

Department of Economics  
Working Paper No. 262

# Dealing with heterogeneity in panel VARs using sparse finite mixtures

Florian Huber

April 2018



# Dealing with heterogeneity in panel VARs using sparse finite mixtures

Florian Huber\*

WU Vienna University of Economics and Business

## Abstract

In this paper, we provide a parsimonious means of estimating panel VARs with stochastic volatility. We assume that coefficients associated with domestic lagged endogenous variables arise from a finite mixture of Gaussian distribution. Shrinkage on the cluster size is introduced through suitable priors on the component weights and cluster-relevant quantities are identified through novel normal-gamma shrinkage priors. To assess whether dynamic interdependencies between units are needed, we moreover impose shrinkage priors on the coefficients related to other countries' endogenous variables. Finally, our model controls for static interdependencies by assuming that the reduced form shocks of the model feature a factor stochastic volatility structure. We assess the merits of the proposed approach by using synthetic data as well as a real data application. In the empirical application, we forecast Eurozone unemployment rates and show that our proposed approach works well in terms of predictions.

**Keywords:** multi country models, density predictions, hierarchical modeling, factor stochastic volatility models.

**JEL Codes:** C11, C30, C53, E52.

---

\*Correspondence to: Florian Huber, WU Vienna University of Economics and Business. Welthandelsplatz 1, 1020 Vienna, Austria. E-mail: [fhuber@wu.ac.at](mailto:fhuber@wu.ac.at).

## 1 Introduction

In the present paper, we combine the literature on Bayesian panel VAR (PVAR) models (see [Canova and Ciccarelli, 2004; 2009](#); [Koop and Korobilis, 2016](#); [Korobilis, 2016](#)) with the literature on finite mixture models (see [Allenby et al., 1998](#); [Lenk and DeSarbo, 2000](#); [Frühwirth-Schnatter et al., 2004](#); [Frühwirth-Schnatter and Kaufmann, 2008](#); [Malsiner-Walli et al., 2016](#)). Our proposed framework efficiently alleviates overparameterization concerns by assuming that coefficients associated with domestic macroeconomic fundamentals come from a common, flexible mixture distribution. By contrast, coefficients related with other countries' lagged endogenous variables are pushed to zero by means of a global-local shrinkage prior in the spirit of [Griffin and Brown \(2010\)](#) and [Huber and Feldkircher \(2017\)](#). In addition, we assume that the errors feature a factor structure, providing a parsimonious representation of the variance-covariance matrix.

The existing literature on PVARs names three important dimensions of model uncertainty. The first one is whether to allow for lagged dependencies between countries (henceforth labeled as dynamic interdependencies, DIs) while the second dimension is concerned with modeling contemporaneous relations across the shocks in the system (called static interdependencies, SIs). The final dimension centers on the question whether coefficients associated with lagged domestic variables are homogenous across countries. Thus, if such "domestic" coefficients tend to be similar, so-called homogeneity restrictions might be imposed, effectively introducing the same set of coefficients for several countries and thus reducing the number of free parameters.

Considering the recent literature on model specification and selection in PVAR models suggests two commonly used approaches to deal with the aforementioned issues. The first strand of the literature suggests using shrinkage priors to stochastically select an appropriate model specification (see [Koop and Korobilis, 2016](#); [Korobilis, 2016](#)). In light of the huge number of potential restrictions, however, mixing issues typically arise, leading to weak convergence properties of existing algorithms ([Bhattacharya et al., 2015](#)). The second strand considers additional restrictions that reduce the dimension of the parameter space. For instance, [Canova and Ciccarelli \(2009\)](#) assume that the (time-varying) coefficients of the PVAR model feature a factor structure. This translates into statistical and computational gains since the dimension of the state space is massively reduced. Another prominent example are global VAR models ([Pesaran et al., 2004](#); [Dees et al., 2007](#); [Feldkircher and Huber, 2016](#); [Crespo Cuaresma et al., 2016](#); [Huber, 2016](#)) that introduce

parametric restrictions on the coefficients associated with other countries' endogenous variables, assuming that only weighted averages enter a given country-specific submodel.

In this contribution, we propose a way to link both approaches mentioned above. To control for cross-country homogeneity of the coefficients associated with lagged domestic quantities, we introduce a finite mixture model. The key assumption is that domestic coefficients arise from a flexible mixture of Gaussian distribution. A set of latent country-specific indicators enable us to quantitatively assess whether a given economy belongs to a certain country group. For example, our approach could be used to group countries endogenously in emerging and developed economies. Since this directly gives rise to the important issue of selecting an appropriate number of clusters (see [Richardson and Green, 1997](#); [Biernacki et al., 2000](#); [Dellaportas and Papageorgiou, 2006](#)), we follow [Malsiner-Walli et al. \(2016\)](#) and use an overfitting mixture model (i.e. with the number of clusters/groups set to a large value) in combination with a shrinkage prior on the component weights. Moreover, to unveil important driving forces in shaping the group allocation, we use normal-gamma (NG) shrinkage priors on the variances associated with an underlying common distribution that gives rise to the component-specific means. [Yau and Holmes \(2011\)](#) show that a shrinkage prior on the variances of the common distribution implies shrinkage on the (standardized) distance between cluster centers, shrinking component means towards a common location for a given covariate.

To cope with the large number of lagged dependencies across countries, we also use a specific version of a NG shrinkage prior that introduces a country-specific global scaling parameter to decide whether DIs are needed within a given country. We allow for country/variable specific deviations by introducing a set of local scaling parameters that provide additional flexibility in the presence of strong country-specific shrinkage on DIs. Finally, we solve issues associated with SIs through a factor stochastic volatility model on the errors of the system. This provides a parsimonious representation of the variance-covariance matrix and entails significant computational gains (see [Kastner and Huber, 2017](#)).

We illustrate the merits of our approach by carrying out a simulation study. Considering a wide range of different data generating processes (DGPs), we find that our framework works well in simulations when benchmarked against an unrestricted VAR model equipped with suitable shrinkage priors, a model akin to a random coefficient specification (i.e. with cluster size one), and a model estimated using OLS. In the empirical application, we use a dataset for eleven Eurozone member countries and predict unemployment rates over

a hold-out period of ten years. This exercise enables us to assess whether our approach improves upon a set of competing alternatives in terms of predictive accuracy.

The paper is structured as follows. Section 2 introduces the econometric framework proposed, specifies the prior setup adopted and outlines the posterior simulation algorithm. We then proceed to conduct our simulation exercise in Section 3. After providing evidence that our approach works well on synthetic data, we apply the model to real world data in Section 4. Finally, the last section summarizes and concludes the paper.

## 2 Econometric framework

Before proceeding to the model, it proves to be convenient to introduce some generic notation. In what follows, capitalized letters without a time index refer (unless otherwise noted) to full-data matrices, i.e.  $\mathbf{Y} = (\mathbf{y}_1, \dots, \mathbf{y}_T)'$ . The notation  $[\mathbf{Y}]_{j\bullet}$  selects the  $j$ th row of the matrix  $\mathbf{Y}$  while  $[\mathbf{Y}]_{\bullet j}$  selects the  $j$ th column of the matrix concerned. In addition, we let  $\mathbf{y}_{-i,t}$  denote the vector  $\mathbf{y}_t$  with the  $i$ th subvector excluded, i.e.  $\mathbf{y}_{-i,t} = (\mathbf{y}'_{1t}, \dots, \mathbf{y}'_{i-1t}, \mathbf{y}'_{i+1t}, \dots, \mathbf{y}'_{Nt})'$ . Finally, we let  $\bullet$  be a generic notation that indicates conditioning on all remaining coefficients in the model as well as the data.

### 2.1 The panel vector autoregressive model

In this paper, we aim to model a set of  $M$  macroeconomic and financial indicators across a set of  $N$  countries. For each country, the domestic quantities are stored in an  $M$ -dimensional vector  $\mathbf{y}_{it}$  for  $t = 1, \dots, T$  and  $i = 1, \dots, N$ , and consequently stacked vector  $\mathbf{y}_t = (\mathbf{y}'_{1t}, \dots, \mathbf{y}'_{Nt})'$  of dimension  $K = MN$ .

The panel VAR for a given economy  $i$  then reads,

$$\mathbf{y}_{it} = \boldsymbol{\beta}_i + \mathbf{A}_{i1}\mathbf{y}_{it-1} + \dots + \mathbf{A}_{iP}\mathbf{y}_{it-P} + \mathbf{B}_{i1}\mathbf{y}_{-i,t-1} + \dots + \mathbf{B}_{iP}\mathbf{y}_{-i,t-P} + \boldsymbol{\varepsilon}_{it}, \quad (2.1)$$

where  $\boldsymbol{\beta}_i$  is a  $M$ -dimensional intercept vector and  $\mathbf{A}_{ij}$  ( $j = 1, \dots, P$ ) denotes a set of  $M \times M$ -dimensional coefficient matrices associated with the  $P$  lags of  $\mathbf{y}_{it}$ . In what follows, we are going to label these parameters the domestic VAR coefficients. The impact of other countries' lagged dependent variables  $\mathbf{y}_{-i,t-p}$  is measured through the matrices  $\mathbf{B}_{ij}$ , which are of dimension  $M \times (N-1)M$ . Finally,  $\boldsymbol{\varepsilon}_{it} \sim \mathcal{N}(\mathbf{0}_M, \boldsymbol{\Sigma}_{it})$  is a Gaussian vector white noise process with a time-varying variance-covariance matrix  $\boldsymbol{\Sigma}_{it}$ .

Equation (2.1) can be cast in the usual regression form,

$$\mathbf{y}_{it} = \mathbf{C}_i \mathbf{x}_{it} + \mathbf{B}_i \mathbf{x}_{-i,t} + \varepsilon_{it}, \quad (2.2)$$

with  $\mathbf{x}_{it} = (1, \mathbf{y}'_{it-1}, \dots, \mathbf{y}'_{it-P})'$ ,  $\mathbf{C}_i = (\beta_i, \mathbf{A}_{i1}, \dots, \mathbf{A}_{iP})$ ,  $\mathbf{x}_{-i,t} = (\mathbf{y}'_{-i,t-1}, \dots, \mathbf{y}'_{-i,t-P})'$  and  $\mathbf{B}_i = (\mathbf{B}_{i1}, \dots, \mathbf{B}_{iP})'$ . The matrix  $\mathbf{B}_i$  establishes dynamic interdependencies (DI) between countries  $i$  and  $j$ . In the literature on PVAR models (see [Canova and Ciccarelli, 2013](#), for a recent survey), an important modeling decision is whether to set certain sub-matrices of  $\mathbf{B}_i$  to zero, shutting off dynamic relations between country pairs. An extreme version of the model would set the whole matrix  $\mathbf{B}_i$  to zero, ruling out lagged relations between country  $i$  and the remaining economies.

Up to this point, we remained silent on the specific assumptions on the error covariances across countries. In what follows we are going to stack the country-specific errors  $\varepsilon_{it}$  in a  $K$ -dimensional vector  $\varepsilon_t$ ,

$$\varepsilon_t \sim \mathcal{N}(\mathbf{0}, \Sigma_t), \quad (2.3)$$

where  $\Sigma_t$  is a full  $K \times K$ -dimensional variance covariance matrix.

Another important question is how to parsimoniously model static interdependencies between countries. In this paper, we address this important issue by introducing a factor stochastic volatility structure ([Pitt and Shephard, 1999](#); [Aguilar and West, 2000](#)) on  $\Sigma_t$ ,

$$\Sigma_t = \mathbf{L} \mathbf{H}_t \mathbf{L}' + \Omega_t. \quad (2.4)$$

$\mathbf{L}$  is a  $K \times q$  matrix of factor loadings (with  $q \ll K$ ),  $\mathbf{H}_t = \text{diag}(e^{h_{1t}}, \dots, e^{h_{qt}})$  is a diagonal matrix of the variances of a set of  $q$  common factors  $\mathbf{f}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{H}_t)$ , and  $\Omega_t = \text{diag}(e^{\omega_{1t}}, \dots, e^{\omega_{Kt}})$  is a diagonal variance-covariance matrix of idiosyncratic shocks  $\boldsymbol{\eta}_t \sim \mathcal{N}(\mathbf{0}, \Omega_t)$ .

An equivalent representation of [Eq. \(2.4\)](#) is the regression form,

$$\varepsilon_t = \mathbf{L} \mathbf{f}_t + \boldsymbol{\eta}_t.$$

Hereby, the key feature from an computational point of view is that, conditional on  $\mathbf{L} \mathbf{f}_t$ , the PVAR is simply a system of unrelated regression models. This leads to substantial computational gains relative to full system estimation (see [Kastner and Huber, 2017](#), for more details and an efficient algorithm)

We assume that the (log) of the main diagonal elements of  $\mathbf{H}_t$  and  $\mathbf{\Omega}_t$  follow independent AR(1) processes,

$$h_{jt} = \phi_{hj} + \rho_{hj}(h_{jt-1} - \phi_{hj}) + \sigma_{hj}\zeta_{hj,t}, \text{ for } i = 1, \dots, q, \quad (2.5)$$

$$\omega_{jt} = \phi_{\omega j} + \rho_{\omega j}(\omega_{jt-1} - \phi_{\omega j}) + \sigma_{\omega j}\zeta_{\omega j,t}, \text{ for } j = 1, \dots, K. \quad (2.6)$$

We let  $\phi_{sj}$  for  $s \in \{h, \omega\}$  denote the unconditional mean of the log-volatility,  $\rho_{sj}$  the autoregressive parameter and  $\sigma_{sj}^2$  the process innovation variance. Moreover,  $\zeta_{sj,t} \sim \mathcal{N}(0, 1)$  is a serially uncorrelated white noise shock.

Notice that opposed to  $\frac{K(K+1)}{2}$  total parameters in the case of an unrestricted  $\Sigma_t$ , the structure in Eq. (2.4) implies that we only have to estimate  $(K+1)q + K$  parameters, a massive reduction in the number of parameters relative to an unrestricted variance-covariance matrix, if  $q$  is small. One important implication of Eq. (2.4) is that the covariance structure of the errors is driven by relatively few latent factors that summarize the joint dynamics of  $\varepsilon_t$ .

## 2.2 A sparse finite mixture model specification

It is worth noting that the total number of parameters of the PVAR model outlined in the previous section is  $K(pK + 1) + (K + 1)q + K$ , and thus rises rapidly with  $M$  and  $N$ . Since the typical macroeconomic dataset includes time series with a few hundred observations, some form of regularization is needed. To cope with this issue, the Bayesian literature suggested various means of achieving parsimony in the panel VAR framework. One strand of the literature uses shrinkage priors on several parts of the parameter space (Koop and Korobilis, 2016; Korobilis, 2016; Koop and Korobilis, 2018). This approach conceptually treats the panel VAR as a large VAR with asymmetric shrinkage on the different coefficients in  $\mathbf{A}_i$ ,  $\mathbf{B}_i$  and the free elements in  $\Sigma_t$ . Another strand (Canova and Ciccarelli, 2004; 2009; Jarociński, 2010) exploits the observation that countries do not differ much in terms of their macroeconomic dynamics, implying that the matrices  $\mathbf{A}_i$  tend to be similar across countries. This literature often pools information across countries by shrinking towards a common mean of  $\mathbf{A}_i$ , but neglects dynamic or static interdependencies.

Here we assume that the domestic coefficients (including the intercept)  $\mathbf{c}_i = \text{vec}\{\mathbf{C}_i\}$  arise from a  $G$ -component mixture of Gaussian distribution. A variant of this model has been proposed in the marketing literature (Allenby et al., 1998; Lenk and DeSarbo, 2000; Frühwirth-Schnatter et al., 2004) and is commonly referred to as the heterogeneity model.

In the present framework, the mixture distribution for  $\mathbf{c}_i$  is given by,

$$p(\mathbf{c}_i|\mathbf{w}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_G, \mathbf{V}) = \sum_{g=1}^G w_g f_{\mathcal{N}}(\mathbf{c}_i|\boldsymbol{\mu}_g, \mathbf{V}). \quad (2.7)$$

$\mathbf{w} = (w_1, \dots, w_G)'$  is a vector of component weights that satisfy  $\sum_{g=1}^G w_g = 1$  and  $w_g \geq 0$ . Additionally,  $f_{\mathcal{N}}$  is the density of the multivariate Gaussian distribution,  $\boldsymbol{\mu}_g$  is a  $m = M(Mp + 1)$ -dimensional component-specific mean vector, and  $\mathbf{V}$  is a common variance-covariance matrix. This specification assumes that coefficients of countries within a given country group tend to be similar, with potential deviations from  $\boldsymbol{\mu}_g$  being driven by  $\mathbf{V}$ .

Introducing a set of  $N$  binary indicators  $\delta_i$  allows stating Eq. (2.7) as follows,

$$p(\mathbf{c}_i|\delta_i = g, \boldsymbol{\mu}_g, \mathbf{V}) = f_{\mathcal{N}}(\mathbf{c}_i|\boldsymbol{\mu}_g, \mathbf{V}) \quad (2.8)$$

with  $Prob(\delta_i = g) = w_g$ . In what follows we are going to exploit this auxiliary representation for estimation of the mixture model. Notice that ergodic averages of the posterior draws of  $\delta_i$  can be used to obtain the probability that country  $i$  is located within a specific country group.

On the main diagonal elements of  $\mathbf{V}$ , we use a set of independent inverted Gamma priors,

$$v_j \sim \mathcal{G}^{-1}(w_0, w_1), \text{ for } j = 1, \dots, m, \quad (2.9)$$

with  $w_0$  and  $w_1$  being hyperparameters typically set to small values, i.e.  $w_0 = w_1 = 0.01$ . This leads to a weakly informative prior on the common variances.

Another key assumption is that each mixture component, again, comes from a common distribution,

$$\boldsymbol{\mu}_g|\boldsymbol{\mu}_0, \mathbf{Q}_0 \sim \mathcal{N}(\boldsymbol{\mu}_0, \mathbf{Q}_0) \text{ for } g = 1, \dots, G. \quad (2.10)$$

We let  $\boldsymbol{\mu}_0$  denote a common mean and  $\mathbf{Q}_0$  is a diagonal variance-covariance matrix that can be decomposed as (Malsiner-Walli et al., 2016)

$$\mathbf{Q}_0 = \boldsymbol{\Lambda} \mathbf{R}_0 \boldsymbol{\Lambda},$$

where the matrix  $\boldsymbol{\Lambda} = \text{diag}(\sqrt{\lambda_1}, \dots, \sqrt{\lambda_m})$  contains the standard deviations and  $\mathbf{R}_0 = \text{diag}(R_1^2, \dots, R_m^2)$  constitutes an additional scaling matrix with  $R_j^2$  denoting the range of  $\mathbf{c} = (\mathbf{c}_1, \dots, \mathbf{c}_N)$  along the  $j$ th dimension.



## Selecting cluster-relevant quantities

To select the driving forces behind the cluster allocation, we follow [Yau and Holmes \(2011\)](#) and consider the standardized distance between cluster centers for a given element  $j$  of  $\boldsymbol{\mu}_i$  for clusters  $g$  and  $s$ ,

$$\frac{(\mu_{gj} - \mu_{sj})}{\sqrt{2R_j^2}} \sim \mathcal{N}(0, \lambda_j) \text{ for } j = 1, \dots, m. \quad (2.11)$$

By specifying a suitable mixing density on  $\lambda_j$ , we can flexibly shrink the distance between cluster centers to zero and thus are able to identify cluster relevant variables. As an example, consider a situation where the conditional mean of output growth strongly differs across countries while the remaining quantities (i.e. the coefficients associated with the lags of  $\mathbf{y}_{it}$ ) display only minor differences. In such a situation, a shrinkage prior would strongly pull the cluster centers together for elements in  $\boldsymbol{\mu}$  not related to the intercept, while at the same time allowing for large differences between the cluster means for the intercept terms.

Following [Malsiner-Walli et al. \(2016\)](#), we introduce a Gamma prior on  $\lambda_j$ , leading to a variant of the NG prior ([Griffin and Brown, 2010](#)). More specifically, we set

$$\lambda_j \sim \mathcal{G}(\nu_1, \nu_2). \quad (2.12)$$

$\nu_1$  and  $\nu_2$  are hyperparameter specified by the researcher. Notice that if  $\nu_1 = 1$ , we obtain the Bayesian Lasso ([Park and Casella, 2008](#)) used in [Yau and Holmes \(2011\)](#). The NG prior improves upon the Lasso by having a marginal prior that features heavier tails than the Laplace distribution in the case of the Lasso. In fact, the marginal prior of the NG prior is available in closed form ([Frühwirth-Schnatter, 2011](#)),

$$p(\mu_{1j}, \dots, \mu_{Gj} | \boldsymbol{\mu}_0) = \frac{\nu_2^{\nu_1}}{(2\pi)^{G/2} \Gamma(\nu_1)} 2K_{p_G}(\sqrt{d_j} e_j) \left(\frac{e_j}{d_j}\right)^{p_G/2}, \quad (2.13)$$

with  $d_j = 2\nu_2$ ,  $p_G = \nu_j - G/2$ ,  $e_j = \sum_{g=1}^G (\mu_{gj} - \mu_{0j})^2 / R_j^2$  and  $\Gamma(\star)$  denotes the Gamma function. In addition,  $K_\alpha(\star)$  denotes the modified Bessel function of the second kind and  $\mu_{0j}$  denotes the  $j$ th element of  $\boldsymbol{\mu}_0$ . [Griffin and Brown \(2010\)](#) show that the excess kurtosis of the NG prior is given by  $3/\nu_1$  and thus rises with smaller values of  $\nu_1$ . If  $\nu_1$  is close to zero, more mass is placed on zero while at the same time maintaining heavy tails of the marginal prior. In the applications, we specify  $\nu_1 = \nu_2 = 1/2$  to strongly push the standardized distance between cluster centers to zero.

The prior on  $\mu_0 \sim \mathcal{N}(\mathbf{m}_0, \mathbf{M}_0)$  is improper with  $\mathbf{m}_0$  denoting the median over the columns of  $\mathbf{c}$  and  $\mathbf{M}_0^{-1} = \mathbf{0}$ . Here, one alternative would be to use a Minnesota prior (Doan et al., 1984) at the top level of the hierarchy, assuming that  $\mu_0$  again features a normally distributed prior centered on a multivariate random walk with a known prior variance-covariance matrix. For several datasets, however, we found that this choice only exerts a minor impact on the actual results.

### Choosing the number of mixture components

To select the number of components  $G$ , we follow Malsiner-Walli et al. (2016) and introduce a symmetric Dirichlet prior on the mixture component weights  $\mathbf{w}$ ,

$$\mathbf{w} \sim \text{Dir}(p_0, \dots, p_0), \quad (2.14)$$

where  $p_0$  denotes the intensity parameter of the Dirichlet distribution. In the framework of overfitting mixture models (i.e. models that set  $G$  greater than the true number of clusters,  $G^{true}$ ), the parameter  $p_0$  plays an important role in shaping the way the posterior distribution treats redundant mixture components.<sup>1</sup>

In what follows, we are going to place yet another Gamma prior on  $p_0$ . We follow Ishwaran et al. (2001) and Malsiner-Walli et al. (2016) and choose a Gamma prior with expectation  $E(p_0) = 1/G$ , implying that

$$p_0 \sim \mathcal{G}(c_0, c_0 G). \quad (2.15)$$

Hereby, we let  $c_0$  be a hyperparameter that controls the variance of the prior  $1/(c_0 G^2)$ . This prior choice will handle irrelevant mixture components by shrinking the associated weights to zero and emptying superfluous components. Consistent with simulation evidence provided in Malsiner-Walli et al. (2016), we set  $c_0 = 10$ .

### 2.3 Priors on dynamic interdependencies

To decide on whether DIs for a given country  $i$  are needed, we use a NG shrinkage prior similar to the one discussed above. While the prior on  $\mu_0$  introduces local shrinkage parameters that push the differences between cluster centers to zero, the standard implemen-

---

<sup>1</sup>For a discussion, see Frühwirth-Schnatter (2006) and Rousseau and Mengersen (2011)

tation of the NG prior combines local shrinkage parameters with a global shrinkage factor that pulls all coefficients concerned to zero.

To illustrate the problem of selecting DIs, we partition the matrices  $\mathbf{B}_{ip}$  for  $p = 1, \dots, P$  and stack them to obtain,

$$\mathbf{B}_p = \begin{pmatrix} \mathbf{B}_{1p} \\ \mathbf{B}_{2p} \\ \vdots \\ \mathbf{B}_{Np} \end{pmatrix} = \begin{pmatrix} \mathbf{B}_{12,p} & \mathbf{B}_{13,p} & \dots & \mathbf{B}_{1N,p} \\ \mathbf{B}_{21,p} & \mathbf{B}_{23,p} & \ddots & \vdots \\ \vdots & \ddots & \vdots & \mathbf{B}_{N-1N,p} \\ \mathbf{B}_{N1,p} & \dots & \mathbf{B}_{NN-2,p} & \mathbf{B}_{NN-1,p} \end{pmatrix}, \quad (2.16)$$

where the submatrix  $\mathbf{B}_{ij,p}$  measures the DIs between countries  $i$  and  $j$  for lag  $p$ . Model specification boils down to deciding whether a given  $\mathbf{B}_{ij,p}$  equals zero, ruling out DIs between countries  $i$  and  $j$ . [Koop and Korobilis \(2016\)](#) use a stochastic search variable selection (SSVS) prior that is based on a set of auxiliary indicators that determine whether different sub-matrices of  $\mathbf{B}_p$  are pushed to zero. While this approach is conceptually straightforward to implement, it implies that a high dimensional model space needs to be explored. Using MCMC techniques helps to circumvent this issue by using a stochastic model specification search that only explores a fraction of the full model space. However, in large dimensions (i.e. if the number of countries is large), the possible number of DI restrictions is huge, even for a moderate number of countries included. In that case, even SSVS priors manage to exploit only a tiny fraction of the model space, leading to weak convergence ([Bhattacharya et al., 2015](#)).

In this paper, we assume that each element of  $\text{vec}(\mathbf{B}_i)$ , labeled  $b_{ij}$ , features a normally distributed prior,

$$b_{ij} | \tau_{ij}, \xi_i \sim \mathcal{N}\left(0, \frac{2\tau_{ij}^2}{\xi_i}\right), \quad \tau_{ij}^2 \sim \mathcal{G}(\vartheta_i, \vartheta_i), \quad \xi_i \sim \mathcal{G}(\mathbf{c}_0, \mathbf{c}_1), \quad (2.17)$$

for  $j = 1, \dots, k = PM^2(N - 1)$  and  $i = 1, \dots, N$ .  $\xi_i$  denotes a country-specific global scaling parameter that pushes all elements in  $\mathbf{B}_i$  (or equivalently  $\mathbf{B}_{ip}$  for all  $p$ ) to zero, effectively shutting off DIs between a given country and all remaining countries, if necessary. Overall shrinkage is then driven by the hyperparameters  $\mathbf{c}_0, \mathbf{c}_1$ , with small values for both translating into heavy overall shrinkage.

Since shutting of all DIs within a given country is overly restrictive, we introduce a set of local scaling parameters  $\tau_{ij}^2$ . The local scaling parameters allow for non-zero  $b_{ij}$ 's, even in the presence of strong global shrinkage due to a heavy tailed marginal prior (see

Eq. (2.13)), with excess kurtosis depending on  $\vartheta_i$ . This enables flexible selection of restrictions of the form whether country  $i$ 's output depends on country  $c$ 'th lagged output while turning off dependencies between output in country  $i$  and, say, lagged interest rates in country  $c$ . We set  $\epsilon_0 = \epsilon_1 = 0.01$  and  $\vartheta_i = 0.1$ . Both hyperparameter values are based on evidence in [Huber and Feldkircher \(2017\)](#), who integrate out  $\vartheta_i$  in a Bayesian fashion and find values between 0.1 to 0.3, depending on the size of the model involved.

## 2.4 Priors on the remaining coefficients in the model

For the remaining coefficients we utilize standard priors. On the elements of  $\mathbf{L}$ , we use a standard normally distributed prior  $l_{ij} \sim \mathcal{N}(0, 1)$  for  $i = 1, \dots, K; j = 1, \dots, q$ . On the parameters of the state equations for the log-volatility processes, we follow [Kastner and Frühwirth-Schnatter \(2014\)](#) and use a normally distributed prior on the unconditional mean  $\mu_{sj} \sim \mathcal{N}(0, 10^2)$  for all  $s, j$ , a Gamma prior on the process innovation variances  $\sigma_{sj}^2 \sim \mathcal{G}(1/2, 1/2)$ , and a Beta prior on the (transformed) autoregressive parameter  $\frac{\rho_{sj}+1}{2} \sim \mathcal{B}(25, 5)$ .

## 2.5 Identification issues

The model described above is econometrically not identified. Identification issues stem from two sources. First, the factor model in [Eq. \(2.4\)](#) is not identified unless suitable restrictions are introduced. Here, we take a simple stance and assume that the upper  $q \times q$  block of  $\mathbf{L}$  is a lower uni-triangular (i.e. with unit diagonal) matrix ([Aguilar and West, 2000](#)). This assures that both, the sign and the scale of the latent factors  $\mathbf{f}_t$  are identified.

The second source stems from the well known label switching problem.<sup>2</sup> This issue comes from the invariance of the mixture likelihood function in [Eq. \(2.7\)](#) with respect to relabeling the components,

$$p(\mathbf{c}_i | \mathbf{w}, \boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_G, \mathbf{V}) = \sum_{g=1}^G w_g f_{\mathcal{N}}(\mathbf{c}_i | \boldsymbol{\mu}_g, \mathbf{V}) \quad (2.18)$$

$$= \sum_{g=1}^G w_{\varrho(g)} f_{\mathcal{N}}(\mathbf{c}_i | \boldsymbol{\mu}_{\varrho(g)}, \mathbf{V}), \quad (2.19)$$

---

<sup>2</sup>For a discussion, see [Frühwirth-Schnatter \(2006\)](#).

with  $\varrho$  being a random permutation of  $\{1, \dots, G\}$ . We obtain identification by applying the random permutation sampler outlined in [Frühwirth-Schnatter \(2001\)](#) and then performing ex-post identification of the model. In our case, and since  $N$  is typically a moderate number of countries, we can easily identify different country groups by using economic reasoning. In the empirical application, for instance, we introduce an ordering constraint on the conditional mean in the inflation equation. Notice that if interest centers exclusively on functionals of the coefficients in [Eq. \(2.1\)](#), like impulse response functions or predictive densities, obtaining explicit identification is not necessary. However, it is worth emphasizing that if unbalanced label switching takes place (i.e. the posterior simulator jumps only between a small number of the  $G!$  potential modes), inference could be distorted. Using the random permutation sampler in that situation thus leads to balanced label switching, ensuring that the algorithm visits all modes.

## 2.6 Posterior simulation

Our Markov chain Monte Carlo (MCMC) algorithm consists of several blocks. In what follows we briefly summarize the full algorithm proposed with all full conditional posterior distributions involved. Conditional on a set of starting values, the algorithm cycles through the following steps.

### 1. Simulation of VAR coefficients, factor loadings and stochastic volatility components

- (a) Sample  $\mathbf{A}_i$  and  $\mathbf{B}_i$  from their Gaussian conditional posterior distributions on an equation-by-equation basis. Conditional on  $\mathbf{L}\mathbf{f}_t$ , the conditional posterior for each equation of [Eq. \(2.1\)](#) is given by

$$\begin{pmatrix} [\mathbf{C}_i]_{j\bullet}' \\ [\mathbf{B}_i]_{j\bullet}' \end{pmatrix} | \bullet \sim \mathcal{N}(\bar{\mathbf{c}}_{ij}, \bar{\mathbf{M}}_{ij}) \quad (2.20)$$

for  $i = 1, \dots, N$  and  $j = 1, \dots, M$ . The posterior mean and variance are given by

$$\bar{\mathbf{M}}_{ij} = (\tilde{\mathbf{X}}_i' \tilde{\mathbf{X}}_i + \mathbf{W}_i^{-1})^{-1}, \quad (2.21)$$

$$\bar{\mathbf{c}}_{ij} = \bar{\mathbf{M}}_{ij} (\tilde{\mathbf{X}}_i' [\tilde{\mathbf{Y}}_i]_{j\bullet} + \mathbf{W}_i^{-1} \boldsymbol{\psi}_{ij}), \quad (2.22)$$

with  $\tilde{\mathbf{X}}_i$  being a full-data matrix with typical  $t$ th row given by  $(\mathbf{x}'_{it}, \mathbf{x}'_{-i,t}) \exp(-\omega_{tn}/2)$ . The index  $n$  selects the element of  $\Omega_t$  associated with the  $j$ th equation in country

$i$  and  $[\tilde{\mathbf{Y}}_i]_{\bullet j}$  has typical element  $y_{ij,t} - [\mathbf{L}]_{n\bullet} \mathbf{f}_t$ . In addition,  $\mathbf{W}_i = \text{diag}(\mathbf{V}_j, \mathbf{\Delta}_{ij})$  with  $\mathbf{\Delta}_{ij}$  being a diagonal prior variance-covariance matrix for the  $j$ th equation constructed using Eq. (2.17), and  $\psi_{ij}$  is a prior mean matrix that consists of the elements in  $\boldsymbol{\mu}_g$  associated with the  $j$ th equation, for  $\delta_i = g$ , and the remaining elements are set equal to zero. The matrix  $\mathbf{V}_j$  is constructed by selecting the variance parameters in  $\mathbf{V}$  that relate to the  $j$ th equation.

- (b) We simulate the free elements in  $\mathbf{L}$  on an equation-by-equation basis by running a set of  $K$  regression models with heteroscedastic shocks. The conditional posterior of  $[\mathbf{L}]_{j\bullet}$  is Gaussian,

$$[\mathbf{L}]_{j\bullet} | \bullet \sim \mathcal{N}(\bar{\mathbf{l}}_j, \bar{\mathbf{S}}_j), \quad (2.23)$$

where

$$\bar{\mathbf{S}}_j = (\tilde{\mathbf{F}}' \tilde{\mathbf{F}} + \mathbf{I}_q)^{-1}, \quad (2.24)$$

$$\bar{\mathbf{l}}_j = \bar{\mathbf{S}}_j \tilde{\mathbf{F}}' \tilde{\boldsymbol{\varepsilon}}_{j\bullet}. \quad (2.25)$$

Here,  $[\tilde{\mathbf{F}}]_{t\bullet} = \mathbf{f}'_t \exp(-\omega_{jt}/2)$  and  $\tilde{\boldsymbol{\varepsilon}}_{j\bullet}$  is the  $j$ th column of the rescaled structural shock vector with typical element given by  $\varepsilon_{jt} \exp(-\omega_{jt}/2)$ . Note that for equations  $j \leq q$ , the quantities need to be adjusted to be consistent with the identifying assumptions described in Section 2.5.

- (c) Simulate the full history of latent factors  $\{\mathbf{f}_t\}_{t=1}^T$  independently for each  $t$  from a Gaussian distribution,

$$\mathbf{f}_t | \bullet \sim \mathcal{N}(\bar{\mathbf{f}}_t, \mathbf{P}_t) \quad (2.26)$$

with

$$\mathbf{P}_t = \mathbf{H}_t - \mathbf{W}_t \boldsymbol{\Sigma}_t \mathbf{W}_t' \quad (2.27)$$

$$\bar{\mathbf{f}}_t = \mathbf{W}_t \boldsymbol{\varepsilon}_t, \quad (2.28)$$

and  $\mathbf{W}_t = (\mathbf{H}_t \mathbf{L}' \boldsymbol{\Sigma}_t^{-1})$ .

- (d) Simulate the full history of log-volatilities  $\{h_{st}\}_{t=1}^T$  ( $s = 1, \dots, q$ ) and  $\{\omega_{jt}\}_{t=1}^T$  ( $j = 1, \dots, K$ ) and the parameters of the state equation using the algorithm outlined in [Kastner and Frühwirth-Schnatter \(2014\)](#); [Kastner \(2016\)](#).

## 2. Simulation of quantities associated with the mixture model

(a) Sample the mixture probabilities  $\boldsymbol{w}$  from a Dirichlet distribution given by

$$\boldsymbol{w}|\bullet \sim \text{Dir}(p_1, \dots, p_G), \quad (2.29)$$

with  $p_g = p_0 + N_g$  and  $N_g = \#\{i : \delta_i = g\}$  denoting the number of countries located within cluster  $g$ .

(b) The regime indicators  $\delta_i$  are simulated from a multinomial distribution with

$$\text{Prob}(\delta_i = k) \propto w_k f_{\mathcal{N}}(\boldsymbol{c}_i | \boldsymbol{\mu}_g, \mathbf{V}). \quad (2.30)$$

(c) We obtain draws for the group-specific means from a multivariate Gaussian distribution,

$$\boldsymbol{\mu}_g|\bullet \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_g, \bar{\mathbf{V}}_g), \quad (2.31)$$

with

$$\bar{\mathbf{V}}_g = (N_g \mathbf{V}^{-1} + \mathbf{Q}_0^{-1})^{-1}, \quad (2.32)$$

$$\bar{\boldsymbol{\mu}}_g = \bar{\mathbf{V}}_g (N_g \mathbf{V}^{-1} \bar{\boldsymbol{c}}_g + \mathbf{Q}_0^{-1} \boldsymbol{\mu}_0). \quad (2.33)$$

$\bar{\boldsymbol{c}}_g = \frac{\sum_{i=1}^N \boldsymbol{c}_i \delta_i}{N_g}$  denotes the mean of the domestic quantities associated with group  $g$ .

(d) The common variance-covariance matrix  $\mathbf{V}$  is obtained by independently sampling  $v_j$  ( $j = 1, \dots, m$ ) from

$$v_j|\bullet \sim \mathcal{G}^{-1} \left( w_0 + \frac{N}{2}, w_1 + \frac{\sum_{n=1}^N (c_{nj} - \mu_{nj})^2}{2} \right), \quad (2.34)$$

where  $\mu_{nj} = \mu_{gj}$  if  $\delta_n = g$ .

(e) We simulate  $\lambda_j$  from a generalized inverted Gaussian (GIG) distribution,<sup>3</sup>

$$\lambda_j|\bullet \sim \text{GIG}(p_G, d_j, e_j). \quad (2.35)$$

After simulating all  $\lambda_j$ s we construct  $\mathbf{Q}_0 = \boldsymbol{\Lambda} \mathbf{R}_0 \boldsymbol{\Lambda}$ , with  $\mathbf{R}_0$  being based on the most recent Gibbs draw of  $\boldsymbol{c}$ .

---

<sup>3</sup>We assume that  $x$  follows a GIG distribution if its density is proportional to  $x^{a-1} \exp\{-(bx + c/x)/2\}$  with  $a \in \mathbb{R}$  and  $b, c > 0$ .

(f) The full conditional posterior of  $\boldsymbol{\mu}_0$  is Gaussian with

$$\boldsymbol{\mu}_0|\bullet \sim \mathcal{N}(\bar{\boldsymbol{\mu}}_0, \bar{\mathbf{Q}}_0), \quad (2.36)$$

whereby  $\bar{\boldsymbol{\mu}}_0 = \frac{1}{G} \sum_{g=1}^G \boldsymbol{\mu}_g$  and  $\bar{\mathbf{Q}}_0 = \frac{1}{G} \mathbf{Q}_0$ .

(g) Simulate the intensity parameter of the Dirichlet prior  $p_0$  using a random walk Metropolis Hastings algorithm on the log scale. The full conditional posterior density of  $p_0$  is given by

$$p(p_0|\mathbf{w}) \propto p(\mathbf{w}|p_0) p(p_0). \quad (2.37)$$

We propose a value  $p_0^*$  from  $p_0^* \sim p_0^{(a)} e^z$  with  $z \sim \mathcal{N}(0, \mathfrak{c})$ . Here we let  $\mathfrak{c}$  be a tuning parameter specified such that the acceptance rate lies between 20 and 40 percent and  $p_0^{(a)}$  denotes the last accepted draw. The probability of accepting a new draw is then

$$\alpha(p_0^*, p_0^{(a)}) = \min \left[ \frac{p(\mathbf{w}|p_0^*) p(p_0^*) p_0^*}{p(\mathbf{w}|p_0^{(a)}) p(p_0^{(a)}) p_0^{(a)}}, 1 \right]. \quad (2.38)$$

### 3. Simulation of shrinkage parameters on dynamic interdependencies

(a) For each country  $i = 1, \dots, N$ , simulate the global shrinkage parameters  $\xi_i$  from a Gamma distribution,

$$\xi_i|\bullet \sim \mathcal{G} \left( \mathfrak{c}_0 + \vartheta_i k, \mathfrak{c}_0 + \frac{\vartheta_i}{2} \sum_{i=1}^k \tau_{ij} \right). \quad (2.39)$$

(b) Sample the local shrinkage parameters from their GIG distributed posteriors

$$\tau_{ij}|\bullet \sim \mathcal{GIG} \left( \vartheta_i - \frac{1}{2}, \vartheta_i \xi_i, b_{ij}^2 \right), \quad (2.40)$$

for  $i = 1, \dots, N$  and  $j = 1, \dots, k$ .



4. Apply a random permutation step by simulating one of  $G!$  possible permutations of  $\{1, \dots, G\}$ , labeled  $\varrho$ ,

$$\begin{aligned}(w_1, \dots, w_G)' &= (w_{\varrho(1)}, \dots, w_{\varrho(G)}), \\ (\boldsymbol{\mu}_1, \dots, \boldsymbol{\mu}_G)' &= (\boldsymbol{\mu}_{\varrho(1)}, \dots, \boldsymbol{\mu}_{\varrho(G)}), \\ \boldsymbol{\delta} &= \varrho(\boldsymbol{\delta}).\end{aligned}$$

Step (4) ensures that the algorithm visits all  $G!$  symmetric modes of the posterior distribution. In general, we repeat this algorithm 30,000 times and discard the first 15,000 as burn-ins. In terms of computational intensity, Step (1) proves to be challenging, especially in the presence of a large number of countries and endogenous variables. Considering mixing and convergence properties suggests that our algorithm performs quite well, with inefficiency factors below 30 for most parameters.

### 3 Simulation results

In this section we evaluate the merits of our approach by means of an extensive simulation exercise. To this end, we consider a range of alternative DGPs and scenarios which differ in terms of the implied sparsity on the dynamic interdependencies in  $\mathbf{B}_i$  as well as the length of the time period.

The DGP we consider assumes that  $N = 26$ ,  $M = 2$ ,  $P = 1$  and  $T \in \{80, 150, 250\}$ . The  $M$  variables per country are labeled  $UN$  and  $DP$ . Here, and in the empirical application, these acronyms refer to unemployment and inflation, respectively. Moreover, we assume that the domestic coefficients (including the intercept) come from a two component mixture of Gaussians (i.e.  $G^{true} = 2$ ) with mean vectors given by  $\boldsymbol{\mu}_1^{true} = (0.6, 0.2, 2.0, 0.3, 0.6, -3.0)'$ ,  $\boldsymbol{\mu}_2^{true} = (-0.6, 0.2, 5, -0.8, 0.6, 0.0)'$ , variance-covariance matrix  $\mathbf{V}^{true} = \frac{1}{10^3} \times \mathbf{I}$  and  $\mathbf{w}^{true} = (0.4, 0.6)'$ . Notice that across clusters, the coefficients associated with lagged DP in the UN equations as well as the first, own lag of DP are equal across clusters. This serves as a simple test whether our NG shrinkage prior successfully shrinks the corresponding differences in cluster centers to zero.

Coefficients measuring lagged interdependencies are constructed by drawing from univariate Gaussian distributions  $b_{ij}^{true} \sim \mathcal{N}(0, \frac{1}{10^2})$ . To control the degree of actual sparsity, we zero out all coefficients that are below a fraction, denoted by  $\varpi \in \{0.15, 0.30, 0.60, 0.90\}$ , of the maximum absolute value across all  $b_{ij}$ s. Finally, for the factor model in Eq. (2.4), we

Sparsity	$T = 80$				$T = 150$				$T = 250$			
	0.15	0.30	0.60	0.90	0.15	0.30	0.60	0.90	0.15	0.30	0.60	0.90
PVAR $G = 8$	0.026	0.026	0.025	0.024	0.021	0.021	0.019	0.019	0.017	0.016	0.015	0.014
PVAR $G = 1$	0.031	0.031	0.030	0.030	0.025	0.025	0.023	0.023	0.019	0.019	0.018	0.017
VAR-NG	0.042	0.041	0.041	0.041	0.031	0.031	0.030	0.030	0.022	0.022	0.021	0.020
VAR-OLS	0.179	0.177	0.179	0.178	0.102	0.100	0.101	0.101	0.061	0.061	0.060	0.060

**Table 1:** Root mean square errors of posterior median of the VAR coefficients and the true values. Median across 50 replications.

use  $q = 2$ , sample the loadings from zero mean Gaussian distributions with variances given by  $1/(1000^2)$  and specify  $\phi_{hj} = 0$ ,  $\rho_{hj} = 0.9$  and  $\sigma_{hj} = 0.1$  for all  $j$ . For the measurement errors we set  $\phi_{\omega j} = -10$ ,  $\rho_{\omega j} = 0.9$  and  $\sigma_{\omega j} = 0.01$ . For each of the different DGPs, we run a set of 50 simulations and focus on root mean square errors (RMSEs).

Before showing the actual results, a brief word on the specification and identification of the underlying mixture model is necessary. We assume that  $G = 10$ , implying that  $G \gg G^{true}$ , the number of factors equals the true number  $q = 2$ , and set the lag length to  $P = 1$ . Using higher lag orders would strongly favor models with shrinkage priors relative to simple OLS.

In this simulation exercise, we identify the model by applying the permutation sampler outlined in the previous section and identify the mixture model by assuming that  $w_1 < \dots < w_G$ . The hyperparameters are specified as described in Section 2.

As competing alternatives we include a model estimated with a single regime ( $Q = 1$ ), a flat prior VAR (labeled VAR-OLS), and a VAR with a NG shrinkage prior (VAR-NG) that treats the PVAR as a large VAR without discriminating, a priori, between domestic and foreign variables. All models, except the flat prior VAR, feature a factor stochastic volatility specification in the reduced form errors. For the VAR-NG model, we use the lag-wise prior specification described in [Huber and Feldkircher \(2017\)](#) with the choice of the hyperparameters closely mirroring the ones used here on the DIs.

Table 1 shows the median across replications of the RMSEs for the different models. A few findings are worth emphasizing. First, notice that irrespective of the DGP adopted, our proposed overfitting mixture model performs particularly well, outperforming all competing models. The misspecified model with  $G = 1$  also tends to perform quite well. Unsurprisingly, the VAR estimated by OLS performs worst. Second, considering different lengths of the sample reveals that if more information is available, the accuracy gap between the unrestricted VARs (i.e. estimated using OLS and under the lag-wise NG prior) and the two PVAR specifications tends to decline. Nevertheless, OLS still yields estimation errors that

are sizable and about four times as large as the estimation errors obtained by using the PVAR with  $G = 8$  (for  $T = 250$ ). From a practical perspective, and by considering the unrestricted VAR with the NG prior, this simply implies that for large  $T$ , simpler specifications also tend to work well.

Third, and finally, we observe a better performance of our proposed framework for higher levels of sparsity, especially when benchmarked against unrestricted specifications. This finding indicates that separately controlling for dynamic interdependencies evidently improves model performance as compared to a prior that does not discriminate between coefficients associated with foreign quantities and their domestic counterparts.

To assess whether our mixture model successfully detects the correct number of regimes as well as the correct regime allocation, Table 2 shows the (half) quadratic probability score (QPS) defined as,

$$\text{QPS} = \frac{1}{N} \sum_{i=1}^N (\delta_i^{\text{true}} - \bar{\delta}_i)^2, \quad (3.1)$$

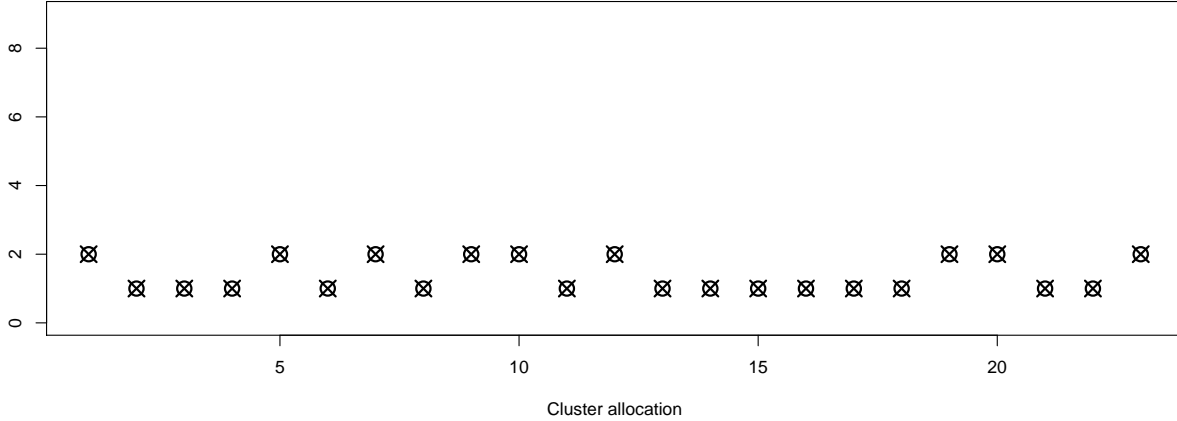
with  $\delta_i^{\text{true}}$  denoting the true cluster allocation for country  $i$  and  $\bar{\delta}_i$  is the posterior mean of  $\delta_i$ . All numbers in the table refer to the median across simulation replications. The QPS score is bounded between zero and unity, with a value of zero indicating perfect accuracy.

The results in Table 2 suggest that irrespective of the DGP, our model appears to work quite well, yielding QPS scores close to zero. Interestingly, differences in QPS scores tend to be quite unsystematic, providing only limited evidence that accuracy improves if the length of the sample is increased (see the final two rows of the table). This finding can be traced back to the fact that our DGP induces quite large differences between  $\mu_1^{\text{true}}$  and  $\mu_2^{\text{true}}$ , implying that the conditional likelihood carries sufficient information (see Eq. (2.7)). In addition, notice that we leave the number of countries  $N = 26$  the same across DGPs, implying that the number of observations from which we infer the correct clustering stays the same. The robustness of the QPS scores with respect to the sparsity level implies that the actual level of sparsity on DIs does not exert a significant feedback effect on the actual cluster allocation.

After providing some evidence that our model performs well, we proceed by showing selected empirical features of the proposed framework for a single simulation run. In Fig. 1, we show the estimated regime allocation across countries alongside the true regime allocation. Black circles represent the posterior mean of the regime indicators while black crosses mark the actual country allocation. Consistent with the findings reported in Table 2,

Sparsity	$T = 80$	$T = 150$	$T = 250$
0.15	0.155	0.124	0.156
0.3	0.124	0.156	0.155
0.6	0.156	0.123	0.126
0.9	0.156	0.154	0.140

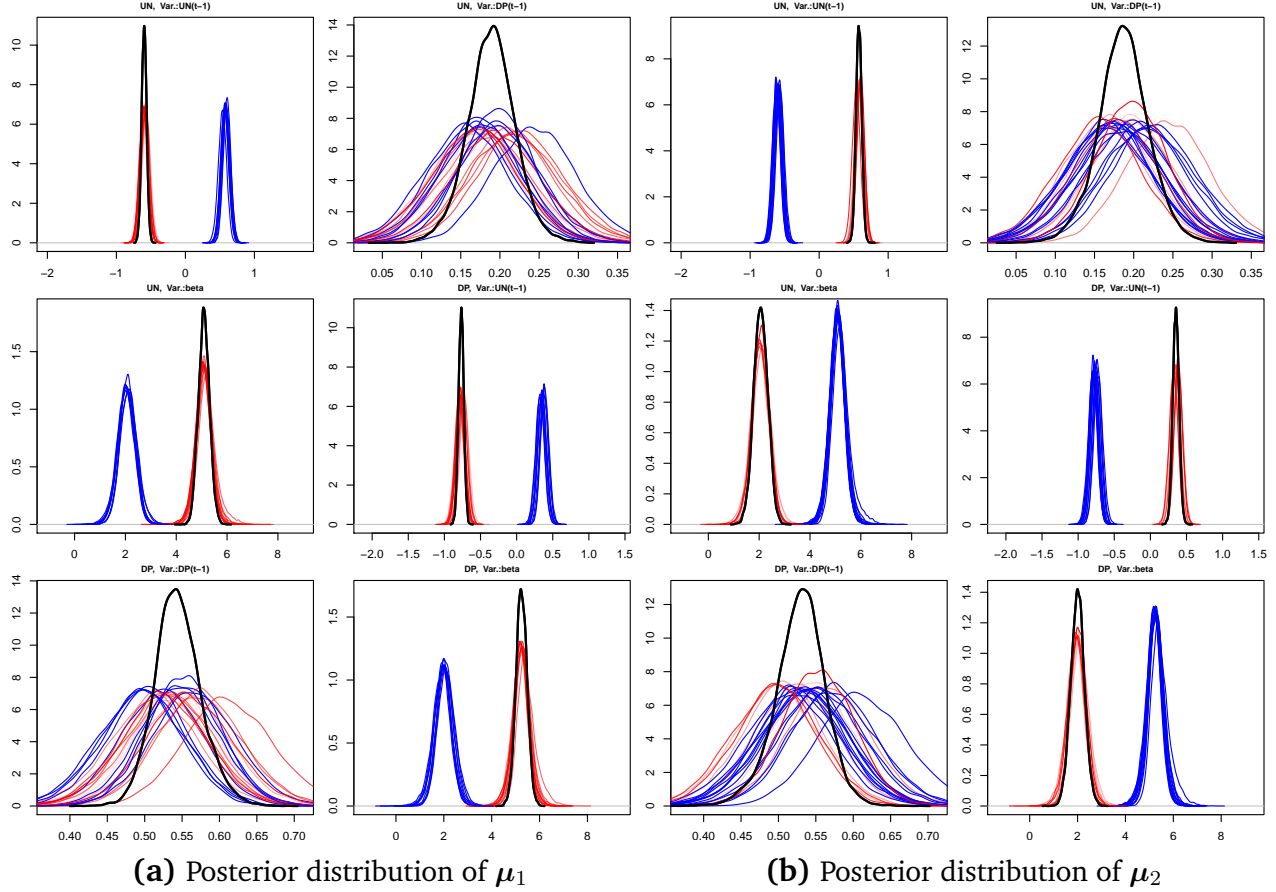
**Table 2:** Quadratic probability score across simulation runs.



**Fig. 1:** Estimated regime allocation across countries (black circles) and true allocation (black crosses)

the figure suggests that our model performs well in detecting country clusters and selecting the true number of regimes.

Figure 1 shows the posterior distribution of  $c_i$  for all  $i$  as well as the posterior density of the cluster centers  $\mu_j$  for  $j = 1, 2$ . For the country-specific coefficients we use red colored densities to indicate group membership of a country within a group, blue densities point towards membership in other groups and the black density is the corresponding element of  $\mu_g$ . In the figure titles, we first state the equation (i.e. the equation for UN or DP) and afterwards the variable within each equation, with beta referring to the intercept. For each coefficient, the country-specific posterior distributions are closely centered on the posterior distribution of the group-specific coefficient. This holds true for both clusters that are not emptied out in MCMC sampling. Considering the coefficients that are homogenous across both clusters reveals that our model approach successfully detects homogeneity, pushing the country-specific estimates on lagged inflation towards the common mean. Comparing

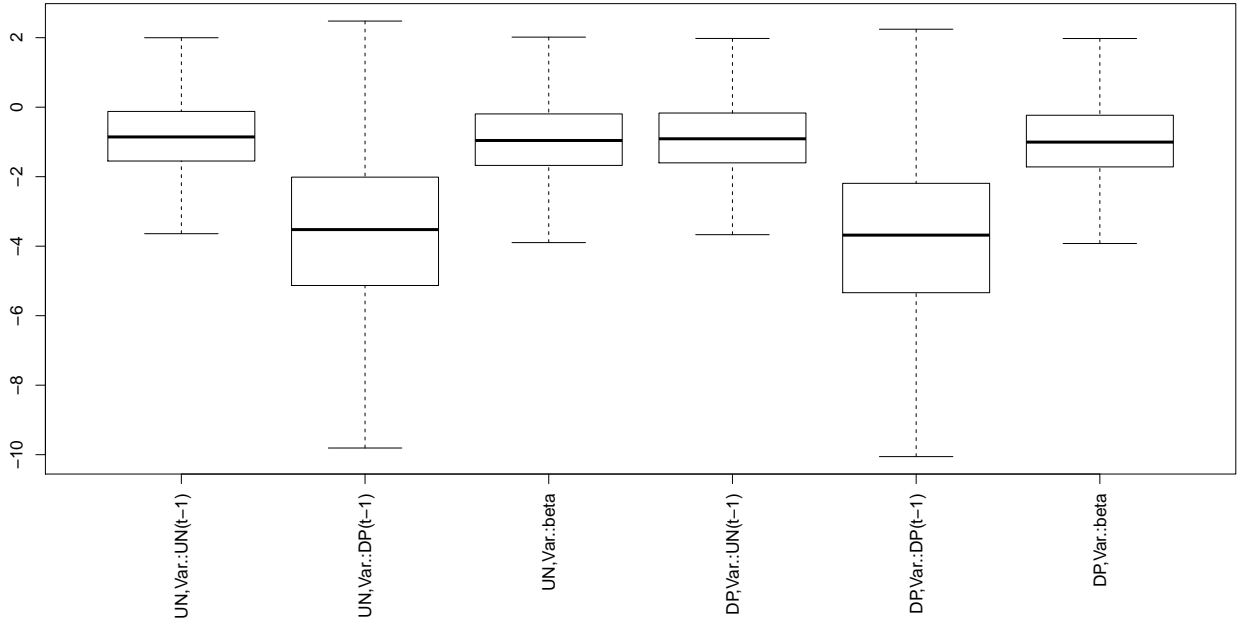


**Fig. 2:** Posterior distribution of  $c_i$  alongside the group specific mean  $\mu_i$  (in solid black) for simulated data. Red lines indicate that a given country is member of the group considered while blue lines indicate membership to other groups.

the posterior densities across regimes for these coefficients also shows that our shrinkage prior successfully pulls the standardized distance between cluster centers to zero. Following [Yau and Holmes \(2011\)](#), this finding is corroborated by considering boxplots of the log posterior distribution of  $\lambda_j$ , depicted in [Fig. 3](#). Here, we observe that for the coefficients specified to be equal across clusters, the corresponding (log) shrinkage factor is much smaller.

#### 4 Empirical application: Modeling Eurozone unemployment

In the empirical application, the modeling approach is applied to a set of eleven Eurozone members. In the next subsection we briefly describe the data and model specification while we show key features of the model comparable to the ones shown in the previous section.

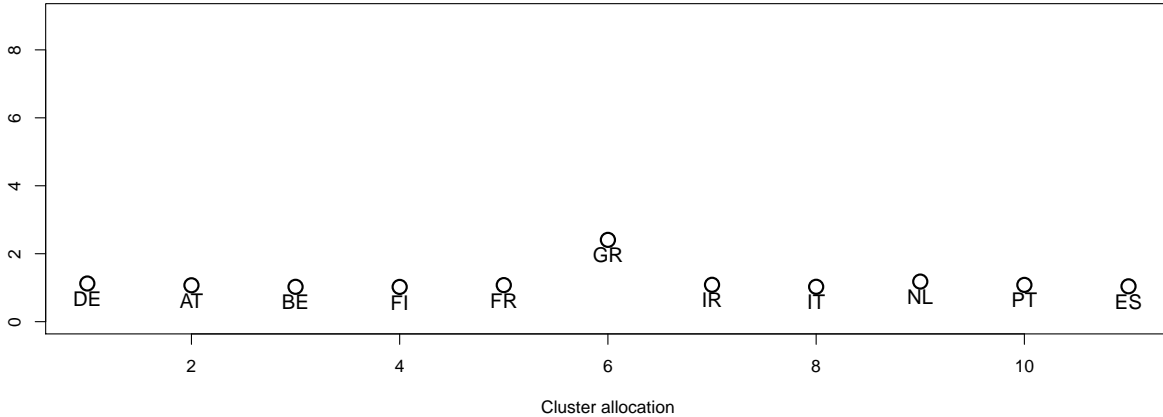


**Fig. 3:** Posterior distribution of  $\log(\lambda_j)$  for a synthetic dataset

We then proceed by performing a forecasting exercise with the goal to assess the merits of our approach relative to a set of nested alternatives.

#### 4.1 Data overview and model specification

The dataset adopted here is a variant of the data used in [Koop and Korobilis \(2018\)](#) and runs from 1999:M01 to 2014:M12. We include data on eleven Eurozone countries, namely Germany (DE), Austria (AT), Belgium (BE), Finland (FI), France (FR), Greece (GR), Ireland (IE), Italy (IT), Netherlands (NL), Portugal (PT), and Spain (ES). For each country, we include  $M = 4$  macroeconomic quantities. These comprise data on inflation, unemployment, and two measures of survey expectations. The first one captures expectations on the financial situation (henceforth labeled FS) and the second measures expectations on the general economic situation (labeled GE), both in terms of one year ahead developments. All data are obtained from Eurostat, seasonally adjusted and transformed to be (approximately) stationary. We set  $G = 8$  which translates into a generous choice of the number of clusters and specify the lag length equal to one  $P = 1$ . Moreover, preliminary factor analysis on the growth rates of the data set indicates that a single factor  $q = 1$  represents the dataset sufficiently well. The prior setup used is the same as described in Section 2.



**Fig. 4:** Posterior mean of estimated regime allocation across countries for Eurozone data

## 4.2 Key model features

In this section we start by investigating the country allocation to the different clusters and the estimated number of clusters. Figure 4 displays the estimated regime allocation across countries. The estimated cluster allocation clearly indicates that within our sample of Eurozone countries, Greece appears to display the largest degree of heterogeneity in terms of domestic dynamics induced by  $c_i$ . When interpreted in terms of inclusion probabilities, Greece is included with around 98 percent in the second group whereas the remaining economies display similar inclusion probabilities in the first group.

Next, we turn to the posterior distribution of the estimated number of regimes, which we compute as follows (Malsiner-Walli et al., 2016),

$$G^* = G - \sum_{g=1}^G I(N_g = 0), \quad (4.1)$$

where  $I(\star)$  denotes the indicator function. The posterior probability of a given number of clusters  $Pr(G^* = g|\bullet)$ , for  $g = 1, \dots, G$ , is then computed by considering the relative frequencies across all MCMC draws.

Table 3 shows the estimated regime allocation. The numbers can be interpreted as the posterior probability of a given number of clusters. The two component mixture specification receives most posterior weight, being selected in around 74 percent of MCMC runs.

A three component mixture also receives considerable posterior support, providing at least some evidence that more than two clusters exist when considering Eurozone countries.

$G =$	1	2	3	4	5	6	7	8
$Pr(G^* = g \bullet)$	0.05	0.74	0.20	0.01	0.00	0.00	0.00	0.00

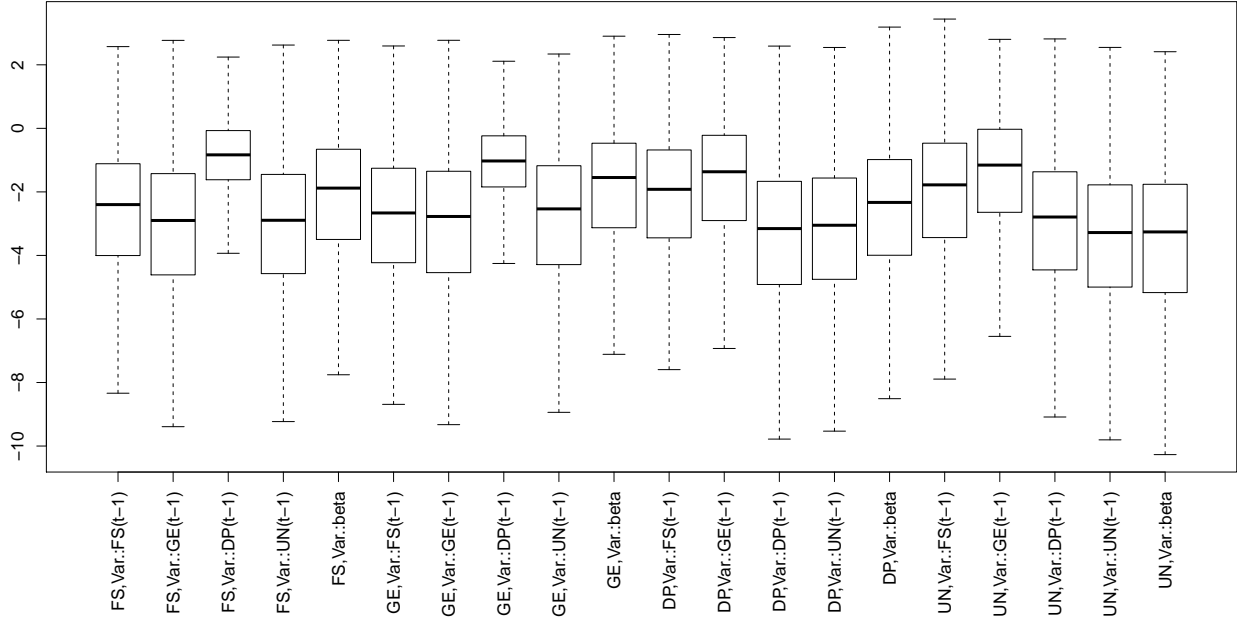
**Table 3:** Posterior probability for a given number of clusters

Similarly to [Fig. 3](#), we use the log of the posterior distribution of  $\lambda_j$  to investigate what variables within a given equation drive the clustering behavior of the model. [Figure 5](#) again presents boxplots of  $\log(\lambda_j)$  across equations and variables. From the figure we see that the shrinkage prior pushes cluster centers to zero for most equations and variables under scrutiny. However, a few variables across equations tend to determine the cluster allocation. Especially the first lag of inflation in the equation for survey information on the expected financial situation (FS) and the expected general economic situation (GE) are less heavily shrunk towards the common mean. Some other quantities also tend to display a limited impact on the clustering behavior of the model. For instance, the coefficients associated with the first lag of GE in both, the inflation and the unemployment equations, feature a somewhat larger  $\lambda_j$ . [Figure 5](#) thus clearly suggests that cluster determination is mainly driven by relatively few variables across equations, also implying that a small number of components might be sufficient for the dataset at hand.

### 4.3 Forecasting results

To assess whether our approach also excels in terms of predictive capabilities, we forecast country-specific unemployment over a time period of ten years. Our forecasting design is recursive. We use the period from 1999:M01 to 2004:M12 as an initial training sample and predict  $h \in \{1, 3\}$  steps ahead. After obtaining the corresponding predictive distributions, we expand the initial estimation period by a single month and repeat the procedure until we reach the final point in the sample. This procedure yields a sequence of 120 predictive densities. Point forecasts are then compared using the RMSE. Since RMSEs neglect gains in forecasting accuracy associated with higher order moments of the predictive density, we also consider average differences in LPS that can be interpreted as log (predictive) Bayes factors (BF), computed using the Gaussian approximation to the predictive likelihood ([Geweke and Amisano, 2010](#)).





**Fig. 5:** Posterior distribution of  $\log(\lambda_j)$  for Eurozone data

The competing models considered are all nested alternatives of the proposed framework and the unrestricted VAR with the NG shrinkage prior described in Section 3. To assess the merits of using a finite mixture model, we estimate different variants of the panel VAR with  $Q = 8$  as well as  $Q = 1$ . In addition, we moreover investigate how country-specific models without DIs perform by setting  $\xi_i = 10^6$  for all  $i$ . This effectively turns off lagged interdependencies across countries but still allows for contemporaneous relations through the factor model in the errors. Moreover, cross-country information is still exploited to estimate the mixture components. Notice that if  $Q = 1$ , our model conceptually resembles the framework proposed in [Jarociński \(2010\)](#). Finally, all models are benchmarked against a simple AR(1) model with SV.

Table 4 shows the results of the forecasting exercise. The left panel of the table refers to relative RMSEs while the right panel of the table shows differences in average marginal LPS, relative to the AR(1) model. Moreover, the upper part of the table displays the results for the one-month-ahead horizon and the lower part shows the findings for the one-quarter-ahead horizon. The final row in each part of the table displays average joint log predictive scores for the unemployment equations across all countries (relative to the AR(1) benchmark), obtained by integrating out the effect of the other quantities.

We start by considering one-step-ahead forecasts. Here, we observe that in terms of point predictions, the accuracy differences against the AR(1) model vary widely across

countries. For some countries (Austria, Belgium, and France) we find that all models considered fail to improve upon the AR(1) benchmark, being outperformed by margins up to twelve percent in RMSE terms. This, however, does not carry over to the remaining countries where we observe a particularly strong performance of our approach, especially for Germany, Greece, Finland, Ireland, as well as Spain. Notice that the accuracy differences between estimating a model with  $G = 1$  and  $G = 8$  tend to be quite small. This, of course, can be related to the estimated regime allocation discussed above, where we find that all countries except Greece are pooled in a single country group, with the latter forming its own group. This could also explain why we observe slightly better predictive capabilities for Greece. The unrestricted VAR with a NG prior also performs well for these countries, showing the best performance in two cases (i.e. for Germany and Spain).

Considering density predictions corroborates the findings discussed for the point forecasts to the extent that our proposed framework is performing well, improving upon all competing models for four countries while being almost on par with the single regime specification in several other cases. The only country where the AR(1) benchmark consistently outperforms all multivariate competitors is Belgium. Again, in the case of Greece, our finite mixture model yields precise density predictions whereas for the remaining countries, the differences between the model with  $G = 1$  and  $G = 8$  appear to be negligible.

Considering the three-months-ahead horizon does not substantially alter the insights gained above. More specifically, we still observe weaker forecasting accuracy of our multivariate models in Austria, Belgium, and France. For the remaining models, all competitors improve upon the AR(1) benchmark with the mixture model outperforming all competing specifications for three countries. This also carries over to density predictions, with accuracy gains from estimating a mixture model being quite pronounced for Greece while being approximately on par with the simpler specifications for most remaining countries.

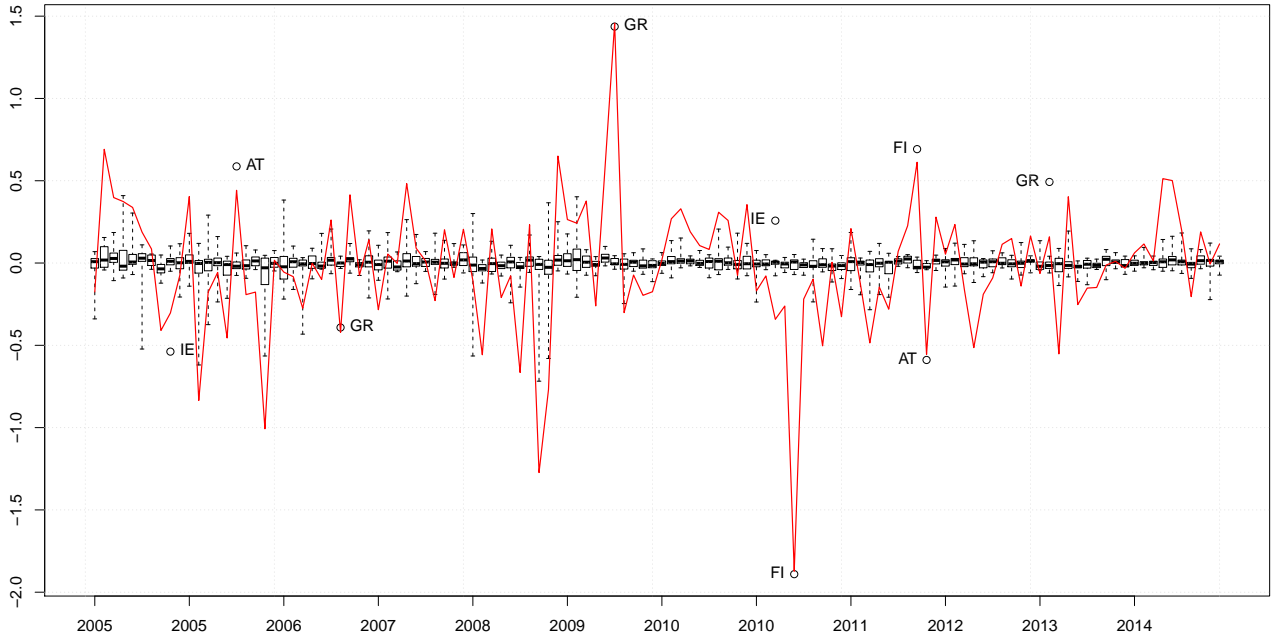
Next, we turn to the joint performance measured by evaluating the joint LPS for unemployment across all countries. This reveals that both PVAR specifications perform extraordinary well, showing no substantial forecast differences between choosing  $G = 1$  and  $G = 8$ . However, we would like to stress that this is specific to the dataset adopted. Mixing in other economies that feature stronger cross-sectional heterogeneity in terms of their domestic dynamics could lead to more pronounced differences in predictive performance between our proposed mixture model and a model with a single cluster. Moreover, notice that in terms of cluster allocation, our framework is already quite close to the single regime specification. The muted accuracy losses observed can thus be traced back to the additional

	panel VAR		VAR		NG	panel VAR		VAR		NG
	$G = 8$	$G = 1$	$G = 8$	$G = 1$		$G = 8$	$G = 1$	$G = 8$	$G = 1$	
One-month-ahead										
DE	0.901	0.901	0.968	0.971	0.870	1.572	1.579	1.504	1.538	1.618
AT	1.114	1.116	1.000	1.000	1.143	6.096	6.098	1.750	1.971	5.881
BE	1.120	1.111	1.003	1.001	1.133	-0.094	-0.088	-0.008	-0.007	-0.116
FI	0.884	0.885	1.042	1.048	0.896	6.046	6.055	4.160	3.973	6.030
FR	1.031	1.028	1.000	1.000	1.009	5.787	5.785	1.391	1.550	5.795
GR	0.718	0.721	1.001	0.987	0.723	6.635	6.615	2.873	2.843	6.561
IE	0.586	0.589	0.834	0.832	0.593	0.263	0.268	0.066	0.063	0.250
IT	0.959	0.955	0.985	0.985	0.969	0.605	0.598	0.430	0.447	0.582
NL	0.942	0.944	1.016	1.016	0.928	1.762	1.751	1.674	1.642	1.760
PT	0.931	0.932	1.009	1.010	0.915	0.032	0.037	-0.018	-0.022	0.056
ES	0.544	0.546	0.964	0.962	0.492	0.361	0.356	0.011	0.006	0.421
Joint log predictive score						29.022	29.048	13.856	14.014	28.837
One-quarter-ahead										
DE3	0.923	0.920	0.968	0.974	0.875	1.605	1.613	1.991	2.021	1.270
AT3	1.197	1.200	1.000	1.000	1.229	7.129	7.138	1.870	2.112	6.360
BE3	1.216	1.209	1.008	1.008	1.261	-0.272	-0.274	-0.027	-0.025	-0.844
FI3	0.844	0.846	1.052	1.070	0.889	6.667	6.677	4.596	4.481	6.117
FR3	1.098	1.102	1.000	1.000	1.089	7.249	7.241	1.903	2.081	6.761
GR3	0.659	0.661	1.045	1.019	0.670	6.869	6.841	2.968	2.985	6.696
IE3	0.599	0.601	0.830	0.829	0.595	0.066	0.065	0.130	0.129	-0.302
IT3	0.866	0.869	0.986	0.986	0.905	0.750	0.750	0.644	0.621	0.253
NL3	0.813	0.810	1.008	1.004	0.806	2.038	2.040	2.302	2.277	1.665
PT3	0.844	0.841	0.994	0.995	0.823	-0.087	-0.083	0.021	0.012	-0.483
ES3	0.592	0.593	0.967	0.963	0.519	-0.405	-0.401	0.094	0.090	-0.344
Joint log predictive score						33.213	33.252	15.247	15.655	31.807

**Table 4:** Relative RMSEs and average LPS against the AR(1) benchmark over the hold-out period: 2005:01 to 2014:12

set of parameters to be estimated. Nevertheless, the findings indicate that our approach successfully selects an appropriate model specification, yielding a situation where accuracy differences for most countries are small while being somewhat larger for some other countries (like Greece).

Finally, to gain a deeper understanding on how model performance changes over time, [Fig. 6](#) displays boxplots of the differences of the marginal LPS for our proposed model with  $G = 8$  versus the marginal LPS of a specification with a single cluster. In addition, the red line represents the difference in joint LPS between both models. At least two findings stand out. First, we observe pronounced differences in forecasting performance over time, being



**Fig. 6:** Boxplots of the marginal log predictive scores for the PVAR with  $G = 8$  relative to the PVAR with  $G = 1$ , and the evolution of the joint LPS between both models (in red).

negative during the run-up to the financial crisis. Especially in the first half of 2008, we observe a weaker performance of our mixture model. We conjecture that this is mainly driven by the fact that during this particular business cycle upturn, business cycles have been rather synchronized and domestic dynamics appeared homogeneous across countries. By contrast, the figure suggests that during the great financial crisis in the second half of 2008 and the first half of 2009, our model improves upon the single regime specification. Second, notice that cross-country differences tend to be small over time with a few exceptions. For instance, the comparatively better performance of the finite mixture specification during the crisis is mainly driven by Greece. The fact that the marginal LPS and the joint LPS almost coincide provides evidence that the difference is almost exclusively due to a better forecasting performance for Greece unemployment. During the second part of the hold-out sample, we observe that the sharp decline in model evidence in May 2010 is largely caused by a particularly weak performance in terms of predicting unemployment in Finland.

## 5 Concluding remarks

In this paper, we develop a panel VAR model that efficiently pools information across countries to inform country-specific estimates on lagged domestic quantities. We control for dynamic interdependencies by constructing a shrinkage prior that pulls all coefficients associated with other countries' lagged endogenous variables to zero while at the same time provides additional flexibility due to local scalings that enables variable and country-specific testing on whether lagged relations exist. Static interdependencies are parsimoniously modeled through a factor stochastic volatility model on the error variance-covariance matrix.

We assess the merits of the proposed framework by carrying out a simulation exercise. Using synthetic data, we find that our model successfully recovers the actual regime allocation and increases the accuracy of the point estimators of the VAR coefficients. In a real data application, we forecast Eurozone unemployment rates over the last ten years. We find that using a finite mixture models yields precise point and density forecasts, performing well relative to a range of commonly used benchmark models in the literature.

## References

- Aguilar O and West M (2000) Bayesian dynamic factor models and portfolio allocation. *Journal of Business & Economic Statistics* 18(3), 338–357
- Allenby GM, Arora N and Ginter JL (1998) On the heterogeneity of demand. *Journal of Marketing Research* , 384–389
- Bhattacharya A, Pati D, Pillai NS and Dunson DB (2015) Dirichlet–Laplace priors for optimal shrinkage. *Journal of the American Statistical Association* 110(512), 1479–1490
- Biernacki C, Celeux G and Govaert G (2000) Assessing a mixture model for clustering with the integrated completed likelihood. *IEEE transactions on pattern analysis and machine intelligence* 22(7), 719–725
- Canova F and Ciccarelli M (2004) Forecasting and turning point predictions in a Bayesian panel VAR model. *Journal of Econometrics* 120(2), 327–359
- Canova F and Ciccarelli M (2009) Estimating multicountry VAR models. *International Economic Review* 50(3), 929–959
- Canova F and Ciccarelli M (2013) Panel Vector Autoregressive Models: A Survey. In *VAR Models in Macroeconomics–New Developments and Applications: Essays in Honor of Christopher A. Sims*. Emerald Group Publishing Limited, 205–246
- Crespo Cuaresma J, Feldkircher M and Huber F (2016) Forecasting with global vector autoregressive models: A Bayesian approach. *Journal of Applied Econometrics* 31(7), 1371–1391
- Dees S, Di Mauro F, Pesaran HM and Smith LV (2007) Exploring the international linkages of the euro area: A global VAR analysis. *Journal of Applied Econometrics* 22(1), 1–38
- Dellaportas P and Papageorgiou I (2006) Multivariate mixtures of normals with unknown number of components. *Statistics and Computing* 16(1), 57–68

- Doan TR, Litterman BR and Sims CA (1984) Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews* 3(1), 1–100
- Feldkircher M and Huber F (2016) The international transmission of US shocks?evidence from Bayesian global vector autoregressions. *European Economic Review* 81, 167–188
- Frühwirth-Schnatter S (2001) Markov chain Monte Carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association* 96(453), 194–209
- Frühwirth-Schnatter S (2006) *Finite mixture and Markov switching models*. Springer Science & Business Media
- Frühwirth-Schnatter S (2011) Label switching under model uncertainty. *Mixtures: Estimation and Application* , 213–239
- Frühwirth-Schnatter S and Kaufmann S (2008) Model-based clustering of multiple time series. *Journal of Business & Economic Statistics* 26(1), 78–89
- Frühwirth-Schnatter S, Tüchler R and Otter T (2004) Bayesian analysis of the heterogeneity model. *Journal of Business & Economic Statistics* 22(1), 2–15
- Geweke J and Amisano G (2010) Comparing and evaluating Bayesian predictive distributions of asset returns. *International Journal of Forecasting* 26(2), 216–230
- Griffin JE and Brown PJ (2010) Inference with normal-gamma prior distributions in regression problems. *Bayesian Analysis* 5(1), 171–188
- Huber F (2016) Density forecasting using Bayesian global vector autoregressions with stochastic volatility. *International Journal of Forecasting* 32(3), 818–837
- Huber F and Feldkircher M (2017) Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business & Economic Statistics* , 1–13
- Ishwaran H, James LF and Sun J (2001) Bayesian model selection in finite mixtures by marginal density decompositions. *Journal of the American Statistical Association* 96(456), 1316–1332
- Jarociński M (2010) Responses to monetary policy shocks in the east and the west of Europe: a comparison. *Journal of Applied Econometrics* 25(5), 833–868
- Kastner G (2016) Dealing with Stochastic Volatility in Time Series Using the R Package stochvol. *Journal of Statistical Software* 69(5), 1–30
- Kastner G and Frühwirth-Schnatter S (2014) Ancillarity-sufficiency interweaving strategy (ASIS) for boosting MCMC estimation of stochastic volatility models. *Computational Statistics & Data Analysis* 76, 408–423
- Kastner G and Huber F (2017) Sparse Bayesian vector autoregressions in huge dimensions. *arXiv preprint arXiv:1704.03239*
- Koop G and Korobilis D (2016) Model uncertainty in panel vector autoregressive models. *European Economic Review* 81, 115–131
- Koop G and Korobilis D (2018) Forecasting with High-Dimensional Panel VARs. *Essex Finance Centre Working Papers* (31)
- Korobilis D (2016) Prior selection for panel vector autoregressions. *Computational Statistics & Data Analysis* 101, 110–120
- Lenk PJ and DeSarbo WS (2000) Bayesian inference for finite mixtures of generalized linear models with random effects. *Psychometrika* 65(1), 93–119
- Malsiner-Walli G, Frühwirth-Schnatter S and Grün B (2016) Model-based clustering based on sparse finite Gaussian mixtures. *Statistics and computing* 26(1-2), 303–324
- Park T and Casella G (2008) The Bayesian Lasso. *Journal of the American Statistical Association*

103(482), 681–686

Pesaran MH, Schuermann T and Weiner SM (2004) Modeling regional interdependencies using a global error-correcting macroeconomic model. *Journal of Business and Economic Statistics* 22(2), 129–162

Pitt M and Shephard N (1999) Time varying covariances: a factor stochastic volatility approach. *Bayesian statistics* 6, 547–570

Richardson S and Green PJ (1997) On Bayesian analysis of mixtures with an unknown number of components (with discussion). *Journal of the Royal Statistical Society: series B (statistical methodology)* 59(4), 731–792

Rousseau J and Mengersen K (2011) Asymptotic behaviour of the posterior distribution in overfitted mixture models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 73(5), 689–710

Yau C and Holmes C (2011) Hierarchical Bayesian nonparametric mixture models for clustering with variable relevance determination. *Bayesian Analysis* 6(2), 329