterior of ϕ_3.

Fig. 3. Explorative Bayesian analysis for a switching model with $K_1 = 4, K_2 = 2, r = 3$.

the AR(r)-parameters is carried out in a similar fashion as in Engle and Kim (1999) from the regression model $c_t = c_{t-1}\phi_1 + \cdots + c_{t-r}\phi_r + (\theta_{I_t^1}^{1,I})^{0.5} \cdot \varepsilon_t$, where ε_t is iid standard normal. However, as samples of $c_0, \ldots, c_{1-r}$ are available from our Gibbs sampler, t is running from 1 to N rather than from $r+2$ to N. Within one iteration, sampling $\phi_1, \ldots, \phi_r$ is repeated until the stationarity condition on the AR(r)-process is fulfilled.

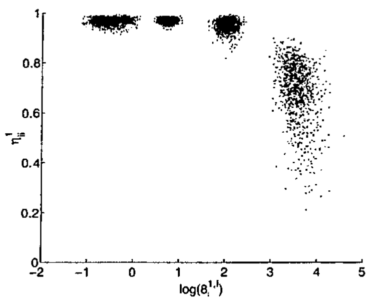
Estimation is based on the symmetric priors $\theta_i^{1,I} \sim \mathrm{IG}(3,8)$, $i = 1, \ldots, K_1$, and $\theta_j^{2,I} \sim \mathrm{IG}(3,2)$, $j = 1, \ldots, K_2$. The prior for all conditional transition probabilities $\eta_{i\cdot}^1$, and $\eta_{j\cdot}^2$ is chosen to be D(1,. . . ,1).

We start with explorative Bayesian analysis of a model with $K_1 = 4, K_2 = 2$ and $r = 3$. Although this model is not identified, a lot of interesting information is available from the MCMC output of the random permutation sampler. Part (a) and (b) of Fig. 3, for instance, show a scatter plot of (stationary) MCMC simulations $(\theta_i^{1,I})^{(m)}$ versus $(\eta_{ii}^1)^{(m)}$ and $(\theta_j^{2,I})^{(m)}$ versus $(\eta_{jj}^2)^{(m)}$ for all possible states $i \in \{1, \ldots, K_1\}$ and $j \in \{1, \ldots, K_2\}$, respectively. For I_t^1 we have allowed for four states and there are actually four distinct groups; for I_t^2, however, we have allowed for two states, and there are no separate groups. This provides empirical evidence in favour of a homogeneous rather than a switching variance of the permanent component. This hypothesis is further supported by part (c) of the figure where $(\theta_i^{1,I})^{(m)}$ is plotted versus $(\theta_j^{2,I})^{(m)}$ for all possible combinations (i,j) of states. Finally, part (d) of the same figure plots the posterior of the AR-parameter ϕ_3 which may be estimated directly from the output of the random permutation sampler as ϕ_3 is state independent. The mode of the posterior is close to 0 providing evidence for the hypothesis that ϕ_3 is equal to zero. To sum up, from our explorative analysis we obtain evidence in favour of a model with $K_1 = 4$, $K_2 = 1$ and $r = 2$ rather than

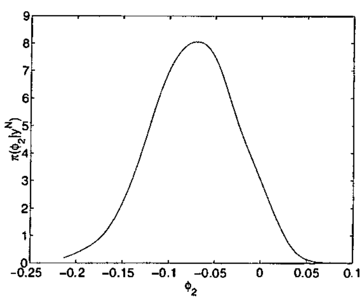
Table 3. Formal model selection for exchange rate data.

(Standard error given in parenthesis).

Model	$\log L(y^N \mid \text{Model})$
$K_1 = 4$, $K_2 = 2$, $r = 3$	−2562.4 (0.217)
$K_1 = 4$, $K_2 = 1$, $r = 2$	−2515.5 (0.083)
$K_1 = 4$, $K_2 = 1$, $r = 1$	−2612.5 (0.124)
$K_1 = 3$, $K_2 = 1$, $r = 2$	−2605.9 (0.076)
$K_1 = 5$, $K_2 = 1$, $r = 2$	−2880.2 (0.120)
No switching, $r = 2$	−2914.4 (0.081)



(a) $\log(\theta_i^{1,I})$ versus η_{ii}^1 for all possible i.



(b) Posterior of ϕ_2.

Fig. 4. Explorative Bayesian analysis for a switching model with $K_1 = 4, K_2 = 1, r = 2$.

$K_1 = 4$, $K_2 = 2$ and $r = 3$.

In Table 3 both models are compared by formal Bayesian model selection, showing a much higher model likelihood for the simpler model. Explorative Bayesian analysis of the new model is reported in Fig. 4. The scatter plot of $(\theta_i^{1,I})^{(m)}$ versus $(\eta_{ii}^1)^{(m)}$ for all possible states $i \in \{1, \ldots, K_1\}$ still shows four groups. The posterior of ϕ_2 is shifted away from 0, but there is some evidence for the hypothesis that ϕ_2 actually is 0. The model likelihood of the model with $K_1 = 4$, $K_2 = 1$ and $r = 1$ reported in Table 3, however, clearly favours the model with $K_1 = 4$, $K_2 = 1$ and $r = 2$ which is our final choice. This model differs from the one selected in Engle and Kim (1999) by the number of states of the variance of the transitory component. The model likelihoods in Table 3 clearly favour our "best" model rather than the one selected by Engle and Kim (1999). Increasing the number of states from four to five, however, reduces the model likelihood drastically. For completeness, we report the model likelihood for a model without switching, showing that this model is the most unlikely of all. Computation of all model likelihoods reported in Table 3 is based on the random permutation sampler.

We can draw further interesting inference from the output of the random permutation sampler without the need to identify the model. This is especially true for the smoothed permanent component $\hat{p}_{t|N}$ which is compared in Fig. 5 with the observed time series. Our estimate of the permanent component is much smoother than the rather noisy estimate published in Engle and Kim (1999), being nearly constant till the end of the fifties and increasing afterwards. Another interesting picture is obtained, if we plot the

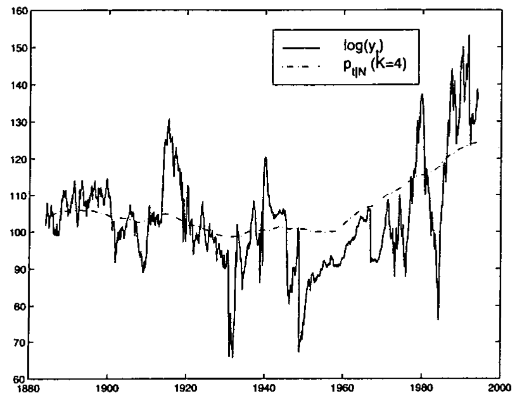


Fig. 5. Smoothed real exchange rate $\hat{p}_{t|N}$ for a four state model ($K_1 = 4$, $K_2 = 1$, $r = 2$).

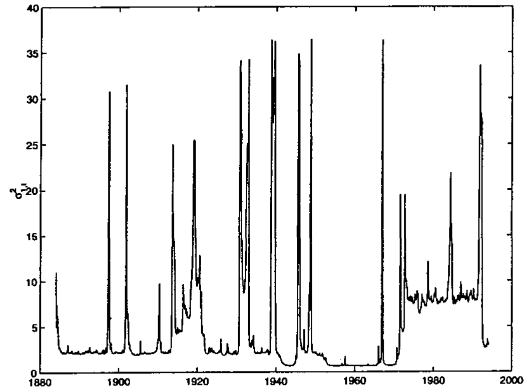


Fig. 6. Estimated time varying variance $\hat{\sigma}^2_{1,t}$ for a four state model ($K_1 = 4$, $K_2 = 1$, $r = 2$).

time varying variance $\sigma^2_{1,t}$ estimated from (5.1) by

$$\hat{\sigma}^2_{1,t} = \frac{1}{M} \sum_{m=1}^{M} \left(\theta_s^{1,I}\right)^{(m)},$$

where $s = (I_t^1)^{(m)}$ over time t (Fig. 6).

To draw state specific inference on the variances of the different states as well as to obtain the smoothed posterior state probabilities, however, we have to identify the model. For comparison, we will identify both the "best" model and the three state model selected by Engle and Kim (1999). An identifiability constraint for the "best" model is obvious from Fig. 4. The main difference between the states is the associated variance of the transitory component leading to the constraint $\theta_1^{1,I} < \theta_2^{1,I} < \theta_3^{1,I} < \theta_4^{1,I}$. If this constraint is included into the permutation sampler under identifiability constraints, no label switching occurs. Table 4 reports point estimates as well as 95%-H.P.D.-regions for all model parameters, including estimates of the state specific variances as well as estimates of the transition probabilities. Similarly, the identifiability constraint for the

Table 4. Estimation results for $K_1 = 4, K_2 = 1, r = 2$.

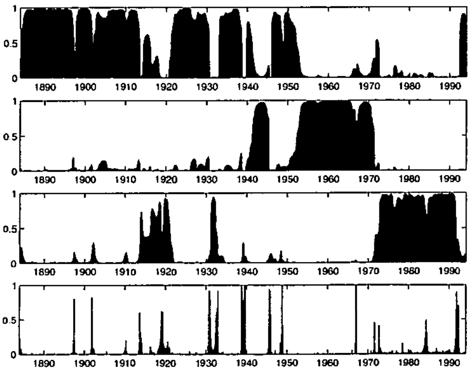
Parameter	Mean	Std.dev.	95%-H.P.D. regions	
θ_1^I	0.634	0.151	0.371	0.93
θ_2^I	2.05	0.196	1.67	2.42
θ_3^I	7.63	1.07	5.9	9.88
θ_4^I	36.4	9.13	20.7	53.9
σ_2^2	0.366	0.132	0.121	0.608
ϕ_1	1.06	0.0474	0.967	1.14
ϕ_2	−0.0729	0.046	−0.158	0.0139
η_{11}	0.968	0.0132	0.943	0.991
η_{12}	0.0091	0.00861	2.84e-006	0.0256
η_{13}	0.00639	0.00586	2.87e-006	0.0189
η_{14}	0.0162	0.00987	0.000231	0.0341
η_{21}	0.00855	0.00576	0.000165	0.0205
η_{22}	0.973	0.00853	0.957	0.988
η_{23}	0.00587	0.0057	6.19e-006	0.0155
η_{24}	0.0123	0.00697	0.000484	0.0246
η_{31}	0.00498	0.00489	1.24e-005	0.0144
η_{32}	0.0139	0.0123	9.59e-006	0.0373
η_{33}	0.956	0.0222	0.916	0.992
η_{34}	0.0248	0.0161	0.00129	0.0562
η_{41}	0.039	0.0338	0.000159	0.103
η_{42}	0.147	0.0691	0.024	0.288
η_{43}	0.123	0.0934	0.00108	0.309
η_{44}	0.691	0.116	0.438	0.865

three state model is given by: $\theta_1^{1,I} < \theta_2^{1,I} < \theta_3^{1,I}$.

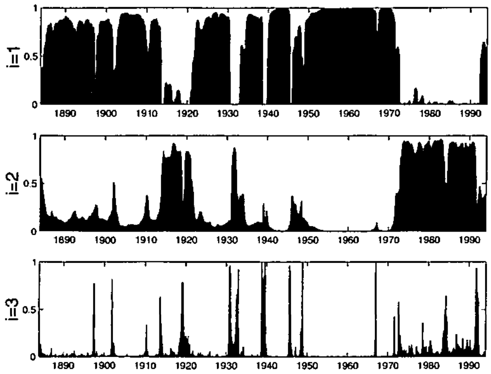
Figure 7 plots the estimated smoothed posterior state probabilities $\Pr(I_t^1 = i \mid y^N)$ of being in a certain state $i \in \{1,2,3,4\}$ over time t, for a four state switching model, and compares them with the probabilities obtained from the three state model. The probabilities $\Pr(I_t^1 = i \mid y^N)$, $i = 1, \dots, K_1$ are estimated from the constrained MCMC output by:

$$\hat{\Pr}(I_t^1 = i \mid y^N) = \frac{1}{M}\#\{(I_t^1)^{(m)} = i\}.$$

For the three state model Engle and Kim (1999) found that periods with low-state variance correspond to periods in which the nominal exchange rate was fixed, whereas periods with medium-state variance correspond to periods of floating nominal exchange rates. Periods of high-state variance are rather singular events and can be identified with specific historical events. It is interesting to observe that for the four state model the third and the fourth state actually correspond to the second and the third state of the three state model, whereas the first two states are a further split of the low variance state of the three state model. The most quiet state occured during the first half of the forties and then from about 1952 to the end of the seventies.



(a) "Best" model ($K_1 = 4, K_2 = 1, r = 2$).



(b) Model selected by Engle and Kim (1999) ($K_1 = 3, K_2 = 1, r = 2$).

Fig. 7. Smoothed probabilities.

References

Carlin, B. P. and Chib, S. (1995). Bayesian model choice via Markov chain Monte Carlo methods, *J. Roy. Statist. Soc. Ser. B*, **57**, 473–484.

Carter, C. K. and Kohn, R. (1994). On Gibbs sampling for state space models, *Biometrika*, **81**, 541–553.

Carter, C. K. and Kohn, R. (1996). Markov chain Monte Carlo in conditionally Gaussian state space models, *Biometrika*, **83**, 589–601.

Celeux, G. (1998). *Bayesian Inference for Mixture: The Label Switching Problem*, Preprint Inria.

Chib, S. (1995). Marginal likelihoods from the Gibbs output, *J. Amer. Statist. Assoc.*, **90**, 1313–1321.

Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models, *J. Econometrics*, **75**, 79–97.

De Jong, P. and Shephard, N. (1995). The simulation smoother for time series models, *Biometrika*, **82**, 339–350.

De Jong, P. and Singfat, Chu-Chun-Lin. (1994). Stationary and non-stationary state space mdoles. *J. Time Ser. Anal.*, **15**, 151–166.

DiCiccio, T. J., Kass, R., Raftery, A. and Wasserman, L. (1997). Computing Bayes factors by combining simulations and asymptotic approximations, *J. Amer. Statist. Assoc.*, **92**, 903–915.

Engle, C. and Kim, C.-J. (1999). The long-run U.S./U.K. real exchange rate, *Journal of Money, Credit, and Banking*, **31**, 335–356.

Frühwirth-Schnatter, S. (1994). Data augmentation and dynamic linear models, *J. Time Ser. Anal.*, **15**, 183–202.

Frühwirth-Schnatter, S. (1995). Bayesian model discrimination and Bayes factors for linear Gaussian state space models, *J. Roy. Statist. Soc. Ser. B*, **57**, 237–246.

Frühwirth-Schnatter, S. (1999). Model likelihoods and Bayes factors for switching and mixture models, Tech. Report, 1999–71, Department of Statistics, Vienna University of Economics and Business Administration.

Frühwirth-Schnatter, S. (2001). MCMC estimation of classical and dynamic switching and mixture models, *J. Amer. Statist. Assoc* (forthcoming).

Gelfand, A. E., and Dey, D. K. (1994). Bayesian model choice: Asymptotics and exact calculations, *J. Roy. Statist. Soc. Ser. B*, **56**, 501–514.

Godsill, S. J. (1997). Bayesian enhancement of speech and audio signals in the presence of both impulsive and background noise, *International Statistical Review*, **65**, 1–21.

Grilli, V. and Kaminsky, G. (1991). Nominal exchange rate regimes and the real exchange rate: Evidence from the United States and Great Britain, 1885–1986, *Journal of Monetary Economics*, **27**, 191–212.

Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle, *Econometrica*, **57**, 357–84.

Harrison, P. J. and Stevens, C. F. (1976). Bayesian forecasting (with discussion), *J. Roy. Statist. Soc. Ser. B*, **38**, 205–247.

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

Kaufmann, S. (2000), Measuring business cycles with a dynamic Markov switching factor model: An assessment using Bayesian simulation methods, *Econometrics Journal*, **3**, 39–65.

Kim, C.-J. (1993). Unobserved-Component Time Series Models with Markov-Switching Heteroscedasticity, *J. Bus. Econom. Statist.*, **11**, 341–349.

Kim, C.-J. (1994). Dynamic linear models with Markov switching, *Journal of Econometrics*, **60**, 1–22.

Kim, C.-J. and Nelson, Ch. R. (1998). Business cycle turning points, a new coincident index, and tests of duration dependence based on a dynamic factor model with regime-switching, *Review of Economics and Statistics*, **80**, 188–201.

Kim, C.-J. and Nelson, Ch. R. (1999). *State-Space Models with Regime Switching*, MIT Press, Cambridge Massachasetts.

Kim, C.-J. and Nelson, Ch. R. (2001). A Bayesian approach to testing for Markov switching in univariate and dynamic factor models, *International Economic Review* (forthcoming).

Kim, S., Shephard, N. and Chib, S. (1998), Stochastic volatility: Likelihood inference and comparison with ARCH models, *Review of Economics Studies*, **65**, 361–393.

Meng, X. L. and Wong, W. H. (1996). Simulating ratios of normalising constants via a simple identity, *Statistica Sinica*, **6**, 831–60.

Peña, D. and Guttman, I. (1988). Bayesian approach to robustifying the Kalman filter, *Bayesian Analysis of Time Series and Dynamic Linear Models* (ed. J.C. Spall), 227–254, Marcel Dekker, New York.

Richardson, S. and Green, P. (1997), On Bayesian analysis of mixtures with an unknown number of components (with discussion), *J. Roy. Statist. Soc. Ser. B*, **59**, 731–92.

Shephard, N. (1994). Partial Non-Gaussian state space models, *Biometrika*, **81**, 115–31.

Shumway, R. H. and Stoffer, D. S. (1991). Dynamic linear models with switching, *J. Amer. Statist. Assoc.*, **86**, 763–9.

Smith, A. F. M. and Roberts, G. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods, *J. Roy. Statist. Soc. Ser. B*, **55**, 3–23.

Stephens, M. (1997). Bayesian methods for mixtures of normal distributions, Ph. D.Thesis, University of Oxford.