

Viefers, Paul

Working Paper

Bayesian inference for the mixed-frequency VAR model

DIW Discussion Papers, No. 1172

Provided in Cooperation with:

German Institute for Economic Research (DIW Berlin)

Suggested Citation: Viefers, Paul (2011) : Bayesian inference for the mixed-frequency VAR model, DIW Discussion Papers, No. 1172, Deutsches Institut für Wirtschaftsforschung (DIW), Berlin

This Version is available at:

<https://hdl.handle.net/10419/61331>

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

1172

Discussion
Papers

Bayesian Inference for the Mixed-Frequency VAR Model

Opinions expressed in this paper are those of the author(s) and do not necessarily reflect views of the institute.

IMPRESSUM

© DIW Berlin, 2011

DIW Berlin
German Institute for Economic Research
Mohrenstr. 58
10117 Berlin

Tel. +49 (30) 897 89-0
Fax +49 (30) 897 89-200
<http://www.diw.de>

ISSN print edition 1433-0210
ISSN electronic edition 1619-4535

Papers can be downloaded free of charge from the DIW Berlin website:
<http://www.diw.de/discussionpapers>

Discussion Papers of DIW Berlin are indexed in RePEc and SSRN:
<http://ideas.repec.org/s/diw/diwwpp.html>
<http://www.ssrn.com/link/DIW-Berlin-German-Inst-Econ-Res.html>

Bayesian inference for the mixed-frequency VAR model

Paul Viefers*

Fall 2011

Abstract

In this paper a mixed-frequency VAR à la Mariano & Murasawa (2004) with Markov regime switching in the parameters is estimated by Bayesian inference. Unlike earlier studies, that used the pseudo-EM algorithm of Dempster, Laird & Rubin (1977) to estimate the model, this paper describes how to make use of recent advances in Bayesian inference on mixture models. This way, one is able to surmount some well-known issues connected to inference on mixture models, e.g. the label switching problem. The paper features a numerical simulation study to gauge the model performance in terms of convergence to true parameter values and a small empirical example involving US business cycles.

JEL-Classification: C32, C38, E32, E37, E51

Keywords: Markov mixture models, Label switching, Bayesian VAR, Mixed frequencies.

*DIW Graduate Center, Mohrenstr. 58, D-10117 Berlin (email: pviefers@diw.de). I would like to thank Richard Munchinger for many useful comments and for getting me interested in Bayesian statistics in the first place. For comments on early versions of this paper I am also indebted to Vladimir Kuzin and participants of the Humboldt University Macro seminar and participants of the TU Berlin Econometrics Seminar 2011, Tornow. Comments and suggestions welcome.

1 Introduction

In this paper we reconsider estimation of a mixed-frequency VAR à la Mariano & Murasawa (2004) with Markov regime switching in the parameters. So far, both standard mixed-frequency VARs (Mariano & Murasawa, 2003, 2004) as well as mixed-frequency factor models that allow for regime switching (Kholodilin, 2001) have been estimated using mode finding algorithms, e.g. MLE or the EM algorithm of Dempster et al. (1977). However, once Markov switching or more generally a mixture distribution is assumed, studies in the field of inference on mixture models point at potential pitfalls when using MLE or an EM algorithm. In this paper we propose (i) how to suitably extend the Markov switching VAR to allow for regime switching and (ii) to use a Bayesian MCMC algorithm to simulate the model. Even though, using Bayesian inference provides a way to surmount some of the issues connected to standard approaches to the researcher, it has to be carefully designed in order to avoid some other well known caveats.¹ For example, it was illustrated by Frühwirth-Schnatter (2001a,b) and Celeux, Hurn & Robert (2000) that identifiability issues related to label switching can have severe consequences for posterior simulation (see Marin, Mengersen & Christian, 2005, for a survey).

To be more precise, consider a (finite and discrete) n -component mixture distribution, which is a weighted sum of distributions $f_i(\cdot | \theta)$

$$\sum_{i=1}^n \eta_i f_i(\cdot | \theta_i) ; \sum_{i=1}^n \eta_i = 1 . \quad (1.1)$$

The component distributions $f_i(\cdot | \theta)$ are assumed to belong to a parametric family of distributions, indexed by a set of unknown parameters θ_i . For most applications both the mixture weights and the parameters of the component densities are unknown. In a time series setting, for example, it is often assumed that the sampling density of the observations is a mixture distribution. Note that the general class of mixture models nests the iid mixture model and the Markov mixture model. A Markov mixture arises when the weights η_i are the ergodic probabilities of a Markov chain on the state space $\{1, \dots, n\}$.

Despite its conceptual simplicity, mixture distributions are extremely flexible and accommodate a wide array of distributional features.² However, mixture models are often challenging to estimate. Presumably one of the most prominent problems is related to the fact that we cannot distinguish the labels i or j for two sets of parameters θ_i and θ_j . Permuting labels yields the same likelihood. This exchangeability implies that the likelihood function has $n!$ identical posterior modes differing only in the labeling scheme. The simulation procedure put forward here addresses the label switching problem for MCMC

¹See Ryden (2008) for an in-depth treatment of numerical aspects of EM and MCMC algorithms in Markov mixture models.

²See Schwaab, Creal, Koopman & Lucas (2009) for a nice example.

simulation of mixed-frequency models as put forward by Frühwirth-Schnatter (2001a).

Moreover, it is easy to construct simple examples where the mixture components are identified, but the likelihood surface has a spurious mode that is not located anywhere near the true parameter vector.³ This, of course, poses a challenge for iterative maximization procedures, especially if starting values happen to fall into the domain of attraction of the minor mode. What is more is that MLE or EM involves fairly hard analytical problems to compute the mixture likelihood or to handle the proliferating state space.

Because the mixed-frequency nature of the data is framed as a missing value problem (Harvey & Pierse, 1984), some of the filtered state variables have a direct structural meaning, e.g. the monthly growth rate of real GDP. We are of the opinion that it is an imperative to carefully gauge the estimation error associated with these key variables and other unknowns in the system. Classical or frequentist applications of the Kalman filter neglect the fact that parameters of the state space models carry over some estimation error into the filtering stage. Bayesian inference, namely Gibbs sampling, provides the possibility to fully account for any estimation error uncertainty in the latent series that stems from other unknowns in the system.

Last, several recent empirical studies highlight the effectiveness of Bayesian shrinkage (through classical priors such as the Minnesota or Litterman prior) as a valid alternative to factor models (e.g. Banbura, Giannone & Reichlin, 2010; De Mol, Giannone & Reichlin, 2008; Koop & Korobilis, 2010; Banbura et al., 2010).

The paper is organized as follows. In section two we introduce the mixed-frequency VAR á la Mariano & Murasawa (2004) and outline how regime switches may be modelled. We then describe how the model may be simulated using Bayesian MCMC methods. In section three we apply the model to two data sets. The first data set is artificially generated data from a known data-generating process. This gives us the possibility to gauge the model performance, i.e. convergence towards the true parameters of the process. The second data set is supposed to illustrate how in-sample inference on the stance of the business cycle may be drawn. Toward this end, we employ a set of four US macroeconomic variables (GDP, industrial production, M2 money stock and the PPI).

2 Econometric methodology

In this section we address the econometric methodology. First, we review how mixed-frequency models can be understood as a state space model and discuss how missing observations are handled. It is then straightforward to introduce time-varying coefficients via a latent index following a binary Markov chain.

³In figure 1, we borrowed the example of a two-component normal mixture from Marin et al. (2005) and depicted the likelihood surface for a simulated data set.

2.1 Mixed frequency VARs

Second, we discuss how estimation via Gibbs sampling is implemented. The only additional complications compared to standard Markov switching VARs are: (i) to add an imputation step to generate missing values and (ii) to permute values either randomly or to meet an identification constraint in every sweep of the sampler. Our algorithm works very much along the lines of (see Frühwirth-Schnatter, 2001b,a, 2006).

2.1 Mixed frequency VARs

Historically, Harvey & Pierse (1984) and Zdrozny (1988, 1990) were the first to apply the idea of mixed frequencies to enhance timeliness in inferring on the movements of aggregate economic time series. More recently, the mixed-frequency literature is receiving significant attention (see Mariano & Murasawa, 2004; Clements & Galvao, 2008; Schwaab et al., 2009; Aruoba, Diebold & Scotti, 2009; Hamilton, 2010). Theoretically, the state space framework is general enough to bridge any frequency mismatch. For example, consider financial data. It is sampled at extremely high frequency and it would be desirable to preserve the valuable variation between periods at the highest frequent grid. Even though theoretically bridging greater frequency mismatches preserves more variation in the higher frequent variable, inference becomes increasingly fuzzy. This is because the amount of noise in the series used for estimation usually becomes larger. Hence, it becomes harder to separate systematic from unsystematic movements in the data. This illustrates a general trade-off between timeliness and accuracy. For example consider a quarterly business cycle model compared to a monthly model. On the one hand, mixture inference based on quarterly figures suffers less from unsystematic movements in the data. This leads to relatively smooth and well-behaved inference, e.g. on regime shifts. On a real-time basis, however, one usually wants a regime switch to be confirmed by arriving data, e.g. the probability of a nationwide recession should not be below 50% for at least three periods. On a quarterly grid this means one has to wait for three quarters, while on a monthly grid this reduces to waiting a single quarter.

Based on the work by Harvey & Pierse (1984); Zdrozny (1988, 1990) a mixed frequency time series model is formulated as a state-space model with missing observations.⁴ This approach is widely applied to macroeconomic time series, e.g. by Mariano & Murasawa (2003, 2004). The classical linear state space model without regression effects reads

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{f}_t + \mathbf{G}_t \boldsymbol{\epsilon}_t \quad (2.1)$$

$$\mathbf{f}_t = \mathbf{c}_t + \mathbf{A}_t \mathbf{f}_{t-1} + \mathbf{D}_t \boldsymbol{\eta}_t \quad (2.2)$$

⁴Another prominent approach to mixed frequency models is given by the MIDAS regression models due to Ghysels, Santa-Clara & Valkanov (2004). See Kuzin, Marcellino & Schumacher (2010) for a comparison. Only with Guérin & Marcellino (2011) were MIDAS regressions extended to allow for Markov switching.

2.1 Mixed frequency VARs

where $\mathbf{Y}_t = [y_{1,t}, \dots, y_{N,t}]'$ is a $N \times 1$ vector of observations with potentially missing entries, \mathbf{f}_t is a $M \times 1$ vector of latent state variables and $\boldsymbol{\epsilon}_t$ and $\boldsymbol{\eta}_t$ are $N \times 1$ and $M \times 1$ vectors of error terms. Finally \mathbf{H}_t , \mathbf{G}_t , \mathbf{A}_t , \mathbf{c}_t and \mathbf{D}_t are conformable matrices of coefficients (potentially time-varying). The chief attraction of state-space models is that they encompass a very wide class of data generating processes and make it easy to model missing observations, time-varying coefficients and unobservable (latent) factors that drive a dynamic system (Kim & Nelson, 1999, provide a detailed discussion).

Let the autoregressive order of the transition equation (2.2) be denoted by p . Mariano & Murasawa (2004) understand $y_{1,t}^*$ as being the month-on-month growth rate of some macroeconomic variable. To make exposition viable, suppose that $p = 1$, $N = 2$, i.e. $\mathbf{Y}_t = [y_{1,t}, y_{2,t}]'$ and further suppose that it is $y_{1,t}$ that we do not observe every period $t = 1, \dots, T$, whereas for $y_{2,t}$ we do. Following Mariano & Murasawa (2004) we assume

$$\mathbf{H}_t = H(L) = \begin{bmatrix} 1/3 & 0 \\ 0 & 1 \end{bmatrix} + \begin{bmatrix} 2/3 & 0 \\ 0 & 0 \end{bmatrix} L + \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} L^2 \quad (2.3)$$

$$+ \begin{bmatrix} 2/3 & 0 \\ 0 & 0 \end{bmatrix} L^3 + \begin{bmatrix} 1/3 & 0 \\ 0 & 0 \end{bmatrix} L^4 \quad (2.4)$$

$$= \mathbf{H}_0 + \mathbf{H}_1 L + \mathbf{H}_2 L^2 + \mathbf{H}_3 L^3 + \mathbf{H}_4 L^4, \quad (2.5)$$

where

$$\mathbf{f}_t = \begin{bmatrix} y_{1,t}^* & y_{2,t} \end{bmatrix}'. \quad (2.6)$$

Whenever $y_{1,t}^*$ is not observable, we set the first row of \mathbf{H}_t equal to zero, i.e. the latent state variable $y_{1,t}^*$ is not matched with an actual observation. Denote this matrix with first row equal zero as \mathbf{H}^* . Whenever $\mathbf{H}_t = \mathbf{H}^*$ the error term $\boldsymbol{\epsilon}_t$ enters the measurement equation (2.1), since the first element of \mathbf{G}_t is assumed to be non-zero whenever $y_{1,t}^*$ is not observed.⁵ Then the missing observations may be filtered from the system given by (2.1) and (2.2) by means of the Kalman filter and smoother (Meinhold & Singpurwalla, 1983). The specific form of \mathbf{H}_t implies that the month-on-month growth rate is filtered as to match

$$y_{1,t} = 1/3y_{1,t}^* + 2/3y_{1,t-1}^* + y_{1,t-2}^* + 2/3y_{1,t-3}^* + 1/3y_{1,t-4}^*, \quad (2.7)$$

⁵Using the prediction error decomposition, Mariano & Murasawa (2004) show that as long as the distribution of the error $\boldsymbol{\epsilon}_t$ is independent of the data and the parameters, its distribution is irrelevant for MLE or an EM algorithm. This is because it is simply a constant multiplied with the iid likelihood of the observations. In our setting, the random values imputed only serve as starting values and given ergodicity of the Markov chain constructed through the Gibbs sampler, the results are independent of the starting values after some burn-in period.

2.1 Mixed frequency VARs

which is the geometric average of the monthly growth rates.⁶

That being said, it becomes clear that the state space form of our simple mixed-frequency VAR is nearly equivalent to the companion form of a standard VAR(p). In particular, we have

$$\mathbf{f}_t = \begin{bmatrix} y_{1,t}^* & y_{2,t} & \cdots & y_{1,t-p}^* & y_{2,t-p} \end{bmatrix}' , \quad (2.8)$$

$$\mathbf{H}_t = \begin{cases} \mathbf{H}^* & \text{if } [t/3] \notin \mathbb{N} \\ \mathbf{H} & \text{else} \end{cases} , \quad (2.9)$$

$$\mathbf{G}_t = \begin{cases} \begin{bmatrix} \mathbf{0}_N \end{bmatrix}' & \text{if } [t/3] \in \mathbb{N} \\ \begin{bmatrix} 1 & \mathbf{0}_{N-1} \end{bmatrix}' & \text{else} \end{cases} , \quad (2.10)$$

$$\mathbf{A}_t = \begin{bmatrix} \Phi_{1,t} & \cdots & \Phi_{p,t} \\ \mathbf{I}_{N(p-1) \times N(p-1)} & & \mathbf{0}_{N(p-1) \times N} \end{bmatrix} , \quad (2.11)$$

$$\mathbf{c}_t = \begin{bmatrix} c_1 & \cdots & c_N & \mathbf{0}_{N(p-1)} \end{bmatrix}' , \quad (2.12)$$

$$\mathbf{D}_t = \begin{bmatrix} \Sigma_{\eta,t}^{1/2} & \mathbf{0}_{N(p-1) \times N(p-1)} \end{bmatrix}' , \quad (2.13)$$

$$\boldsymbol{\epsilon}_t \sim N(\mathbf{0}, \mathbf{I}_N) \quad ; \quad \boldsymbol{\eta}_t \sim N(\mathbf{0}, \Sigma_{\eta,t}) . \quad (2.14)$$

where $[t/m]$ denotes the largest integer value that is less than or equal to t/m . Given the above specification, the system is a simple linear Gaussian state space model with missing observations and time-varying coefficients (see Hamilton, 1994, pp. 373 for a primer). We may construct the vector of parameters for later use by making use of the vectorization and half-vectorization operator

$$\Theta_t = [\mathbf{c}_t, \text{vec}(\Phi_1), \dots, \text{vec}(\Phi_p), \text{vech}(\Sigma_{\eta,t})]' . \quad (2.15)$$

Conditional on the state vector, the transition equation (2.2) may be estimated using standard methods for Bayesian VAR analysis spearheaded by Christopher Sims (Doan, Litterman & Sims, 1984; Sims, 1993; Sims & Zha, 1998). Studies, e.g. by Banbura et al. (2010); De Mol et al. (2008), confirm that Bayesian VARs are valid alternatives to factor models even for a large number of variables.

⁶The arguably more intuitive aggregation by means of an arithmetic average leads to a non-linear state-space representation that makes signal extraction more cumbersome and requires to use the extended Kalman filter or particle filtering (Proietti & Moauro, 2006). Since we are working with relatively small growth rates, the difference between both accounting rules would be negligible (see Camacho & Perez-Quiros, 2010, for an example).

2.2 Bayesian inference & MCMC simulation

2.1.1 Markov regime switching

So far we have not described the probabilistic model that governs the behavior of the parameters in (2.1) to (2.2). Following the seminal paper by Hamilton (1989), we assume that the distribution of the stochastic process $\{\mathbf{f}_t\}_{t=1}^T$ depends on the realization of an unobservable binary stochastic process $S_t \in \{0, 1\}$.⁷ S_t is assumed to be an irreducible, aperiodic Markov chain with transition kernel $\mathbf{P} \in (\mathcal{E}_2) \times (\mathcal{E}_2)$ where \mathcal{E}_2 is the two-dimensional unit simplex. The random variables \mathbf{f}_t are then said to arise from a Markov mixture of distributions. The vector of parameters Θ_t governs the differences in the component distributions for \mathbf{f}_t and is allowed to be distinct on the state space $\{0, 1\}$. Intuitively, the binary state space is understood as the realization of the business cycle, which is partitioned into contractions and expansions. Let $\mathcal{F}(\Theta)$ be a (continuous) parametric family of distributions with density $p(\mathbf{y} | \Theta)$ which is indexed by $\Theta^{(S_t)}$. Then the random variables $\{\{\mathbf{f}_1\}, \dots, \{\mathbf{f}_T\}\}$ are assumed to be independent conditional on $\mathbf{S} = \{S_1, \dots, S_T\}$ and their own past. Depending on the realization of S_t , the distribution of \mathbf{f}_t arises from one of the members of $\mathcal{F}(\Theta^{S_t})$, i.e.

$$\mathbf{f}_t | S_t = i \sim \mathcal{F}(\Theta^{(i)}) . \quad (2.16)$$

For our application let $\mathcal{F}(\Theta)$ be the family of normal distributions. We allow all parameters of the transition equation (2.2) to depend on S . In the notation of Krolzig (1997), we estimate a MSIAH-VAR(p) model. It is worth noting that even though in our model all parameters are allowed to be time-varying the state space model (2.1) to (2.2) is linear and Gaussian *conditional* on \mathbf{S} .

2.2 Bayesian inference & MCMC simulation

As already mentioned, we use Bayesian methods to simulate the model and its parameters. Thus, inference in our model is based on the joint posterior density of all unknowns. More precisely, we aim to take samples from

$$\begin{aligned} p(\mathbf{f}, \mathbf{S}, \mathbf{P}, \Theta, | \mathbf{Y}) &\propto p(\mathbf{Y} | \mathbf{f}, \Theta, \mathbf{S}, \mathbf{P})p(\mathbf{f} | \Theta, \mathbf{S}, \mathbf{P}) \\ &\times p(\Theta)p(\mathbf{S} | \mathbf{P})p(\mathbf{P}) . \end{aligned} \quad (2.17)$$

This is a standard hierarchical probabilistic model, e.g. it reflects the assumption that the regime process S_t and the unknown parameters in Θ are *a priori* independent. Note that the form of the above joint density is unknown and hence there is no simple way to sample from it directly. Therefore we use the Gibbsian paradigm (Geman & Geman,

⁷Historically, the idea of time-varying coefficients and mixture models in economics dates back to Goldfeld & Quandt (1973a,b).

2.2 Bayesian inference & MCMC simulation

1984; Gelfand, Hills, Racine-Poon & Smith, 1990; Gelfand & Smith, 1990) to decompose sampling into blocks of full conditional distributions that have known form. As is illustrated below, our sampling scheme involves four such blocks. This way we construct a Markov chain that has the distribution (2.17) as its limiting distribution. Gibbs sampling and other simulation techniques are proliferating based on (i) its theoretical appeal, (ii) the fact that it is comparably easy to implement and (iii) the fast-paced development of computational power available on standard workstations. Concerning the first point, it is often theoretically attractive to be able to work with a full probabilistic model for *all* unknowns in the system. In the present case, this comprises missing observations, the latent regime process and the unknown parameters. More specifically, by being able to draw samples from the *joint* posterior distribution of all unknowns via Gibbs sampling, we are able to explicitly take into account the joint dependence of the various unknowns to be estimated. For example, standard algorithms that use the Kalman filter to generate the latent series, filter values conditional on the parameter estimates, i.e. they treat parameters in Θ as if they were known with certainty. Similarly, when the likelihood function is evaluated using the filtered state variables, maximum likelihood treats the imputed data as if it was perfectly observable. Quite contrary, after a suitable number of burn-in iterations, the Gibbs sampler generates the latent series and all other unknowns from the joint distribution (2.17). This fully accounts for any estimation error uncertainty within the specified model.⁸ Concerning the second point, by making use of the principle of data-augmentation (Tanner & Wong, 1987; van Dyk & Meng, 2001), we may treat latent variables (here: missing observations and the regime process) as missing data upon which we may condition in the other blocks of the sampler. This leads to relatively easy conjugate sampling schemes and avoids for example expensive computation of the mixture likelihood. Last but not least, this method is approximation-free.⁹

To be more precise, the Gibbs sampler implemented iterates over the following steps:

1. Sampling the latent observations

Given $\{\mathbf{Y}_t\}_{t=1}^T$ and conditional on $\{\mathbf{H}_t, \mathbf{G}_t, \mathbf{c}_t, \mathbf{A}_t, \boldsymbol{\Sigma}_t, S_t\}_{t=1}^T$, the model is a simple Gaussian linear state-space model where \mathbf{f}_t can be generated using the Kalman filter. Thus, in the first block we condition upon $\boldsymbol{\Psi} = \{\mathbf{S}, \mathbf{P}, \Theta\}$. The full conditional posterior of \mathbf{f}_t then

⁸What, of course, remains is the model uncertainty itself. In principle, the model specification could be treated fully Bayesian as well, thereby further endogenizing the estimation error (see Richardson & Green, 1997, for example).

⁹As mentioned in the seminal paper by Kim & Nelson (1998) this claim is not entirely correct, since “the key approximation in the Gibbs framework is associated with declaring the Gibbs chain to have converged to its steady state.”

2.2 Bayesian inference & MCMC simulation

reads

$$p(\mathbf{f} \mid \Psi, \mathbf{Y}) \propto \prod_{t=1}^T p(\mathbf{Y}_t \mid S_t, \mathbf{f}_t, \mathbf{P}, \Theta) p(\mathbf{f}_t \mid S_t, \mathbf{f}_{t-1}, \mathbf{P}, \Theta) p(\mathbf{f}_0 \mid \mathbf{P}, \Theta) \quad (2.18)$$

For a given t , however, most of the above factors are part of the normalizing constant, leaving only¹⁰

$$p(\mathbf{f}_t \mid \mathbf{f}_{-t}, \Psi, \mathbf{Y}) \propto p(\mathbf{Y}_t \mid \mathbf{f}_t, \Psi) p(\mathbf{f}_{t+1} \mid \mathbf{f}_t, \Psi) p(\mathbf{f}_t \mid \mathbf{f}_{t-1}, \Psi) . \quad (2.19)$$

The first two factors constitute the likelihood, which is defined by the underlying state space model. We assume a normal distribution for our application. Given gaussian errors, $\mathbf{f}_{t|t-1} \mid \{\mathbf{Y}_\tau\}_{\tau=1}^{t-1} \sim N(\hat{\mathbf{f}}_{t|t-1}, \mathbf{R}_{t|t-1})$ may be viewed as prior distribution of \mathbf{f}_t . Since this prior is conjugate for the normal likelihood, the posterior is also of normal form. The Kalman recursions then give the mean $\hat{\mathbf{f}}_{t|t}$ and variance $\mathbf{R}_{t|t}$ of the conditional posterior distribution for $t = 1, \dots, T$ (Meinhold & Singpurwalla, 1983)

$$\hat{\mathbf{f}}_{t|t} = \hat{\mathbf{f}}_{t|t-1} + \mathbf{A}_t^{-1} \mathbf{K}_t (\mathbf{Y}_t - \mathbf{H}_t \hat{\mathbf{f}}_{t|t-1}) \quad (2.20)$$

$$\mathbf{R}_{t|t} = \mathbf{R}_{t|t-1} - \mathbf{R}_{t|t-1} \mathbf{H}_t (\mathbf{H}_t' \mathbf{R}_{t|t-1} \mathbf{H}_t + \mathbf{G}_t \mathbf{G}_t')^{-1} \mathbf{H}_t' \mathbf{R}_{t|t-1} \quad (2.21)$$

where

$$\mathbf{K}_t = \mathbf{A}_t \mathbf{R}_{t|t-1} \mathbf{H}_t (\mathbf{H}_t' \mathbf{R}_{t|t-1} \mathbf{H}_t + \mathbf{G}_t \mathbf{G}_t')^{-1} . \quad (2.22)$$

In a backward-sampling step, these moments are smoothed to enhance efficiency of the sampler, i.e.

$$\mathbf{f}_{t|T} \mid \mathbf{Y}^T \sim N(\hat{\mathbf{f}}_{t|T}, \mathbf{R}_{t|T}) , \quad (2.23)$$

$$\hat{\mathbf{f}}_{t|T} = \hat{\mathbf{f}}_{t|t} + \mathbf{J}_t \left(\hat{\mathbf{f}}_{t+1|T} - \hat{\mathbf{f}}_{t+1|t} \right) , \quad (2.24)$$

$$\mathbf{R}_{t|T} = \mathbf{R}_{t|t} + \mathbf{J}_t \left(\mathbf{R}_{t+1|T} - \mathbf{R}_{t+1|t} \right) \mathbf{J}_t' , \quad (2.25)$$

$$\mathbf{J}_t = \mathbf{R}_{t|t} \mathbf{A}_t \mathbf{R}_{t+1|t}^{-1} . \quad (2.26)$$

The final estimate of \mathbf{f}_t is sampled from (2.23) backwards through the sample. It is important to note that the latent series is not subject to identification problems due to relabeling of the regimes. For every labeling scheme, the filtered series is equivalent. Thus any reordering (randomly or to meet a constraint) will leave the series unchanged (see Frühwirth-Schnatter, 2001a).

¹⁰Since the elements of Θ are independent of the transition probabilities \mathbf{P} , we may regard is part of the normalizing constant as well.

2. Sampling the coefficients of the transition equation

Given the missing observations from the previous step and the latent regimes \mathbf{S} , we may use the complete-data likelihood to derive the full conditional posterior of the parameters in the transition equation. There are two things worth noting: (i) conditional on \mathbf{f}_t and \mathbf{S} , the likelihood of the model is equal to that of an intervention model (Lütkepohl, 2005) and (ii) only the first N rows of the transition equation have to be estimated, whereas the remaining rows are identities. For the AR parameters and the covariance matrix of the transition equation, we set a conjugate Minnesota-type prior. Toward that end, write the transition equation in form of a matrix regression by stacking the variables, i.e.

$$\mathbf{Y}_{T \times N} = \mathbf{X}_{T \times (Np+1)} \boldsymbol{\beta}_t + \mathbf{U}_{T \times N} \quad (2.27)$$

and $\mathbf{U}_t \sim N(\mathbf{0}_{(T \times N)}, \boldsymbol{\Sigma}_{\eta,t} \otimes I_T)$. The first column in \mathbf{X} is a column of ones, representing the intercept of the model. The prior distribution for $\boldsymbol{\beta}_t$ and $\boldsymbol{\Sigma}_{\eta,t}$ is of multivariate normal-inverse Wishart form, following Doan et al. (1984); Kadiyala & Karlsson (1997); Sims & Zha (1998); Robertson & Tallman (2001); Banbura et al. (2010).

$$\boldsymbol{\beta}_t \mid \boldsymbol{\Sigma}_{\eta,t} \sim N(\boldsymbol{\beta}^{(0)}, \boldsymbol{\Psi}^{(0)}), \quad (2.28)$$

$$\boldsymbol{\Sigma}_{\eta,t} \sim IW(\mathbf{S}_0, \kappa), \quad (2.29)$$

$$\boldsymbol{\beta}^{(0)} = [\delta \mathbf{I}_{(N) \times N}, \mathbf{0}_{(N \times (p-1) + 1) \times N}]'; \quad \boldsymbol{\Psi}^{(0)} = \text{diag}(\boldsymbol{\Psi}_0^{(0)}, \boldsymbol{\Psi}_1^{(0)}, \dots, \boldsymbol{\Psi}_p^{(0)}) \quad (2.30)$$

$$\boldsymbol{\Psi}_0^{(0)} = (\lambda_0 \lambda_4)^2 \mathbf{I}_N; \quad \boldsymbol{\Psi}_i^{(0)} = \text{diag}\left((\lambda_0 \lambda_1 / s_j i^{\lambda_3})^2\right), \quad (2.31)$$

$$\mathbf{S}_0 = \text{diag}(\lambda_0^2 / s_i^2); \quad \kappa = N + 1. \quad (2.32)$$

The parameters $\delta, \lambda_0, \lambda_1, \lambda_3$ and λ_4 are hyperparameters in the discretion of the researcher. While δ sets the prior mean of the AR(1) coefficients, $\lambda_0, \lambda_1, \lambda_3$ and λ_4 control the tightness of the prior belief, which is proportional to the value of the variance of series j which we estimate as s_j^2 . We follow the usual approach and set s_j^2 equal to the variance estimated from a univariate AR(p) model. Also note that with increasing lag length i , the prior shrinks the coefficients toward δ more strongly. The constant in the model has a separate prior, which is independent of the error variances. The prior for the coefficients is identical across regimes, i.e. the prior belief assumes no regime switches in the parameters. Due to the conjugacy of the prior, the full conditional posterior of the parameters is also of multivariate normal-inverse Wishart form with posterior moments given by the Bayesian

2.2 Bayesian inference & MCMC simulation

regression:

$$\boldsymbol{\beta}_t \mid \boldsymbol{\Sigma}_{\eta,t}, \mathbf{Y} \sim N(\boldsymbol{\beta}^{(1)}, \boldsymbol{\Psi}^{(1)}) \quad (2.33)$$

$$\boldsymbol{\beta}^{(1)} = (\mathbf{X}'\mathbf{X} + (\boldsymbol{\Psi}^{(0)})^{-1})^{-1}(\mathbf{X}'\mathbf{Y} + (\boldsymbol{\Psi}^{(0)})^{-1}\boldsymbol{\beta}^{(0)}) , \quad (2.34)$$

$$\boldsymbol{\Psi}^{(1)} = T^{-1}(\mathbf{Y}'_t\mathbf{Y}_t - \boldsymbol{\beta}^{(1)'}(\mathbf{X}'\mathbf{X} + (\boldsymbol{\Psi}^{(0)})^{-1})\boldsymbol{\beta}^{(1)} + \boldsymbol{\beta}^{(0)' }(\boldsymbol{\Psi}^{(0)})^{-1}\boldsymbol{\beta}^{(0)} + \mathbf{S}_0) , \quad (2.35)$$

$$\boldsymbol{\Sigma}_{\eta,t} \mid \mathbf{Y} \sim IW(\tilde{\boldsymbol{\Sigma}}_{\eta,t}, T + \kappa) , \quad (2.36)$$

$$\tilde{\boldsymbol{\Sigma}}_{\eta,t} = (\mathbf{Y}_t - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y})'(\mathbf{Y}_t - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}) . \quad (2.37)$$

Both parameters are then generated from the above distributions for each iteration of the sampler.

3. Filtering the regimes

Given $\{\mathbf{Y}_t\}_{t=1}^T$ and conditional on $\{\mathbf{H}_t, \mathbf{G}_t, \mathbf{c}_t, \mathbf{A}_t, \boldsymbol{\Sigma}_t, \mathbf{f}_t\}_{t=1}^T$, the hidden Markov component may be filtered from the transition equation using the BHLK filter Baum, Petrie, Soules & Weiss (1970); Lindgren (1978); Hamilton (1989); Kim (1994); Chib (1996). Let $\mathbf{Y}^{(a,\tau)}$ denote $\{\mathbf{Y}_t, \mathbf{f}_t\}_{t=1}^\tau$. We suppress conditioning on other unknowns for notational convenience. The filtered regime probabilities are given via Bayes' rule as

$$p(S_t = i \mid \mathbf{Y}^{(a,t)}, \boldsymbol{\nu}) = \frac{p(\mathbf{Y}_t^{(a)} \mid S_t = i, \mathbf{Y}^{(a,t-1)}, \boldsymbol{\nu})p(S_t = i \mid \mathbf{Y}^{(a,t-1)}, \boldsymbol{\nu})}{p(\mathbf{Y}_t^{(a)} \mid \mathbf{Y}^{(a,t-1)}, \boldsymbol{\nu})} \quad (2.38)$$

Information contained in $\mathbf{Y}^{(a,\tau)}$ for $\tau > t$ is utilized in a post-processing step. Thus the filter starts from $t = T, \dots, 1$ iterating through

$$p(S_t = i \mid \mathbf{Y}^{(a,T)}, \boldsymbol{\nu}) = \sum_{j=0}^1 \frac{\mathbf{P}_{i,j} p(S_t = i \mid \mathbf{Y}^{(a,t)}, \boldsymbol{\nu}) p(S_{t+1} = j \mid \mathbf{Y}^{(a,T)}, \boldsymbol{\nu})}{\sum_{l=0}^1 \mathbf{P}_{l,j} p(S_t = l \mid \mathbf{Y}^{(a,t)}, \boldsymbol{\nu})} . \quad (2.39)$$

Finally, draw regimes from the joint conditional posterior as

$$p(S_t \mid S_{t+1}) = \frac{p(S_t, S_{t+1} \mid \mathbf{Y}_t)}{p(S_{t+1} \mid \mathbf{Y}_t)} . \quad (2.40)$$

Krolzig (1997) provides a detailed description.

4. Generating the transition probabilities

Given $\mathbf{S} = \{S_t\}_{t=1}^T$, the transition matrix \mathbf{P} may be estimated counting the number of transitions from $S_{t-1} = i$ to $S_t = j$ for $i, j \in \{0, 1\}$. The likelihood for the discrete random variable S_t is multinomial. We set a conjugate Dirichlet prior for each row $p_{i,\cdot}$ of \mathbf{P} , i.e.

$$p(p_{i,\cdot}) = \prod_{j=0}^1 p_{i,j}^{\alpha_{i,j}-1} \quad (2.41)$$

2.2 Bayesian inference & MCMC simulation

where $\alpha_{i,\cdot}$ is a hyperparameter. Setting $\alpha_{i,\cdot} = 1$ constitutes a diffuse prior. Since we only consider two regimes in this paper, the full conditional posterior for $p_{i,\cdot}$ reduces to a Beta distribution

$$p_{11} \sim \text{Beta}(\alpha_{1,1} + N(S_{t-1} = 1, S_t = 1), \alpha_{1,2} + N(S_{t-1} = 1, S_t = 2)) , \quad (2.42)$$

$$p_{22} \sim \text{Beta}(\alpha_{2,1} + N(S_{t-1} = 2, S_t = 1), \alpha_{2,2} + N(S_{t-1} = 2, S_t = 2)) . \quad (2.43)$$

where $N(A, B)$ is the operator counting the number of instances where event A and B occur.

5. Permutation

Introducing regime switching into a model as above is fairly straightforward conceptually. At the same time, however, one has to be careful when estimating such models. On the one hand finite mixture models, either as standard mixture models or Markov mixture models, provide an easy tool to accommodate a wide variety of distributional features in the data observed. They comprise a finite or even infinite number of component distributions, possibly discrete and continuous (see e.g. Schwaab et al., 2009).

On the other hand, the extreme flexibility of mixture distributions comes at the cost of several issues, e.g. identification of the mixture components. This is also known as relabeling problem (Redner & Walker, 1984; Xu & Jordan, 1996; Celeux et al., 2000). To illustrate the label switching problem, recall that a parametric family of distributions is identified in a strict sense, iff

$$p(\mathbf{y} \mid \boldsymbol{\nu}) = p(\mathbf{y} \mid \tilde{\boldsymbol{\nu}}) \text{ almost everywhere} \Rightarrow \tilde{\boldsymbol{\nu}} = \boldsymbol{\nu} . \quad (2.44)$$

This is clearly given for any single normal distribution with $\boldsymbol{\nu} = (\mu, \sigma^2)$, but it is not guaranteed for a normal mixture with more than a single component. It is easy to show and also quite intuitive, that permuting the labels of the regimes by some function

$$\rho_S : \{1, \dots, K\} \mapsto \{1, \dots, K\} \quad (2.45)$$

renders a mixture of two or more normals unidentified in the strict sense above. This means that for a K -component mixture distribution, there exist $K!$ subspaces of the posterior likelihood that differ only in their labeling of the states, but yield the same likelihood of the data.¹¹ While this is not a severe statistical problem, because the relation between parameters is preserved among different labeling schemes, it is something that has to be handled carefully when estimating the model. An immediate implication for MCMC sampling from the joint posterior, is that one cannot have any indication from which

¹¹There exists only a $K!/L!$ such subspaces in case L of the parameters in $\boldsymbol{\nu}$ are identical across regimes.

2.2 Bayesian inference & MCMC simulation

labelling subspace the sample arises. The algorithm will most likely jump between different subspaces and lead to fuzzy inference. This is aggravated by the fact that unconstrained Gibbs sampling tends to stay within current labeling subspaces and only jump between them occasionally. Thus the algorithm not only yields unidentified samples, but is also slow mixing.

There are a number of workarounds proposed to surmount identification and efficiency issues. Because it is widely noticed in the literature that the Gibbsian paradigm provides a formidable approach to estimation of such models, most of these approaches adopt a Bayesian estimation technique (see Marin et al., 2005, for a survey).

The two easiest workarounds for this problem are (i) permutation sampling or re-ordering (Frühwirth-Schnatter, 2001b; Marin et al., 2005), and (ii) constrained or naive Gibbs sampling (see Albert & Chib, 1993; Kim & Nelson, 1998, among many others). The latter approach achieves identification by *a priori* imposing an identification constraint, e.g. through truncated sampling. Celeux et al. (2000) demonstrate, however, that in case of a poor constraint – poor in the sense that it does not provide a clear-cut hyperplane between the multiple modes of the posterior – constrained sampling may approximate the posterior distribution very poorly.

Permutation sampling, in contrast, works along the lines of the Gibbs sampling procedure outlined above. It appends an additional step at the end of each iteration, however, in which regime-dependent parameters are permuted. This is done via the function ρ_S , which selects a permutation $\rho_S(1), \dots, \rho_S(K)$ of the current labeling $1, \dots, K$ and applies the permutation to the parameters. The permuted parameters are then handed over to the next iteration of the Gibbs sampler. The chief attraction of this sampling scheme, compared to simple unconstrained Gibbs sampling, is that it induces balanced label switching and thus the algorithm visits different areas of the parameter space with equal probability. In a post-processing step, clustering algorithms on subsets of the parameter vector may be used to pin down suitable identification constraints. In case such constraints are chosen carefully to separate the posterior modes, they do not run the risk of introducing an artificial bias.

Finally, the issue of model selection is important. In the Bayesian framework the marginal data density $f(\mathbf{Y})$ after integrating out the unknowns is used as a selection device. However, approximation of the marginal data density is far from trivial in practice. Chib (1995) proposes a widely-used method for approximating the marginal data density in a Bayesian setting, but e.g. Frühwirth-Schnatter (2004) shows that the approximation may often be rather poor due to the label switching problem. While this is an interesting problem for applications, we view this as a separate topic largely ignore this problem here and instead refer the reader to Frühwirth-Schnatter (2004); Marin & Robert (2008).

3 Applications

In this section we illustrate the posterior simulation algorithm with two examples.

First, we conduct a simulation study. This provides first and foremost a consistency check for our algorithm. The finding suggests that inference on the latent series and the regime process are fairly precise and converge reasonably. As one could have expected, however, the fact that different augmented data set is used for estimating the remaining parameters in every sweep of the sampler leads to some pronounced imprecision with respect to the VAR parameters. This illustrates the point raised earlier that the measurement error connected to the latent series is not negligible. Interestingly, this does not have any serious repercussions on the filtered series and the regime process, which both converge to the true values.

3.1 A numerical simulation study

We conduct a small-scale simulation study, to provide a consistency check for the algorithm and to gauge the precision of the sampler for a given number of iterations. We simulated a trivariate VAR(2) where one series, $y_{1,t}$ say, was constructed such that

$$y_{1,t} = 1/3y_{1,t}^* + 2/3y_{1,t-1}^* + y_{1,t-2}^* + 2/3y_{1,t-3}^* + 1/3y_{1,t-4}^* , \quad (3.1)$$

and $y_{1,t}^*$ were not observable. The number of observations was set to $T = 500$, which is a number that seems reasonable for empirical applications. To initialize estimation, we provided a crude estimate $\hat{y}_{1,t}^* = 1/2(y_{2,t} + y_{3,t})$ for the latent series. We made all parameters of the transition equation regime-dependent. Regimes are indicated as shaded areas, e.g. in figure 2.

To find initial values for the parameters and regimes in the transition equation, we employ an approximate EM procedure. This pseudo-EM algorithm assumes away the mixed-frequency nature of the data and finds a posterior mode given the approximation $\hat{y}_{1,t}^*$. We then take 10,000 burn-in and 20,000 posterior samples using the Gibbs sampler outlined in the previous section.

For posterior simulation we set the following hyperparameters:

$$\delta = 0 ; \lambda_0 = 0.95 ; \lambda_1 = 0.95 ; \lambda_3 = 1.1 ; \lambda_4 = 10 . \quad (3.2)$$

These hyperparameters reflect the initial belief that series are white noise ($\delta = 0$) and stationary ($|\delta| < 1$). The choice for λ_0 to λ_4 results in fairly little prior precision for the AR parameters. Since these values are paired with an estimate of the individual variances, the prior precision depends on the data. We adopt an uninformative prior for the transition probabilities, i.e. we set $\alpha_i = 1$ for $i = 1, 2$.

3.1 A numerical simulation study

Figures 2 to 10 show some results from the exploratory Bayesian analysis after running a random permutation sampler with 2,000 burn-in and 8,000 posterior sampling iterations. As can be seen, the sampler finds two clearly separated clusters for the intercepts and the transition probabilities. For the variances, clusters are less obviously separated. This matches the true variance-covariance matrices quite well, which also differ only slightly. Since the latent series filtered from the state space are not subject to the relabeling problem (see Frühwirth-Schnatter, 2001a), we may interpret the sample of filtered series as a valid sample from its posterior. This in turn also means we may take its posterior mean as starting value for the identified permutation sampler in order to enhance sampling efficiency. Based on the clear separation visible in fig. 3, we choose to identify the model in terms of the intercept of $y_{2,t}$. Hence, for each iteration of the sampler regime 1 is identified by $c_{2,1} < c_{2,2}$. All other regime dependent variables are permuted accordingly.

With this identification constraint defined, we run an identified permutation sampler with 4,000 burn-in and 6,000 posterior sampling iterations. To assess convergence within the Markov chain, we employed Geweke's (1991) z-score. Figure 9 shows the z-score for all AR parameters and the intercepts. The results confirm the impression that the Markov chain has converged, which we also conclude from visually inspecting fig. 6, 7 and 8. Table 1 shows summary statistics for the posterior of the parameters VAR parameters. It reveals that the parameters are sampled with adequate precision. While for some parameters, most notably for the intercepts, posterior means closely match the true parameters, it is less true for some AR parameters. One also sees from table 1 that posterior distributions for the parameters are fairly wide compared to standard Bayesian VAR models without mixed frequencies. Given the mixed-frequency nature of the data, this is not a particularly surprising result. However, it is important to note that standard or frequentist models compute confidence regions under the assumption of no estimation error in the latent data. The results from table 1 indicate that for the VAR parameters this might give a false sense of security. Table 2 shows the results for the covariance matrix of the errors. Again, posterior means match the true values quite closely, while HPD regions are fairly wide.

Finally, the latent series in our model are also sampled with satisfying precision. Juxtaposing the posterior mean of the filtered series to the true series reveals that the closely tracks the true series (cf. fig. 10). The correlation between the true series and the filtered posterior mean is 0.83. However, it also shows that point estimates should be taken with great care. While visual inspection leads to the impression that there is a close match between the true series and the filtered mean, figure 11 shows, the estimation uncertainty associated with the latent series as measured by the HPD intervals. Figure 11 also shows that credibility regions are not uniform across the sample and sometimes widen considerably. Where posterior intervals widen, the posterior mean usually shows a higher deviation from the true value. Consulting figure 12, however, we see that the 68%-HPD

3.2 The US Business Cycle

region overlaps the true series over almost over the entire sample.

Concerning the latent regime indicator, inference is similarly reliable. Figure 13 shows the filtered and smoothed estimates of the regime probabilities. As can be seen, filtered and smoothed probabilities are in complete concordance with the true underlying regimes.

We take away several key insights from this exercise. First, posterior inference on mixed-frequency Markov mixture models is precise in terms of the latent series and the regime process. Both can be assumed to be approximated to a satisfying degree. Second, inference on the VAR parameters is more difficult. While intercept terms show a high degree of precision, autoregressive parameters are more delicate to estimate. This is what one could have expected, given the fact that in our example one series out of three series changes with each sweep of the sampler.

3.2 The US Business Cycle

In this section we apply the model to a real data set. We consider monthly data on industrial production, M2 money stock and the PPI (all converted into monthly percentage changes) as well as quarterly data on GDP growth. On the monthly grid, this amounts to 523 observations. The source of the data is the FRED data service of the Reserve Bank of St. Louis. In the spirit of Hamilton (1989), we fit a mixed-frequency VAR(6) to the data with two regimes representing economic expansions and contractions.¹²

We then follow the same steps as above. First, we set a prior distribution for the VAR parameters and the transition probabilities. The prior for the AR coefficients and the variances are again

$$\delta = 0 ; \lambda_0 = 0.95 ; \lambda_1 = 0.95 ; \lambda_3 = 1.1 ; \lambda_4 = 10 . \quad (3.3)$$

This time we adopt an informative prior for the transition probabilities, i.e. we set α_i for $i = 1, 2$ such that the prior weights roughly match the empirical frequencies of NBER cycles over the past 50 years.¹³

As we illustrated above, in order to choose a suitable identification constraint for the final model, we use the explorative methodology outlined in Frühwirth-Schnatter (2001b). Thus, we first run a random permutation sampler with 1,000 burn-in iterations and 2,000 posterior sampling iterations. We then use K -means clustering to sort sampled values into two groups (see Hartigan, 1975). Figure 14 shows a clear separation in terms of the variances of all series. Clustering on the intercepts in a similar fashion shows that no

¹²As previously mentioned, model selection in mixture models is not trivial in practice. Hence, we refrain from a data-driven model selection step here, but note instead that our empirical results are not particularly sensitive to lag selection.

¹³Since we define a Dirichlet prior for the transition probabilities, the hyperparameters α_i may be interpreted as number of observations for each category i .

3.2 The US Business Cycle

separation along this line is possible. This seems at odds with intuition at first sight, because business cycle regimes are thought to be identified in terms of levels of growth. It is worth keeping in mind, however, that equal intercept terms do not necessarily translate into equal unconditional means of the series across regimes. Thus, we first impose an identification constraint in terms of the variances of the M2 money stock

$$\text{Var}(\text{M2} \mid \text{regime 1}) \leq \text{Var}(\text{M2} \mid \text{regime 2}) .$$

We then check the meaning of the regimes associated with this constraint in a postestimation step.

With this identification constraint, we run the permutation sampler 6,000 times and discard the first 2,000 iterations as the burn-in sample. As noted above, we use two key outputs that we use in the post-estimation analysis phase: (i) the recession probabilities and (ii) the filtered GDP growth series. Eye-balling figure 15, we see that the Kalman filter and smoother returns plausible values for GDP growth rates. The series conforms with stylized facts, such as the Great Moderation, i.e. the observation that volatility in macroeconomic growth rates is smaller during the years following 1984. Considering the behavior of the filtered growth rates around NBER recession dates, we can see that growth rates settle down to zero or become negative around these dates. As expected, the 1990-91 and the 2001 recession are seen as far less severe than those before 1985 or the most recent recession. This is confirmed by other studies, which rank the 2001 recession the mildest and shortest of all 11 post-World War II recessions (see Jeliazkov & Liu, 2010). Turning to the 2008 Global Financial Crisis, it becomes clear that this was among the worst economic recessions in the post-war period.

It then has to be clarified that the regimes we identify are representing periods of negative growth in real economic variables, i.e. GDP and industrial production. Toward this end, we computed the unconditional mean $\boldsymbol{\mu}$ of the series as

$$\boldsymbol{\mu} = (\mathbf{I}_n - \boldsymbol{\Phi}(L))^{-1} \mathbf{c} .$$

Table 3 shows the results for all four series. Obviously, the regimes represent periods of economic expansions (regime 2) and economic contractions (regime 1).

Because other studies using mixed-frequency models are mostly silent about confidence regions for the filtered values, we construct the highest-posterior density (HPD) interval for our series. Figure 19 depicts the filtered and standardized values for GDP growth together with the 95% and the 68% HPD interval. It is worth stressing once more that by treating missing values in a fully Bayesian or probabilistic fashion, uncertainty surrounding the estimation error of the missing values is resolved. This means that in the given model, with the given sample size and the given number of iterations inference is exact. Our

3.2 The US Business Cycle

results show that the sampled values fluctuate surprisingly little, with a width of the 69% HPD interval roughly equal to 0.18 percentage points on average. Thus, point forecasts and end-of-sample values filtered from mixed-frequency VARs can be regarded as valuable and fairly precise on average. However, it must be kept in mind that end-of-sample values suffer from the well-known revision bias, and therefore should be treated carefully.

Figure 16 shows the filtered and the full-sample smoothed recession probabilities. There are several features of the recession probabilities worth mentioning. First, they show a relatively high concordance with NBER recession dates, i.e. they peak around recessions from 1976 to 2008. As expected the probabilities behave more erratic than recession probabilities filtered from quarterly models. The posterior means for the transition probabilities are

$$\begin{bmatrix} 0.899 & 0.101 \\ 0.283 & 0.717 \end{bmatrix}. \quad (3.4)$$

Hence, both regimes are found to be highly persistent and expansions have a longer average duration compared to recessions. Considering the overall picture, we see that it matches NBER dates quite well, but perform significantly worse between 2001 and 2008. During this time there are several peaks in the series of recession probabilities that differ from NBER recession dates. Figure 18 depicts the recession probabilities around this time. There are three peaks that are in discordance with official NBER dates: (i) January to May 2003, (ii) April to July 2004 and (iii) September to December 2005. The first period can be associated with the ongoing downturn on US stock markets and weak growth figures. Further, the collapse of Enron and increasing concerns about widespread misreporting of firms, e.g. by WorldCom, put pressure on overall conditions through the end of 2002. In terms of our model, the sudden drop of producer prices is the harbinger of a recessionary period. The second period unfortunately is completely unrelated to any obvious periods of slow growth. While both growth figures for GDP and industrial production are somewhat weak, this seems to be within the range of natural fluctuation. This is less so for the peak around the end 2005. Economic growth weakened unexpectedly in the fourth quarter of 2005. Figure 17 illustrates this. Growth of GDP and industrial production slowed down suddenly in September and October 2005. However, in terms of a nationwide economic recession, the downturn did not last long enough and was not pronounced enough to represent a recession. This illustrates already a potential weakness of the mixed-frequency VAR model: It sometimes fails to filter unsystematic movements in the data from systematic movements. Dynamic factor models that allow for regime switching in the factor equation might be less prone to this problem, since they smooth outliers across series.

4 Conclusion

In this paper we reconsider estimation of mixed-frequency VAR models. While standard mixed-frequency VAR models can be suitably estimated using versions of the EM algorithm (Dempster et al., 1977), introducing Markov regime switching or more generally a mixture sampling density potentially leads to convergence to spurious modes of the likelihood (or posterior) surface. Similarly, it is shown by Frühwirth-Schnatter (2001a,b) and Celeux et al. (2000) that classical MCMC simulation of mixture models suffers from the label switching problem.

The estimation technique put forward here surmounts both problems by using a random permutation Gibbs sampler (Frühwirth-Schnatter, 2001a) to explore the posterior distribution. Moreover, the latent series in the mixed-frequency VAR model usually has a direct structural meaning, e.g. it represents latent monthly GDP growth. Standard approaches to estimation, i.e. EM or MLE, do not account for the joint dependence of the latent series and the estimated parameters. Instead they treat as mutually independent. The extent to which this neglects estimation error uncertainty depends on the application, but MCMC simulation enables us to explicitly take into account the joint dependence of all unknowns in the model.

We illustrate above how the algorithm works, both in a controlled environment of a simulation study and a small empirical example. The simulation study reveals that the algorithm delivers satisfactory approximations of the data generating process. Applying the model to macroeconomic data for the US also produced reasonable results

References

- Albert, J. & Chib, S. (1993). Bayes inference via Gibbs sampling of autoregressive time series subject to Markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1), 1–15.
- Aruoba, S. B., Diebold, F. X., & Scotti, C. (2009). Real-time measurement of business conditions. *Journal of Business and Economic Statistics*, 27(4), 417–427.
- Banbura, M., Giannone, D., & Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1), 71–92.
- Baum, L., Petrie, T., Soules, G., & Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of Markov chains. *The Annals of Mathematical Statistics*, 41(1), 164–171.
- Camacho, M. & Perez-Quiros, G. (2010). Introducing the Euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics*, 25(4), 663–694.
- Celeux, G., Hurn, M., & Robert, C. (2000). Computational and inferential difficulties with mixture posterior distributions. *Journal of the American Statistical Association*, 957–970.
- Chib, S. (1995). Marginal Likelihood from the Gibbs Output. *Journal of the American Statistical Association*, 90(432), 1313.
- Chib, S. (1996). Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75(1), 79–97.
- Clements, M. & Galvao, A. (2008). Macroeconomic Forecasting With Mixed-Frequency Data. *Journal of Business and Economic Statistics*, 26(4), 546—554.
- De Mol, C., Giannone, D., & Reichlin, L. (2008). Forecasting using a large number of predictors: Is Bayesian shrinkage a valid alternative to principal components? *Journal of Econometrics*, 146(2), 318–328.
- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 1–38.
- Doan, T., Litterman, R., & Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3(1), 1–100.
- Frühwirth-Schnatter, S. (2001a). Fully Bayesian analysis of switching Gaussian state space models. *Annals of the Institute of Statistical Mathematics*, 53(1), 31–49.
- Frühwirth-Schnatter, S. (2001b). Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453), 194–209.

References

- Frühwirth-Schnatter, S. (2004). Estimating marginal likelihoods for mixture and Markov switching models using bridge sampling techniques. *The Econometrics Journal*, 7(1), 143–167.
- Frühwirth-Schnatter, S. (2006). *Finite mixture and Markov switching models*. New York: Springer.
- Gelfand, A., Hills, S., Racine-Poon, A., & Smith, A. (1990). Illustration of Bayesian inference in normal data models using Gibbs sampling. *Journal of the American Statistical Association*, 972–985.
- Gelfand, A. & Smith, A. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American statistical association*, 85(410), 398–409.
- Geman, S. & Geman, D. (1984). Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6, 721–741.
- Geweke, J. (1991). Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. Bernardo, J. Berger, A. Dawid, & A. Smith (Eds.), *Bayesian Statistics 4*. Oxford, UK.: Clarendon Press.
- Ghysels, E., Santa-Clara, P., & Valkanov, R. (2004). The MIDAS touch: MIXed DATA Sampling regression models. mimeo, Chapel Hill.
- Goldfeld, S. & Quandt, R. (1973a). A Markov model for switching regressions. *Journal of Econometrics*, 1(1), 3–15.
- Goldfeld, S. M. & Quandt, R. E. (1973b). The estimation of structural shifts by switching regressions. *Annals of Economic and Social Measurement*, 2(4), 475–485.
- Guérin, P. & Marcellino, M. (2011). Markov-Switching MIDAS Models. *EUI Working Paper Working Papers*.
- Hamilton, J. (2010). Calling recessions in real time. *International Journal of Forecasting*, (forthcoming).
- Hamilton, J. D. (1989). A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica*, 57(2), 357–384.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton University Press.
- Hartigan, J. (1975). *Clustering algorithms*. John Wiley & Sons, Inc. New York, NY, USA.
- Harvey, A. & Pierse, R. (1984). Estimating missing observations in economic time series. *Journal of the American Statistical Association*, 79(385), 125—131.
- Jeliazkov, I. & Liu, R. (2010). A model-based ranking of US recessions. *Economics Bulletin*, 30(3), 2289–2296.

References

- Kadiyala, K. & Karlsson, S. (1997). Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12(2), 99–132.
- Kholodilin, K. (2001). Markov-Switching Common Dynamic Factor Model with Mixed-Frequency Data. *Discussion Papers (IRES-Institut de Recherches*
- Kim, C. J. (1994). Dynamic linear models with Markov-switching. *Journal of Econometrics*, 60(1-2), 1–22.
- Kim, C. J. & Nelson, C. 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(2), 188–201.
- Kim, C. J. & Nelson, C. R. (1999). *State-space Model with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press.
- Koop, G. & Korobilis, D. (2010). Bayesian Multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics*, 3(4), 267–358.
- Krolzig, H. (1997). *Markov-switching vector autoregressions: modelling, statistical inference, and application to business cycle analysis*. Berlin: Springer.
- Kuzin, V., Marcellino, M., & Schumacher, C. (2010). MIDAS vs. mixed-frequency VAR: Nowcasting GDP in the Euro Area. *International Journal of Forecasting*.
- Lindgren, G. (1978). Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics*, 5(2), 81–91.
- Lütkepohl, H. (2005). *New Introduction to Multiple Time Series Analysis* (1st ed.). Berlin: Springer.
- Mariano, R. & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of Applied Econometrics*, 18, 427–443.
- Mariano, R. & Murasawa, Y. (2004). Constructing a coincident index of business cycles without assuming a one-factor model.
- Marin, J., Mengersen, K., & Christian, R. (2005). Bayesian modelling and inference on mixtures of distributions. *Handbook of Statistics: Bayesian Thinking, Modeling and Computation*, 25, 459–483.
- Marin, J. & Robert, C. (2008). Approximating the marginal likelihood in mixture models. *Arxiv preprint arXiv:0804.2414*.
- Meinhold, R. J. & Singpurwalla, N. D. (1983). Understanding the Kalman filter. *American Statistician*, 37(2), 123–127.

References

- Proietti, T. & Moauro, F. (2006). Dynamic factor analysis with non-linear temporal aggregation constraints. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 55(2), 281–300.
- Redner, R. & Walker, H. (1984). Mixture densities, maximum likelihood and the EM algorithm. *SIAM Review*, 26(2), 195–239.
- Richardson, S. & Green, P. (1997). On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society Series B*, 59, 731–792.
- Robertson, J. & Tallman, E. (2001). Improving federal-funds rate forecasts in VAR models used for policy analysis. *Journal of Business and Economic Statistics*, 19(3), 324–330.
- Ryden, T. (2008). EM versus Markov chain Monte Carlo for estimation of hidden Markov models: A computational perspective. *Bayesian Analysis*, 3(4), 659–688.
- Schwaab, B., Creal, D., Koopman, S. J., & Lucas, A. (2009). A dynamic factor model for mixed-measurement mixed-frequency panel data, with application to credit and recovery risk. *VU Amsterdam and Tinbergen Institute Job Market Paper*, 1–38.
- Sims, C. (1993). A Nine-Variable Probabilistic Macroeconomic Forecasting Model. In J. Stock & M. Watson (Eds.), *Business cycles, indicators, and forecasting* chapter 7, (pp. 179–204). Chicago: University Of Chicago Press for the NBER.
- Sims, C. & Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4), 949–968.
- Tanner, M. & Wong, W. (1987). The calculation of posterior distributions by data augmentation. *Journal of the American Statistical Association*, 82(398), 528–540.
- van Dyk, D. & Meng, X. L. (2001). The art of data augmentation. *Journal of Computational and Graphical Statistics*, 10, 1–111.
- Xu, L. & Jordan, M. (1996). On convergence properties of the EM algorithm for Gaussian mixtures. *Neural computation*, 8(1), 129–151.
- Zadrozny, P. A. (1988). Gaussian-Likelihood of countinuous-time ARMAX models when data are stocks and flows at different frequencies. *Econometric Theory*, 4(1), 108–124.
- Zadrozny, P. A. (1990). Estimating a multivariate ARMA model with mixed-frequency data: An application to forecating U.S. GNP at monthly intervals. *Federal Reserve Bank of Atlanta Working Paper, Series No.*

A Figures

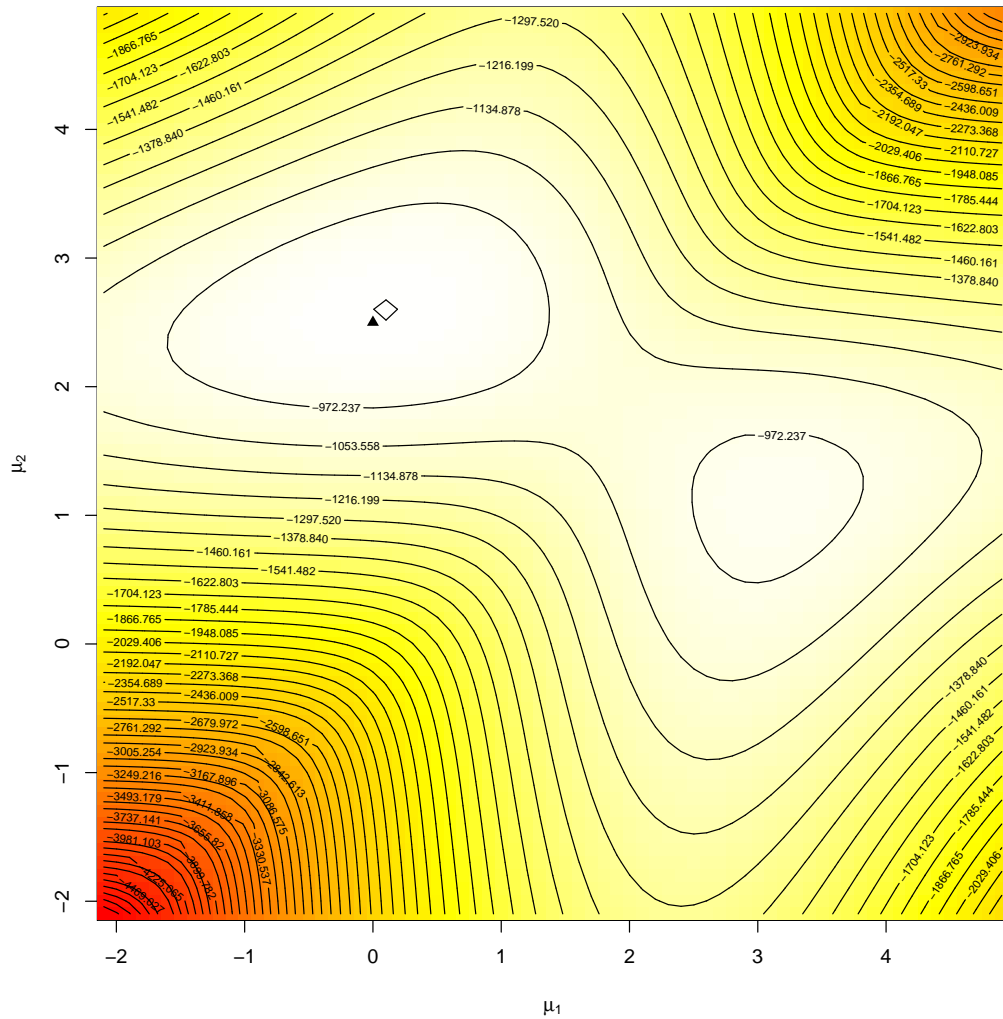


Figure 1: Shows the log-likelihood surface of the mixture $0.7 \cdot N(2.5, 1) + 0.3 \cdot N(0, 1)$ for a simulated data set of 500 observations. The mode in the lower right corner of the plot is a spurious mode that is not located at the true parameter vector $(\mu_1, \mu_2) = (0, 2.5)$ (indicated by the little black triangle).

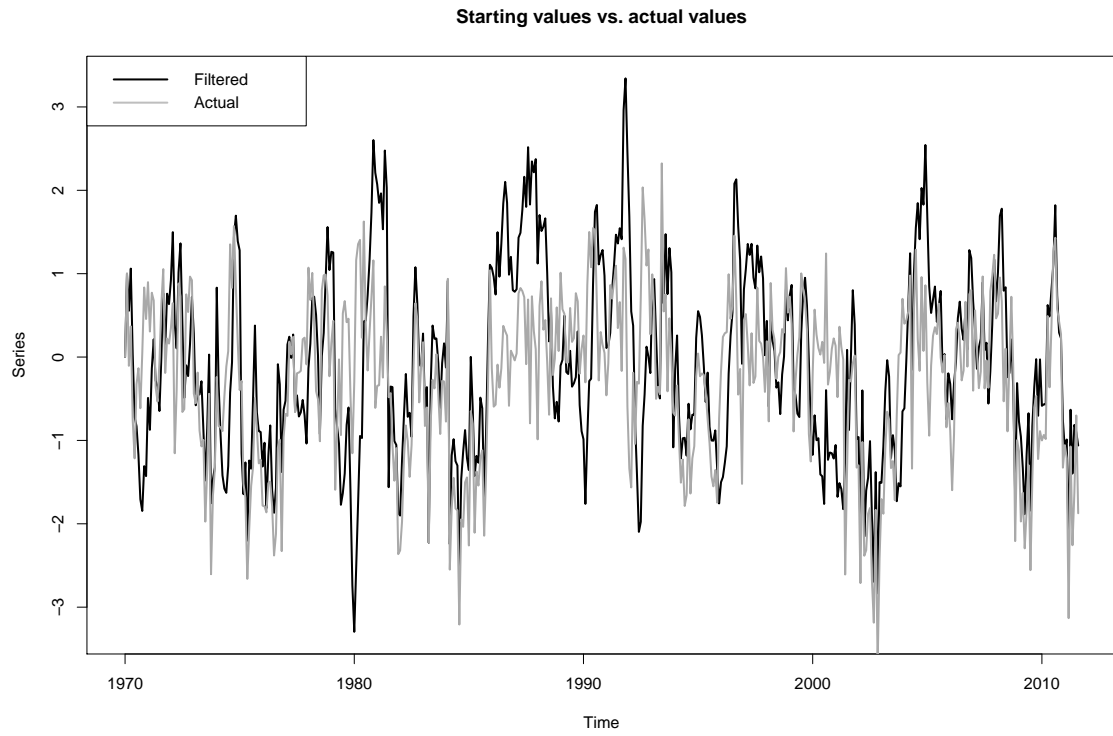


Figure 2: Shows the starting value for the latent series (mean of two higher frequent series) and the actual latent series (RMSE = 1.082676). Shaded areas indicate regimes.

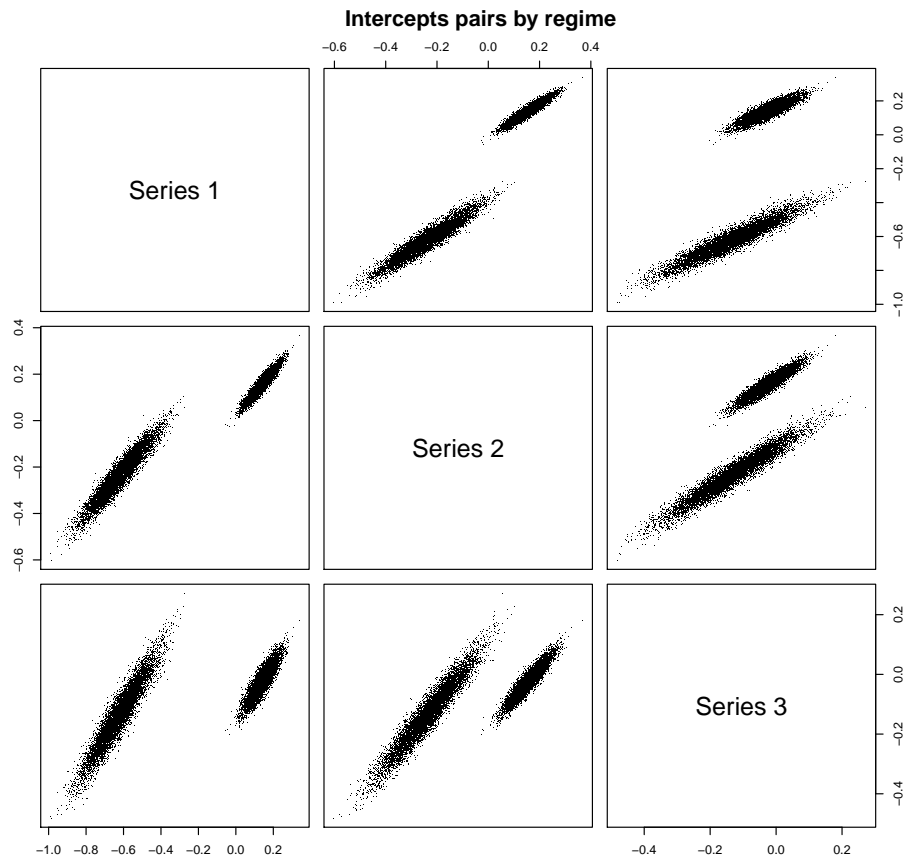


Figure 3: Scatterplot for intercepts by regime. Each box depicts the sampled values of intercepts for the two variables mentioned on the main diagonal of the plot.

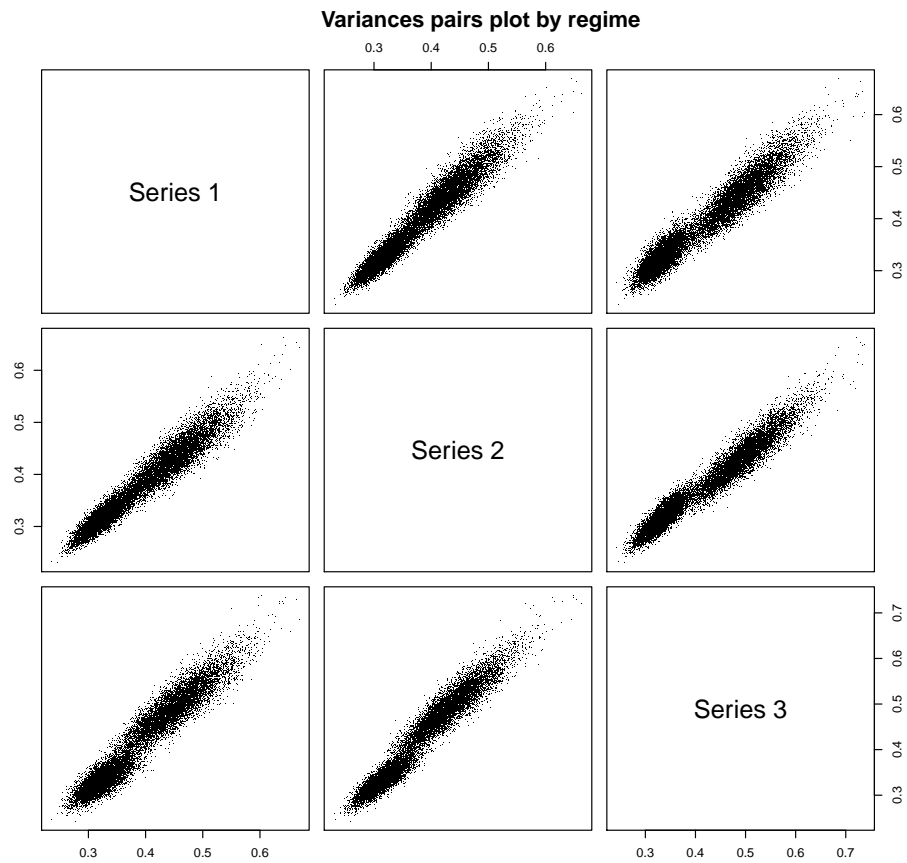


Figure 4: Scatterplot for variances by regime. Each box depicts the sampled values of sampled variances for the two variables mentioned on the main diagonal of the plot.

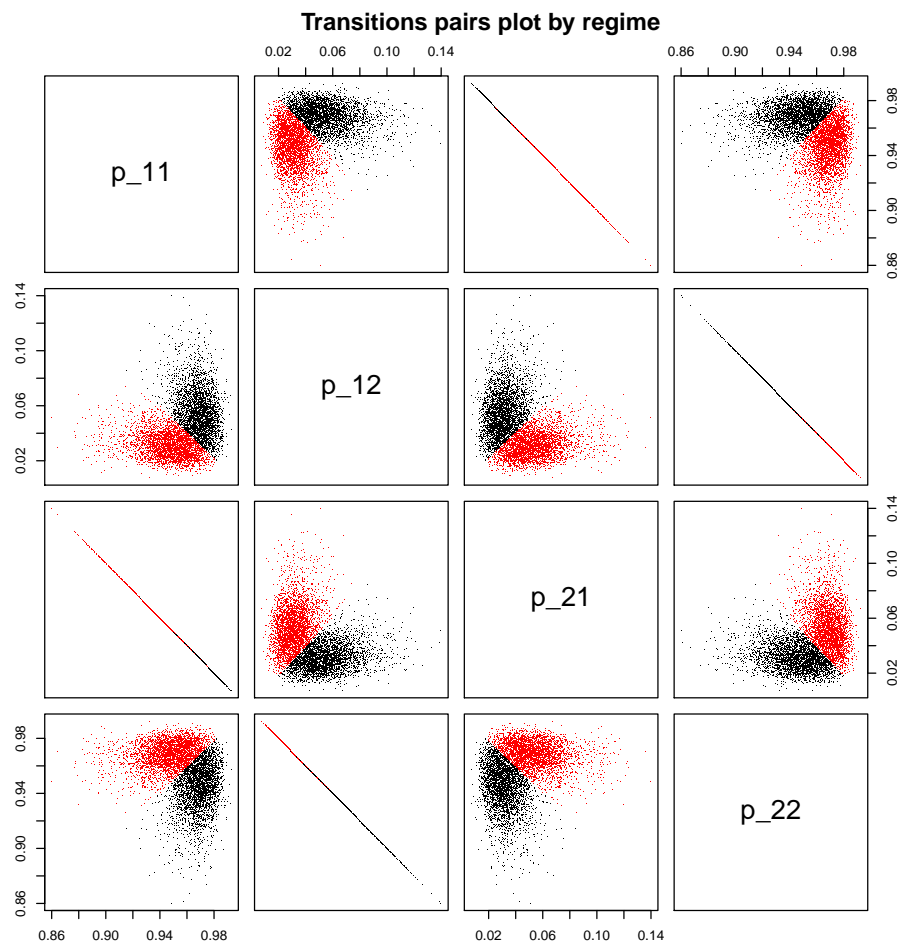


Figure 5: Scatterplot for transition probabilities by regime. Each box depicts the sampled values of sampled transition probabilities mentioned on the main diagonal of the plot.

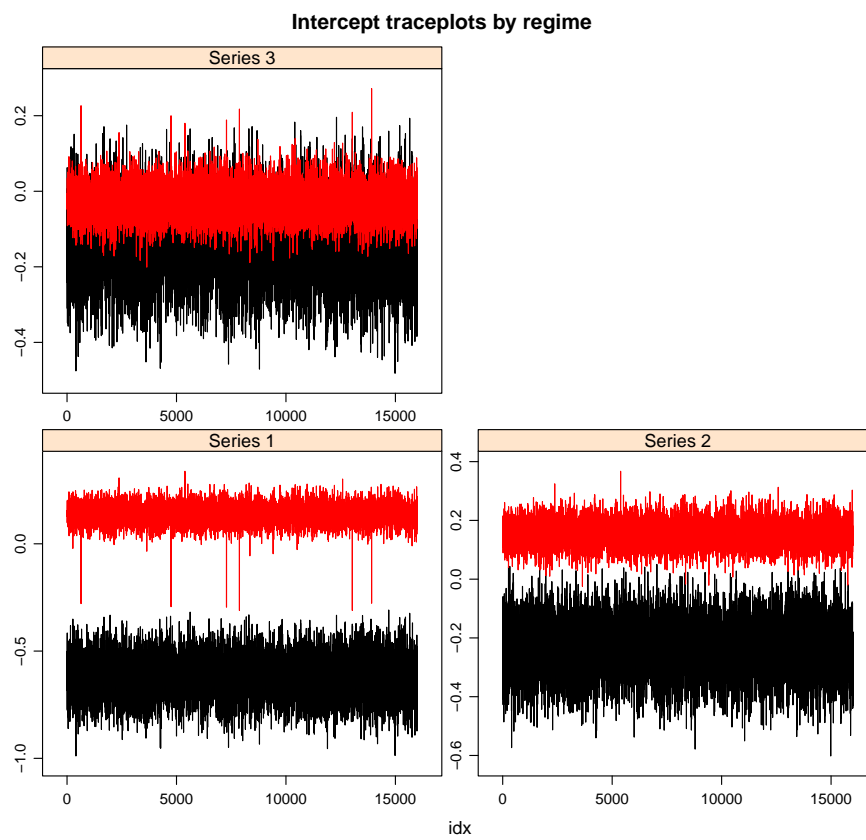


Figure 6: Traces for intercepts by regime.

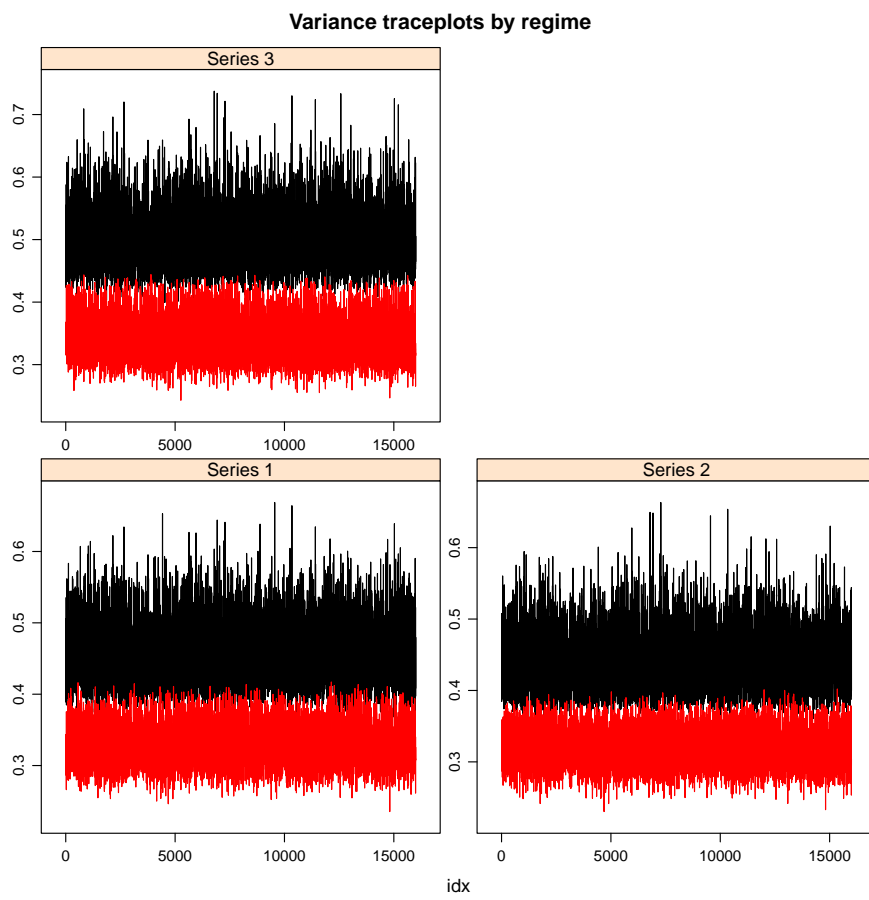


Figure 7: Traces for variances by regime.

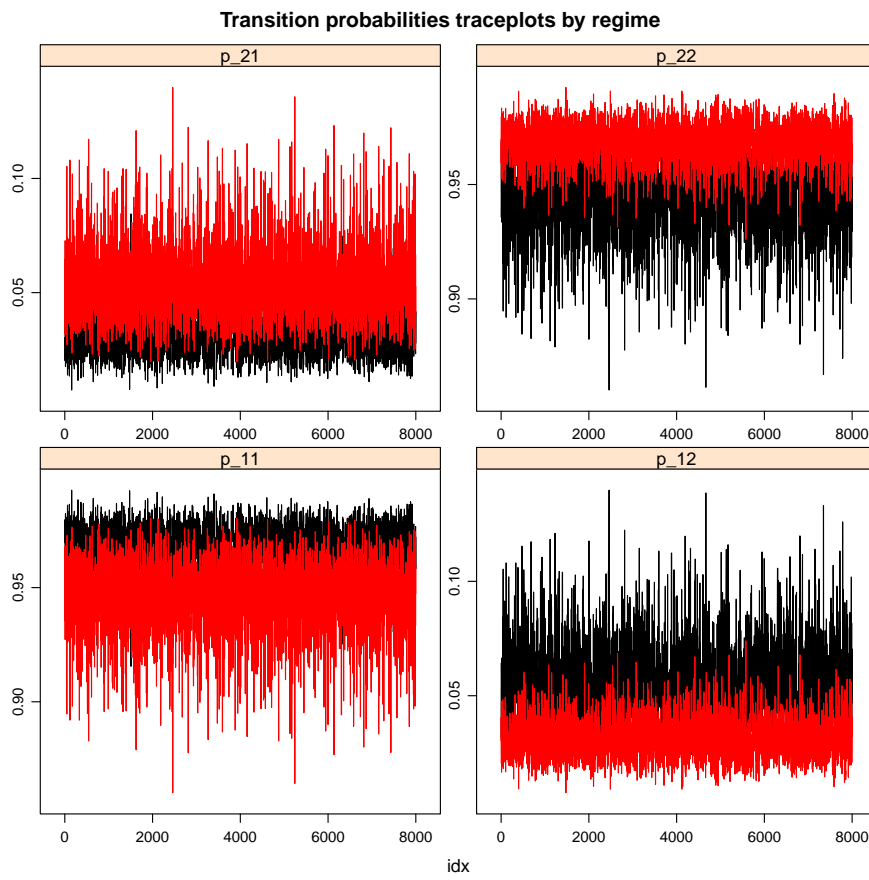


Figure 8: Traces for transition probabilities by regime.

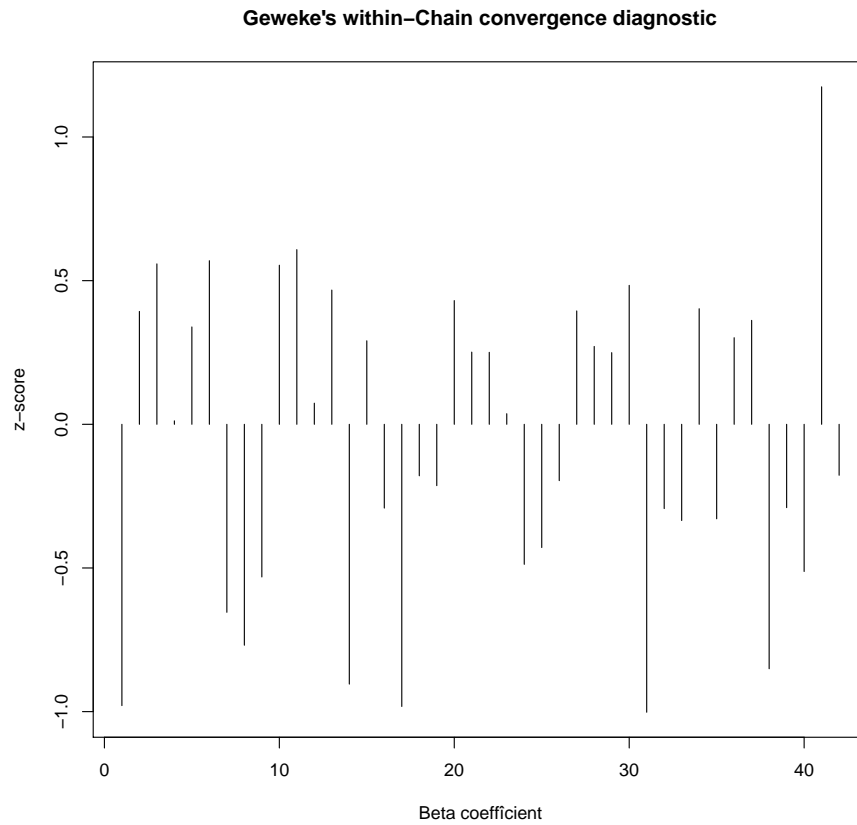


Figure 9: This figure shows the results from a test for equality of the means of the first and last part of a Markov chain (here the first 10% and the last 50%). Under the hypothesis that the samples are drawn from the limiting distribution of the chain, the two means are equal and Geweke's statistic is a standard z-statistic.

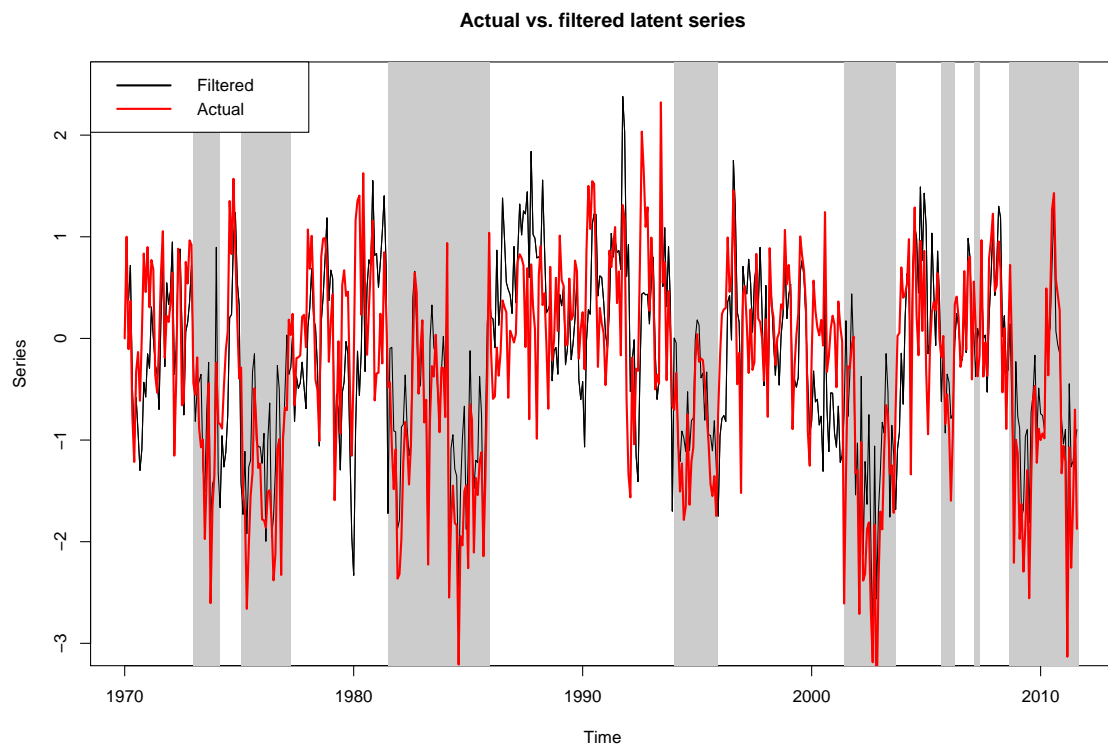


Figure 10: Shows the posterior mean of filtered values for the latent series after running a random permutation sampler on the posterior distribution with 4,000 burn-in iterations and 6,000 sampling iterations. Shaded areas indicate regime shifts.

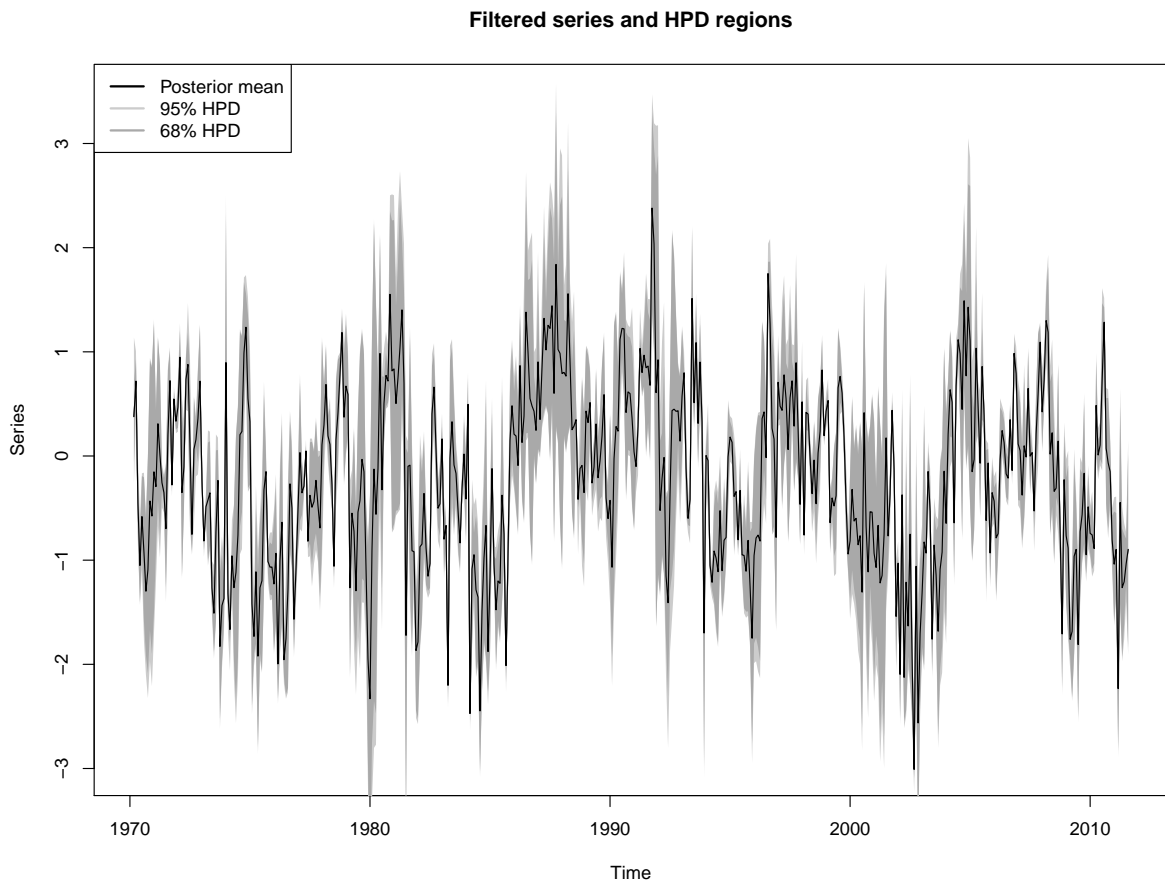


Figure 11: Shows the posterior mean of filtered values for the latent series after running a permutation sampler under the identification constraint $c_2^1 < c_2^2$ with 4,000 burn-in iterations and 6,000 sampling iterations and highest posterior density regions (shaded areas).

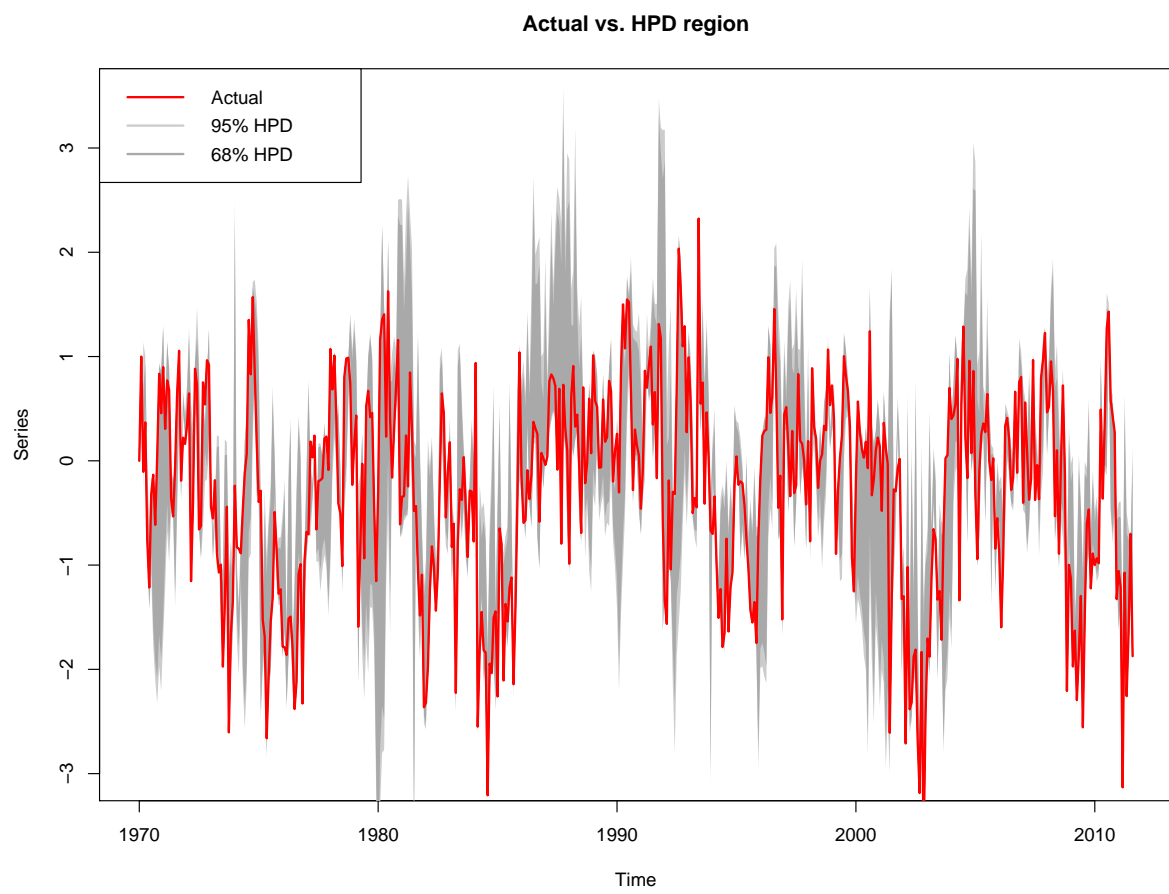


Figure 12: Shows the actual latent series and highest posterior density regions of the filtered series (shaded areas).

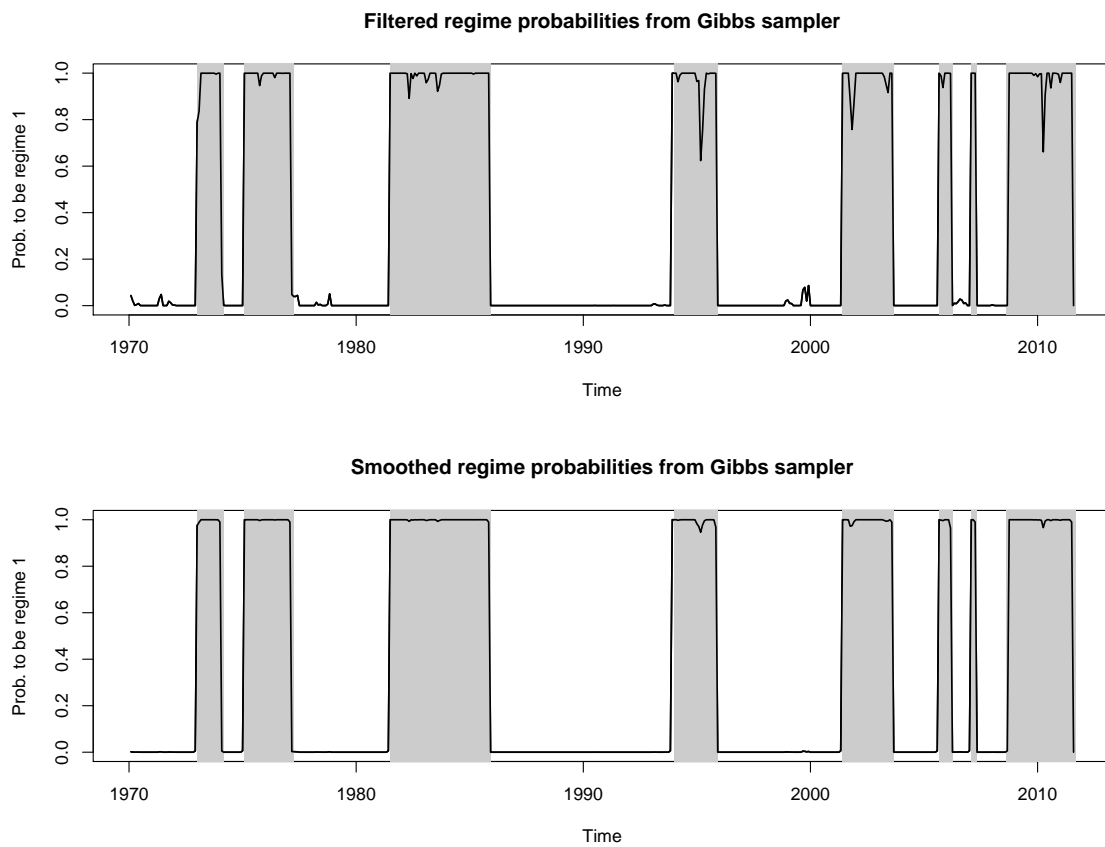


Figure 13: Shows filtered and smoothed regime probabilities from the Gibbs sampler. Shaded areas indicate the true underlying regimes.

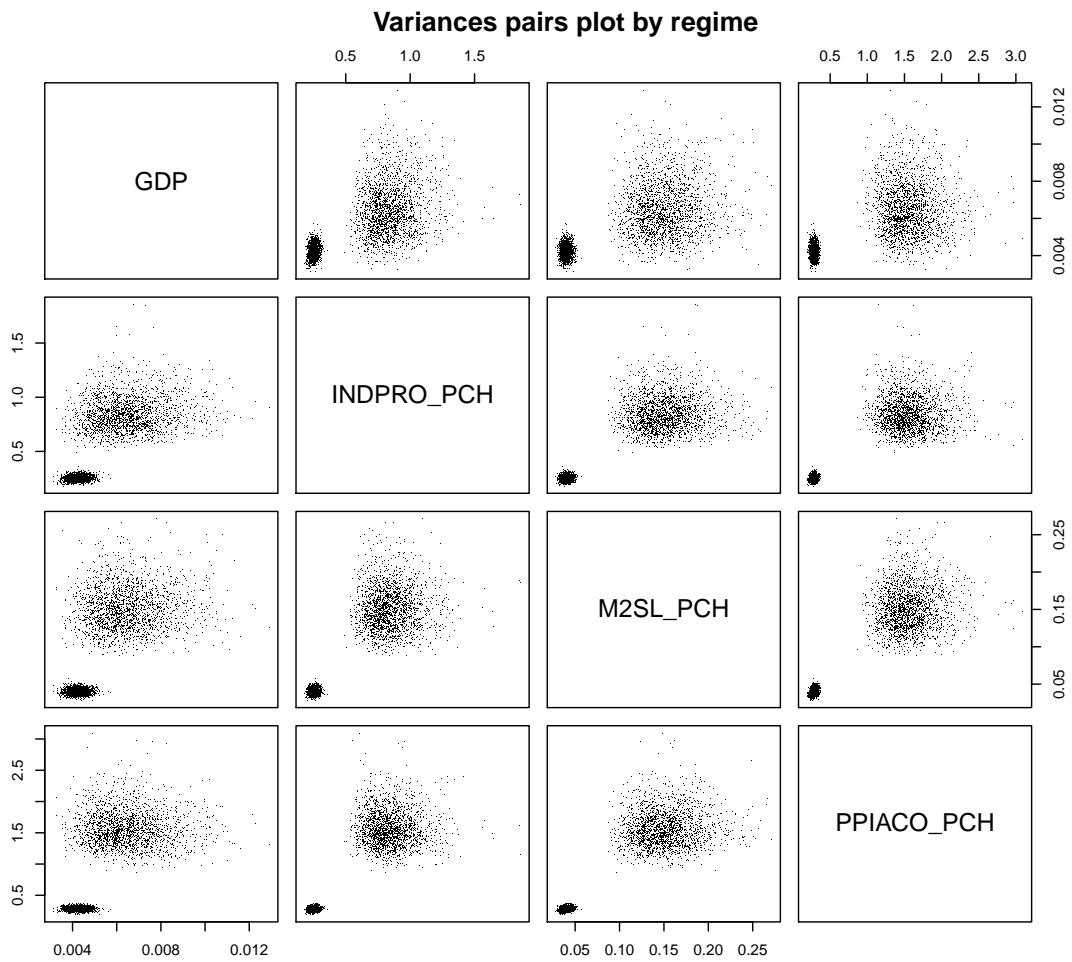


Figure 14: Scatterplot for variances by regime from an identified permutation sampler with 2,000 burn-in and 4,000 posterior sampling iterations. Each box depicts the sampled values of sampled variances for the two variables mentioned on the main diagonal of the plot.

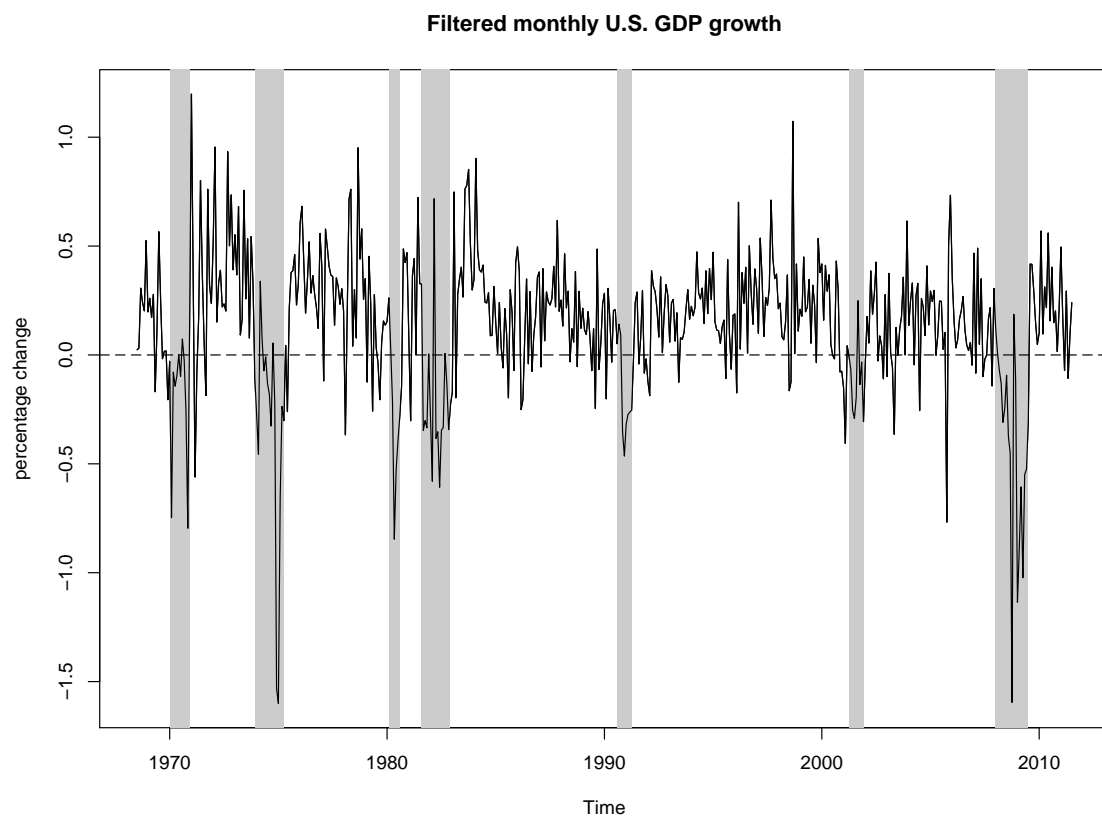


Figure 15: Figure shows the posterior sample mean of the imputed GDP values for each period. Shaded areas indicate NBER recession dates.

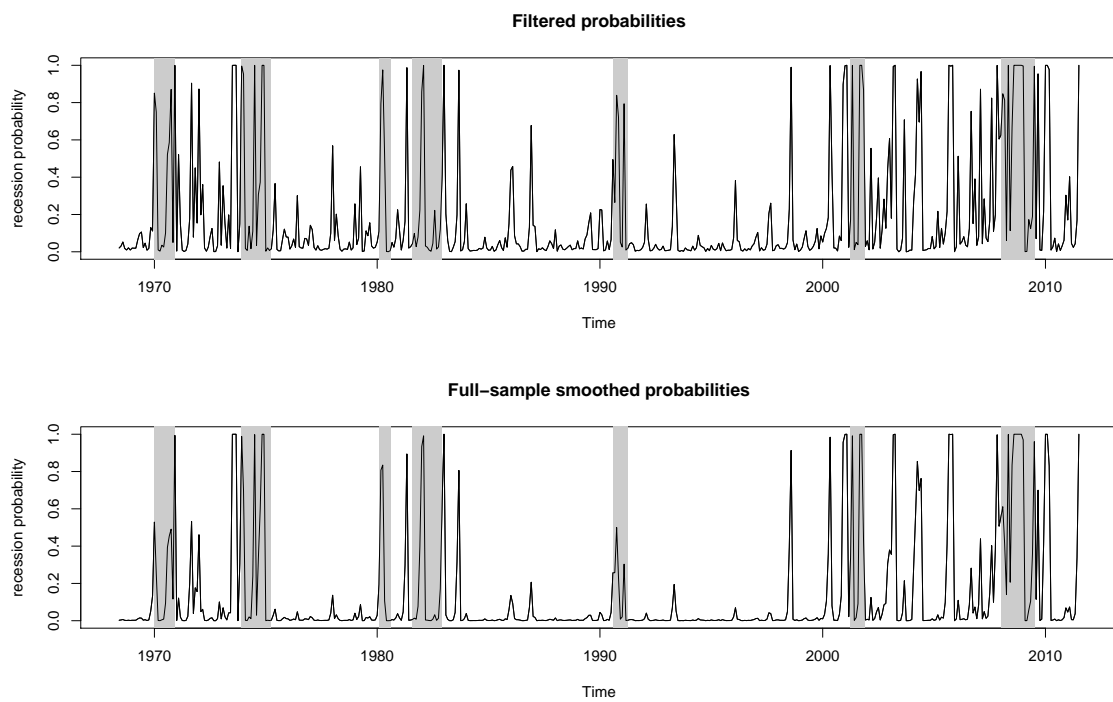


Figure 16: Shows the filtered and full-sample smoothed probabilities of being in a recession. Probabilities are based on data as of August 2011. Shaded areas indicate NBER recession dates.

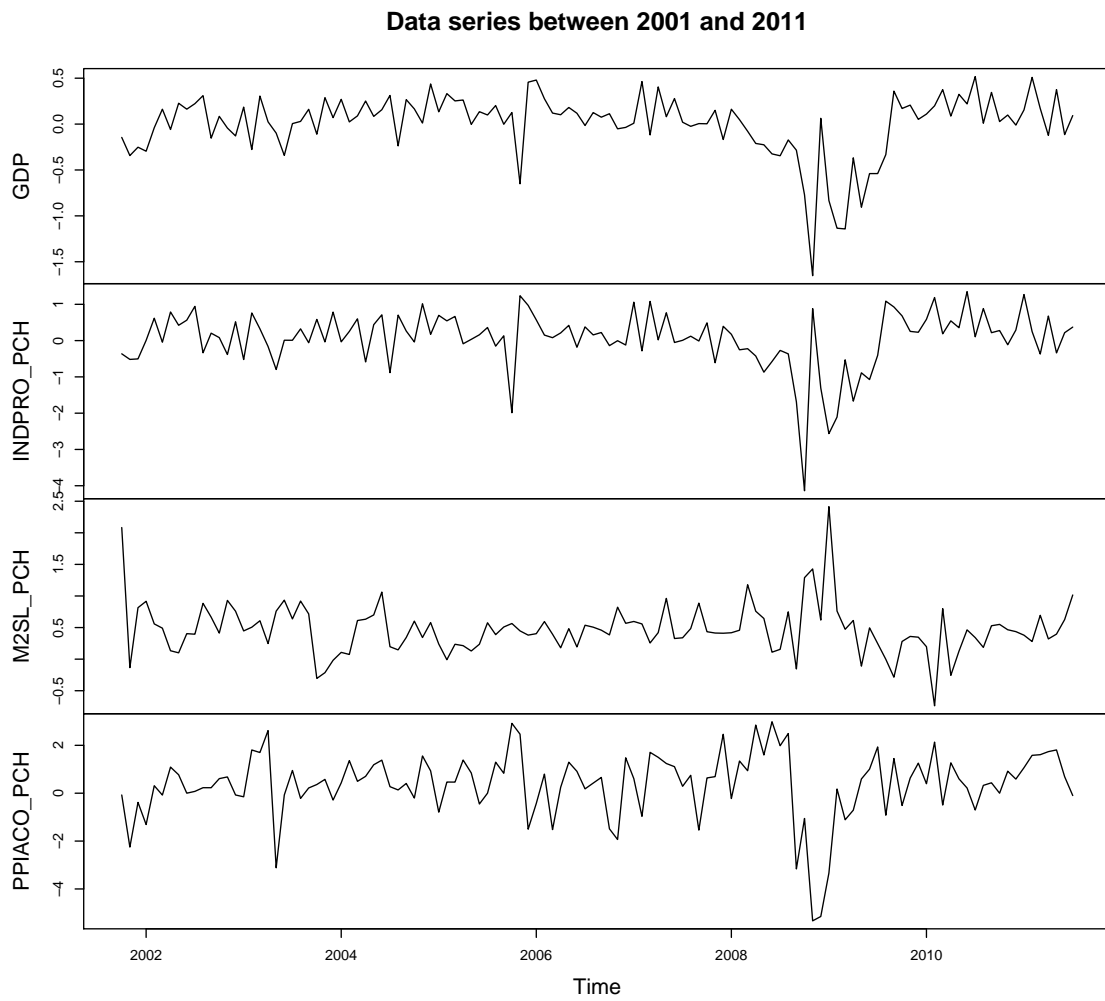


Figure 17: Shows the data series from 2001 to 2011.

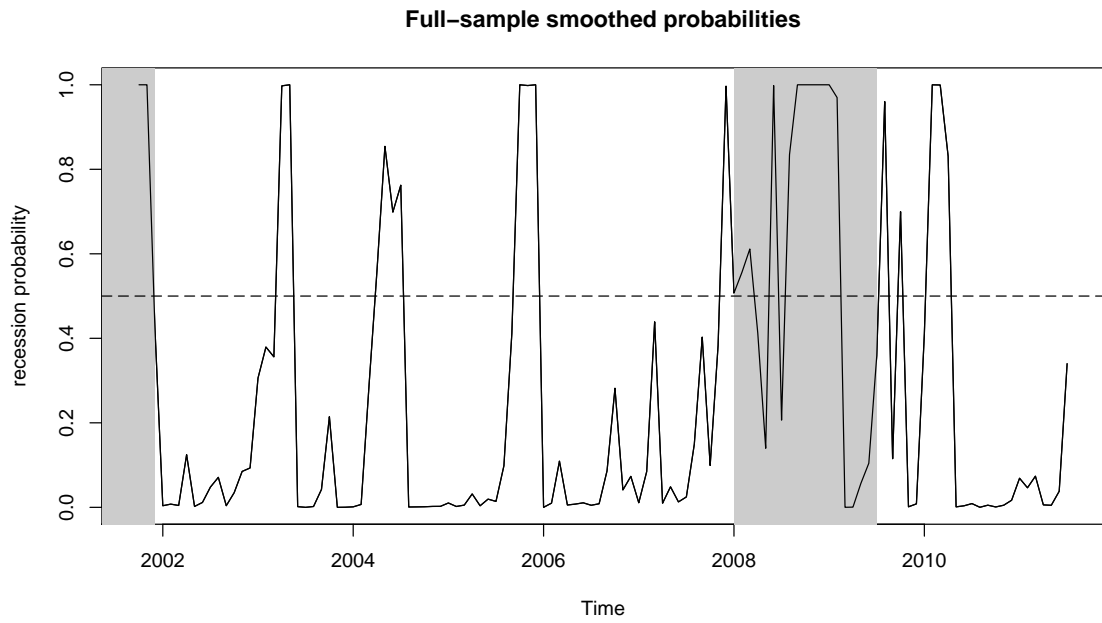


Figure 18: Shows the full-sample smoothed probabilities of being in a recession from 2001 to 2011. Probabilities are based on data as of August 2011. Shaded areas indicate NBER recession dates.

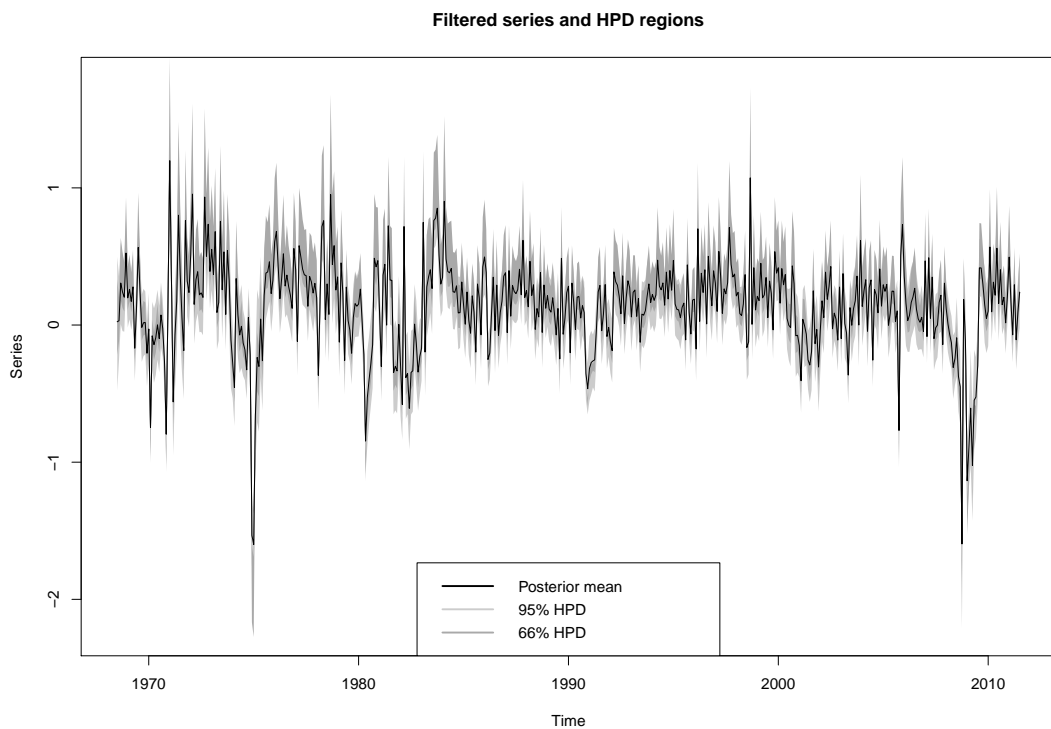


Figure 19: Figure shows the posterior sample mean of the imputed GDP values for each period (solid line) together with highest posterior density regions for the filtered series (shaded areas).

B Tables

Table 1: True parameter values and Bayesian prior and posterior distributions from a Gibbs sampler with 5,000 burn-in and 15,000 posterior sampling iterations.

	True			Posterior		
	Regime 1	Regime 2	Mean: Regime 1	68% HPD int.: Regime 1	Mean: Regime 2	68% HDP int.: Regime 2
a_{11}	-0.15	0.25	0.04	(-0.1051,0.1694)	0.2012	(0.0908,0.336)
a_{12}	0.53	-0.13	0.4484	(0.341,0.5538)	-0.1539	(-0.3101,-0.005)
a_{13}	0.02	0.12	-0.0157	(-0.1351,0.1004)	0.1745	(0.067,0.2885)
b_{11}	-0.05	0.05	-0.0194	(-0.147,0.1032)	0.0799	(0.0088,0.1467)
b_{12}	0.065	0.13	-0.0704	(-0.1743,0.0319)	0.1463	(0.0061,0.2844)
b_{13}	0.15	-0.35	0.1727	(0.0887,0.2547)	-0.3545	(-0.4614,-0.2455)
c_1	-0.61	0.1	-0.6206	(-0.7097,-0.5181)	0.1384	(0.0905,0.1803)
a_{21}	0.15	0.15	0.3246	(0.1863,0.4572)	0.1124	(-0.0088,0.2382)
a_{22}	0.15	0.8	0.0673	(-0.0396,0.1683)	0.8073	(0.6422,0.9515)
a_{23}	0.22	-0.02	0.2105	(0.0893,0.3212)	-8e-04	(-0.1059,0.112)
b_{21}	0.65	-0.065	0.7195	(0.6058,0.8487)	0.0036	(-0.0589,0.0764)
b_{22}	0.059	0.059	-0.0868	(-0.182,0.0202)	0.0363	(-0.1126,0.1692)
b_{23}	0.02	0	0.0208	(-0.0655,0.0986)	0.0648	(-0.0393,0.1727)
c_2	-0.23	0.13	-0.236	(-0.3226,-0.1357)	0.1545	(0.1101,0.199)
a_{31}	0.04	0.034	0.1904	(0.0469,0.3326)	-0.0158	(-0.1356,0.1156)
a_{32}	0.16	0.056	0.0718	(-0.0365,0.1874)	-0.0076	(-0.1814,0.1375)
a_{33}	-0.05	0.8	-0.0085	(-0.1423,0.1032)	0.8855	(0.7775,1.0039)
b_{31}	-0.023	0.23	0.0815	(-0.0428,0.2167)	0.2973	(0.2284,0.3678)
b_{32}	-0.23	-0.13	-0.3812	(-0.4831,-0.2675)	-0.0846	(-0.2181,0.0738)
b_{33}	0	0	-0.0202	(-0.108,0.0663)	-0.0233	(-0.13,0.0895)
c_3	-0.15	-0.05	-0.1345	(-0.2321,-0.0332)	-0.0261	(-0.0756,0.016)

Table 2: True parameter values and Bayesian prior and posterior distributions from a Gibbs sampler with 5,000 burn-in and 15,000 posterior sampling iterations.

	True		Posterior			
	Regime 1	Regime 2	Mean: Regime 1	68% HPD int.: Regime 1	Mean: Regime 2	68% HPD int.: Regime 2
σ_{11}	0.3953	0.2616	0.3631	(0.3226,0.3952)	0.2637	(0.243,0.2857)
σ_{12}	0.3724	0.2582	0.3611	(0.3163,0.3914)	0.2693	(0.2418,0.2873)
σ_{13}	0.3623	0.2755	0.3332	(0.2921,0.3626)	0.2765	(0.2508,0.2974)
σ_{22}	0.4168	0.3001	0.4187	(0.374,0.4587)	0.3262	(0.2961,0.3486)
σ_{23}	0.3681	0.3036	0.3583	(0.3178,0.3937)	0.3156	(0.2861,0.3381)
σ_{33}	0.3891	0.3433	0.3659	(0.3251,0.3989)	0.3433	(0.315,0.3697)

Table 3: Posterior mean for unconditional means of series. 68% HPD intervals in parantheses.

Series	Unconditional mean	
	Regime 2	Regime 1
GDP	0.133 (-0.154,0.384)	-0.735 (-0.934,0.148)
IND	0.182 (-0.165, 0.436)	-1.385 (-1.444,0.117)
M2	0.136 (-0.136, 0.209)	0.970 (-0.054,1.349)
PPI	0.209 (-0.168, 0.564)	0.690 (-0.169,0.746)