

# Prior selection for panel vector autoregressions

Dimitris Korobilis\*  
University of Glasgow

February 20, 2016

## Abstract

Bayesian shrinkage priors have been very popular in estimating vector autoregressions (VARs) of possibly large dimensions. Many of these priors are not appropriate for multi-country settings, as they cannot account for the type of restrictions typically met in panel vector autoregressions (PVARs). With this in mind, new parametric and semi-parametric priors for PVARs are proposed, which perform valuable shrinkage in large dimensions and also allow for soft clustering of variables or countries which are homogeneous. The implication of these new priors for modelling interdependencies and heterogeneities among different countries in a panel VAR setting, is discussed. Monte Carlo evidence and an empirical forecasting exercise show clear and important gains from the new priors compared to existing popular priors for VARs and PVARs.

**Keywords:** Bayesian model selection; shrinkage; spike and slab priors; forecasting; large vector autoregression

---

\*Adam Smith Business School, University of Glasgow, Room 204c Gilbert Scott building, Glasgow, G12 8QQ, United Kingdom. Tel: +44 (0)141 330 2950. e-mail: Dimitris.Korobilis@glasgow.ac.uk

# 1 Introduction

Most issues that economists have to deal with when evaluating macroeconomic policies or forecasting economic trends are inherently multivariate, involving analysis of variables such as inflation, GDP, interest rate, and unemployment rate. Since the seminal work of Sims (1980), possibly the most popular econometric tool for analyzing multivariate time series data has been the vector autoregressive (VAR) model; see Koop and Korobilis (2010) for a recent review. In an increasingly globalized world characterized by a post-financial crisis quagmire of elevated economic and political risk for several individual countries (e.g. Iceland's banking sector collapse) and unions (e.g. the Eurozone debt crisis which peaked in 2010-2012), turbulence in global oil markets, and unprecedented exchange rate fluctuations, economists are faced with the challenge of having to monitor and model the global rather than the local economy. Such events have given rise to recent research which develops econometric methods for panel vector autoregressive (PVAR) models; see Canova and Ciccarelli (2013) for a recent review. PVAR models extend vector autoregressions for macroeconomic variables of a single country, to a setting with many macroeconomic and/or financial variables for several countries. This feature allows one to examine interactions, interdependencies, and linkages between different variables of different countries. Considering that the VAR has been a powerful tool that allows macroeconomists to link data to economic theories, measure impulse responses, and forecast, the panel VAR setting can allow us to generalize such useful econometric exercises to the global dimension.

In this paper, novel Bayesian prior specifications for panel VARs are proposed which allow for examining the existence (or absence) of certain dependencies and homogeneities across countries. The starting point is a setting where the researcher is faced with a possibly large number of macroeconomic variables  $G$  for a large number of countries  $N$ . The definition "large" here means that the model is big enough in order to possibly have a sparse structure. Note that if a VAR for a single country has  $G = 10$  variables, then this would be of medium size. Once we consider only, say,  $N = 5$  such countries then the PVAR has 50 variables in total and can be considered large dimensional. It is important to clarify, following ideas in Canova and Ciccarelli (2013) and Koop and Korobilis (2015), that sparsity in a PVAR is expected to be of a very specific form which has to be reflected when designing priors for such models. For example, it might be the case that homogeneities exist between certain countries such that some groups of PVAR coefficients are similar among these countries. Similarly, lags of macroeconomic variables of one country may not affect the macroeconomic variables of some other country, a case which reflects the absence of dynamic interdependencies from one country to the other. As explained in detail in this paper, this type of restriction is different in nature from typical variable shrinkage/selection procedures which rely on finding zero restrictions on the coefficients of a certain regression model (e.g. VAR). Additionally, priors should be specified in such a way that reflect our desire to be agnostic about which (groups of) countries are homogeneous and which countries lack dynamic interdependencies. For example, a researcher might impose interdependencies based on experience and come up with premises such as "large economies (e.g. US) affect smaller countries, but small countries do not affect larger countries". The recent experience of the European debt crisis, where the default risk of smaller countries such as Greece, Ireland and Portugal kept the global economy in agony for several years, shows that in a

complex, globalized economy the econometrician cannot be certain apriori about the nature of interdependencies that may hold in the data. In extreme cases, for the sake of parsimony and simplicity, many researchers decide to estimate a single VAR for each country, thus, ignoring the possibility of linkages between countries. An alternative extreme case is to allow unrestricted estimation of the large PVAR - such strategy will inevitably lead to poor estimates due to shortage of degrees of freedom.

There are numerous examples of meticulously developed priors that have been used in order to impose restrictions on single-country VARs. For example, Banbura et al. (2010) consider VARs with 130 macroeconomic variables for the US, leading to more than 200,000 autoregressive coefficients, estimated using 700 monthly time series observations for each variable. Banbura et al. (2010), as well as several other papers, such as Carriero, Clark and Marcellino (2011) and Carriero, Kapetanios and Marcellino (2009), rely on a natural conjugate prior which features hyperparameters similar to the traditional Minnesota prior (Littermann, 1986). While natural conjugate priors imply certain restrictions (Koop and Korobilis, 2010), they lead to fast computation in large VAR systems. Giannone, Lenza and Primiceri (forthcoming) propose a full Bayes treatment of the natural conjugate version of the Minnesota prior by estimating its shrinkage hyperparameter from the likelihood, rather than fine-tuning it subjectively. George, Sun and Ni (2008) and Korobilis (2008, 2013) develop Bayesian model selection priors which find elements of the autoregressive coefficients and/or the VAR covariance matrix which are zero; see also Koop (2013) for an application. Villani (2010) and Giannone, Lenza and Primiceri (2014) develop priors for the long-run/steady-state VAR, where both priors have shrinkage properties. One could argue that all these approaches could be readily used in the PVAR setting in order to impose restrictions. Nevertheless, all these types of shrinkage priors developed for the VAR model completely ignore the panel dimension of a PVAR and the existence of possible homogeneities between countries. This means that all the priors above will treat each of the  $N \times G$  equations of the PVAR with equal weight a-priori, ignoring that there are  $N$  copies of the same  $G$  variables in such a VAR, and that many times macroeconomic and financial variables such as GDP, inflation, and asset prices for several countries tend to comove.

Following the contribution of Koop and Korobilis (2015), parametric and semiparametric Bayesian model selection priors are defined, carefully tailored to incorporate panel restrictions. In particular, these restrictions involve finding groups of homogeneous coefficients by testing their equality, and examining the lack of dynamic interdependencies between countries by restricting certain coefficients to zero. A Markov Chain Monte Carlo (MCMC) algorithm is proposed, which stochastically examines the most probable combinations of these panel restrictions, and samples from the posterior distribution of the restricted PVAR coefficients. The set of priors specified have different properties and reflect a varying degree of trade-off between flexibility and computational tractability. A detailed Monte Carlo exercise shows that both parametric and semi-parametric priors find with high accuracy the correct panel restrictions in sparse PVARs of large dimensions. Additionally, in a forecasting exercise which involves modeling three variables for ten Eurozone countries (i.e. 30-variable panel VAR), it is shown that the priors proposed in this paper can significantly improve mean and density forecasts compared to a Minnesota-type prior and an automatic Bayesian model selection prior for VARs, as well as existing competing priors for PVARs.

Therefore, the main contribution of this paper is to show that when panel structure is explicit in the data, and interdependencies and heterogeneities are present, there is clear empirical evidence that the proposed priors will significantly improve inference. This result cannot generalize, of course, to settings without panel structure (e.g. typical VAR for one country), in which case less computationally complex priors such as the Minnesota prior implementation of Giannone, Lenza and Primiceri (forthcoming) are expected to be more efficient and possibly more accurate.

In the next section, the Panel VAR framework and the type of restrictions a researcher is interested in examining, are defined. Then, in Section 3, the relevant parametric and semi-parametric priors are specified. Section 4 details a Monte Carlo exercise where the panel VAR priors are contrasted to typical shrinkage priors for large VARs (without panel structure). Finally, Section 5 evaluates the same set of priors on real data, and Section 6 concludes.

## 2 Vector autoregressions for panels of countries

Let  $y_{it}$  denote a vector of  $G$  dependent variables for country  $i$  observed at time  $t$ ,  $i = 1, \dots, N$ ,  $t = 1, \dots, T$ . The VAR for country  $i$  can be written as:

$$y_{it} = A_{i1}y_{1,t-1} + \dots + A_{ii}y_{i,t-1} + \dots + A_{iN}y_{N,t-1} + \varepsilon_{it}, \quad (1)$$

where  $A_{i,j}$  are  $G \times G$  matrices for each  $i, j = 1, 2, \dots, N$ , and  $\varepsilon_{it} \sim N(0, \Sigma_{ii})$  with  $\Sigma_{ii}$  covariance matrices of dimension  $G \times G$ . The collection of such  $N$  country-specific VARs, which is of the form

$$Y_t = AY_{t-1} + \varepsilon_t, \quad (2)$$

is equivalent to a multivariate regression model for the  $NG \times 1$  vector of endogenous variables  $Y_t = (y'_{1t}, \dots, y'_{Nt})'$ . One lag of each dependent variable is assumed throughout this paper for notational simplicity, since all higher order VARs can be written in VAR(1) companion form. An additional assumption is that  $\varepsilon_t \sim N(0, \Sigma)$  with  $\Sigma$  the full  $NG \times NG$  covariance matrix, meaning that  $cov(\varepsilon_{it}, \varepsilon_{jt}) = E(\varepsilon_{it}, \varepsilon_{jt}) = \Sigma_{ij} \neq \mathbf{0}$  where  $\Sigma_{ij}$  is a the  $ij$ -th  $G \times G$  block of the matrix  $\Sigma$  that denotes the covariance matrix between the errors in the VARs of country  $i$  and country  $j$ . If no further assumptions are made about the model coefficients, this specification is referred to as the *unrestricted* PVAR.

Just by working with moderate values of  $N$  and  $G$ , the dimension of the PVAR will grow quickly and shrinkage may be desirable. For instance, an application of the PVAR methodology for the currently 19 Eurozone countries using, say, three macroeconomic/financial variables for each country, means that the VAR has  $NG = 57$  endogenous variables and we have to estimate  $3249 \times p$  autoregressive coefficients, for some choice of lag length  $p$ . Canova and Ciccarelli (2013) and Koop and Korobilis (2015) argue that it is not optimal to treat the PVAR in equation (2) as a large VAR, and shrink uniformly the  $NG \times NG$  coefficient matrix  $A$ . This is because typical shrinkage priors for VARs would ignore the panel structure of the PVAR model. Looking at equation (1) we should expect that lags of own variables for country  $i$  have little probability of being zero. In that respect, there is more probability that one or more of the remaining  $N - 1$  countries' variables might not be relevant for the

equation of country  $i$ , that is, one or more of the matrices  $A_{i1}, \dots, A_{ii-1}, A_{ii+1}, \dots, A_{iN}$  is zero. When such a restriction exists, e.g.  $A_{ij} = 0$ , then there are no dynamic interdependencies from country  $j$  to country  $i$ ,  $i, j = 1, \dots, N$ ,  $i \neq j$ . Similarly, due to the panel structure of the data, one would also expect that some coefficients are homogeneous. Koop and Korobilis (2015) note that such cross-sectional homogeneities might exist in the own lags of different countries, that is,  $A_{ii} = A_{jj}$ ,  $i, j = 1, \dots, N$ ,  $i \neq j$ . Such a restriction might not shrink parameters to zero, but also saves degrees of freedom and has very important and interesting structural implications (it is a direct test for heterogeneities among countries).

There are  $N-1$  dynamic interdependency restrictions for each country  $i$ , meaning that in the PVAR of equation (2) we can impose a maximum of  $N(N-1)$  such restrictions. Given that different combinations of these restrictions can hold, the total number of dynamic interdependencies one can have in the PVAR is  $2^{N(N-1)}$ . Additionally, a maximum of  $\frac{N(N-1)}{2}$  cross-sectional homogeneity restrictions can be imposed. Koop and Korobilis (2015) develop a stochastic search algorithm, that explicitly tests all possible  $2^{N(N-1)}$  dynamic interdependency restrictions, and all the possible  $2^{N(N-1)/2}$  cross-sectional homogeneity restrictions. In their application of just 10 Euro-Area countries, the total number of restrictions they search using the Gibbs sampler is  $2^{90} \times 2^{45}$  which is a very large number. It is clear that such interdependency and homogeneity restrictions take into account explicitly the panel structure of the VAR.

The **Stochastic Search Specification Selection (SSSS)** algorithm of Koop and Korobilis (2015) builds on the Stochastic Search Variable Selection prior of George and McCulloch (1993) and George et al (2008) for VARs, but it takes into account the panel restrictions described above. The  $S^4$  prior for the dynamic interdependency (denoted with the superscript DI) restrictions is

$$vec(A_{ij}) | \gamma_{ij}^{DI} \sim (1 - \gamma_{ij}^{DI}) N(0, \underline{\tau}_1^2 \times I) + \gamma_{ij}^{DI} N(0, \underline{\tau}_2^2 \times I), \quad (3)$$

$$\gamma_{ij}^{DI} \sim \text{Bernoulli}(\pi_{ij}^{DI}), \quad \forall j \neq i, \quad (4)$$

$$\pi_{ij}^{DI} \sim \text{Beta}(1, \underline{\varphi}), \quad (5)$$

where  $\underline{\tau}_1^2$  is “small” and  $\underline{\tau}_2^2$  “large” so that, if  $\gamma_{ij}^{DI} = 0$ ,  $A_{ij}$  is shrunk to be near zero and, and if  $\gamma_{ij}^{DI} = 1$ , a relatively uninformative prior is used. The  $S^4$  prior for the cross-sectional homogeneity (denoted with the superscript CSH) restrictions is

$$vec(A_{ii}) | \gamma_{ij}^{CSH} \sim (1 - \gamma_{ij}^{CSH}) N(A_{jj}, \underline{\xi}_1^2 \times I) + \gamma_{ij}^{CSH} N(A_{jj}, \underline{\xi}_2^2 \times I), \quad (6)$$

$$\gamma_{ij}^{CSH} \sim \text{Bernoulli}(\pi_{ij}^{CSH}), \quad \forall j \neq i, \quad (7)$$

$$\pi_{ij}^{CSH} \sim \text{Beta}(1, \underline{\varphi}), \quad (8)$$

where  $\underline{\xi}_1^2$  is “small” and  $\underline{\xi}_2^2$  is “large” so that, if  $\gamma_{ij}^{CSH} = 0$ ,  $A_{ii}$  is shrunk to be near  $A_{jj}$ , and if  $\gamma_{ij}^{CSH} = 1$ , a relatively uninformative prior is used.

While application of the DI restrictions is relatively simple, application of the CSH restrictions is non-trivial. This is because with the CSH prior we seek to test equality of two matrices ( $A_{ii} = A_{jj}$ ) and do so for all possible combinations of  $i$  and  $j$ ,  $i, j = 1, \dots, N$ . The authors provide a novel solution to this sampling problem, and more details can be found in

their Appendix. At the same time, there are two important limitations of this prior. First, the kind of restrictions that we want to look at involve matrices with  $G^2$  elements which are either zero ( $A_{ij} = 0$ ) or equal to each other ( $A_{ii} = A_{jj}$ ). This prior cannot account for the fact that, say, only some elements of  $A_{ii}$  could be equal to zero. Additionally, because of this group model selection procedure, it is very hard to test the actual restrictions. This would be equivalent to setting  $\underline{\tau}_1^2 = \underline{\xi}_1^2 = 0$ . However, for computational reasons the authors set  $\underline{\tau}_1^2, \underline{\xi}_1^2 = \delta$  for some positive  $\delta$  being very small but not exactly zero. But that means that the DI and CSH restrictions will only hold approximately, in particular these priors will allow to test the hypothesis  $A_{ij} \approx 0$  and  $A_{ii} \approx A_{jj}$ . In the next Section, model selection priors for PVARs are motivated, that do not suffer from these two shortcomings of the SSSS prior.

### 3 Flexible model selection priors

In order to develop the relevant priors proposed in this paper, consider an alternative form of the PVAR which is

$$Y_t = Z_t \alpha + \varepsilon_t, \quad (9)$$

where  $Z_t = I_{NG} \otimes Y_{t-1}$ ,  $\alpha = \text{vec}(A')$  is the  $K \times 1$  vector of all PVAR coefficients,  $K = NG^2$ . The models in equations (2) and (9) are observationally equivalent; the difference in their specifications serves as a means of using alternative expressions for posterior estimation. Finally, in the hierarchical priors introduced in this Section, some prior hyperparameters have to be selected by the researcher and some prior hyperparameters will have their own priors. I use an underscore in order to distinguish between these two sets of prior hyperparameters, that is,  $\underline{m}$  is a fixed hyperparameter selected by the researcher, and  $m$  showing up in a prior is a hyperparameter which is a random variable.

#### 3.1 A parametric PVAR prior

The first prior I consider is inspired by Canova and Ciccarelli (2009) who, in the context of time-varying parameter VARs, extract latent factors from the VAR coefficients. These factors are lower dimensional representation of the coefficient and also serve the purpose of grouping relevant coefficients. For example, Canova and Ciccarelli (2009) show that we might want to extract one factor from each of the  $G^2$  coefficients of the own lags for each country  $i$  - these are the coefficients in the matrix  $A_{ii}$  in equation (1). Similarly, we can cluster all  $(N-1)G^2$  in the matrices  $(A_{i1}, \dots, A_{i,i-1}, A_{i,i+1}, \dots, A_{iN})$  into a separate factor for each country  $i$ . Finally, we can extract a single factor from all  $K = NG^2$  coefficients. This structure can be represented using the following equation

$$\alpha = \Xi \theta + v,$$

where  $\Xi$  is a  $K \times s$  matrix of predetermined loadings,  $\theta$  is an  $s \times 1$  lower dimensional parameter vector (“factors”) with  $s \ll K$ , and  $v \sim N(0, \Sigma \otimes \sigma^2 I)$ . The loadings  $\Xi$  have entries zero and one, and are defined in such a way that coefficients of the same country, or of the same variable for different countries are grouped together; see Canova and Ciccarelli (2009) for a detailed description and an example of how the matrix  $\Xi$  looks like in an

example with  $G = 2$ ,  $N = 2$ . The equation above means that the implied prior for  $\alpha$  is  $\alpha|\Sigma \sim N(\Xi\theta, \Sigma \otimes \sigma^2 I)$ , and is indeed a conjugate prior since the error variance  $\Sigma$  shows up in the prior variance term.

Canova and Ciccarelli (2009) do not consider the possibility that a coefficient might be zero, so that their prior can be quite restrictive: it assumes that a single coefficient  $\alpha_k$  is always clustered with some other non-zero coefficient  $\alpha_l$ , even if the “true” value of  $\alpha_k$  is zero. In order to deal this culprit of the Canova and Ciccarelli (2009) prior, I propose a modification based on spike and slab priors leading to a **Bayesian Factor Clustering and Selection (BFCS)** prior which is of the form

$$\alpha_k|\gamma_k, \theta, \Sigma \sim (1 - \gamma_k) \delta_0(\alpha) + \gamma_k \Delta_k, \quad (10)$$

$$\Delta|\theta, \Sigma \sim N(\Xi\theta, \Sigma \otimes \sigma^2 I), \quad (11)$$

$$\theta \sim N(0, \underline{c}), \quad (12)$$

$$\gamma_k \sim \text{Bernoulli}(\underline{\pi}), \quad (13)$$

where  $\Delta_k$  is the  $k$ -th row of the matrix  $\Delta$ . Therefore, with probability  $1 - \pi$  the coefficient  $\alpha_k$  has prior a point mass at zero, denoted using the Dirac delta  $\delta_0$ . With probability  $\pi$  the same coefficient might come from the clustering/factor structure a-la Canova and Ciccarelli, which is fully described in equation (11).

### 3.2 A nonparametric PVAR prior

Following ideas from Dunson et al. (2008) we can use infinite mixtures, by means of Dirichlet process priors, in order to generalize spike and slab priors and at the same time allow for soft clustering of similar coefficients. Dunson et al. (2008) and MacLehose et al. (2007) propose the specification

$$\alpha_k|\gamma_k \sim (1 - \gamma_k) \delta_0(\alpha) + \gamma_k DP(\theta F_0),$$

where  $DP(\theta F_0)$  is a Dirichlet process with base measure  $F_0$ , typically a Gaussian distribution  $N(b, c)$ . The formulation above allows a coefficient either to shrink to zero or belong in one of many (infinite) other Gaussian mixture components. Note, however, that all non-zero coefficients will be clustered in  $N(b, c)$  components for some choice of  $b$  and  $c$ . That is, this prior does not allow to obtain more information about common prior locations for homogeneous coefficients, and enhance posterior inference when the information in the likelihood is weak. To solve this issue, I propose a prior which allows similar coefficients to be shrunk towards a common prior mean, which can be different for different groups of similar coefficients. In particular, I define the following **Bayesian Mixture Shrinkage (BMixS)** prior

$$\alpha_k|\mu_k, \tau_k^2 \sim N(\mu_k, \tau_k^2), \quad (14)$$

$$(\mu_k, \tau_k^{-2})|\pi \sim \pi \delta_0(\alpha) \times \delta_{10^{10}}(\tau^{-2}) + (1 - \pi) F \quad (15)$$

$$F \sim DP(\underline{\theta} F_0), \quad (16)$$

$$F_0 \sim N(0, \underline{\lambda}) \times \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right), \quad (17)$$

$$\pi \sim \text{Beta}(1, \underline{\varphi}). \quad (18)$$

The coefficients  $\alpha_k$  have a typical Normal prior, but now there is a multitude of prior location and scale parameters, which are defined by the Dirichlet process in equation (15). Therefore, this prior can achieve shrinkage towards multiple prior locations - one being the point zero which is of interest for model selection, but other locations  $\mu_k \neq 0$  can exist. The fact that  $\tau_k^{-2}$  has a Gamma prior implies that it can obtain a range of values that will allow to achieve such shrinkage towards the prior location parameter  $\mu_k$ . In fact, the particular choice  $\tau_k^{-2} \sim \text{Gamma}(\frac{1}{2}, \frac{1}{2})$  induces a heavy-tailed Cauchy prior marginally for coefficient  $\alpha_k$ . Therefore, this prior is more flexible than the Dunson et al. (2008) prior as it can achieve more complex patterns of clustering of relevant parameters. At the same time it can help decrease estimation error in the PVAR model by providing more informative prior means and variances.

### 3.3 Prior elicitation, and sensitivity issues

The priors specified in the previous subsections have multiple hierarchical layers and are, thus, more complicated than typical priors used in the VAR literature. The reason for this is that these priors have both a high degree of flexibility in capturing panel restrictions, and require minimal input by the user. While it is not possible to be completely uninformative when doing model selection (the marginal likelihood doesn't even exist for very flat priors), assigning further prior distributions on prior hyperparameters means that the likelihood will update them. The alternative would be for the researcher to preselect these hyperparameters, which might not be so straightforward thing to do in a large model with thousands of coefficients.

The BFCS is a tweak on an already established prior for panel VARs; see Canova and Ciccarelli (2013) and the numerous references therein. Its major contribution over the original prior is to allow for the possibility of model selection, so it follows the tradition of the SSVS prior of George and McCulloch (1993). Note that for identification issues we cannot place a prior on the loadings  $\Xi$ , as we would do in a factor model, so selection of this matrix is of paramount importance. Canova and Ciccarelli (2013) show in several instances that their proposed specification for  $\Xi$  works very well for panel data, something that is verified in the empirical section as well. For the variance hyperparameter  $\underline{c}$  one can simply be noninformative and choose a "large" value (e.g.  $\underline{c} = 100$ ), since it is hard to assume that in practical situations there is any prior information on the factors  $\theta$ .

Similar arguments apply to the choice of hyperparameters of the BMixS prior. That is, hyperparameter which are harder to choose have their own prior distributions, while others are chosen by the researcher. For example,  $\underline{\theta}$  is the hyperparameter controlling the concentration of draws from the Dirichlet process prior. Draws from a Dirichlet process are discrete distributions, where the base measure in this case is  $F_0$ , thus higher values of  $\underline{\theta}$  mean that such draws will be more spread out. In practical situations, with potentially large models where the likelihood might not be that informative, it is wise to fix this coefficient to  $\underline{\theta} = 1$  so that these draws are more concentrated. Note that for computational stability issues, detailed in the Appendix, I do not allow the Dirichlet process to move among infinite mixture components rather I fix the maximum number of mixture components to four. Therefore, choice of the coefficient  $\underline{\theta}$  becomes less relevant in this case. The variance parameter  $\underline{\lambda}$  can admit any uninformative value, and in this paper I set  $\underline{\lambda} = 4$ .

Both proposed priors share a common hyperparameter, that is the prior variable selection probabilities  $\underline{\pi}$ . Simply setting  $\underline{\pi} = 0.5$  is not a noninformative choice since it implies that half the coefficients should be zero, and half the coefficients should be unrestricted, but this has come to be a default choice in the empirical literature. In the case of BFCS prior, which is parametric and has a straightforward structure, I simply set  $\underline{\pi} = 0.5$ . In cases of very large models, where one expects to find a very sparse structure in the data, or simply coefficients are much more than the number of observations and there are degrees of freedom concerns, one can choose to set e.g.  $\underline{\pi} = 0.2$  which restricts 80% of the coefficients. In the case of the BMixS prior selection of  $\underline{\pi}$  is more complicated because it controls the mixing of a point mass distribution with an infinite mixture implied by the Dirichlet process. In this case, I allow the data to select the optimal value of  $\underline{\pi}$  by assigning a Beta prior on this hyperparameter. For  $\underline{\varphi} = 1$  the Beta distribution becomes noninformative, which is the choice used to produce empirical results in the subsequent sections.

## 4 Monte Carlo simulations

In this section I evaluate the ability of the two newly-developed priors to pick up the correct restrictions in PVARs. I compare these priors to unrestricted least squares, two priors for panel vector autoregressions and two popular priors for general vector autoregressions. The PVAR priors are the ones by Koop and Korobilis (2015) and Canova and Ciccareli (2009), both of which are described above in detail. The first general VAR prior is the popular Minnesota prior which Banbura, Giannone and Reichlin (2010) and Koop and Korobilis (2013) have used to estimate large VAR systems. The Minnesota prior is based on a shrinkage hyperparameter, which these two studies optimize on a grid based on goodness-of-fit measures. Here I use the algorithm of Giannone, Lenza and Primiceri (forthcoming) who develop a full-Bayes approach to estimating the Minnesota shrinkage hyperparameter. The second prior for imposing restrictions on the VAR is the one by George, Sun and Ni (2008). This algorithm is a generalization of the popular Stochastic Search Variable Selection (SSVS) algorithm of George and McCulloch (1993) for univariate regressions. Note that the SSSS of Koop and Korobilis (2015) also builds on the SSVS of George, Sun and Ni (2008). The SSSS algorithm takes into account possible panel restrictions in the VAR and is computationally efficient in very high dimensions. In contrast, the SSVS examines all possible  $2^K$  restrictions in VAR coefficients and, as a result, it can only be used in VARs of moderate dimensions. Therefore, we compare the performance of the following priors proposed in this paper:

1. **BFCS**: Bayesian Factor Clustering and Selection,
2. **BMixS**: Bayesian Mixture Shrinkage,

with the following priors which are specifically developed for PVARs:

3. **CC**: Factor shrinkage prior of Canova and Ciccareli (2009),
4. **SSSS**: Stochastic Search Specification Selection prior of Koop and Korobilis (2015),

and, finally, the following priors which are developed for general large VARs:

6. **SSVS**: Stochastic Search Variable Selection as in George, Sun and Ni (2008),
7. **GLP**: Hierarchical Minnesota-type prior with data-based estimation of shrinkage factor, as in Giannone, Lenza and Primiceri (forthcoming),
8. **OLS**: Unrestricted estimator, equivalent to a diffuse prior for VARs; see Kadiyala and Karlsson (1997).

I implement two Monte Carlo experiments: one using a small panel VAR where we impose specific interdependency and homogeneity restrictions among different countries; and one using a larger system with exactly the same VAR structure for each country (full homogeneity imposed). In both experiments I use the same default hyperparameters for all priors (uninformative, where possible). For the BFCS and Canova and Ciccarelli (2009) priors I specify  $\Xi$  following Canova and Ciccarelli (2013, page 22), and I set  $\underline{c} = 4$ . For the additional hyperparameter of the BFCS prior I set  $\underline{\pi} = 0.5$ . For the BMixS prior I set  $\underline{\lambda} = 4$  and  $\underline{\varphi} = 1$ , which are also fairly uninformative choices. For the GLP and S<sup>4</sup> priors I use the default settings described by the authors. Finally, for the SSVS of George, Sun and Ni (2008) I set  $\underline{\tau}_1^2 = 0.0001$ ,  $\underline{\tau}_2^2 = 4$  and  $\underline{\pi} = 0.5$ ; see the Technical Appendix for more details. I also simplify estimation by plugging in the OLS estimate of the PVAR covariance matrix, which allows to reduce uncertainty regarding covariance matrix estimates. This is a typical thing to do in Bayesian analysis of large systems, and has been extensively used in the first Bayesian VAR applications of the Minnesota prior; see Kadiyala and Karlsson (1997) for more details and references. In this Monte Carlo exercise interest lies in the large dimensional vector of coefficients  $\alpha$  so I use the OLS estimate of the covariance matrix in order to control for uncertainty regarding (MCMC) sampling of  $\Sigma$ .

Performance of each of the eight estimators/priors is assessed using the Mean Absolute Deviation (MAD). In particular, if  $\hat{\alpha}$  is an estimate of  $\alpha$  based on any of the eight priors, and  $\tilde{\alpha}$  is its true value from the DGP, then I define

$$MAD = \frac{1}{K} \sum_{i=1}^K |Z_i \hat{\alpha}_i - Z_i \tilde{\alpha}_i|,$$

where  $K$  denotes the number of VAR coefficients and  $Z_i$  denotes the  $i$ -th column of  $Z$ . For each of the exercises below I generate  $S = 500$  datasets and, therefore, I calculate 500 such *MAD* statistics which I summarize using boxplots.

#### 4.1 Simulation 1: small panel VAR

I generate data from a panel VAR with  $N = 3$  countries and  $G = 2$  series for each country,  $p = 1$  lags, and  $T = 50$  observations. Therefore, we have 36 autoregressive coefficients to estimate with just 50 time series observations. The model I generate has the following

parameters

$$A = \begin{bmatrix} a_1 & 0 & d_1 & 0 & e_1 & 0 \\ 0 & a_2 & 0 & d_2 & 0 & e_2 \\ b_1 & 0 & a_3 & 0 & d_3 & 0 \\ 0 & b_2 & 0 & a_4 & 0 & d_4 \\ c_1 & 0 & b_3 & 0 & a_5 & 0 \\ 0 & c_2 & 0 & b_4 & 0 & a_6 \end{bmatrix}, \quad \Psi = \begin{bmatrix} 1 & .5 & .5 & .5 & .5 & .5 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

where  $a_i \sim U(0.5, 0.9)$ ,  $b_j, d_j, c_k, e_k \sim U(-0.5, 0.5)$ ,  $i = 1, \dots, 6$ ,  $j = 1, \dots, 4$ ,  $k = 1, 2$ , and  $\Sigma = \Psi^{-1}\Psi^{-1\prime}$ . The structure for the VAR coefficients  $A$  does not imply any consistent pattern of cross-sectional homogeneities or absence of dynamic interdependencies. Nevertheless, this specific configuration for the VAR coefficients  $A$  is used in order to test the general shrinkage performance of the various priors compared in this simulation, regardless of whether heterogeneities and interdependencies occur or not in the (P)VAR model.

Figure 1 presents boxplots of the  $MAD$  statistic over the 100 samples. All six Bayesian shrinkage priors (BFCS, BMS, CC, SSSS, GLP and SSVS) introduce some bias in order to achieve a larger reduction in variance, based on the expectation that many coefficients are zero. The four panel priors introduce a much larger bias since they incorporate the expectation that groups/clusters of parameters are zero or identical to each other, and their performance is suboptimal, based on the  $MAD$ , compared to unrestricted OLS. In fact, the shrinkage GLP and SSVS priors only marginally improve over OLS, showing that in small systems there are no substantial benefits from shrinkage.

## 4.2 Simulation 2: large panel VAR

In the second DGP I consider the case with  $N = 10$ ,  $G = 4$ ,  $p = 1$  and  $T = 100$ . There are 1600 autoregressive coefficients to estimate in this model. This model has true parameters

$$A_{ij} = 0.3 \times d^{|i-j|}, \quad d \sim U(0, 0.5),$$

$$\Psi_{ij} = \begin{cases} 1, & \text{if } i = j \\ 0.5, & \text{if } i < j \\ 0, & \text{if } i > j \end{cases},$$

where  $i, j = 1, \dots, N \times G$ . This DGP does not have an explicit panel structure, but a closer look reveals that several panel restrictions can hold under this form. The VAR coefficient matrix  $A$  has a form similar to a correlation matrix, where elements which have more distance from the diagonal are essentially zero (thus implying dynamic interdependencies). At the same time, several coefficients around the main diagonal, that is, coefficients which describe the evolution of the own VAR for each country, will inevitably be similar even when  $d$  is generated randomly from a Uniform distribution (thus implying cross-sectional homogeneities). Finally, the factor 0.3 is chosen so as to ensure that the generated PVAR is stationary.

Figure 2 clearly shows that the reduction from the panel VAR priors is substantial. The best performance on average is obtained from the BFCS prior, although the BMS prior has much smaller standard deviation of  $MAD$ s over the 500 Monte Carlo samples. The

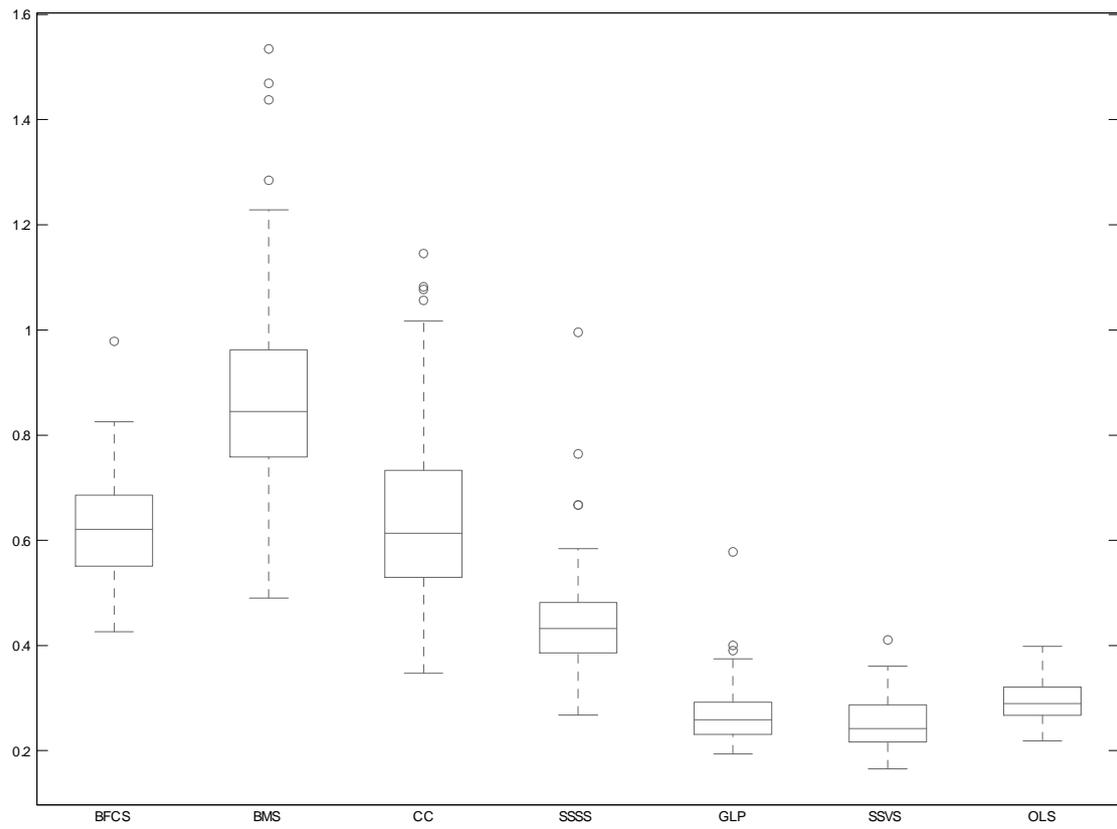


Figure 1: Boxplots of MAD statistics in the first Monte Carlo exercise (small VAR model)

GLP and SSVS priors are also performing well relative to OLS. In fact the SSVS turns out to have less uncertainty around posterior estimates compared to the SSSS. This is to be expected given the nature of the two algorithms. The SSSS only examines prespecified groups of restrictions, so unless such groupings hold exactly, the SSVS will tend to do better. However, given the quite similar performance of the two algorithms, the SSSS is to be preferred from a computational point of view. In this example with  $N = 10$ ,  $G = 4$ , the SSSS algorithm stochastically examines  $2^{90+45}$  possible DI and CSH restrictions, while the more flexible SSVS examines a whopping  $2^{1600}$  possible restrictions.

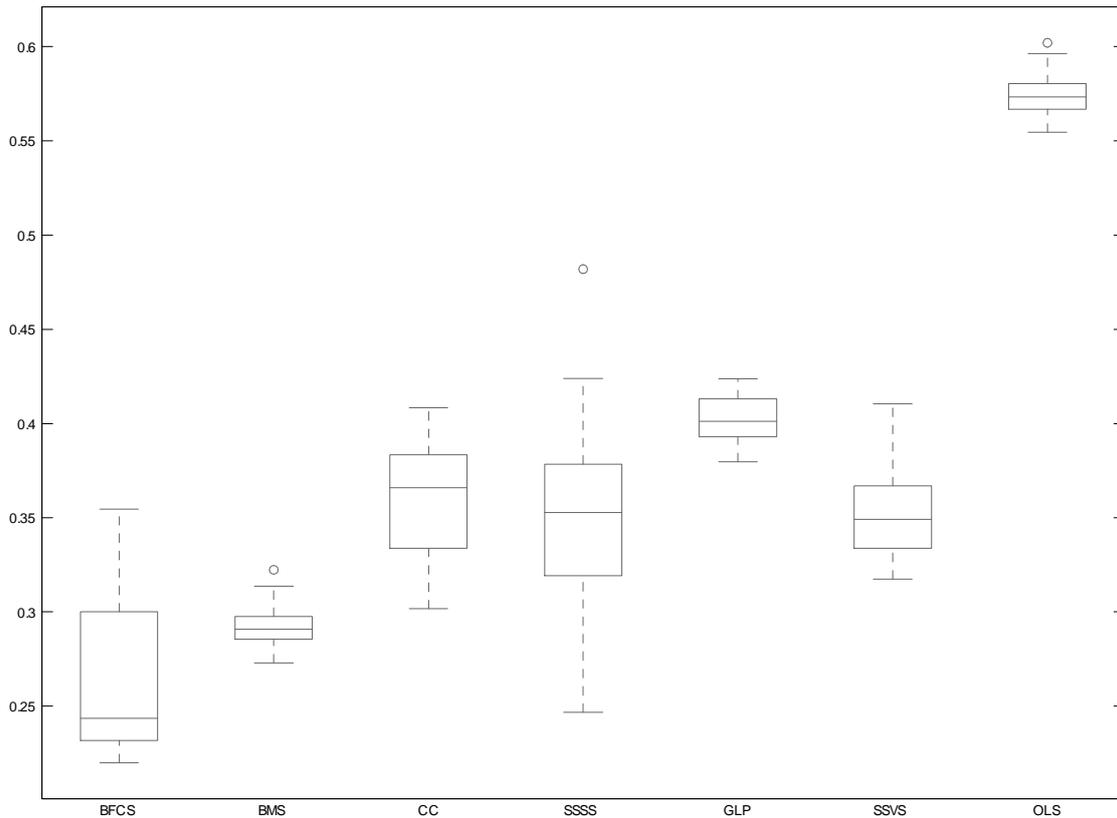


Figure 2: Boxplots of MAD statistics in the second Monte Carlo exercise (large VAR model)

## 5 Forecasting EuroZone bond yields

In this Section I present evidence on the ability of the priors suggested in this paper to provide a parsimonious representation of the PVAR, prevent overfitting and give superior forecasts. For this reason, I work with  $G = 3$  monthly variables for  $N = 10$  Eurozone countries for the period 1999M1-2012M12. The series I use are the 10 year bond yields (variable of interest during the EuroZone crisis), total industrial production (a macro

fundamental), and the average bid-ask spread (a liquidity measure), for Austria, Belgium, Finland, France, Greece, Ireland, Italy, Netherlands, Portugal and Spain. All series are expressed as spreads from the respective series of Germany. Additionally, in order to ensure that the VARs and PVARs estimated with these data are stationary, the 10 year spread and the bid-ask spread are further expressed in first differences. In this exercise the variable of interest is the first difference of the spread of the 10 year bond yields of each country compared to the yield of the 10 year German bund. These spreads have been the focus of popular press and academic research for the duration of the Eurozone debt crisis.

For the purpose of this paper, a more important aspect is that this dataset is a representative example of panel structure, that is, of possible existence or absence of homogeneities and interdependencies, along with other random groupings between countries. For example, many analysts and policy-makers when looking at these data have been using a grouping between core (Austria, Belgium, Finland, France, Netherlands) and periphery (Greece, Ireland, Italy, Portugal, Spain), in order to show that peripheral countries were exposed to higher sovereign default risk. The kind of comovements in these data can be seen in Figure 3 which plots the original data, before these are transformed. The priors suggested in this paper could be used to provide a formal data-based grouping of countries and variables, rather than relying on arbitrary groupings.

Forecasts are generated iteratively for horizons  $h = 1, \dots, 12$  and evaluated recursively for the period 2007M1-2012M12, starting with the estimation sample 199M1-2006M12 and adding one observation at a time. Here, I follow Korobilis (2013) and rely on the mean square forecast error (MSFE) and the average predictive likelihood (APL), the former being a measure of accuracy of point forecasts and the latter being a measure of accuracy of the whole predictive distribution (thus, incorporating parameter and estimation uncertainty). Here I consider the exact same priors/estimators I defined in the Monte Carlo Section, namely BFCS and BMixS proposed in this paper, the CC and SSSS panel VAR priors, the GLP and SSVS priors for VARs, and finally the unrestricted OLS estimator (noninformative prior). Note that comparisons should be straightforward and meaningful since all models have exactly the same likelihood, and any differences in posterior predictions are coming from the specification of prior distributions.

Table 1 presents MSFEs for each of the six priors relative to the MSFE of the OLS. Values lower (higher) than one mean that a method is performing better (worse) compared to OLS. Results are presented for the representative horizons  $h = 1, 3, 6, 12$ , in order to evaluate monthly, quarterly, bi-annual and annual forecasts. The results are quite clear and give full support for the following observations

1. All panel priors other than the SSSS (i.e. BFCS, BMix and CC) are consistently better than the Minnesota-type prior for the VAR.
2. The other VAR prior, the SSVS, seems to be performing relatively well, but it has the lowest MSFEs only in 10% of the cases. Additionally, whenever the SSVS is performing well the BFCS and BMixS priors are very close in terms of performance (only exception is Greece for  $h = 3$ ). In contrast, in many of the cases that either the BFCS or the BMixS priors are performing well, this performance is far more superior

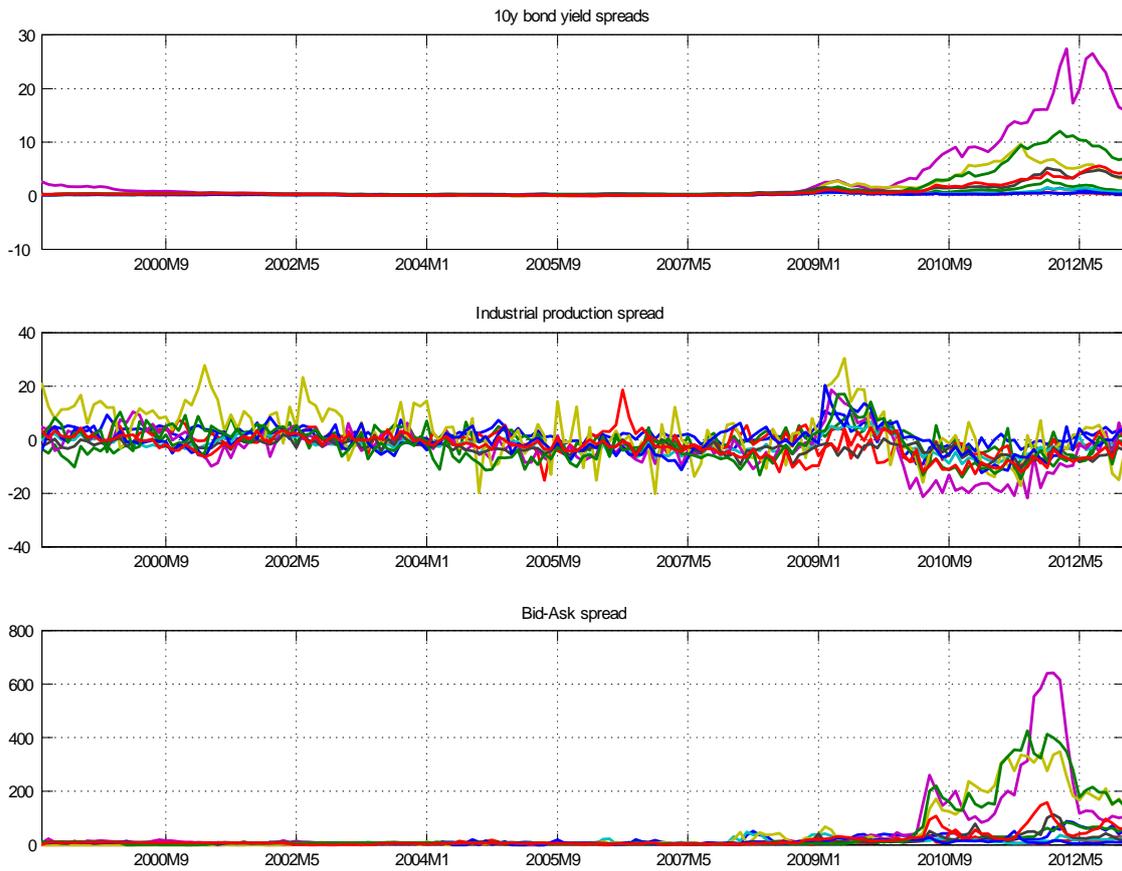


Figure 3: The original (untransformed) data used in the empirical forecasting exercise. Bond yields and the bid-ask spreads are transformed by using first differences, due to their explosive behavior towards the end of the sample.

than the SSVS (e.g. Ireland for  $h = 1$ ). This shows that there is sparsity in the data, which the three model selection priors capture, but at the same time there are homogeneities that the SSVS prior cannot capture (and the two priors suggested in this paper do capture).

3. The CC and SSSS priors can do well in some cases, but they are very volatile and unreliable in the sense that for some countries and horizons the forecasts can be extremely bad (this problem is more pronounced in the case of the SSSS prior). Note that both these priors specify in advance grouping possibilities between countries. In contrast, the priors proposed in this paper can allow the possibility for more complex groupings. The BFCS prior does this by generalizing the CC prior and allowing for sparsity, thus, elements which the CC prior might have wrongly grouped can now be zero (if the relevant variable is not important). The BMixS prior allows for both zero restrictions to occur, as well as data-based clustering of coefficients through the Dirichlet process.

Table 1. rMSFEs of Eurozone 10-y bond yield spreads forecasts

<u>FORECAST HORIZON <math>h = 1</math>:</u>										
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	0.56	0.77	<b>0.40</b>	0.76	0.72	<b>0.57</b>	0.70	0.48	<b>0.83</b>	0.61
BMixS	<b>0.50</b>	0.81	0.41	0.76	<b>0.64</b>	0.60	<b>0.69</b>	0.62	0.84	0.60
CC	0.59	0.76	0.44	0.81	0.72	<b>0.57</b>	0.70	<b>0.47</b>	<b>0.83</b>	0.63
SSSS	0.75	2.16	1.72	1.07	1.01	0.88	0.72	0.63	0.84	0.62
GLP	0.83	0.97	0.78	0.97	0.90	0.94	1.00	0.89	0.93	0.96
SSVS	0.58	<b>0.75</b>	0.45	<b>0.74</b>	0.78	0.76	0.75	1.00	1.01	<b>0.59</b>
<u>FORECAST HORIZON <math>h = 3</math>:</u>										
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	0.64	0.67	<b>0.61</b>	<b>0.57</b>	1.31	0.74	<b>0.71</b>	<b>0.53</b>	0.83	0.70
BMixS	<b>0.62</b>	<b>0.62</b>	0.62	0.61	1.22	<b>0.67</b>	0.73	0.57	0.84	<b>0.67</b>
CC	0.78	0.72	1.09	0.70	1.31	<b>0.67</b>	0.74	1.07	<b>0.82</b>	0.68
SSSS	0.79	0.83	1.12	1.26	1.41	0.76	0.80	2.74	0.88	1.41
GLP	0.90	0.80	0.85	0.76	1.11	0.93	0.90	0.78	0.93	0.97
SSVS	0.65	0.71	0.66	0.63	<b>1.03</b>	0.78	0.76	0.54	0.88	0.83
<u>FORECAST HORIZON <math>h = 6</math>:</u>										
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	<b>0.86</b>	0.94	<b>0.68</b>	0.91	0.88	0.83	0.78	0.80	<b>0.59</b>	<b>0.79</b>
BMixS	<b>0.86</b>	<b>0.93</b>	0.69	<b>0.89</b>	0.89	0.83	0.78	<b>0.74</b>	0.62	0.80
CC	1.11	1.10	1.51	0.81	<b>0.87</b>	0.78	<b>0.77</b>	1.58	0.60	0.79
SSSS	1.05	1.35	1.34	1.58	1.04	0.76	0.87	2.94	0.71	0.92
GLP	0.96	<b>0.93</b>	0.87	0.90	0.95	1.02	0.97	0.84	0.88	1.03
SSVS	0.91	<b>0.93</b>	0.76	0.94	1.01	<b>0.74</b>	0.80	0.87	0.65	0.80
<u>FORECAST HORIZON <math>h = 12</math>:</u>										
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	<b>0.72</b>	<b>0.78</b>	<b>0.58</b>	0.75	<b>0.83</b>	0.74	0.68	<b>0.46</b>	0.79	0.83
BMixS	<b>0.72</b>	<b>0.78</b>	0.61	0.75	0.84	0.77	0.69	0.52	<b>0.72</b>	0.84
CC	0.95	0.89	1.13	<b>0.65</b>	<b>0.83</b>	<b>0.73</b>	<b>0.65</b>	0.70	0.75	<b>0.80</b>
SSSS	0.93	1.15	0.81	1.51	0.86	0.77	0.78	1.62	0.92	1.11
GLP	0.97	0.91	0.93	0.89	0.94	0.96	0.95	0.86	0.95	0.90
SSVS	0.83	0.85	0.67	0.83	0.89	0.82	0.76	0.56	0.74	0.85

Notes: Entries are MSFEs for each model relative to OLS, and lower values signify better performance. Entries should be compared column-wise, that is, for each country compare the best performing model.

Table 2 shows average predictive likelihoods (APLs), which are obtained by evaluating the posterior predictive density at the true observation  $y_{t+h}$ , hence, higher values signify better performance. This table says exactly the same story and provides further support for the MSFE results. In particular, the APL results favor the BFCS and BMixS priors for producing accurate density forecasts. Note that the BFCS prior has consistently higher APL compared to the BMixS. This is to be expected, because the former prior has fewer parameters since it is a mixture of two distributions, while the BMixS is based on an “infinite” mixture which implies higher parameter uncertainty that feeds in the posterior predictive density.

Table 2. APLs of Eurozone 10-y bond yield spreads forecasts

	<u>FORECAST HORIZON <math>h = 1</math>:</u>									
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	<b>5.06</b>	4.11	<b>6.33</b>	<b>7.56</b>	<b>1.48</b>	2.97	<b>3.35</b>	<b>7.19</b>	<b>2.99</b>	<b>3.40</b>
BMixS	4.87	<b>4.17</b>	5.66	7.10	1.20	3.09	3.01	6.64	2.61	3.15
CC	4.06	3.61	3.90	5.24	1.36	2.42	3.13	4.16	2.45	3.36
SSSS	3.88	4.01	4.50	4.37	1.12	2.68	2.48	4.99	2.54	2.79
GLP	3.75	3.88	4.06	5.47	1.37	2.65	2.53	4.97	2.17	2.81
SSVS	4.72	3.80	5.40	7.11	1.22	<b>3.12</b>	2.80	6.14	2.61	2.80
	<u>FORECAST HORIZON <math>h = 3</math>:</u>									
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	4.53	3.92	<b>6.25</b>	<b>7.22</b>	<b>1.35</b>	2.45	<b>3.19</b>	<b>7.24</b>	<b>2.49</b>	<b>3.16</b>
BMixS	4.59	3.84	5.87	6.85	1.24	2.50	3.10	6.62	2.40	<b>3.16</b>
CC	3.76	3.24	3.99	5.42	1.33	2.35	2.79	4.85	2.22	3.06
SSSS	3.76	3.30	4.47	4.39	1.17	2.02	2.78	4.67	2.28	2.58
GLP	3.94	3.47	4.55	5.64	1.02	2.39	2.55	5.27	2.17	2.73
SSVS	<b>4.65</b>	<b>4.20</b>	5.86	6.65	1.24	<b>2.59</b>	3.09	6.45	2.27	2.96
	<u>FORECAST HORIZON <math>h = 6</math>:</u>									
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	<b>3.96</b>	<b>3.51</b>	<b>6.10</b>	<b>6.62</b>	<b>1.18</b>	<b>2.04</b>	<b>2.94</b>	<b>6.79</b>	2.08	<b>2.82</b>
BMixS	3.90	3.42	5.73	6.58	1.04	1.92	2.81	6.54	1.93	2.74
CC	3.22	2.85	3.80	4.81	1.13	1.90	2.55	4.34	1.94	2.64
SSSS	3.04	2.86	4.82	4.12	0.89	1.70	2.20	4.45	1.95	2.65
GLP	3.00	2.64	4.15	5.06	0.85	1.73	2.30	4.75	1.67	2.43
SSVS	3.77	3.40	5.40	6.24	0.97	1.90	2.86	6.20	<b>2.15</b>	2.47
	<u>FORECAST HORIZON <math>h = 12</math>:</u>									
	AT	BE	FI	FR	GR	IE	IT	NL	PT	ES
BFCS	<b>3.61</b>	3.14	<b>5.59</b>	<b>5.57</b>	0.91	1.41	<b>2.48</b>	<b>6.29</b>	1.32	2.29
BMixS	3.59	3.10	5.22	<b>5.57</b>	0.83	<b>1.43</b>	2.35	6.08	1.39	2.17
CC	2.99	2.71	3.52	4.37	<b>0.92</b>	1.36	2.10	4.43	1.35	2.18
SSSS	2.96	2.69	4.27	3.19	0.73	1.31	1.75	4.74	1.57	1.58
GLP	2.78	2.83	3.57	4.43	0.85	1.19	2.02	4.37	1.27	1.93
SSVS	3.58	<b>3.26</b>	4.74	5.52	0.83	1.39	2.16	5.64	<b>1.52</b>	<b>2.30</b>

Note: Entries are Average Predictive Likelihoods (APLs), and higher values signify better performance. Entries should be compared column-wise, that is, for each country compare the best performing model.

## 6 Conclusions

Given the increased need to model interactions among different economies or different financial markets (e.g. for stocks, exchange rates, or other assets), panel VARs are meant to become a major tool of empirical analyses and a very natural extension of the benchmark single-country VAR framework. There are, of course, other models for multi-country data such as factor models (Kose, Otrok and Whiteman, 2003) or Global VARs (Dees et al, 2007). However, such alternative methods impose shrinkage by projecting the data into a lower dimensional space. Factor models do this in a data-based way, while GVARs model weakly exogenous variables using weights obtained from bilateral trades between the countries involved in the dataset.

In contrast, the panel VAR approach is the only one that allows one to potentially

uncover all possible interdependencies and homogeneities among countries since all the original  $N \times G$  series ( $N$  countries,  $G$  macroeconomic variables) enter the VAR specification. The culprit of this increased flexibility is that panel VARs can be heavily parametrized. So instead of shrinking the dimension of the original data (as is the case with factor or GVAR models), in this paper I propose shrinkage priors for the autoregressive coefficients of the panel VAR. The kind of relationships that may hold among different countries motivate my choices of priors. In particular, I propose priors which restrict coefficients to be zero, while allowing unrestricted coefficients to be clustered in different directions. The empirical results clearly suggest the benefits of the proposed approach compared to traditional prior choices such as the Minnesota prior.

## References

- [1] Banbura, M., Giannone, D., Reichlin, L., 2010. Large Bayesian vector autoregressions. *Journal of Applied Econometrics* 25, 71-92.
- [2] Belmonte, M, Koop, G., Korobilis, D., 2014. Hierarchical Shrinkage in Time-Varying Parameter Models. *Journal of Forecasting* 33 (1), 80-94.
- [3] Canova, F., Ciccarelli, M., 2009. Estimating multicountry VAR models. *International Economic Review* 50, 929-959.
- [4] Canova, F., Ciccarelli, M., 2013. Panel vector autoregressive models: A survey. European Central Bank Working Paper 1507.
- [5] Carriero, A., Clark, T., Marcellino, M., 2011. Bayesian VARs: Specification choices and forecast accuracy. Federal Reserve Bank of Cleveland, Working Paper 1112.
- [6] Carriero, A., Kapetanios, G., Marcellino, M., 2009. Forecasting exchange rates with a large Bayesian VAR. *International Journal of Forecasting* 25, 400-417.
- [7] Dees, S., Di Mauro, F., Pesaran, M.H., Smith, V., 2007. Exploring the international linkages of the Euro area: A global VAR analysis. *Journal of Applied Econometrics* 22, 1-38.
- [8] Dunson, D. B., Herring, A. H. and Engel, S. M., 2008. Bayesian selection and clustering of polymorphisms in functionally related genes. *Journal of the American Statistical Association* 103, 534-546.
- [9] Gefang, D., 2013. Bayesian doubly adaptive elastic-net Lasso for VAR shrinkage. *International Journal of Forecasting* 30,1-11.
- [10] Gelman, A., 2006. Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis* 1, 515-533.
- [11] George, E., McCulloch, R., 1993. Variable selection via Gibbs sampling. *Journal of the American Statistical Association* 88, 881-889.
- [12] George, E., McCulloch, R., 1997. Approaches for Bayesian variable selection. *Statistica Sinica* 7, 339-373.
- [13] George, E., Sun, D., Ni, S., 2008. Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics* 142, 553-580.
- [14] Giannone, D., Lenza, M., Primiceri, G., 2014. Priors for the long run. manuscript
- [15] Giannone, D., Lenza, M., Primiceri, G., forthcoming. Prior selection for vector autoregressions. *Review of Economics and Statistics*.
- [16] Kadiyala, K. R., Karlsson, S., 1997. Numerical methods for estimation and inference in Bayesian VAR-models. *Journal of Applied Econometrics*, 12, 99-132.

- [17] Koop, G., 2013. Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics* 28, 177-203.
- [18] Koop, G., Korobilis, D., 2010. Bayesian multivariate time series methods for empirical macroeconomics. *Foundations and Trends in Econometrics* 3, 267-358.
- [19] Koop, G., Korobilis, D., 2015. Model uncertainty in panel vector autoregressions. *European Economic Review*, doi:10.1016/j.euroecorev.2015.09.006.
- [20] Korobilis, D., 2008. Forecasting in vector autoregressions with many predictors. *Advances in Econometrics* 23, 403-431.
- [21] Korobilis, D., 2013. VAR forecasting using Bayesian variable selection. *Journal of Applied Econometrics* 28, 204-230.
- [22] Kose, M. A., Otrok, C., Whiteman, C. H., 2003. International business cycles: World, region, and country-specific factors. *American Economic Review* 93(4), 1216-1239.
- [23] Litterman, R., 1986. Forecasting with Bayesian vector autoregressions – Five years of experience. *Journal of Business and Economic Statistics* 4, 25-38.
- [24] MacLehose, R. F., Dunson, D. B., Herring, A. H. and Hoppin, J. A., 2007. Bayesian methods for highly correlated exposure data. *Epidemiology* 18, 199-207.
- [25] Raftery, A., Madigan, D., Hoeting, J., 1997. Bayesian model averaging for linear regression models. *Journal of the American Statistical Association* 92, 179-191.

## A. Technical Appendix

Consider the parametrization of the PVAR of the form

$$Y_t = Z_t \alpha + \varepsilon_t, \quad (\text{A.1})$$

where  $Z_t = I_{NG} \otimes X_t$ ,  $X_t = Y_{t-1}$ ,  $\alpha = \text{vec}(A')$  is the  $K \times 1$  vector of all PVAR coefficients,  $K = 1, \dots, NG^2$ . The parameter vector of interest is now  $\alpha$ , but once we know this vector we can easily rearrange its elements to construct the original PVAR matrix  $A$ .

### A.1 Posterior inference in the PVAR using the Bayesian Factor Clustering and Selection (BFCS)

The Bayesian Factor Clustering and Selection prior has the following structure

$$\alpha_k | \gamma_k, \theta, \Sigma \sim (1 - \gamma_k) \delta_0(\alpha) + \gamma_k \Delta_k, \quad (\text{A.2})$$

$$\Delta | \theta, \Sigma \sim N(\Xi \theta, \Sigma \otimes \sigma^2 I) \quad (\text{A.3})$$

$$\theta \sim N(0, \underline{c} I) \quad (\text{A.4})$$

$$\gamma_k \sim \text{Bernoulli}(\pi), \quad (\text{A.5})$$

$$\pi \sim \text{Beta}(1, \underline{\varphi}). \quad (\text{A.6})$$

However, this structure implies the following specification for the vector of PVAR coefficients  $\alpha$

$$\alpha = \Gamma \times (\Xi \theta) + v, \quad (\text{A.7})$$

where  $v \sim N(0, \Sigma \otimes \sigma^2 I)$  and  $\Gamma$  is a  $K \times K$  diagonal matrix with element  $\Gamma_{ii} = \gamma_i$ ,  $i = 1, \dots, K$ . Here I follow the recommendation of Canova and Ciccarelli (2009) and use the exact decomposition for  $\alpha$ , observed without error. This is the case where  $\sigma^2 = 0$ .

#### Gibbs sampling algorithm for the BFCS algorithm

1. Sample  $\theta$  from

$$(\theta | -) \sim N(E_\theta, V_\theta), \quad (\text{A.8})$$

where  $E_\theta = V_\alpha \left[ \tilde{Z}' (I \otimes \tilde{\Sigma})^{-1} Y \right]$  and  $V_\theta = \left[ \underline{c}^{-1} I + \tilde{Z}' (I \otimes \tilde{\Sigma})^{-1} \tilde{Z} \right]^{-1}$ , where  $\tilde{Z} = Z \times \Gamma \times \Xi$  and  $\tilde{\Sigma} = (I + \sigma^2 Z' Z) \Sigma$ .

2. Recover  $\alpha$  from

$$(\alpha | -) \sim N(\Gamma \times (\Xi \theta), \Sigma \otimes \sigma^2 I). \quad (\text{A.9})$$

3. Sample  $\gamma_k | \gamma_{-k}$ , where  $\gamma_{-k}$  denotes the vector  $\gamma$  with its  $k$ -th element removed, from

$$(\gamma_k | \gamma_{-k}, -) \sim \text{Bernoulli}(\omega_k), \quad (\text{A.10})$$

where  $\omega_k = \frac{l_{1k}}{l_{0k} + l_{1k}}$ , and  $l_{0k} = p(Y | \alpha_k, \gamma_{-k}, \gamma_k = 0) \pi$ ,  $l_{1k} = p(Y | \alpha_k, \gamma_{-k}, \gamma_k = 1) (1 - \pi)$ . Note

that evaluation of  $p(Y|\alpha_k, \gamma_{-k}, \gamma_k = 0)$  and  $p(Y|\alpha_k, \gamma_{-k}, \gamma_k = 1)$  is computationally costly, and can also be subject to overflow/underflow problems. In very large models, one could use an approximate algorithm and update all  $\gamma_k$  at once (not conditional on  $\gamma_{-k}$ ), but such an approximation should be applied with caution. In this case one can evaluate all  $l_{0k} \approx N(\alpha_k|0, 1e-8)\pi$  and  $l_{1k} \approx N(\alpha_k|0, \underline{c})(1-\pi)$  for all  $k = 1, \dots, K$  at once (i.e. without the need of a for loop), where  $N(x|a, b)$  denotes the Normal density with mean  $a$  and variance  $b$  evaluated at the observations  $x$ .

4. Sample  $\pi$  from

$$(\pi|-) \sim \text{Beta}\left(1 + \sum \gamma_k, \underline{\varphi} + \sum (1 - \gamma_k)\right). \quad (\text{A.11})$$

5. Sample  $\Sigma$  conditional on  $\alpha$  using standard expressions (see e.g. Koop, 2003).

## A.2 Posterior inference in the PVAR using the Bayesian Mixture Shrinkage (BMixS) prior

The Bayesian Mixture Shrinkage (BMixS) prior has the following hierarchical structure

$$\alpha_k | \mu_k, \tau_k^2 \sim N(\mu_k, \tau_k^2), \quad (\text{A.12})$$

$$(\mu_k, \tau_k^{-2}) | \pi \sim \pi \delta_0(\alpha) \times \delta_{1e+10}(\tau^{-2}) + (1 - \pi) F, \quad (\text{A.13})$$

$$F \sim DP(\underline{\theta} F_0), \quad (\text{A.14})$$

$$F_0 \sim N(0, \underline{\lambda}^2) \times \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right), \quad (\text{A.15})$$

$$\pi \sim \text{Beta}(1, \underline{\varphi}). \quad (\text{A.16})$$

Given  $C_\alpha$  mixture components, the equivalent stick breaking representation of this prior is

$$\alpha_k \sim N(\tilde{\mu}_l, \tilde{\tau}_l^2), \quad k = 1, \dots, K, \quad l = 1, \dots, C_\alpha, \quad (\text{A.17})$$

$$(\tilde{\mu}_l, \tilde{\tau}_l^{-2}) \sim w_0 \delta_0(\alpha) \times \delta_{1e+10}(\tau^{-2}) + \sum_{l=2}^{C_\alpha} w_l N(0, \underline{\lambda}^2) \times \text{Gamma}\left(\frac{1}{2}, \frac{1}{2}\right), \quad (\text{A.18})$$

where  $w_0 = \pi$  and  $w_l = \omega_l \prod_{h < l} (1 - \omega_h)$  with  $\omega_l \sim \text{Beta}(1, \underline{\varphi})$ ,  $l = 2, \dots, C_\alpha$ . Here it greatly simplifies computation if we pre-fix the maximum number of clusters  $C_\alpha$ ; otherwise a Metropolis-Hastings step is required in order to sample the number of cluster configurations. We don't need to be very informative and set  $C_\alpha$  to a very low value (e.g. one or two clusters), but it generally helps if  $C_\alpha \ll K$ .

### Gibbs sampling algorithm for the BMixS algorithm

1. Sample  $\alpha$  from

$$(\alpha|-) \sim N(E_\alpha, V_\alpha), \quad (\text{A.19})$$

where  $E_\alpha = V_\alpha \left[ T^{-1} M + Z' (I \otimes \Sigma)^{-1} Y \right]$  and  $V_\alpha = \left[ T^{-1} + Z' (I \otimes \Sigma)^{-1} Z \right]^{-1}$ , with  $T = \text{diag}(\tau_1^2, \dots, \tau_K^2)$  and  $M = (\mu_1, \dots, \mu_K)'$ .

2. Sample  $\tilde{\mu}_l$ ,  $l = 1, \dots, C_\alpha$ , from

$$(\tilde{\mu}_l | -) \sim \begin{cases} \delta_0(\tilde{\mu}_l), & \text{if } l = 1 \\ N(E_\mu, V_\mu), & \text{otherwise} \end{cases}, \quad (\text{A.20})$$

where  $\delta_0(\tilde{\mu}_l)$  is the Dirac delta at zero for parameter  $\tilde{\mu}_l$ ,  $E_\mu = V_\mu \left( \sum_{j=1, j \in l}^K \alpha_j \tau_j^{-2} \right)$ , and  $V_\mu = \left( 1/\lambda^2 + \sum_{j=1, j \in l}^K \tau_j^{-2} \right)^{-1}$ .

3. Sample  $\tilde{\tau}_l^2$ ,  $l = 1, \dots, C_\alpha$ , from

$$(\tilde{\tau}_l^2 | -) \sim \begin{cases} \delta_{10^{10}}(\tilde{\tau}_l), & \text{if } l = 1 \\ iGamma\left(\frac{1}{2} + n_l, \frac{1}{2} + \sum_{k=2, k \in l}^{C_\alpha} (\alpha_k - \mu_l)^2\right), & \text{otherwise} \end{cases}, \quad (\text{A.21})$$

where  $n_l$  is the number of coefficients (elements) that belong in cluster  $l$ .

4. Sample  $w_l$  from

$$(w_l | -) \equiv \omega_l \prod_{h < l} (1 - w_h), \quad (\text{A.22})$$

where  $\omega_l$  is sampled from

$$(w_l | -) \sim Beta\left(n_l + 1, C_\alpha - \sum_{j=1}^l n_j + \underline{\varphi}\right). \quad (\text{A.23})$$

5. Sample  $\Sigma$  conditional on  $\alpha$  using standard expressions (see e.g. Koop, 2003).

### A.3 Posterior inference in the PVAR using the Stochastic Search Specification Selection (S<sup>4</sup>) prior of Koop and Korobilis (2015)

Following the main text, the VAR for country  $i$  is

$$y_{it} = A_{i1}y_{1,t-1} + \dots + A_{ii}y_{i,t-1} + \dots + A_{iN}y_{N,t-1} + \varepsilon_{it}, \quad (\text{A.24})$$

and the compact form of the PVAR (in matrix form) is

$$Y = XA + \varepsilon,$$

where  $Y = (Y_1', \dots, Y_T)'$ ,  $X = (X_1', \dots, X_T)'$  and  $\varepsilon = (\varepsilon_1', \dots, \varepsilon_T)'$ . Note that for notational simplicity I have defined  $X_t = Y_{t-1}$ , however, the formulae below remain the same if we generalize to  $X_t = (I, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}, W_{t-1}, G_{t-1})$  where  $W_t$  are country-specific exogenous variables and  $G_t$  are global exogenous variables.

We pre-specify two groups of panel-type restrictions: dynamic interdependencies (DI), and cross-sectional homogeneity (CSH). The existence (or absence) of DI can be tested using the prior

$$vec(A_{ij}) | \gamma_{ij}^{DI} \sim (1 - \gamma_{ij}^{DI}) N(0, \underline{\tau}_1^2 \times I_{G^2}) + \gamma_{ij}^{DI} N(0, \underline{\tau}_2^2 \times I_{G^2}), \quad (\text{A.25})$$

$$\gamma_{ij}^{DI} | \pi_{ij}^{DI} \sim \text{Bernoulli}(\pi_{ij}^{DI}), \quad \forall j \neq i, \quad (\text{A.26})$$

$$\pi_{ij}^{DI} \sim \text{Beta}(1, \underline{\varphi}), \quad (\text{A.27})$$

while the existence (or absence) of cross-sectional homogeneity can be tested using the prior

$$vec(A_{ii}) | \gamma_{ij}^{CSH} \sim (1 - \gamma_{ij}^{CSH}) N(A_{jj}, \underline{\xi}_1^2 \times I_{G^2}) + \gamma_{ij}^{CSH} N(A_{jj}, \underline{\xi}_2^2 \times I_{G^2}), \quad (\text{A.28})$$

$$\gamma_{ij}^{CSH} | \pi_{ij}^{DI} \sim \text{Bernoulli}(\pi_{ij}^{CSH}), \quad \forall j \neq i, \quad (\text{A.29})$$

$$\pi_{ij}^{CSH} \sim \text{Beta}(1, \underline{\varphi}). \quad (\text{A.30})$$

We take the hyperparameters with an underscore ( $\underline{\varphi}, \underline{\tau}_1^2, \underline{\tau}_2^2, \underline{\xi}_1^2, \underline{\xi}_2^2$ ) as given, that is, prespecified by the researcher. Additionally, as explained in detail in Koop and Korobilis

(2015) we define a matrix  $\Gamma = \prod_{i=1}^{N-1} \prod_{j=i+1}^N \Gamma_{i,j}$ , where  $\Gamma_{i,j}$  are  $K \times K$  matrices constructed

using the CSH restriction indicators  $\gamma_{ij}^{CSH}$ . First note that  $\gamma_{ij}^{CSH} = 0$  implies that countries  $i$  and  $j$  have similar coefficients (i.e. the homogeneity restriction  $A_{ii} \approx A_{jj}$  holds), and the opposite is true when  $\gamma_{ij}^{CSH} = 1$ . The matrix  $\Gamma_{i,j}$  can be created starting from the identity matrix (i.e. ones on the diagonal zeros elsewhere) with the restriction that its  $\{i, i\}$  diagonal element is equal to  $\gamma_{ij}^{CSH}$  and its  $\{i, j\}$  non-diagonal element is equal to  $(1 - \gamma_{ij}^{CSH})$  for  $i < j$ . Therefore, each of the possible  $N(N-1)/2$  matrices  $\Gamma_{i,j}$  allow us

to impose on the PVAR coefficients the CSH restriction between countries  $i$  and  $j$ , and their product, which is the matrix  $\Gamma = \prod_{i=1}^{N-1} \prod_{j=i+1}^N \Gamma_{i,j}$ , allows us to index all  $2^{N(N-1)/2}$

possible CSH restrictions among the  $N$  countries. Therefore, if  $\mu_\alpha$  denotes the posterior mean of the *unrestricted* vectorized PVAR coefficients (i.e. using a noninformative prior),

then  $\tilde{\mu}_\alpha = \Gamma \mu_\alpha = \prod_{i=1}^{N-1} \prod_{j=i+1}^N \Gamma_{i,j} \mu_\alpha$  is simply the  $K \times 1$  vector of posterior means of the

PVAR coefficients with the cross-sectional homogeneity restrictions imposed; see Koop and Korobilis (2015) for further details.

### Gibbs sampler algorithm for the SSSS algorithm

1. Sample  $vec(A)$  from

$$(vec(A) | -) \sim N(\Gamma \times \mu_\alpha, D_\alpha), \quad (\text{A.31})$$

where  $D_\alpha = (\Sigma^{-1} \otimes X'X + V'V)^{-1}$  and  $\mu_\alpha = D_\alpha [(\Sigma^{-1} \otimes X'X) \alpha_{OLS}]$ , where  $\alpha_{OLS}$  is the OLS estimate of  $\alpha$ , and  $V$  is a diagonal matrix which has its respective diagonal block of  $G^2$  elements equal to  $\underline{\tau}_1^2 \times \mathbf{1}$  if  $\gamma_{ij}^{DI} = 0$  or equal to  $\underline{\tau}_2^2 \times \mathbf{1}$  if  $\gamma_{ij}^{DI} = 1$ , where  $\mathbf{1}$  is a  $G^2 \times 1$  vector of ones.

2. Sample  $\gamma_{ij}^{DI}$  from

$$(\gamma_{ij}^{DI} | -) \sim \text{Bernoulli}(\omega_{ij}^{DI}), \quad (\text{A.32})$$

where  $\omega_{ij}^{DI} = \frac{u_{2,ij}}{u_{1,ij} + u_{2,ij}}$  with  $u_{1,ij} = N(\text{vec}(A_{ij}) | \mathbf{0}, \underline{\tau}_1^2 I_{G^2}) \pi_{ij}^{DI}$  and  $u_{2,ij} = N(\text{vec}(A_{ij}) | \mathbf{0}, \underline{\tau}_2^2 I_{G^2}) (1 - \pi_{ij}^{DI})$ , and  $N(x|a, b)$  denotes the Normal density with mean  $a$  and variance  $b$  evaluated at the observations  $x$ .

3. Sample  $\pi_{ij}^{DI}$  from

$$(\pi_{ij}^{DI} | -) \sim \text{Beta} \left( 1 + \sum \gamma_{ij}^{DI}, \underline{\varphi} + \sum (1 - \gamma_{ij}^{DI}) \right). \quad (\text{A.33})$$

4. Sample  $\gamma_{ij}^{CSH}$  from

$$(\gamma_{ij}^{CSH} | -) \sim \text{Bernoulli} (\omega_{ij}^{CSH}), \quad (\text{A.34})$$

where  $\omega_{ij}^{CSH} = \frac{v_{2,ij}}{v_{1,ij} + v_{2,ij}}$  with  $v_{1,ij} = N(\text{vec}(A_{ii}) | \text{vec}(A_{jj}), \underline{\xi}_1^2 I_{G^2}) \pi_{ij}^{CSH}$  and  $v_{2,ij} = N(\text{vec}(A_{ii}) | \text{vec}(A_{jj}), \underline{\xi}_2^2 I_{G^2}) (1 - \pi_{ij}^{CSH})$ , and  $N(x|a, b)$  denotes the Normal density with mean  $a$  and variance  $b$  evaluated at the observations  $x$ .

5. Sample  $\pi_{ij}^{CSH}$  from

$$(\pi_{ij}^{CSH} | -) \sim \text{Beta} \left( 1 + \sum \gamma_{ij}^{CSH}, \underline{\varphi} + \sum (1 - \gamma_{ij}^{CSH}) \right). \quad (\text{A.35})$$

6. Sample  $\Sigma$  conditional on  $A$  using standard expressions (see e.g. Koop, 2003).

#### A.4 Posterior inference in other models examined in this paper

For the SSVS prior for VAR developed by George, Sun and Ni (2008), see the Appendix of their paper. This prior is similar to the SSSS prior with the exception that it does not distinguish between DIs and CSHs, rather it treats restrictions on each VAR coefficient uniformly (meaning that each VAR coefficient has equal prior weight of importance and only the data will determine which coefficients should be shrunk to zero). This prior can be written as

$$\alpha_k | \gamma_k \sim (1 - \gamma_k) N(0, \underline{\tau}_1^2) + \gamma_k N(0, \underline{\tau}_2^2), \quad (\text{A.36})$$

$$\gamma_k \sim \text{Bernoulli}(\underline{\pi}_k) \quad (\text{A.37})$$

where in this paper I set  $\underline{\tau}_1^2 = 0.001$ ,  $\underline{\tau}_2^2 = 4$  and  $\underline{\pi}_k = 0.5$  for all  $k$ .

In the case of the Giannone et al.(forthcoming) prior, I use the code provided by D. Giannone (<http://homepages.ulb.ac.be/~dgiannon/GLPreplicationWeb.zip>) and I work with their default settings. Note that this code by default allows to work with posterior medians and analytical formulas. In order to have better comparability with all other priors in this paper, I allow MCMC updates for this prior in order to account for approximation error when using the Gibbs sampler.

The prior of Canova and Ciccareli (2009) can be obtained as a special case of the BFCS prior, by setting  $\gamma_k = 1$  for all  $k = 1, \dots, K$  and by not updating this parameter from its posterior.