

TI 2016-107/III
Tinbergen Institute Discussion Paper



A Bayesian Infinite Hidden Markov Vector Autoregressive Model

Revision: 13-10-2017

Didier Nibbering¹

Richard Paap²

Michel van der Wel²

¹Econometric Institute, Erasmus University Rotterdam

²Erasmus University Rotterdam; Tinbergen Institute, The Netherlands

Tinbergen Institute is the graduate school and research institute in economics of Erasmus University Rotterdam, the University of Amsterdam and VU University Amsterdam.

Contact: discussionpapers@tinbergen.nl

More TI discussion papers can be downloaded at <http://www.tinbergen.nl>

Tinbergen Institute has two locations:

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 598 4580

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900

A Bayesian Infinite Hidden Markov Vector Autoregressive Model*

Didier Nibbering[†] Richard Paap Michel van der Wel

Econometric Institute, Tinbergen Institute, Erasmus University Rotterdam

October 13, 2017

Abstract

We propose a Bayesian infinite hidden Markov model to estimate time-varying parameters in a vector autoregressive model. The Markov structure allows for heterogeneity over time while accounting for state-persistence. By modelling the transition distribution as a Dirichlet process mixture model, parameters can vary over potentially an infinite number of regimes. The Dirichlet process however favours a parsimonious model without imposing restrictions on the parameter space. An empirical application demonstrates the ability of the model to capture both smooth and abrupt parameter changes over time, and a real-time forecasting exercise shows excellent predictive performance even in large dimensional VARs.

Keywords: Time-Varying Parameter Vector Autoregressive Model, Semi-parametric Bayesian Inference, Dirichlet Process Mixture Model, Hidden Markov Chain, Monetary Policy Analysis, Real-time Forecasting

JEL Classification: C11, C14, C32, C51, C54

*We thank the conference participants at the 3rd Annual Conference of the International Association for Applied Econometrics (2016), the 69th European Meeting of the Econometric Society (2016), the NBER-NSF Time Series Conference (2016), and the 7th Rhenich Multivariate Time Series Econometrics Meeting (2017).

[†]Correspondence to: Didier Nibbering, Econometric Institute, Erasmus University Rotterdam, P.O. Box 1738, NL-3000 DR Rotterdam, The Netherlands, e-mail: nibbering@ese.eur.nl

1 Introduction

Many researchers study estimation methods for time-varying parameters in vector autoregressive (VAR) models. The large sets of variables and hence parameters considered in these models compared to the number of available observations, increase the complexity of estimating time-varying parameters. Therefore, feasible estimation methods rely invariably on a set of model restrictions. To manage the number of time-varying parameter estimates, Cogley and Sargent (2005) impose the instantaneous relations among the VAR variables to be time-invariant. Chib et al. (2006) assume a factor structure for the covariance matrix. Primiceri (2005) imposes parameters to evolve smoothly over time, by modelling the evolution of the parameters in the coefficient and covariance matrices as random walks. Koop and Korobilis (2013) make use of forgetting factors to model time-variation in the parameters.

Parameter breaks in the models proposed by abovementioned papers are drawn from the same distribution. However, there is a wide variety in shapes and magnitudes of shocks to the economy, from abrupt shocks following rapid shifts in policy to smoother changes due to learning of economic agents. The model of Primiceri (2005), as a well-known example, imposes a break in each time period by modelling the law of motion by random walks. This approach does not allow for the presence of occasional jumps in parameter values, and the continuous changes imply a linear increase in parameter uncertainty over time. Alternatively, Sims and Zha (2006) (among others) capture time variation with a finite number of regimes in a Markov switching framework. These discrete break models are able to model shifts in policy but cannot account for smoother changes. Moreover, the number of regimes needs to be arbitrarily fixed before parameter estimation, ignoring the uncertainty in the number of breaks. Univariate models that do account for different types of breaks, such as Pesaran et al. (2006) and Giordani et al. (2007), do not account for state-persistence, and cannot easily be scaled up to (large dimensional) multivariate models.

In this paper we contribute to the literature of time-varying parameter vector autoregressive models by proposing a semi-parametric Bayesian model which accounts for heterogeneous parameters. Both the autoregressive parameters and the covariances of the innovations in the model are allowed to change over time

without imposing any restrictions on the parameter space to make estimation feasible. We employ a hidden Markov chain in combination with a Dirichlet process to allow for time-varying parameters. The Dirichlet process mixture encourages parameters to cluster in regimes with similar values. This feature favours a parsimonious model, which is a huge advantage in modelling parameter heterogeneity in structural time series models which, due to large sets of variables, already suffer from the curse of dimensionality. Moreover, the Dirichlet process mixture allows parameters to be drawn from different distributions over time by a potentially infinite number of regimes, which makes it possible to model abrupt breaks together with smoother changes. In contrast to parameter estimation in models with a fixed finite number of switching regimes, we can estimate parameter values, state values, and the number of regimes along with their uncertainty, together in one round. Furthermore, the hidden Markov structure accounts for state-persistence, as is often encountered in macroeconomic data.

We illustrate the performance of the model in an extensive empirical application on a monetary VAR. We show the ability of the model to capture heterogeneity over time together with both abrupt shocks and smooth changes in a structural analysis. We especially find posterior evidence for time-varying volatility. In a real-time forecasting exercise we compare the forecast performance of a small VAR (3 variables) and a large VAR (10 variables) infinite hidden Markov model to time-varying parameter VAR benchmarks. The infinite hidden Markov model with a time-varying covariance matrix shows for most forecast horizons and variables the best performance based on various evaluation measures. The relative performance of the infinite hidden Markov model benefits from the combination of both smooth and abrupt changes in parameter estimates, and its ability to switch to a stable regime from an explosive state, both in-sample and out-of-sample. Policy making based on models that only allow for smooth changes may be delayed in intervening after an abrupt shock. Discrete break models tend to be more prone to overreact to smooth changes or ignore small changes. For in-sample analysis, the infinite hidden Markov model allows for a more balanced deployment of policy instruments. Out-of-sample, it enables policy makers to anticipate more accurately to future changes in the economy.

Next to the empirical contribution, the technical contribution of our approach is threefold. First, we generalize the infinite hidden Markov model to a multivari-

ate setting, and construct a novel alternative to existing restrictive time-varying parameter VARs. The infinite hidden Markov VAR model builds upon work of Jochmann (2015), Song (2014), and Bauwens et al. (2017). They bring a semi-parametric Bayesian model, developed by Fox et al. (2011) for speaker diarization, to the univariate time series literature. This results in autoregressive models with an infinite number of regimes. Bauwens et al. (2017) show the superiority in forecast performance on macroeconomic time series relative to univariate models with fixed parameters. We extend the result of Bauwens et al. (2017) to analyse not only the predictive performance compared to multivariate fixed parameter models, but also to often used time-varying parameter VAR models.

Second, we contribute to the growing literature on estimating large time-varying parameter VARs. Recent studies show that increasing the dimensions of the VAR model improves forecasting and structural analysis (Carriero et al., 2015b). Bańbura et al. (2010), Koop (2013), Carriero et al. (2015a), and Giannone et al. (2015) estimate large VARs but do not account for parameter change, while Cogley and Sargent (2005), Primiceri (2005), Chib et al. (2006), Clark (2012), and Clark and Ravazzolo (2015) find convincing evidence for time-varying parameters in small VAR models. Only a few papers try to bridge the gap between large and time-varying systems. Koop and Korobilis (2013) use a semi-Bayesian approach which imposes restrictions on the parameter space and is unsuitable for policy analysis as parameters are not estimated but selected from a small grid of different values using forgetting factors. Carriero et al. (2015b) model time-varying volatility by only a single common unobserved factor, and for high-dimensional models they have to rely on a misspecified model. Since the infinite hidden Markov model estimates time-variation relatively parsimoniously, it can handle high-dimensional VAR systems without restricting the parameter space.

Third, the infinite hidden Markov model accounts for uncertainty in the underlying break processes but reduces the parameter uncertainty relative to other time-varying parameter models (Song, 2014). Traditional regime-switching models capture time-variation by a fixed finite number of regimes (Hamilton, 1989), which ignores the uncertainty around the number of regimes. Chopin and Pelgrin (2004) take this uncertainty into account by jointly estimating the parameters and the number of in-sample regimes. Moreover, traditional Markov switching models assume that future states are always equal to one of the estimated in-sample

regimes, which results in inaccurate forecasts in case of new out-of-sample regimes. The infinite hidden Markov model estimates the number of regimes and allows for new regimes out-of-sample. Other researchers model heterogeneity over time by change-point models (Chib, 1998; Koop and Potter, 2007; Liu et al., 2017) or impose parameters to change each time period, for example Primiceri (2005). In these models, different states cannot reoccur over time, which inevitably results in a loss of estimation efficiency. The infinite hidden Markov model reduces the parameter uncertainty by estimating parameters on data over all similar states, also when observations are separated from each other by break points.

The remainder of this paper is as follows. Section 2 discusses the model specification and explains parameter inference by Bayesian methods. Section 3 explains the empirical application of the methods on a monetary VAR. It introduces the data, discusses how we use the model for monetary policy analysis, and performs a forecasting exercise. We conclude with a discussion in Section 4.

2 Methods

This section discusses the specification and parameter estimation of the infinite hidden Markov VAR model. Section 2.1 introduces the baseline specification of the reduced form of a time-varying parameter VAR. From here, we explain how we capture the parameter heterogeneity over time by constructing regimes with homogeneous parameter values. The regimes and parameter values are estimated by Bayesian methods. In Section 2.2, we specify the prior distributions and set up a Markov Chain Monte Carlo (MCMC) sampler. Moreover, we show how we can sample from the predictive density.

2.1 Model Specification

Consider the reduced form of a time-varying vector autoregressive model of order l

$$y_t = B_t x_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_t), \quad t = 1, \dots, T, \quad (1)$$

where y_t is a $p \times 1$ vector of observed endogenous time series, B_t is a $p \times (k = 1 + pl)$ matrix with time-varying coefficients, and ε_t are heteroskedastic independent disturbances with covariance matrix Σ_t . The $k \times 1$ vector $x_t = [1, y'_{t-1}, \dots, y'_{t-l}]'$

includes an intercept and the endogenous variables up to lag l as explanatory variables.

Both the coefficient matrix B_t and the covariance matrix Σ_t in (1) contain time-varying parameters. Equivalently, we can say that the parameters in B_t and Σ_t vary over an infinite number of regimes, where the number of regimes equals the number of time periods T when each time period has a different parameter value. Within the regimes the parameters are assumed to be time-invariant but across regimes the parameters are allowed to be different. We can write (1) as

$$y_t = B_{s_t} x_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_{s_t}), \quad (2)$$

where $s_{1:T} = \{s_1, \dots, s_T\}$ takes integer values indicating the regime at time t . While there is strong evidence that the behavior of macroeconomic variables changes over time, it is implausible that the economy changes in each time period with probability one. Therefore, we can specify a potentially more parsimonious model by modelling the transition probability of moving from one state to another (Hamilton, 1989).

We let the regime indicators $s_{1:T}$ follow a first-order Markov chain, where π_{ij} denotes the transition probability of moving from state i to state j under the constraint that $\sum_{j=1}^J \pi_{ij} = 1$ for all i , where J possibly goes to infinity. So each state i has a state-specific transition distribution π_i over J states; $s_t \sim \pi_{s_{t-1}}$, where $\pi_i = (\pi_{i1}, \dots, \pi_{iJ})$. When the number of states is possibly infinite, we potentially have infinitely many parameters in the state transition matrix $\pi = (\pi'_1, \dots, \pi'_J)'$. Since estimating all these parameters is infeasible, we follow the framework of Teh et al. (2012), and implicitly integrate out the transition parameters by specifying the transition distributions π_i as a Dirichlet process mixture model,

$$\pi_i | a, H \sim DP(a, H_i),$$

where DP denotes a Dirichlet process distribution (Ferguson, 1973), the scalar $a = \alpha + \kappa$ is the concentration parameter, $H_i = \frac{\alpha \beta + \kappa \delta_i}{\alpha + \kappa}$ the base distribution, δ_i denotes a unit-mass measure concentrated at i , and α captures dispersion. The base distribution is constructed by the global transition distribution $\beta = (\beta_1, \dots, \beta_J)$ and scaled by the persistence parameter κ to account for state-dependence (Fox et al., 2011). When $\kappa = 0$, a standard Dirichlet process mixture model is recovered, that does not take state-persistence into account.

The Dirichlet process can be seen as a mixture over the transition probability distributions of J current states, where the state-specific transition probability distribution runs over J states in the next period, and J possibly goes to infinity (Escobar and West, 1995). So, not only the global transition distribution runs over an infinite number of states, the Dirichlet process is also an infinite discrete distribution over state-specific transition distributions. Since the expectation of the Dirichlet process equals the base distribution, states tend to have similar transition distributions; $E[\pi_{ij}] = \frac{\alpha\beta_j + \kappa 1(i=j)}{\alpha + \kappa}$, where $1(A)$ denotes an indicator variable that equals one if event A occurs and zero otherwise. An amount $\kappa > 0$ is added to the i th component of $\alpha\beta$, such that the expected probability of self-transition is increased by an amount proportional to κ . Moreover, element j of H_i , that is H_{ij} , can be interpreted as prior mean for the transition probabilities into state j . The variance of the Dirichlet process equals $H_i(1 - H_i)/(a + 1)$, from which we infer that α indeed controls the dispersion around the prior mean across rows of the transition matrix.

Conditional on the regimes in the previous time periods, the regime indicator s_t can be equal to the current regime of time period $t - 1$, an existing regime realized more than one period back in time, or switch to a new regime. In the latter case, new parameters values in the added row and column to the transition matrix are generated by a base distribution. The base distribution of the transition parameters is the scaled global transition distribution β , defined as

$$\beta_j = \nu_j \prod_{l=1}^{j-1} (1 - \nu_l), \quad \nu_j | \gamma \sim \text{Beta}(1, \gamma), \quad j = 1, 2, \dots,$$

where $\beta = \{\beta_j\}_{j=1}^{\infty}$ is defined as a probability mass function on a countably infinite set. This is known as a stick-breaking construction, which can also be written as $\beta \sim \text{Stick}(\gamma)$. The expected number of represented hidden states is governed by γ , by controlling how concentrated the probability mass will be across the columns of the transition matrix. Switching to a new regime also implies that new values of the model parameters $\theta_{s_t} = \{B_{s_t}, \Sigma_{s_t}\}$ have to be generated. The base distribution of the model parameters is denoted by

$$\{B_{s_t}, \Sigma_{s_t}\} \sim H_{\theta}(\Theta),$$

where Θ is a set of hyperparameters in the base distribution.

We can summarize the complete model specification by the following equations,

$$y_t = B_{s_t}x_t + \varepsilon_t, \quad (3)$$

$$\varepsilon_t \sim \mathcal{N}(0, \Sigma_{s_t}), \quad (4)$$

$$s_t | s_{t-1}, \{\pi_i\}_{i=1}^{\infty} \sim \pi_{s_{t-1}}, \quad (5)$$

$$\pi_i | \alpha, \kappa, \beta \sim DP \left(\alpha + \kappa, \frac{\alpha\beta + \kappa\delta_i}{\alpha + \kappa} \right), \quad (6)$$

$$\beta_j = \nu_j \prod_{l=1}^{j-1} (1 - \nu_l), \quad (7)$$

$$\nu_j | \gamma \sim \text{Beta}(1, \gamma), \quad j = 1, 2, \dots, \quad (8)$$

$$\{B_{s_t}, \Sigma_{s_t}\} \sim H_{\theta}(\Theta). \quad (9)$$

Equations (3) and (4) specify the reduced form of the time-varying vector autoregressive model, where the parameters $\theta_{s_t} = \{B_{s_t}, \Sigma_{s_t}\}$ vary over an infinite number of regimes. To retrieve the different regimes we use a hidden Markov chain in combination with a Dirichlet process mixture model. Equation (5) specifies the hidden Markov model by introducing a first-order Markov chain with transition probability matrix π . The transition probability distribution of π_i is specified as a Dirichlet process mixture in (6)-(8). The κ parameter captures the persistence in macroeconomic data by controlling the probability that parameters remain constant between time periods. Equation (9) concludes, and provides the base distribution of the model parameters, H_{θ} , parameterized by the hyperparameters Θ .

For ease of notation we follow the Markov-switching literature and specify one regime switching process for all model parameters. To obtain a potential efficiency gain we can easily extend to a model with different regime-switching processes for the parameters in the coefficient matrix and the parameters in the covariance matrix, within the infinite hidden Markov framework. However, a regime switch in this model does not necessarily mean that all parameters change. For example, a regime switch can either be the result of a change in the parameters in the covariance (coefficient) matrix while the coefficient (covariance) parameters remain constant, or a change in all parameter values.

2.2 Bayesian Inference

To estimate the parameters $\theta_{st} = \{B_{st}, \Sigma_{st}\}$ we rely on the Markov Chain Monte Carlo (MCMC) algorithm for the infinite hidden Markov model derived by Fox et al. (2011). Bauwens et al. (2017) apply a variant of this sampler in a univariate econometric time series context.

Although there are sampling algorithms that can deal with an infinite number of regimes (which are also derived and discussed by Fox et al. (2011)), these algorithms suffer in general from slow mixing rates. Therefore, we opt for a sampler which truncates the number of possible states to a fixed degree L , the so called degree L weak limit approximation (Ishwaran and Zarepour, 2002). This sampler fixes the number of states J in the state transition matrix π and the global transition distribution β to L . When L equals the number of time periods T in the sample, the truncated model is in practice equal to the full Dirichlet process mixture model. However, smaller values for L improve computational time significantly and when L is large enough, the approximation error is negligible.

The degree L weak limit approximation fosters models with less than L regimes while allowing for new regimes, bounded by L , when new data are observed. Since the state assignments and the number of different states with nonzero assigned observations can differ over different sample iterations, the posterior distributions of the estimated parameters can be different for each time period while the sampler finds only a small number of different regimes. In each iteration, the sampler draws for each of the L states the transition probabilities, which are used to sample the state assignments for each observation. Due to small transition probabilities, some states can stay empty. Since the state labels may switch over different MCMC iterations (label switching) we cannot identify regime-specific posterior quantities, see, for example, Frühwirth-Schnatter (2001). Therefore, we report observation-specific posterior results which are identified, see Geweke (2007).

2.2.1 Prior Distributions

The degree L weak limit approximation induces finite Dirichlet distribution priors on β and π_i ,

$$\begin{aligned}\beta|\gamma &\sim \text{Dir}(\gamma/L, \dots, \gamma/L), \\ \pi_i|\alpha, \beta, \kappa &\sim \text{Dir}(\alpha\beta_1, \dots, \alpha\beta_i + \kappa, \dots, \alpha\beta_L),\end{aligned}$$

where β and π_i are L -dimensional row vectors and Dir denotes the finite Dirichlet distribution. We let the data determine the number of states with L as an upper bound and the number of regime-switches by treating the hyperparameters of the transition distributions $\{\gamma, \alpha, \kappa\}$ as unknown. We place priors on these hyperparameters,

$$\alpha + \kappa \sim \text{Gamma}(a_\alpha, b_\alpha), \quad \gamma \sim \text{Gamma}(a_\gamma, b_\gamma), \quad \rho = \frac{\kappa}{\alpha + \kappa} \sim \text{Beta}(c_\rho, d_\rho).$$

The parameters in the prior distributions of the concentration parameters, denoted by $\{a_\alpha, b_\alpha, a_\gamma, b_\gamma\}$, control the dispersion with respect to their base distributions and thereby the expected number of states. The prior beliefs about the number of regime-switches are captured by $\{c_\rho, d_\rho\}$. A relatively low value for c_ρ corresponds to rapid switches from one state to another. Increasing c_ρ leads to higher state-persistence.

The prior on the parameters $\theta_{st} = \{B_{st}, \Sigma_{st}\}$ is a Normal-inverse-Wishart,

$$\text{vec}(B_{st})|\Sigma_{st}, \Theta \sim \mathcal{N}(\text{vec}(b_B), V_B \otimes \Sigma_{st}), \quad \Sigma_{st}|\Theta \sim \mathcal{IW}(\nu_\Sigma, S_\Sigma), \quad (10)$$

where the $\text{vec}(A)$ operator stacks the columns of matrix A and Θ is the collection of hyperparameters for θ , $\{b_B, V_B, \nu_\Sigma, S_\Sigma\}$. So the infinite hidden Markov model allows for an elegant conjugate prior structure in which we can put prior beliefs about the model parameters in the coefficient and covariance matrices directly in the prior distribution in (10). For instance, we can control the prior probability mass at stationary VARs, by shrinking the coefficients to zero with values close to zero in b_B and relatively small in V_B .

2.2.2 Posterior Distribution

Fox et al. (2011) derive a sample algorithm applicable for Bayesian parameter inference in the infinite hidden Markov model. We extend the sampling steps to

the multivariate econometric time series context of the time-varying parameter VAR models and present the resulting sampling steps:

Step 1. Set the truncation level L of possible hidden Markov states. Sample an initial draw for the hyperparameters of the transition distributions from their priors and do the same for the hyperparameters in the base distribution H_θ . Initialize the transition distributions β and π_i by drawing from their L -dimensional Dirichlet priors.

Step 2. Sample the regime indicators $s_{1:T}$ using backward messages $m_{t,t-1}(i)$ from the state assignment probabilities in time period t to state i in time period $t - 1$.

(a) First, work sequentially backwards in time. For each $i = 1, \dots, L$, $m_{T+1,T}(i) = 1$ and

$$m_{t,t-1}(i) = \sum_{j=1}^L \pi_{ij} \mathcal{N}(y_t; B_j x_t, \Sigma_j) m_{t+1,t}(j), \quad t = T, \dots, 2,$$

where $\mathcal{N}(y; \mu, \Sigma)$ denotes the probability density function of the multivariate Normal distribution with mean μ and covariance matrix Σ .

(b) Second, work sequentially forward in time and initialize the number of transitions from state i to j observed in the state vector $s_{1:T}$, $n_{ij} = 0$ with $i, j = 1, \dots, L$. For each $j = 1, \dots, L$, compute the probability that observation y_t is assigned to state j

$$f_j(y_t) = \pi_{s_{t-1},j} \mathcal{N}(y_t; B_j x_t, \Sigma_j) m_{t+1,t}(j), \quad t = 1, \dots, T,$$

sample the regime indicators from a discrete distribution,

$$s_t \sim \sum_{j=1}^L f_j(y_t) 1(s_t = j), \quad t = 1, \dots, T,$$

and increment n_{s_{t-1},s_t} .

Step 3. Sample auxiliary variables m , w , and \bar{m} to simplify the resampling of β .

(a) For $i = 1, \dots, L$ and $j = 1, \dots, L$ set $m_{ij} = 0$. For $k = 1, \dots, n_{ij}$ sample $x_k \sim \text{Bernoulli}\left(\frac{\alpha\beta_j + \kappa 1(i=j)}{i-1 + \alpha\beta_j + \kappa 1(i=j)}\right)$ and increment m_{ij} if $x_k = 1$.

- (b) For $i = 1, \dots, L$ sample $w_i \sim \text{Binomial}(m_{ii}, \frac{\rho}{\rho + \beta_i(1-\rho)})$. Set $\bar{m}_{ij} = m_{ij}$ if $i \neq j$ and $\bar{m}_{ij} = m_{ij} - w_i$ if $i = j$.

Step 4. Sample the global transition distribution

$$\beta \sim \text{Dir}(\gamma/L + \sum_i \bar{m}_{i1}, \dots, \gamma/L + \sum_i \bar{m}_{iL}).$$

Step 5. Sample the transition distribution π . For $i = 1, \dots, L$ sample

$$\pi_i \sim \text{Dir}(\alpha\beta_1 + n_{i1}, \dots, \alpha\beta_i + \kappa + n_{ii}, \dots, \alpha\beta_L + n_{iL}).$$

Step 6. Sample the regime parameters θ for $j = 1, \dots, L$. Let x_j be the $t_j \times k$ matrix with rows $x_{s_t=j}$ and t_j the number of observations in state s_t . Define y_j as a $p \times t_j$ matrix. Sample the model parameters

$$\begin{aligned} \bar{B} &= (x'_j x_j + V_B^{-1}), & \bar{b} &= (y_j x_j + b_B V_B^{-1}) \bar{B}^{-1}, \\ \bar{S} &= S_\Sigma + (y'_j - x'_j \bar{b}') (y'_j - x'_j \bar{b}') + (\bar{b} - b_B) V_B^{-1} (\bar{b} - b_B)', \\ \Sigma_j | y_j, \Theta &\sim \mathcal{IW}(\nu_\Sigma + t_j, \bar{S}), & \text{vec}(B_j) | y_j, \Sigma_j, \Theta &\sim \mathcal{N}(\text{vec}(\bar{b}), \bar{B}^{-1} \otimes \Sigma_j). \end{aligned}$$

Step 7. Sample the hyperparameters of the transition distributions γ , α , and κ .

- (a) Sample auxiliary variables $r_i \sim \text{Beta}(\alpha + \kappa + 1, \sum_j n_{ij})$ and $s_i \sim \text{Bernoulli}(\frac{\sum_j n_{ij}}{\sum_j n_{ij} + \alpha + \kappa})$ for $i = 1, \dots, L$ to simplify the posterior distributon of $\alpha + \kappa$.
Sample $\alpha + \kappa \sim \text{Gamma}(a_\alpha + \sum_i \sum_j m_{ij} - \sum_i s_i, (\frac{1}{b_\alpha} - \sum_i \log r_i)^{-1})$.
- (b) Sample $\rho = \frac{\kappa}{\alpha + \kappa} \sim \text{Beta}(c_\rho + \sum_i w_i, d_\rho + \sum_i \sum_j m_{ij} - \sum_i w_i)$.
- (c) Sample auxiliary variables $r \sim \text{Beta}(\gamma + 1, \sum_i \sum_j \bar{m}_{ij})$ and $s \sim \text{Bernoulli}(\frac{\sum_i \sum_j \bar{m}_{ij}}{\sum_i \sum_j \bar{m}_{ij} + \gamma})$. Compute $\bar{K} = \sum_k 1(\sum_i \bar{m}_{ij} > 0)$ and Sample $\gamma \sim \text{Gamma}(a_\gamma + \bar{K} - s, (\frac{1}{b_\gamma} - \log r)^{-1})$.

Step 8. Go to step 2.

2.2.3 Predictive Densities

To construct a predictive density, we again make use of the degree L weak limit approximation. When L is assumed to be much larger than the number of in-sample regimes, the infinite hidden Markov model takes out-of-sample parameter

breaks into account by allowing for new regimes out-of-sample. Here we show how future values are sampled from their predictive densities, together with the potentially new regimes, and the corresponding future model parameter values.

We simulate the predictive densities of y_{T+h} for different horizons h by iterating over the auto-regressive equation in (2), in each iteration of the sampler, using the parameter draws obtained in that sample iteration. In iteration (i) of the sampler we have,

$$y_{T+h}^{(i)} = B_{s_{T+h}^{(i)}}^{(i)} x_{T+h-1}^{(i)} + \varepsilon_{T+h}^{(i)}, \quad \varepsilon_{T+h}^{(i)} \sim \mathcal{N}(0, \Sigma_{s_{T+h}^{(i)}}^{(i)}),$$

$$s_{T+h}^{(i)} \sim \text{Multinomial}(\pi_{s_{T+h-1,1}^{(i)}}, \dots, \pi_{s_{T+h-1,L}^{(i)}}),$$

where $B_{s_{T+h}^{(i)}}^{(i)}$ and $\Sigma_{s_{T+h}^{(i)}}^{(i)}$ are the parameter draws in iteration (i) of the sampler, and $x_{T+h-1}^{(i)}$ is constructed from $y_{T+h-1}^{(i)}, \dots, y_{T+h-l}^{(i)}$, where elements are replaced by in-sample observations when known. The regime indicators $s_{T+h}^{(i)}$ are sampled from a Multinomial distribution using the in-sample draws for the state transition probabilities, and L is the number of clusters under the degree L weak limit approximation. Since the degree L weak limit approximation is assumed to be much larger than the number of estimated in-sample states, future parameter values can be drawn from new regimes which are not present in-sample.

3 Empirical Application

We apply the newly proposed model on a monetary VAR of the U.S. economy consisting of the unemployment rate, inflation rate and federal funds rate, in an in-sample and out-of-sample application. Section 3.1 introduces the data. In Section 3.2 we use the infinite hidden Markov model to study the effects of monetary policy shocks in a structural VAR model. Section 3.3 performs a real-time forecasting exercise with the a small monetary VAR model to assess the out-of-sample performance of the infinite hidden Markov model compared to benchmark models. In 3.3 we also analyse the forecast performance of a large dimensional VAR model.

We follow Primiceri (2005) and consider VAR models with two lags. Posterior results are based on 20,000 iterations of the MCMC sampler, from which the first 10,000 are discarded. Visual inspection shows that this number of iterations is

enough for convergence. The number of possible states is truncated at 20 in the degree L weak limit approximation.

3.1 Data

We use three macroeconomic time series of the U.S. economy, the unemployment rate, inflation rate, and interest rate, to construct a monetary VAR. The unemployment rate is the civilian rate of unemployment, and inflation is calculated as a function of the GDP deflator P_t to obtain the annualized quarterly growth rate of prices; 400 times the first difference of the logarithm of P_t . Since the three month Treasury bill rate is available over a longer period of time than the federal funds rate, the interest rate is represented by the first.

The real-time data for the unemployment rate and the GDP deflator are collected by the Federal Reserve Bank of Philadelphia. The three month Treasury bill rate is not subject to revisions and is available from the Federal Reserve Bank of St. Louis. The GDP deflator is available as quarterly time series and the unemployment rate and interest rate as monthly time series. We follow Cogley and Sargent (2002, 2005); Cogley et al. (2010); D’Agostino et al. (2013), by taking the value at the second month of the quarter for the unemployment and the value at the first month of the quarter for the interest rate, to obtain quarterly series for all three variables.

The first quarter of 1948 is the first time period for which all data is available. We consider data through 2015Q4. When we date a vintage as the last quarter for which all data are available, we have vintages from 1965Q4 to 2015Q4. We use all data in the most recent vintage for in-sample analysis. Figure 1 shows the standardized data series as included in the model.

[Figure 1 about here.]

Table 1 shows the prior parameter values of the model. We follow Fox et al. (2011) in the parameter values in the prior distributions on the hyperparameters of the transition distribution. We opt for a non-informative prior by choosing large scale parameters $\{b_\alpha, b_\gamma\}$ in the Gamma distributions. A relatively low value for c_ρ , that is $c_\rho = 10$, corresponds to rapid switches from one state to another. Setting $c_\rho = 1000$ leads to higher state-persistence.

Since the data is standardized, we can choose non-informative priors for the model parameters. The inverse-Wishart distribution of the covariance parameters has degrees of freedom equal to the number of variables in the model plus two and a scaled identity matrix as scale matrix. The prior mean of the coefficient parameters is set to zero, and we can control the shrinkage of the coefficient estimates towards zero by the shrinkage parameter λ in the prior variance of the coefficient parameters. By scaling the prior variance of coefficient parameters by the lag order of the corresponding variable, we shrink coefficients estimates of higher lag order variables to zero, similar to the Minnesota prior (Doan et al., 1984).

For out-of-sample purposes discussed later, we try to avoid sampling many explosive VAR parameters by choosing a more tighter parametrization of the prior distribution of the coefficients. Although it is known that alternating between explosive and nonexplosive regimes can still produce non-explosive processes in the long-run (Francq and Zakoian, 2001) we want to exclude having long periods of explosive regimes in our prediction sample. In sum, for in-sample analysis we set $\lambda = 1$ and $\lambda = 0.1$ in our forecasting exercise.

[Table 1 about here.]

3.2 Structural VAR model

Since a time varying parameter VAR consists of a large amount of parameters, even in a small model with only three variables and two lags, it takes too much space to present estimation diagnostics of all parameters. Moreover, it is difficult to give an economic interpretation to each posterior distribution. Therefore, we discuss the stability of the VAR model over time, we show the posterior results for the variances of the structural shocks, and we report impulse response functions which summarize the economic implications of the estimated structural coefficients. To identify the structural parameters, we opt for a Cholesky decomposition in our application on a small monetary VAR in which the variables are ordered as {unemployment, inflation, interest rate} (Sims, 1980). However, alternative identification schemes, for instance, long-run restrictions or sign restrictions, can also be applied to the infinite hidden Markov VAR model.

3.2.1 Stability Diagnostics

We find compelling evidence of instability in the parameter estimates over time. The posterior probability for four different regimes equals 70%, for five regimes 28%, and the remaining probability mass is concentrated at six regimes. Figure 2 shows the posterior probability of a regime switch for each time period. Most of the breaks are detected before 1990. Thereafter, there is a more stable period which is followed by higher break probabilities corresponding to the Dot-com bubble in 2000 and the global financial crisis starting in 2007.

[Figure 2 about here.]

The instability is also reflected by the time variation in the posterior probability of an explosive system. When the largest absolute eigenvalue of the companion form of the reduced form VAR is larger than one in a specific time period, the system is in an explosive regime at that point in time. Figure 3 shows the time-varying posterior mean of the largest absolute eigenvalue, together with the 68% and 90% confidence bands. Most of the probability mass is located below one. However, before the mid eighties we observe temporary increases in the probability of an explosive system.

[Figure 3 about here.]

3.2.2 Structural Variance

We define monetary policy shocks as interest rate responses to variables other than unemployment and inflation. The changes in relative importance of these shocks over time are displayed by the time-varying variance of the monetary policy shock. Figure 4 shows the time-varying variance of monetary policy shocks, together with shocks to the unemployment and inflation equation.

[Figure 4 about here.]

The confidence bands around the time-varying variance of monetary policy shocks provide evidence that there is variation over time. The first thing to note beside some peaks in variance in the early sixties and mid seventies, is the long period of high variance running from 1979 to 1983. This feature is well-known and

can be attributed to a period with deviant monetary policy. After this high variance regime, the changes in variance are quite modest with only small exceptions.

The first panel of Figure 4 shows that the variance of unemployment shocks follows a pattern similar to the variance of monetary policy shocks. The confidence bounds also clearly support modelling time variation. The variance of the shocks to inflation, as showed in the second panel of Figure 4, seems to be less volatile over time. Apart from high variance regimes in the seventies, it behaves more stable in the rest of the sample period.

3.2.3 Impulse Response Functions

Figure 5 shows the posterior mean of the impulse response functions of the SVAR to a unit monetary policy shock, conditional on the estimated regimes. The functions trace out the effect of the structural shocks over a time path of four years for each variable, conditional on the estimated states. Because of the time-varying parameters in the model, the time paths are different for each date the shock hits the system. So for each quarter in the estimation sample, we have the time path of the policy shock effect over the sixteen following quarters.

[Figure 5 about here.]

According to the posterior mean, the impulse responses differ in strength over time. In general, the effects in the period around 1970 and 1980 and the period after 2007 seem to be more severe. The initial reaction of unemployment can be both positive or negative, but the result of the shock after four years seems always to be positive. The impulse response of inflation differs also in strength over time, but in general it reacts positive to an interest shock and converges back to zero, which is known as the price puzzle. The interest rate also seems to converge to zero, after a sharp decline following the impulse of magnitude one.

The impulse response functions in Figure 5 suggest some variation over time. However, Figure 5 shows only the mean of the posterior distribution of the impulse responses. To get an idea about the uncertainty around the time variation we arbitrarily choose four different moments at which a monetary policy shock hits the system, and plot the posterior means together with the 68% and 90% of the posterior distributions. Figure 6 shows the impulse response functions to a monetary

policy shock for the different shock dates; the second quarter of 1965, 1980, 1995, and 2010. When taking the whole posterior distribution of the impulse responses into account, we can hardly conclude that the responses are time-varying. Figure 6 shows that, apart from the impulse response of the interest rate in 1980, we can hardly find any differences in the shape of the response functions per variable, but the magnitude of the effect differs over time. However, due to the high uncertainty about the shape and magnitude of the impulse responses, we find no convincing posterior evidence for time-variation in impulse responses functions.

[Figure 6 about here.]

According to the confidence bounds there is little posterior evidence that the impulse responses of unemployment and inflation differ from zero. However, Table 2 shows that after some shock dates the probability mass clearly indicates a positive effect. With a posterior probability of approximately 80 percent, a monetary policy shocks results in a positive effect on unemployment after four years, for all different shock dates. For inflation, there seem to be only a short-term positive effect, also with probabilities close to 80 percent.

[Table 2 about here.]

The impulse response functions are constructed from a large number of coefficients which are allowed to be different over time. To decrease parameter uncertainty, we can impose all coefficients to be constant over time. However, this results also in time-invariant impulse response functions. Alternatively, we restrict only the long term impact to be constant over time and account for variation in short term effects

$$y_t = c_{s_t} + \Pi_{s_t} y_{t-1} + D \Delta y_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_{s_t}). \quad (11)$$

Appendix A discusses the sample steps for this restricted model and Appendix B shows the impulse response functions following from this more parsimonious model. However, despite of the decrease in parameter uncertainty, we still do not find posterior support for time-varying impulse response functions.

3.3 Forecasting Exercise

To assess the out-of-sample performance of the infinite hidden Markov VAR model, we perform a forecasting exercise in which we compare the predictive performance of the infinite hidden Markov model against benchmark models. We adopt a similar real-time forecasting framework as D’Agostino et al. (2013), who iteratively produce forecasts with the time-varying parameter VAR model of Primiceri (2005).

We start the forecasting exercise with an estimation sample running from 1948Q1 up to 1969Q4 of the vintage 1969Q4. We standardize the variables and estimate the model parameters on this sample. We compute with each model forecasts up to five quarters ahead outside the estimation window, from 1970Q1 to 1971Q1. After we have produced the forecasts based on the first estimation sample, we move one quarter ahead and re-estimate the model parameters using the standardized variables based on all the data in vintage 1970Q1. That means that we use an expanding window to estimate the model parameters. Again, we use each model to compute forecasts up to five quarters ahead. We repeat this procedure up to vintage 2014Q1 (as we need later vintages to evaluate forecasts up to five quarters ahead, as we discuss in Section 3.3.2). This exercise results in time series of 178 one-period-ahead forecasts from 1970Q1 to 2014Q2 and a time series of the same length containing five-periods-ahead predictions from 1971Q1 to 2015Q2, since we compare the forecasts to data after 2 revisions.

Since there is evidence that large VAR models can improve in forecast performance upon small models, we also extend the small model to ten variables to construct forecasts of the unemployment rate, inflation rate, and interest rate. We add five variables from the real-time database of the Federal Reserve Bank of Philadelphia; M1 money stock, real gross domestic product, personal consumption expenditures, industrial production index, and imports of goods and services, and we add the S&P 500 index and the total borrowings of depository institutions from the Federal Reserve which are unrevised data from the Federal Reserve Bank of St. Louis. In case of monthly data we take the value of the third month and all data is included in the model as growth rates by taking 400 times the first difference of the logarithm. The flow variables real gross domestic product, personal consumption expenditures, and imports of goods and services are available as quarterly data.

3.3.1 Forecasting Models

We compare the predictive performance of variants of the infinite hidden Markov model against the predictive performance of a time-invariant Bayesian VAR, the time-varying parameter VAR model of Primiceri (2005), and the time-varying parameter VAR model of Koop and Korobilis (2013). The model of Primiceri (2005) is designed for the specific type of small monetary VAR studied in this forecast exercise. To avoid over-fitting and since a parsimonious model could potentially lead to a more efficient model, and therefore an increase in predictive performance, we not only forecast with the unrestricted infinite hidden Markov VAR model, but also with models in which we restrict coefficients or the covariance matrix to be time-invariant. The model in which we restrict both boils down to a linear Bayesian VAR without any time-variation included.

The linear Bayesian VAR model is identical to the model in (1), but now the parameters are fixed over time;

$$y_t = Bx_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma), \quad t = 1, \dots, T, \quad (12)$$

and the prior on the parameters $\theta = \{B, \Sigma\}$ is a Normal-inverse-Wishart as in (10) with values for the hyperparameters given in Table 1. We simulate from the predictive densities of y_{T+h} for different horizons h as described for the infinite hidden Markov model in Section 2.2.3.

We specify and estimate the Primiceri model as outlined in Primiceri (2005). However, for a fair comparison with the other models, we follow Koop and Korobilis (2013) in not using a training sample prior. We generate forecast from the Primiceri model (denoted as P05) in the same way as Koop and Korobilis (2013) and D'Agostino et al. (2013). That means that we use iterated forecasts in the same way as discussed for the linear BVAR, and we allow for out-of-sample parameter change in the VAR.

Note that due to computational constraints and the fact that the estimation algorithm involves taking the inverse of large matrices, the model of Primiceri (2005) cannot be estimated in a stable way for a large VAR model. That is why Koop and Korobilis (2013) propose an alternative model that can handle large dimensional time-varying parameter VAR models by using forgetting factors to model time-variation. We implement this model, which we denote as KK13, as final benchmark. We take for the forgetting factors $\lambda = 0.99$, $\kappa = 0.96$, and

$\gamma = 0.1$. This specification does not involve dynamic model averaging. Comparing averaged forecasts over different model dimensions and prior specifications against the forecasts of the other models, would also require the implementation of the dynamic model averaging technique for the competing models, which is beyond the scope of this paper. Moreover, we set $\hat{\Sigma}_0 = I_p$ instead of using a training sample.

3.3.2 Forecast Evaluation

Following the framework of D’Agostino et al. (2013), we compare the forecasts for a particular time period with the third release of the figures for that time period. So, we evaluate predictions against numbers which may have been revised two times. This means that we evaluate the last one-year-ahead prediction against the numbers in vintage 2015Q4.

We evaluate point forecasts using the root mean squared prediction error (RMSPE) and the mean absolute prediction error (MAPE). For the first (second) metric we set the point forecast equal to the mean (median) of the predictive density. We evaluate the forecast performance of the whole predictive density with the average predictive densities (APD). Beside assessing predictions with the traditional Bayesian forecast performance measures, like the RMSPE and the APD, we follow Groen et al. (2013) in evaluating density forecasts based on alternative measures such as the continuous ranked probability score (CRPS) and the quantile scores.

Where point forecasts emphasis the median or mean of the predictive density, there are often applications in which the tails of the predictive density are of special interest. For instance, in case VAR models are used for constructing impulse response functions to perform policy analysis. The outcomes of policy analysis are heavily affected by the tail behavior of predicted future outcomes. To also evaluate the performance of the tails of the predictive density we employ integrated weighted versions of Gneiting and Raftery (2007) average quantile scores (avQS). With the avQS-C, avQS-R, and avQS-L we evaluate the center, the right tail, and the left tail of the predictive density, respectively. The exact formulas of all evaluation measures are given in Appendix C.

3.3.3 Forecast Performance

The real-time forecasting exercise results in forecasts of five different models¹, for all three variables included in the monetary VAR, over five horizons, for both a small monetary VAR and a large dimensional VAR. Tables 3, 4 and 5 show the values of the predictive performance measures for these forecasts. Underscores indicate the best performing model for a specific horizon and variable, according to a particular evaluation criterion. Except for the average predictive density, the best performing model is the one that produces forecasts with small values for the forecast performance statistics.

[Table 3 about here.]

We find that in general, the infinite hidden Markov model with a time-varying covariance matrix shows the best performance, where the unrestricted infinite hidden Markov model is too flexible and the homogeneous linear VAR too restrictive. Table 3 shows that the unrestricted infinite hidden Markov model outperforms the other models in only one case and the linear BVAR results only once in the lowest RMSPE. Also on the predictive density evaluation measures in Tables 4 and 5, the unrestricted model shows in a few cases the best performance, but the linear BVAR is systematically outperformed.

The restricted infinite hidden Markov model results in all cases in the lowest RMSPE for unemployment. The results for inflation and interest rate vary per horizon and model dimension. The P05 model improves upon forecast accuracy in predicting inflation in the small VAR. Since the estimation procedure of this model cannot be scaled up to high dimensional models, this model is absent in the forecast comparison for the large VAR. The KK13 model replaces the P05 model as best performer for inflation here. The KK13 model and the restricted infinite hidden Markov model are close competitors on accurate point forecasts for interest rate. Table D1 in Appendix D shows the MAPE, another point forecast evaluation criterium, which is more friendly against outliers. This metric shows

¹Although we only present results for the unrestricted infinite hidden Markov model and a variant in which the coefficients are restricted to be time-invariant, Appendix A shows that we can also restrict the covariance matrix, the long term impact matrix, or both. Forecast results for these models do not alter the main findings of our analysis and are available upon request.

results similar to the RMSPE, with the restricted infinite hidden Markov model being superior in forecasting unemployment, P05 consistently better in forecasting inflation in the small model, and in some cases KK13 shows the best performance on inflation and interest rate in the large VAR.

When we take the whole predictive density into account, we find impressive results in favor of the restricted infinite hidden Markov model. Table 4 shows that this model beats all benchmark models for all horizons and model dimensions. The only competing model on the predictive density is the unrestricted infinite hidden Markov model. In the small VAR model, the linear BVAR shows the worst predictive performance of the benchmark models, followed by the KK13 and P05 models. However, unless for the one-period-ahead forecasts, the KK13 model cannot increase in the value of the predictive density relative to the linear BVAR in the large dimensional model.

[Table 4 about here.]

The average predictive density is computed over the forecasts for all variables included in the model. However, out of all variables included in the large VAR model, we focus on unemployment, inflation, and interest rate. The continuous ranked probability score separately evaluates the predictive densities of each variable. Table 5 shows that the strong performance in density forecasts of the restricted infinite hidden Markov model is based on the forecasts for unemployment and interest rates. The P05 model does again a good job in predicting inflation in the small model. In the large dimensional model, the KK13 model only sometimes perform better on short-term density forecasts. Tables D2, D3, and D4 in Appendix D show the quantile scores, which show in which part of the predictive densities which model performs best. In summary, the restricted infinite hidden Markov model in the small VAR performs often better than benchmarks model in the important left-tail.

[Table 5 about here.]

Comparing the forecasts in the small VAR against the forecasts in the large VAR model, we find that in most cases the forecast quality deteriorates with larger model dimensions. Table 3 shows that only for short horizons there is an

improvement in forecast accuracy after adding variables to the small monetary VAR model. In contrast to the average predictive densities, we can compare the density forecast performance between the small and large models using the continuous ranked probability score. Table 5 shows that in the short term, the large model does a better job in density forecasts for unemployment. In general, there is no clear increase in performance by adding more variables.

In sum, the infinite hidden Markov model with a time-varying covariance matrix shows for most forecast horizons and variables the best performance based on various evaluation measures, and is always close to the best performing model if it is not the best one itself. For all considered evaluation measures and forecast horizons, the infinite hidden Markov model outperforms the benchmark models in forecasting unemployment. We find that inflation and interest rate are for some horizons better predicted by the KK13 or P05 models based on point forecast evaluation metrics. However, based on the predictive densities the infinite hidden Markov model shows superior forecast performance. Finally, increasing the number of variables in the VAR does, in general, not lead to an increase in forecast performance for unemployment, inflation rate, and interest rate.

4 Conclusion

In this paper we propose a new method to estimate time-varying parameters in a VAR model. To avoid the curse of dimensionality, we opt for a semi-parametric approach. The infinite hidden Markov model encourages estimation of a parsimonious model by clustering parameter values over time, without restricting the parameter space. To accommodate for persistence in macroeconomic data, we impose the Dirichlet process mixture on the transition probabilities in a hidden Markov-switching framework. Parameter values are assigned to a possibly infinite number of states, with a potentially increased probability of self-transition. Except from the degree L weak limit approximation, which comes with negligible costs, the estimation algorithm of the model does not impose any restrictions on or (linear) approximations to the parameters.

The empirical application shows that the semi-parametric Bayesian framework is a promising alternative for parametric approaches to time-varying parameter VAR modelling. We identify both abrupt and smooth parameter changes in a

structural analysis and find posterior evidence for time-varying volatility. A real-time forecasting exercise shows that over a collection of forecast evaluation criteria the infinite hidden Markov model often outperforms popular benchmark models, even in large VAR models consisting of ten variables.

References

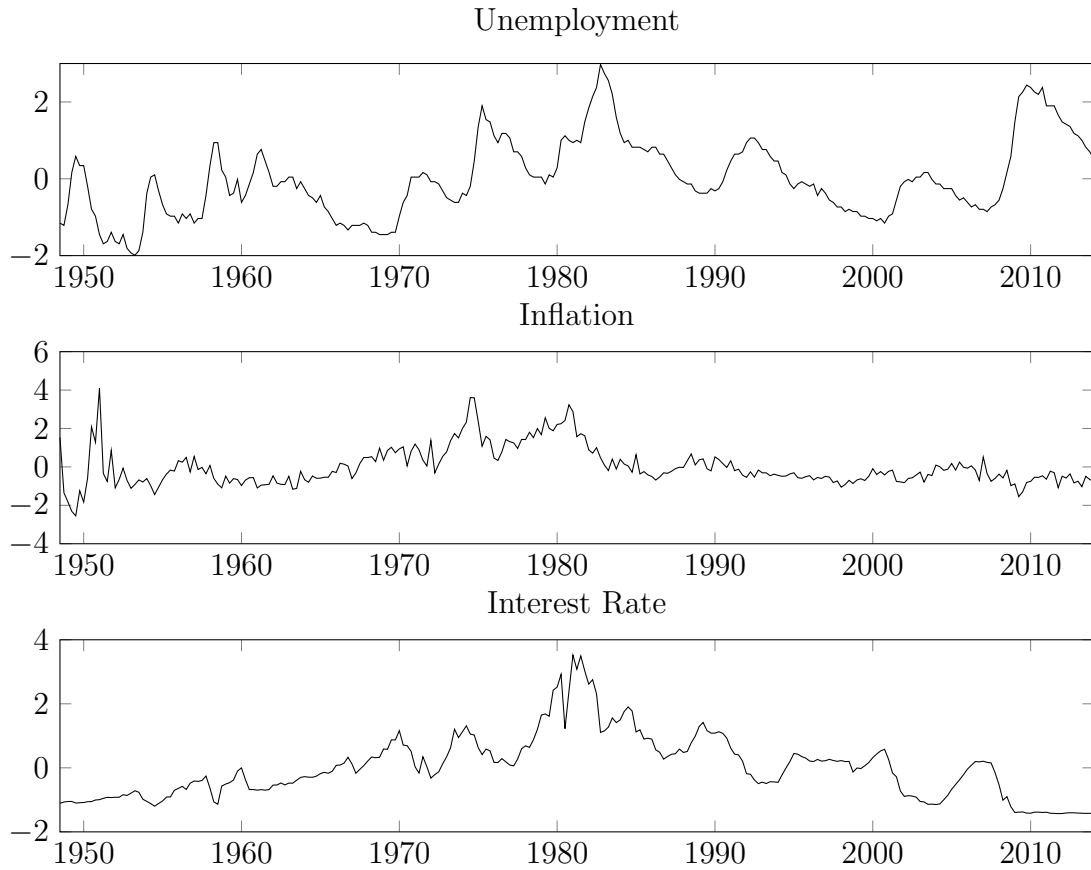
- Bańbura, M., Giannone, D., and Reichlin, L. Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92, 2010.
- Bauwens, L., Carpentier, J.-F., and Dufays, A. Autoregressive moving average infinite hidden markov-switching models. *Journal of Business & Economic Statistics*, 35(2):162–182, 2017.
- Carriero, A., Clark, T. E., and Marcellino, M. Bayesian VARs: specification choices and forecast accuracy. *Journal of Applied Econometrics*, 30(1):46–73, 2015a.
- Carriero, A., Clark, T. E., and Marcellino, M. Common drifting volatility in large Bayesian VARs. *Journal of Business & Economic Statistics*, (forthcoming), 2015b.
- Chib, S. Estimation and comparison of multiple change-point models. *Journal of Econometrics*, 86(2):221–241, 1998.
- Chib, S., Nardari, F., and Shephard, N. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371, 2006.
- Chopin, N. and Pelgrin, F. Bayesian inference and state number determination for hidden Markov models: an application to the information content of the yield curve about inflation. *Journal of Econometrics*, 123(2):327–344, 2004.
- Clark, T. E. Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility. *Journal of Business & Economic Statistics*, 29(3):327–341, 2012.

- Clark, T. E. and Ravazzolo, F. Macroeconomic forecasting performance under alternative specifications of time-varying volatility. *Journal of Applied Econometrics*, 30(4):551–575, 2015.
- Cogley, T. and Sargent, T. J. Evolving post-world WWII US inflation dynamics. *NBER Macroeconomics Annual 2001*, 16:331–388, 2002.
- Cogley, T. and Sargent, T. J. Drifts and volatilities: monetary policies and outcomes in the post WWII US. *Review of Economic Dynamics*, 8(2):262–302, 2005.
- Cogley, T., Primiceri, G. E., and Sargent, T. J. Inflation-gap persistence in the US. *American Economic Journal: Macroeconomics*, 2(1):43–69, 2010.
- D’Agostino, A., Gambetti, L., and Giannone, D. Macroeconomic forecasting and structural change. *Journal of Applied Econometrics*, 28(1):82–101, 2013.
- Doan, T., Litterman, R., and Sims, C. Forecasting and conditional projection using realistic prior distributions. *Econometric reviews*, 3(1):1–100, 1984.
- Escobar, M. D. and West, M. Bayesian density estimation and inference using mixtures. *Journal of the American Statistical Association*, 90(430):577–588, 1995.
- Ferguson, T. S. A Bayesian analysis of some nonparametric problems. *The annals of statistics*, 1(2):209–230, 1973.
- Fox, E. B., Sudderth, E. B., Jordan, M. I., and Willsky, A. S. A sticky hdp-hmm with application to speaker diarization. *The Annals of Applied Statistics*, 5(2A):1020–1056, 2011.
- Francq, C. and Zakoian, J.-M. Stationarity of multivariate markov-switching arma models. *Journal of Econometrics*, 102(2):339–364, 2001.
- Frühwirth-Schnatter, S. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453):194–209, 2001.

- Geweke, J. Interpretation and inference in mixture models: Simple mcmc works. *Computational Statistics & Data Analysis*, 51(7):3529–3550, 2007.
- Giannone, D., Lenza, M., and Primiceri, G. E. Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451, 2015.
- Giordani, P., Kohn, R., and van Dijk, D. A unified approach to nonlinearity, structural change, and outliers. *Journal of Econometrics*, 137(1):112–133, 2007.
- Gneiting, T. and Raftery, A. E. Strictly proper scoring rules, prediction, and estimation. *Journal of the American Statistical Association*, 102(477):359–378, 2007.
- Groen, J. J., Paap, R., and Ravazzolo, F. Real-time inflation forecasting in a changing world. *Journal of Business & Economic Statistics*, 31(1):29–44, 2013.
- Hamilton, J. D. A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2):357–384, 1989.
- Ishwaran, H. and Zarepour, M. Exact and approximate sum representations for the dirichlet process. *Canadian Journal of Statistics*, 30(2):269–283, 2002.
- Jochmann, M. Modeling US inflation dynamics: A Bayesian nonparametric approach. *Econometric Reviews*, 34(5):537–558, 2015.
- Koop, G. and Korobilis, D. Large time-varying parameter VARs. *Journal of Econometrics*, 177(2):185–198, 2013.
- Koop, G. and Potter, S. M. Estimation and forecasting in models with multiple breaks. *The Review of Economic Studies*, 74(3):763–789, 2007.
- Koop, G. M. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203, 2013.
- Liu, P., Mumtaz, H., Theodoridis, K., and Zanetti, F. Changing macroeconomic dynamics at the zero lower bound. *Journal of Business & Economic Statistics*, (just-accepted), 2017.

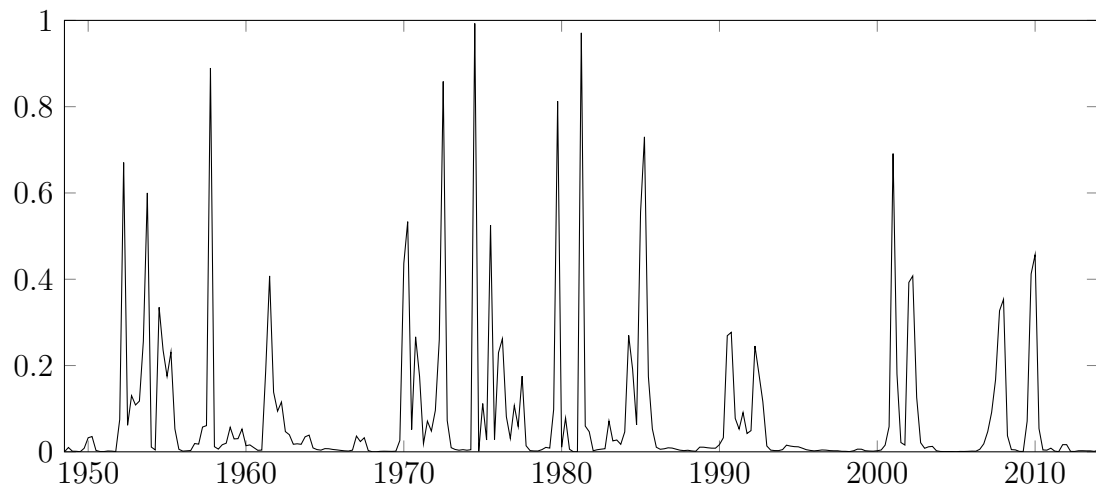
- Pesaran, M. H., Pettenuzzo, D., and Timmermann, A. Forecasting time series subject to multiple structural breaks. *The Review of Economic Studies*, 73(4): 1057–1084, 2006.
- Primiceri, G. E. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.
- Sims, C. A. Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48, 1980.
- Sims, C. A. and Zha, T. Were there regime switches in US monetary policy? *The American Economic Review*, pages 54–81, 2006.
- Song, Y. Modelling regime switching and structural breaks with an infinite hidden Markov model. *Journal of Applied Econometrics*, 29(5):825–842, 2014.
- Teh, Y. W., Jordan, M. I., Beal, M. J., and Blei, D. M. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2012.

Figure 1: Time Series Small Monetary VAR U.S. Economy



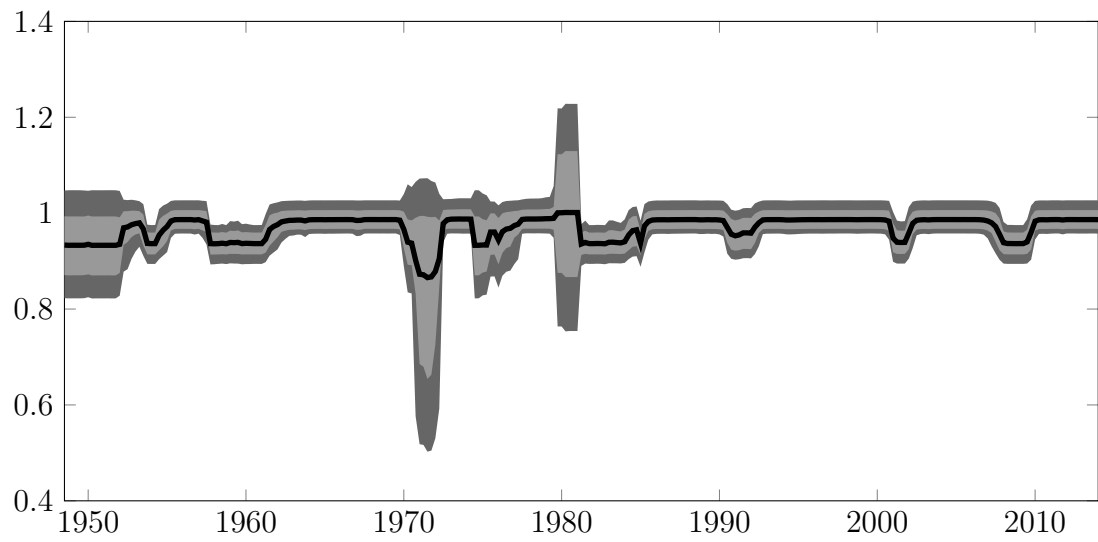
This figure shows the standardized quarterly data series as included in the small monetary VAR of the U.S. economy, with the unemployment rate, inflation rate, and interest rate as variables. The inflation series represents 400 times the first difference of the logarithm of the consumer price index for all urban consumers. The interest rate denotes the effective federal funds rate in percentages. The sample period runs from the first quarter of 1948 to the last quarter of 2015.

Figure 2: Posterior Break Probabilities



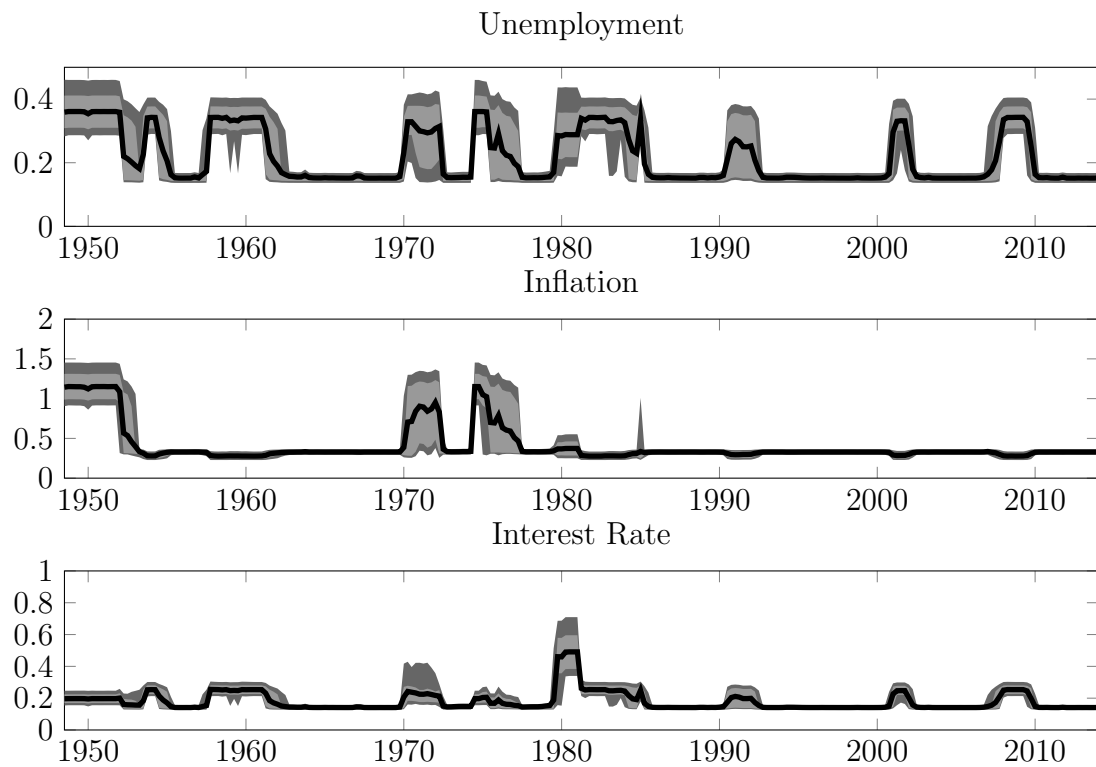
This figure shows for each time period the posterior probability of switching regime in the current period compared to the previous period, computed as the sum of draws in the sampler with a regime-switch in a specific time period divided by the total number of draws.

Figure 3: Largest Absolute Eigenvalues Companion Form



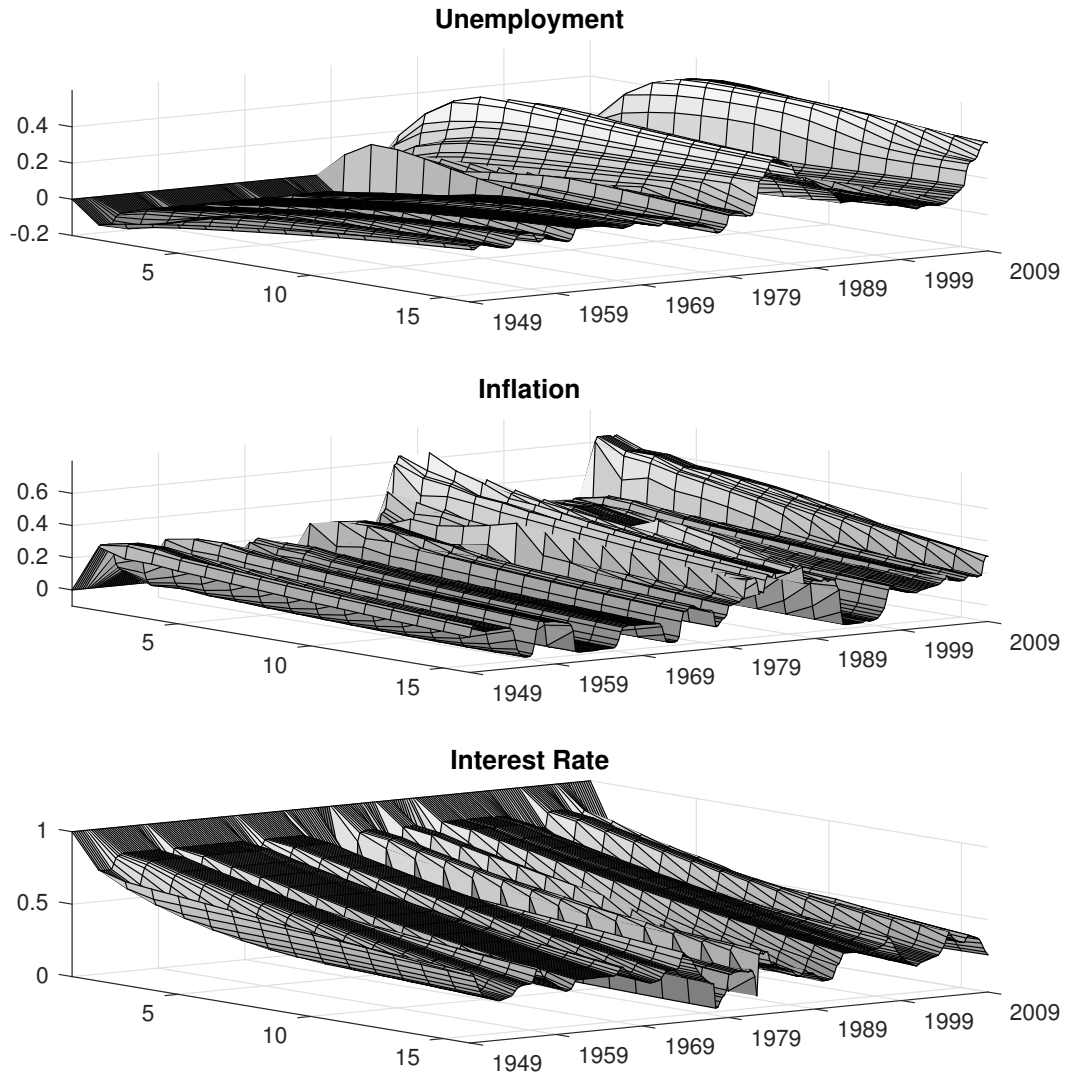
This figure shows the time-varying posterior mean (solid line) of the largest absolute eigenvalue of the companion form of the reduced form VAR model. The shaded areas represent the 68% and 90% confidence bands.

Figure 4: Posterior Means of the Structural Variance Parameters



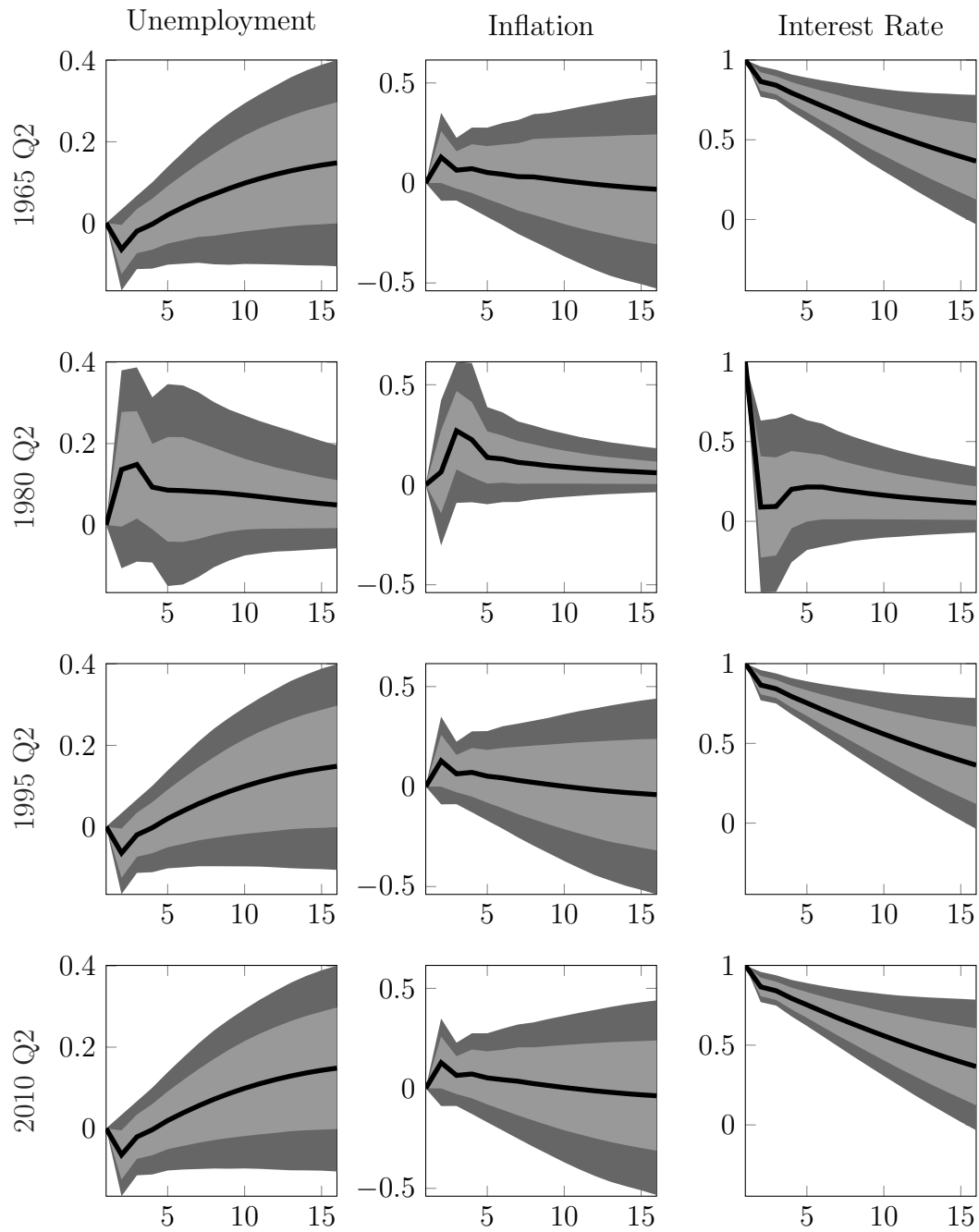
This figure shows the time-varying posterior means (solid line) of the structural variance parameters together with the 68% and 90% confidence bands. The panels show from top to bottom the variance of the residuals in the unemployment equation, inflation equation, and interest equation, respectively.

Figure 5: Impulse Response Functions



This figure shows the posterior means of the impulse response functions to a monetary policy shock. From top to bottom we have the responses in the unemployment equation, inflation equation, and interest equation. The y-axis runs from 0 to 16 and traces out the effect of the shock over a period of four years. The x-axis indicates at which date a shock hits the system. The z-axis shows the magnitude of the response. The monetary policy shock is defined as a one unit shock to interest rate.

Figure 6: Impulse Response Functions



This figure shows the posterior means (solid line) of the impulse response functions together with the 68% and 90% confidence bands for different moments in time. The monetary policy shock is defined as a one unit shock to interest rate.

Table 1: Parameters of Prior Distributions

a_α	b_α	a_γ	b_γ	c_ρ	d_ρ	ν_Σ	S_Σ	b_B	$V_{B[i=j]}$	$V_{B[i \neq j]}$
1	10	1	10	10	1	$p + 2$	$\frac{1}{\nu_\Sigma} I_p$	$0_{p \times k}$	$\left(\frac{\lambda}{\text{lag}^2} \right)$	0

This table shows the parameters of the priors as discussed in Section 2.2, where $0_{p \times q}$ represents a zero matrix of size $p \times q$ and I_q is the identity matrix of dimension q . The diagonal elements of V_B are scaled by the lag order of the corresponding variables in x_t , where the lag length equals λ for the intercept.

Table 2: Posterior Probability of Positive Impulse Response

	Unemployment		Inflation		Interest Rate	
	1 year	4 year	1 year	4 year	1 year	4 year
1965 Q2	49.4	83.9	71.7	46.4	100.0	93.4
1980 Q2	81.0	79.0	88.8	85.7	79.4	86.2
1995 Q2	49.3	83.9	71.6	44.6	100.0	93.1
2010 Q2	49.2	83.8	71.8	45.1	100.0	93.3

This table shows the posterior probabilities that the impulse response functions in Figure 6 are larger than zero, one year and four years after a monetary policy shock.

Table 3: Forecasting Results RMSPE

hor.	var.	small VAR					large VAR			
		IHM		θ	KK13	P05	IHM		KK13	
		θ_{s_t}	Σ_{s_t}		$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{s_t}	Σ_{s_t}	θ	$\theta_t \sim \text{ff}$
1	UR	0.554	<u>0.386</u>	0.444	0.387	0.407	0.398	<u>0.337</u>	0.357	0.363
	Infl.	1.497	1.465	1.516	1.460	<u>1.433</u>	1.486	1.449	1.498	<u>1.375</u>
	IR	1.035	0.869	0.899	<u>0.847</u>	0.864	0.924	0.849	0.871	<u>0.811</u>
2	UR	0.834	<u>0.621</u>	0.723	0.624	0.674	0.612	<u>0.526</u>	0.593	0.560
	Infl.	1.712	1.650	1.756	1.686	<u>1.597</u>	1.680	1.620	1.688	<u>1.549</u>
	IR	1.440	1.225	1.251	<u>1.205</u>	1.242	1.287	<u>1.186</u>	1.242	1.190
3	UR	1.049	<u>0.805</u>	0.939	0.825	0.930	0.847	<u>0.734</u>	0.844	0.763
	Infl.	1.893	1.838	1.982	1.859	<u>1.745</u>	1.890	1.895	1.903	<u>1.720</u>
	IR	1.730	<u>1.435</u>	1.508	1.474	1.551	1.620	<u>1.499</u>	1.571	1.502
4	UR	1.195	<u>0.964</u>	1.109	0.998	1.228	1.092	<u>0.941</u>	1.068	0.962
	Infl.	2.067	2.065	2.222	2.091	<u>2.004</u>	2.269	2.268	2.212	<u>2.036</u>
	IR	2.017	<u>1.713</u>	1.788	1.771	1.899	1.859	1.787	1.848	<u>1.786</u>
5	UR	1.301	<u>1.100</u>	1.236	1.150	1.636	3.474	<u>1.120</u>	1.248	1.131
	Infl.	<u>2.096</u>	2.247	2.383	2.332	2.374	4.191	2.455	2.343	<u>2.235</u>
	IR	2.624	<u>1.961</u>	2.030	2.068	2.313	4.846	2.036	2.054	<u>2.017</u>

This table shows the RMSPE for unemployment (UR), inflation (Infl.), and interest rate (IR) over five different horizons; from one-quarter ahead till five-quarters ahead. Forecasts are produced by the infinite hidden Markov model in which all model parameters $\theta_{s_t} = \{B_{s_t}, \Sigma_{s_t}\}$ change over regimes, a version with only a time-varying covariance matrix Σ_{s_t} , and the linear Bayesian VAR (column θ). The benchmark time-varying parameter VAR models of Koop and Korobilis (2013), in which time-variation is governed by forgetting factors (ff), and Primiceri (2005), wherein time-variation is modelled as random walks (rw), are denoted as KK13 and P05, respectively. The left panel shows results in a small VAR and the right panel in a large VAR model. Underscores indicate the best performing model for a specific horizon, variable and model dimension.

Table 4: Forecasting Results APD

hor.	small VAR					large VAR			
	IHM		θ	KK13	P05	IHM		KK13	
	θ_{s_t}	Σ_{s_t}		$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{s_t}	Σ_{s_t}	θ	$\theta_t \sim \text{ff}$
1	0.569	<u>1.924</u>	0.469	1.352	1.675	0.171	<u>0.335</u>	0.187	0.323
2	0.266	<u>0.678</u>	0.223	0.470	0.555	0.057	<u>0.153</u>	0.060	0.048
3	0.170	<u>0.387</u>	0.143	0.242	0.289	<u>0.069</u>	0.065	0.031	0.010
4	0.119	<u>0.249</u>	0.106	0.137	0.157	0.018	<u>0.042</u>	0.021	0.002
5	0.090	<u>0.166</u>	0.083	0.084	0.095	<u>0.090</u>	0.022	0.015	0.001

This table shows the APD for five different horizons; from one-quarter ahead till five-quarters ahead. For additional information, see the note following Table 3.

Table 5: Forecasting Results avCRPS

		small VAR					large VAR			
		IHM		KK13	P05	IHM			KK13	
hor.	var.	θ_{s_t}	Σ_{s_t}	θ	$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{s_t}	Σ_{s_t}	θ	$\theta_t \sim \text{ff}$
1	UR	0.288	<u>0.199</u>	0.243	0.204	0.215	0.215	<u>0.181</u>	0.199	0.196
	Infl.	0.819	<u>0.793</u>	0.842	0.788	<u>0.775</u>	0.814	0.797	0.822	<u>0.755</u>
	IR	0.512	<u>0.406</u>	0.468	0.420	0.430	0.492	0.442	0.462	<u>0.428</u>
2	UR	0.435	<u>0.314</u>	0.376	0.326	0.346	0.332	<u>0.284</u>	0.324	0.306
	Infl.	0.913	<u>0.869</u>	0.958	0.892	<u>0.846</u>	0.934	0.901	0.937	<u>0.860</u>
	IR	0.737	<u>0.615</u>	0.674	0.631	0.672	0.710	<u>0.643</u>	0.682	0.648
3	UR	0.550	<u>0.408</u>	0.480	0.432	0.466	0.461	<u>0.392</u>	0.455	0.428
	Infl.	1.017	<u>0.965</u>	1.077	0.988	<u>0.938</u>	1.034	1.039	1.049	<u>0.996</u>
	IR	0.922	<u>0.767</u>	0.840	0.817	0.893	0.913	<u>0.846</u>	0.883	0.852
4	UR	0.634	<u>0.498</u>	0.566	0.529	0.591	0.572	<u>0.498</u>	0.572	0.559
	Infl.	1.111	1.068	1.188	1.112	<u>1.059</u>	<u>1.177</u>	1.219	1.197	1.213
	IR	1.086	<u>0.931</u>	1.001	0.998	1.108	1.066	<u>1.019</u>	1.046	1.046
5	UR	0.699	<u>0.575</u>	0.633	0.619	0.718	0.654	<u>0.588</u>	0.661	0.697
	Infl.	1.193	<u>1.168</u>	1.279	1.253	1.216	<u>1.210</u>	1.318	1.276	1.438
	IR	1.233	<u>1.100</u>	1.153	1.181	1.323	1.186	<u>1.167</u>	1.176	1.246

This table shows the avCRPS for unemployment (UR), inflation (Infl.), and interest rate (IR) over five different horizons; from one-quarter ahead till five-quarters ahead. For additional information, see the note following Table 3.

Appendix A Parameter Restrictions

This appendix discusses the sample steps for three different restricted versions of model (1); a model in which either the coefficient matrix, covariance matrix, or long-term impact matrix are set to be constant over time. posterior results of the parameters in these restricted models are obtained with only slight modifications to the sampler in Subsection 2.2. In practice, we adjust Step 6 of the sampler and add an extra step to sample the fixed parameters outside the structure of the mixture model.

The model in (1) can be generalized to

$$y_t = B_t x_t + C \tilde{x}_t + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \Sigma_t), \quad t = 1, \dots, T,$$

from which follows the unrestricted model in (1) by setting the parameters in C equal to zero. Imposing $\Sigma_t = \Sigma$ results in the model with a time-invariant covariance matrix. A model with time-invariant macroeconomic relations is defined by $x_t = 1$, $\tilde{x}_t = [y'_{t-1}, \dots, y'_{t-l}]'$, and $\tilde{y}_t = y_t - B_t x_t$. Finally, the short-term model follows from $x_t = [1, y'_{t-1}]'$, $\tilde{x}_t = [\Delta y'_{t-1}, \dots, \Delta y'_{t-l}]'$, and $\tilde{y}_t = y_t - B_t x_t$.

We sample B_t as in step 6 of the sample algorithm but we take $(y_t - C \tilde{x}_t)$ for y_t . The same holds for Σ_t when the covariance parameters are unrestricted. After the seventh step, when the state assignments in the mixture model are settled down for the current iteration, we sample the time-invariant parameter matrices.

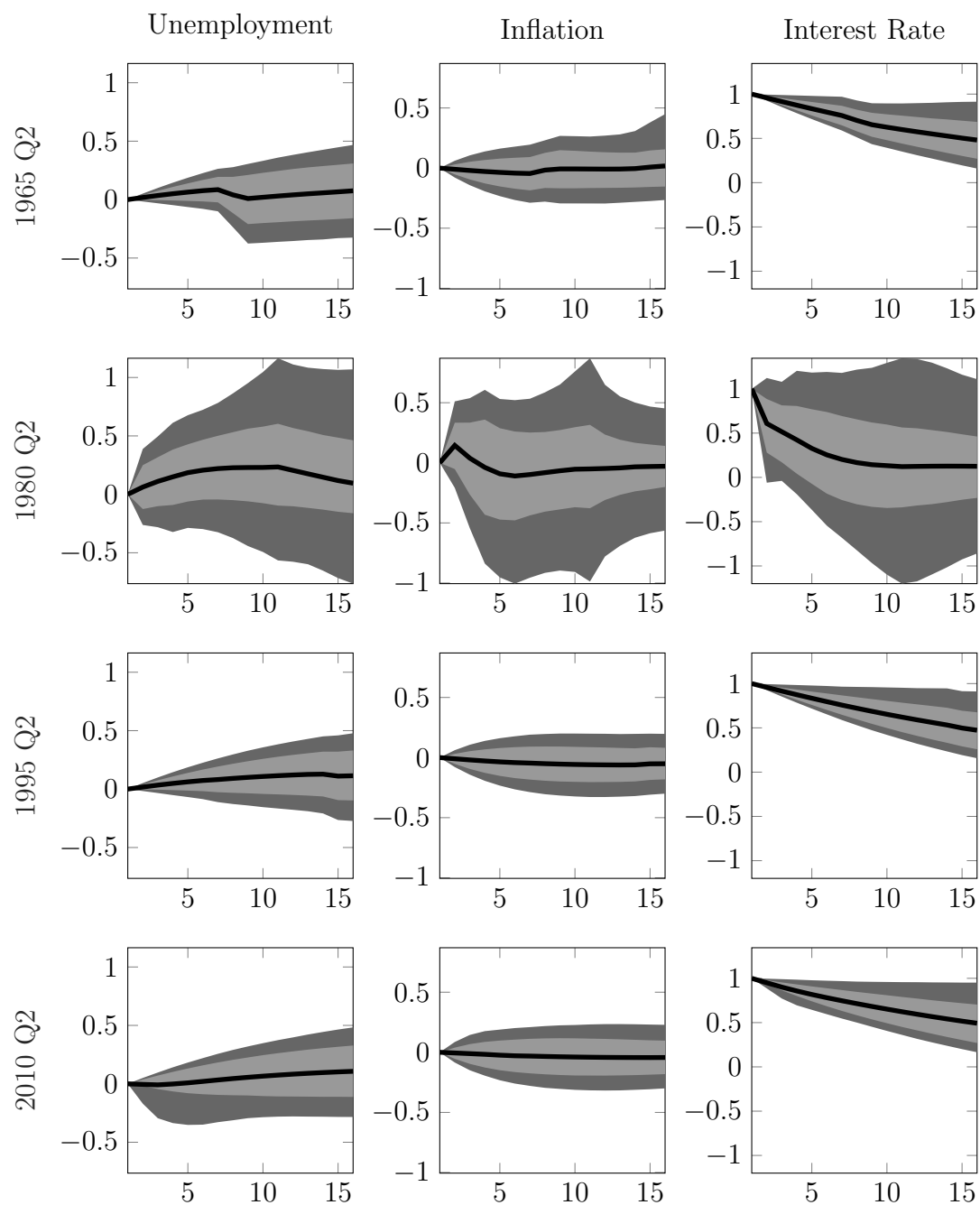
In models with a restricted covariance matrix we compute $\varepsilon = (\varepsilon'_1, \dots, \varepsilon'_T)'$ where $\varepsilon_t = y'_t - x'_t B'_t - \tilde{x}'_t C'$ and add the sampling step

$$\bar{S} = S_\Sigma + \varepsilon' \varepsilon, \quad \Sigma_j | y, \Theta \sim \mathcal{IW}(\nu_\Sigma + T, \bar{S}).$$

For the model with restricted coefficients we compute $\tilde{X} = (\Sigma_1^{-\frac{1}{2}} \tilde{X}'_1, \dots, \Sigma_T^{-\frac{1}{2}} \tilde{X}'_T)'$, where $\tilde{X}_t = I_p \otimes \tilde{x}_t$, $t = 1, \dots, T$, and I_p is the identity matrix of dimension p , and $\tilde{Y} = (\tilde{y}'_1, \dots, \tilde{y}'_T)'$. Now we can perform the sampling steps

$$\begin{aligned} \bar{B} &= (\tilde{X}' \tilde{X} + (I_p \otimes V_B)^{-1})^{-1}, \quad \bar{b} = \bar{B}(\tilde{X}' \tilde{Y} + (I_p \otimes V_B)^{-1} \text{vec}(b_B)), \\ \text{vec}(B) | \tilde{y}, \Sigma, \Theta &\sim \mathcal{N}(\text{vec}(\bar{b}), \bar{B}). \end{aligned}$$

Appendix B Impulse Response Functions



This figure shows impulse response functions constructed from a model with time-invariant long term impact matrix. For additional information, see the note following Figure 6.

Appendix C Forecast Performance

We evaluate point forecasts using the root mean squared prediction error (RMSPE) and the mean absolute prediction error (MAPE). The RMSPE of the forecast produced by model M for variable i at horizon h is

$$\text{RMSPE}_{ih}^M = \sqrt{\frac{1}{P} \sum_{t=T+1}^{T+P} \left(\hat{y}_{t+h}^{(i)}(M) - y_{t+h}^{(i)} \right)^2},$$

where $\hat{y}_{t+h}^{(i)}(M)$ is one of the P point forecast of the i th variable $y_{t+h}^{(i)}$ made by model M . We set the point forecast equal to the mean of the predictive density. The MAPE is defined by

$$\text{MAPE}_{ih}^M = \frac{1}{P} \sum_{t=T+1}^{T+P} \left| \hat{y}_{t+h}^{(i)}(M) - y_{t+h}^{(i)} \right|,$$

with the point forecast equal to the median of the predictive density.

We evaluate the forecast performance of the whole predictive density with the average predictive densities (APD)

$$f^M(y_{t+h}) = \frac{1}{P} \sum_{t=T+1}^{T+P} \left(\frac{1}{S} \sum_{s=1}^S \mathcal{N}(B_{s_{T+h}}^{(i)} x_{T+h-1}^{(i)}, \Sigma_{s_{T+h}}^{(i)}) \right),$$

where S denotes the number of simulations.

The continuous ranked probability score (CRPS) is computed as

$$\text{CRPS}_t(y_{t+h}^{(i)}) = E_f |Y_{t+h}^{(i)} - y_{t+h}^{(i)}| - \frac{1}{2} E_f |Y_{t+h}^{(i)} - Y_{t+h}'^{(i)}|,$$

where f is the predictive density function of model M for prediction $y_{t+h}^{(i)}$, E_f is the expectation operator over the function f , $|\cdot|$ denotes the absolute value, and $Y_{t+h}^{(i)}$ and $Y_{t+h}'^{(i)}$ are independent random variables with sampling density f . The average CRPS across all forecasts is

$$\text{avCRPS}_{ih}^M = \frac{1}{P} \sum_{t=T+1}^{T+P} \text{CRPS}_t(y_{t+h}^{(i)}).$$

The avQS-C, avQS-R, and avQS-L evaluate the center, the right tail, and the left tail of the predictive density, respectively.

$$\begin{aligned} \text{avQS-C}_{ih}^M &= \frac{1}{P} \sum_{t=T+1}^{T+P} \left(\frac{1}{99} \sum_{j=1}^{99} \alpha_j (1 - \alpha_j) \text{QS}(\alpha_j, y_{t+h}^{(i)}, M) \right), \\ \text{avQS-R}_{ih}^M &= \frac{1}{P} \sum_{t=T+1}^{T+P} \left(\frac{1}{99} \sum_{j=1}^{99} \alpha_j^2 \text{QS}(\alpha_j, y_{t+h}^{(i)}, M) \right), \\ \text{avQS-L}_{ih}^M &= \frac{1}{P} \sum_{t=T+1}^{T+P} \left(\frac{1}{99} \sum_{j=1}^{99} (1 - \alpha_j)^2 \text{QS}(\alpha_j, y_{t+h}^{(i)}, M) \right), \end{aligned}$$

where $\alpha_j = j/100$ and

$$\text{QS}(\alpha, y_{t+h}^{(i)}, M) = (I\{y_{t+h}^{(i)} \leq Q_f^\alpha\} - \alpha)(Q_f^\alpha - y_{t+h}^{(i)}),$$

where Q_f^α represents quantile α of the predictive density function f of model M for prediction $y_{t+h}^{(i)}$.

Appendix D Additional Forecasting Results

Table D1: Forecasting Results MAPE

		small VAR					large VAR			
		IHM		KK13	P05	IHM		KK13		
hor.	var.	θ_{s_t}	Σ_{s_t}	θ	$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{s_t}	Σ_{s_t}	θ	$\theta_t \sim \text{ff}$
1	UR	0.377	<u>0.270</u>	0.312	0.276	0.295	0.291	<u>0.251</u>	0.264	0.271
	Infl.	1.159	1.117	1.168	1.115	<u>1.078</u>	1.106	1.083	1.117	<u>1.050</u>
	IR	0.670	<u>0.537</u>	0.599	0.537	0.543	0.645	0.590	0.594	<u>0.564</u>
2	UR	0.575	<u>0.416</u>	0.500	0.437	0.462	0.447	<u>0.387</u>	0.437	0.399
	Infl.	1.255	1.222	1.307	1.235	<u>1.152</u>	1.317	1.269	1.327	<u>1.181</u>
	IR	1.004	<u>0.820</u>	0.924	0.828	0.859	0.992	0.860	0.957	<u>0.855</u>
3	UR	0.728	<u>0.526</u>	0.647	0.567	0.600	0.638	<u>0.527</u>	0.632	0.528
	Infl.	1.404	1.352	1.467	1.327	<u>1.243</u>	1.415	1.410	1.435	<u>1.266</u>
	IR	1.256	<u>1.046</u>	1.156	1.088	1.138	1.290	1.158	1.256	<u>1.137</u>
4	UR	0.836	<u>0.641</u>	0.772	0.689	0.759	0.793	<u>0.679</u>	0.794	0.681
	Infl.	1.544	1.473	1.627	1.476	<u>1.406</u>	1.622	1.663	1.638	<u>1.453</u>
	IR	1.476	<u>1.277</u>	1.382	1.337	1.403	1.500	1.405	1.482	<u>1.377</u>
5	UR	0.923	<u>0.742</u>	0.860	0.788	0.923	0.895	<u>0.790</u>	0.893	0.801
	Infl.	1.666	1.596	1.757	1.628	<u>1.574</u>	1.760	1.785	1.747	<u>1.606</u>
	IR	1.685	<u>1.514</u>	1.598	1.590	1.660	1.695	1.625	1.658	<u>1.592</u>

This table shows the MAPE for unemployment (UR), inflation (Infl.), and interest rate (IR) over five different horizons; from one-quarter ahead till five-quarters ahead. For additional information, see the note following Table 3.

Table D2: Forecasting Results avQS-left

		small VAR					large VAR			
		IHM		KK13	P05	IHM		KK13		
hor.	var.	θ_{st}	Σ_{st}	θ	$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{st}	Σ_{st}	θ	$\theta_t \sim \text{ff}$
1	UR	0.041	<u>0.029</u>	0.036	0.030	0.032	0.032	<u>0.027</u>	0.029	0.030
	Infl.	0.121	0.117	0.124	0.118	<u>0.115</u>	0.121	0.119	0.121	<u>0.116</u>
	IR	0.079	<u>0.064</u>	0.072	0.066	0.069	0.075	0.069	0.070	<u>0.069</u>
2	UR	0.059	<u>0.044</u>	0.053	0.046	0.050	0.049	<u>0.042</u>	0.048	0.047
	Infl.	0.131	0.128	0.138	0.132	<u>0.128</u>	0.135	0.132	0.137	<u>0.131</u>
	IR	0.109	<u>0.095</u>	0.100	0.098	0.104	0.104	0.101	<u>0.100</u>	0.102
3	UR	0.073	<u>0.056</u>	0.066	0.060	0.067	0.068	<u>0.058</u>	0.067	0.067
	Infl.	0.143	<u>0.139</u>	0.152	0.140	0.141	0.151	0.153	0.153	<u>0.151</u>
	IR	0.133	<u>0.116</u>	0.121	0.124	0.138	0.131	0.128	<u>0.126</u>	0.132
4	UR	0.084	<u>0.068</u>	0.077	0.074	0.085	0.083	<u>0.074</u>	0.084	0.089
	Infl.	<u>0.154</u>	0.154	0.168	0.156	0.158	<u>0.169</u>	0.177	0.171	0.181
	IR	0.156	<u>0.138</u>	0.143	0.152	0.171	0.152	0.153	<u>0.148</u>	0.163
5	UR	0.092	<u>0.078</u>	0.085	0.087	0.103	0.095	<u>0.086</u>	0.096	0.113
	Infl.	<u>0.164</u>	0.169	0.180	0.176	0.180	<u>0.179</u>	0.189	0.179	0.216
	IR	0.176	<u>0.164</u>	0.164	0.180	0.205	0.170	0.175	<u>0.165</u>	0.195

This table shows the avQS-1 for unemployment (UR), inflation (Infl.), and interest rate (IR) over five different horizons; from one-quarter ahead till five-quarters ahead. For additional information, see the note following Table 3.

Table D3: Forecasting Results avQS-center

		small VAR					large VAR			
		IHM		KK13	P05	IHM		KK13		
hor.	var.	θ_{st}	Σ_{st}	θ	$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{st}	Σ_{st}	θ	$\theta_t \sim \text{ff}$
1	UR	0.028	<u>0.019</u>	0.023	0.020	0.021	0.021	<u>0.018</u>	0.019	0.019
	Infl.	0.081	0.078	0.083	0.078	<u>0.076</u>	0.080	0.078	0.080	<u>0.074</u>
	IR	0.049	<u>0.039</u>	0.045	0.040	0.041	0.047	0.043	0.044	<u>0.041</u>
2	UR	0.042	<u>0.030</u>	0.036	0.032	0.034	0.032	<u>0.028</u>	0.032	0.030
	Infl.	0.089	0.085	0.093	0.087	<u>0.082</u>	0.092	0.089	0.092	<u>0.084</u>
	IR	0.071	<u>0.060</u>	0.065	0.061	0.064	0.069	<u>0.062</u>	0.067	0.063
3	UR	0.053	<u>0.039</u>	0.046	0.042	0.045	0.045	<u>0.038</u>	0.044	0.041
	Infl.	0.099	0.095	0.105	0.096	<u>0.090</u>	0.102	0.102	0.103	<u>0.096</u>
	IR	0.089	<u>0.075</u>	0.082	0.080	0.085	0.090	<u>0.083</u>	0.087	0.083
4	UR	0.061	<u>0.048</u>	0.055	0.051	0.056	0.056	<u>0.049</u>	0.056	0.053
	Infl.	0.109	0.104	0.116	0.108	<u>0.101</u>	0.116	0.119	0.117	<u>0.115</u>
	IR	0.105	<u>0.091</u>	0.098	0.097	0.105	0.105	<u>0.100</u>	0.103	0.101
5	UR	0.067	<u>0.055</u>	0.061	0.059	0.068	0.064	<u>0.057</u>	0.064	0.065
	Infl.	0.117	<u>0.114</u>	0.125	0.121	0.115	<u>0.125</u>	0.129	0.125	0.134
	IR	0.120	<u>0.108</u>	0.113	0.115	0.125	0.118	<u>0.114</u>	0.116	0.119

This table shows the avCS-c for unemployment (UR), inflation (Infl.), and interest rate (IR) over five different horizons; from one-quarter ahead till five-quarters ahead. For additional information, see the note following Table 3.

Table D4: Forecasting Results avQS-right

		small VAR					large VAR			
		IHM		KK13	P05	IHM		KK13		
hor.	var.	θ_{st}	Σ_{st}	θ	$\theta_t \sim \text{ff}$	$\theta_t \sim \text{rw}$	θ_{st}	Σ_{st}	θ	$\theta_t \sim \text{ff}$
1	UR	0.049	<u>0.032</u>	0.040	0.034	0.034	0.034	<u>0.029</u>	0.033	0.031
	Infl.	0.132	0.127	0.137	0.125	<u>0.125</u>	0.131	0.128	0.133	<u>0.116</u>
	IR	0.081	<u>0.063</u>	0.075	0.065	0.067	0.078	0.069	0.074	<u>0.065</u>
2	UR	0.076	<u>0.053</u>	0.064	0.055	0.057	0.054	<u>0.046</u>	0.053	0.048
	Infl.	0.151	0.140	0.159	0.144	<u>0.136</u>	0.152	0.146	0.151	<u>0.136</u>
	IR	0.121	<u>0.096</u>	0.109	0.099	0.106	0.115	0.099	0.110	<u>0.099</u>
3	UR	0.099	<u>0.071</u>	0.084	0.074	0.079	0.075	<u>0.064</u>	0.073	0.068
	Infl.	0.172	0.159	0.182	0.167	<u>0.154</u>	0.168	0.168	0.171	<u>0.160</u>
	IR	0.153	<u>0.121</u>	0.139	0.129	0.143	0.150	0.134	0.146	<u>0.132</u>
4	UR	0.115	<u>0.088</u>	0.100	0.091	0.100	0.094	<u>0.081</u>	0.093	0.087
	Infl.	0.189	0.177	0.201	0.191	<u>0.175</u>	<u>0.197</u>	0.200	0.199	0.202
	IR	0.182	<u>0.150</u>	0.167	0.158	0.177	0.177	<u>0.162</u>	0.174	0.163
5	UR	0.127	<u>0.101</u>	0.112	0.107	0.123	0.111	<u>0.097</u>	0.110	0.108
	Infl.	0.205	<u>0.193</u>	0.217	0.215	0.203	<u>0.214</u>	0.219	0.215	0.243
	IR	0.208	<u>0.176</u>	0.193	0.187	0.213	0.201	<u>0.185</u>	0.196	0.196

This table shows the avCS-r for unemployment (UR), inflation (Infl.), and interest rate (IR) over five different horizons; from one-quarter ahead till five-quarters ahead. For additional information, see the note following Table 3.