Improving inference and forecasting in VAR models using cross-sectional information*

Jan Prüser[a] and Boris Blagov[b]

[a]TU Dortmund† [b]RWI – Leibniz Institute for Economic Research‡

July 23, 2022

Abstract

We propose a prior for VAR models that exploits the panel structure of macroeconomic time series while also providing shrinkage towards zero to address overfitting concerns. The prior is flexible as it detects shared dynamics of individual variables across endogenously determined groups of countries. We demonstrate the usefulness of our approach via a Monte Carlo study and use our model to capture the hidden homo- and heterogeneities of the euro area member states. Combining pairwise pooling with zero shrinkage delivers sharper parameter inference that improves point and density forecasts over only zero shrinkage or only pooling specifications, and helps with structural analysis by lowering the estimation uncertainty.

Keywords: BVAR, Shrinkage, Forecasting, Structural analysis

JEL classification: C11, C32, C53, E37

---

*Jan Prüser gratefully acknowledges the support of the German Research Foundation (DFG, 468814087).

†Corresponding Author: Fakultät Statistik, 44221 Dortmund, Germany, e-mail: prueser@statistik.tu-dortmund.de

‡Hohenzollernstr. 1–3, 45128 Essen, Germany, e-mail: boris.blagov@rwi-essen.de

1 Introduction

The popularity of Bayesian VAR (BVAR) models for forecasting and structural analysis has risen throughout the past decade. Due to their rich parametrization VARs can fit the data well. However, macroeconomic time series are rather short leading to the risk of overfitting and imprecise inference. BVARs using the Minnesota prior have a long and successful history to address this problem (e.g., Doan et al. (1984), Litterman (1986), Sims and Zha (1998), Bańbura et al. (2010), Giannone et al. (2015), Koop (2013), Carriero et al. (2016), Korobilis and Pettenuzzo (2019), Huber and Feldkircher (2019), Chan (2020) and Cross et al. (2020)). The Minnesota prior shrinks most VAR coefficients towards zero and provides regularization from a frequentist viewpoint. In this paper we consider combining the popular Minnesota prior with additional cross-country information to improve parameter inference.

Assume we have data for a set of countries which could be homogeneous or heterogeneous. If the countries are heterogeneous, we would estimate individual BVARs with each dataset. However, short time series are problematic for precise parameter inference, and shrinkage towards zero is the main strategy to deal with this issue. Alternatively, if we are confident in the similarity of the countries, we could take a panel VAR route by imposing similar dynamics across all variables and thus improving estimation inference by utilizing the cross-country dimension. The more alike countries are, the more increasing their number will compensate for the individual time series' lengths and the sharper the parameter inference will be. However, in practice, as the country dimension grows, heterogeneity naturally increases. It could be that groups of countries share similar characteristics in some way, while others do not. For example, trading partners, nations with similar legal systems or fiscal rules, or countries with cross-border policies such as the common monetary policy in the euro area could all give rise to clustering. Importantly, this clustering does not have to be across all variables for all countries. The euro area is a perfect example of such data set, where individual member states are highly integrated across some dimensions, such as the financial markets, while starkly different in others, such as the labour markets. Furthermore, prices of tradable goods in the common market are expected to share similar dynamics, while at the same time member states that specialize in particular industries could share production booms and busts not present in others. The advanced economies, such as the G-7 group, are another example that could give rise to clustering, while sharing many characteristics. For example, Canada and the US are expected to be more tightly integrated with each other as they are with Germany, France, and/or Italy who themselves could build clusters across different dimensions.

At one end of the spectrum, single-country VARs provide a good setup for capturing individual characteristics and heterogeneity while panel VARs shine when the homogeneity is high. In this paper we provide an alternative that profits from both dimensions in the form of a new prior for BVARs that combines parameter shrinkage towards zero while also exploiting the cross-country dimensions in a flexible manner. We propose to estimate BVARs for each country individually using standard tools such as the Minnesota prior to achieve parsimony while also searching the country space for similar dynamics in the form of pooling pairwise VAR coefficients of individual variables across countries. Thus, our proposed algorithm is able to detect the degree of similarity of individual coefficients across countries and identifies country pairs, and therefore clusters, with similar dynamics without imposing any structure ex-ante.

Taken together our main contribution is to propose a flexible pooling prior which allows for parameter pooling across both a country dimension (e.g. countries have similar dynamics across all variables) and/or variable dimension (dynamics of two or more variables across countries are alike) and at the same time achieves parsimony by providing zero shrinkage of the parameters using the popular Minnesota prior. Hence, from a frequentist view point our prior penalizes both deviations from zero as well as deviations between the coefficients across countries. How much such deviations are penalized is determined by a set of hyperparameters. They play an important role in our model and their choice is not straightforward. High hyperparameter values lead to a low degree of penalization (i.e. to a rather flat prior) which does not offer protection against overfitting. Similarly, low values lead to a high degree of penalization (i.e. to a highly informative prior), which suppresses important signals from the data. We address the problem of choosing these hyperparameters in a hierarchical fashion by providing a two step estimation strategy, which leads to a balanced solution between the two extremes. Furthermore, this estimation lets us detect which countries pairs and/or which variables share similar dynamics without the need to impose any structure ex-ante, making the model highly flexible.

We perform Monte Carlo simulations in order to investigate the frequentist estimation properties of our approach. In total, we consider three different data generating processes (DGPs). The first assumes that three countries are dissimilar to each other, the second assumes that two of the three countries share similar dynamics across all variables and the last assumes that only some individual's variables are similar across countries. We show that our estimation strategy provides sensible estimates for all three different scenarios. Hence, in contrast to panel specifications, we do not need to make assumptions about the structure or homogeneity between economies prior to taking our model to the data as our approach can adapt to different scenarios. Examining the convergence properties of our model, we find that the parameter estimates are perfectly reasonable in normal sample sizes and naturally the parameters are more precisely estimated as the sample size grows.

The model should be well suited to capture the homo- and heterogeneities of the euro area. Our dataset consists of ten of the largest euro area member states that approximate together close to 95% of euro area GDP. We evaluate the forecasting performance of our model against single country VARs with different variants of the Minnesota prior and different panel VARs with pooling priors. The forecasting exercise reveals that combining zero shrinkage with pooling leads to improvements in forecasting performance both in terms of point and density forecasting, outperforming single country VARs with Minnesota prior as well as different panel VAR specifications. In contrast, approaches which provide solely pooling perform relatively worse compared to single country VARs with the Minnesota prior. An exercise with a different dataset of the G-7 advanced economies reveals that relative forecasting performance of our approach is robust.

Furthermore, we showcase how the hierarchical estimation of the hyperparameters may be used for the analysis of the similarities and dissimilarities of the lagged structure of production, prices, and interest rates across the euro area and the G-7 economies. We uncover distinctive production dynamics in Spain and the United Kingdom.

Finally, we demonstrate the usefulness of the proposed method for structural analysis.

The improved inference of our pooling prior provides sharper estimation leading to much narrower probability intervals. For example, a single country setting for the individual euro area member states shows that contractionary monetary policy shock does not lead to fall in production in a statistically meaningful way - the average response of production drops but the uncertainty surrounding the estimates is large and include the zero line, even with 68% error bands. This is true for both a recursive identification as well as a sign restriction approach. In contrast, using the pooling prior combined with zero shrinkage estimates the effects of an interest rate hike more precisely. In this case the probability intervals of the response of output does not include the zero line in all periods.

On the methodological front this paper contributes to the literature on developing pooling prior for VARs. For example Zellner and Hong (1989) and Jarocinski (2010) propose a prior which pools all VAR coefficients towards a common mean. Closely related to this these papers are a factor approach of Canova and Ciccarelli (2009) and a pooled mean group estimator of Pesaran et al. (1999). Our prior is more flexible by allowing that only some subsets of countries share similar dynamics as in Koop and Korobilis (2015). Koop and Korobilis (2015) provide an algorithm which can construct restricted PVARs, which imposes homogeneity between all possible pairs of different countries if empirically warranted. However, their prior cannot account for the case in which only a subset of VAR coefficients is similar between countries. Our prior allows for this. Furthermore, these papers propose priors which do not provide shrinkage towards zero. Korobilis (2016) introduces priors which can shrink the VAR coefficients either towards zero or towards a common cluster. Our prior is more flexible by allowing for homogeneous and heterogeneous dynamics simultaneously and for both pooling and shrinkage at the same time. We find this flexibility to be empirically important. In particular, we show that it is beneficial to have both pooling and shrinkage at the same time and only pooling may not be enough to alleviate overfitting concerns.

Last but not least, the likelihood based approach taken in this paper is flexible and modular. For example, one can extend our approach by allowing for dynamic interactions from one country to another as in e.g. Koop and Korobilis (2015) or Feldkircher et al. (2022). Another interesting extension would be to allow for outliers to handle the sequence of extreme observations due to Covid-19. These outliers can have quite a big influence on the parameter estimates. Our model can be combined with approaches from Carriero et al. (2021) or Prüser (2021) in order to downweigh such observations.

The remainder of this paper is organized as follows. Section 2 lays out and discusses the econometric framework. Section 3 contains a simulation study. Section 4 applies the model to a real dataset of the euro area for forecasting and structural analysis. Section 5 concludes.

2 Econometric Framework

In this section, we discuss how we stack many individual country VARs into one large VAR and, based on this representation, we introduce our prior. Finally, we discuss how to estimate the parameters which determine the strength of our prior - the hyperparameters - as to endogenously determine the similarity across countries.

2.1 The VAR model

Let $\boldsymbol{y}_{nt}$ be a $G \times 1$ vector of endogenous variables for country n at time t. Each country VAR can be written as

$$\boldsymbol{y}_{nt} = \sum_{p=1}^{P} \gamma_{np} \boldsymbol{y}_{n,t-p} + \boldsymbol{\epsilon}_{nt}, \qquad \boldsymbol{\epsilon}_{nt} \sim N(\mathbf{0}, \boldsymbol{\Sigma}_n), \tag{1}$$

and in more compact form

$$\boldsymbol{Y}_n = \boldsymbol{x}_n \boldsymbol{\Gamma}_n + \boldsymbol{U}_n, \tag{2}$$

where $\boldsymbol{Y}_n = (\boldsymbol{y}_{nt}, \ldots, \boldsymbol{y}_{nT})'$, the t-row of $\boldsymbol{x}_n$ is given by $(\boldsymbol{y}'_{n,t-1}, \ldots, \boldsymbol{y}'_{n,t-p})$ and $\boldsymbol{\Gamma}_n = (\boldsymbol{\gamma}_{n1}, \ldots, \boldsymbol{\gamma}_{nP})'$. We omit the constant for simplicity. In practice this is implemented by demeaning our data prior to the estimation and adding back the mean afterwards. By vectorizing equation (2) we get

$$\overline{\boldsymbol{y}}_n = (\boldsymbol{I}_G \otimes \boldsymbol{x}_n) \overline{\boldsymbol{\gamma}}_n + \overline{\boldsymbol{u}}_n, \qquad \overline{\boldsymbol{u}}_n \sim N(\mathbf{0}, \boldsymbol{\Sigma}_n \otimes \boldsymbol{I}_T), \tag{3}$$

with $\overline{\boldsymbol{y}}_n = \text{vec}(\boldsymbol{Y}_n)$ and $\overline{\gamma}_n$ a $K \times 1$ vector containing all VAR coefficients, where $K = G^2P$. Define $\tilde{\boldsymbol{y}} = \text{vec}(\overline{\boldsymbol{y}}_1, \ldots, \overline{\boldsymbol{y}}_N)$ and $\tilde{\boldsymbol{x}} = (\boldsymbol{I}_G \otimes \boldsymbol{x}_1, \ldots, \boldsymbol{I}_G \otimes \boldsymbol{x}_N)$ we can write all country VARs as one large VAR

$$\tilde{\boldsymbol{y}} = (\boldsymbol{I}_N \otimes \tilde{\boldsymbol{x}})\boldsymbol{\beta} + \tilde{\boldsymbol{u}}, \qquad \tilde{\boldsymbol{u}} \sim N(\boldsymbol{0}, \tilde{\boldsymbol{\Sigma}} \otimes \boldsymbol{I}_T). \tag{4}$$

We assume a conditional normal prior for $\boldsymbol{\beta} \sim N(\boldsymbol{0}, \boldsymbol{V}_\beta)$. Now we can use standard linear regression results to get the conditional posterior distribution of $\boldsymbol{\beta}$. The conditional posterior of $\boldsymbol{\beta}$ is $N(\hat{\boldsymbol{\beta}}, \boldsymbol{K}_\beta^{-1})$ with[1]

$$\hat{\boldsymbol{\beta}} = \boldsymbol{K}_\beta^{-1}((\boldsymbol{I}_N \otimes \tilde{\boldsymbol{x}})'(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_T))\tilde{\boldsymbol{y}}), \tag{5}$$

$$\boldsymbol{K}_\beta = \boldsymbol{V}_\beta^{-1} + (\boldsymbol{I}_N \otimes \tilde{\boldsymbol{x}})'(\tilde{\boldsymbol{\Sigma}}^{-1} \otimes \boldsymbol{I}_T)(\boldsymbol{I}_N \otimes \tilde{\boldsymbol{x}}), \tag{6}$$

Would the prior inverse covariance matrix $\boldsymbol{V}_\beta^{-1}$ be a diagonal matrix then the estimation of the large VAR would be identical to the estimation of many small individual country VAR models. Instead, we introduce a prior inverse covariance matrix with non-zero diagonal elements which pool the coefficients across individual country VARs together to exploit the panel structure of the data.

2.2 A combined shrinkage and pooling prior

Our main contribution is to propose a new prior for a set of country VARs. The new prior combines the Minnesota prior with a flexible pooling prior. While the Minnesota prior shrinks the VAR coefficients towards zero, the pooling prior pushes the coefficients of one country VAR towards the coefficients of the other country VARs. The Minnesota prior is given by

$$\overline{\gamma}_{nj} = N(0, V_{nj}^{\text{Min}}), \tag{7}$$

with $j = 1, \ldots, K$ and

[1] We note that draws from the high-dimensional distribution $N(\hat{\boldsymbol{\beta}}, \boldsymbol{K}_\beta^{-1})$ can be obtained efficiently without inverting any large matrices; see, e.g., Chan (2021) for computational details.

$$V_{nj}^{\text{Min}} = \begin{cases} \frac{\kappa_{1,n}^2}{p^2}, & \text{for own lags} \\ \frac{\kappa_{2,n}^2}{p^2}, & \text{for cross-variable lags} \end{cases}. \tag{8}$$

The hyperparameters $\kappa_{1,n}$ and $\kappa_{2,n}$ control the informativeness of the prior. We follow Giannone et al. (2015) and estimate them by using a hierarchical prior which we introduce in the next section. Cross et al. (2020) compare the forecasting performance of the Minnesota prior with a range of other proposed priors in the literature and find that the Minnesota prior remains a solid choice. In our forecasting exercise, we also consider another more flexible variant of the Minnesota prior proposed by Chan et al. (2021).

Our prior provides both shrinkage towards zero as well as pooling towards the other country VARs coefficients:

$$\log p(\boldsymbol{\beta}) \propto \sum_{n=1}^{N}\sum_{j=1}^{K}\frac{\overline{\gamma}_{nj}^2}{V_{nj}^{\text{Min}}} + \sum_{j=1}^{K}\sum_{i=1}^{N-1}\sum_{m=i+1}^{N}\frac{(\beta_{j+K(i-1)} - \beta_{j+K(m-1)})^2}{\lambda_{i,m,j}^2\tau_{i,m}^2}, \tag{9}$$

$$= \sum_{n=1}^{N}\sum_{j=1}^{K}\overline{\gamma}_{nj}^2 z_{j,n} + 2\sum_{j=1}^{K}\sum_{i=1}^{N-1}\sum_{m=i+1}^{N}\beta_{j+K(i-1)}\beta_{j+K(m-1)}\kappa_{i,m,j}, \tag{10}$$

with $z_{j,n} = \left(\frac{1}{V_{nj}^{\text{Min}}} + \frac{1}{\lambda_{1,n,j}^2\tau_{1,n}^2} + \cdots + \frac{1}{\lambda_{n-1,n,j}^2\tau_{n-1,n}^2} + \frac{1}{\lambda_{n+1,n,j}^2\tau_{n+1,n}^2} + \cdots + \frac{1}{\lambda_{N,n,j}^2\tau_{N,n}^2}\right)$ and $\kappa_{i,m,j} = \frac{-1}{\lambda_{i,m,j}^2\tau_{i,m}^2}$. The two hyperparameters $\lambda_{i,m,j}^2$ and $\tau_{i,m}^2$ are crucial as they influence how much the VAR coefficients of the different countries will be pooled together. If $\lambda_{i,m,j}^2\tau_{i,m}^2$ goes to 0, the VAR coefficients will become identical across countries. On the other hand, if $\lambda_{i,m,j}^2\tau_{i,m}^2$ goes to ∞ the prior influence vanishes. We discuss the estimation and further interpretation of these pooling hyperparameters in the next section. Equation (10) reveals that the prior is normal $\boldsymbol{\beta} \sim N(\mathbf{0}, \boldsymbol{V}_\beta)$ and $\boldsymbol{V}_\beta^{-1}$ has the following structure

$$
\begin{pmatrix}
z_{1,1} & \dots & 0 & \kappa_{1,2,1} & \dots & 0 & & \kappa_{1,N-1,1} & \dots & 0 & \kappa_{1,N,1} & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & z_{K,1} & 0 & \dots & \kappa_{1,2,K} & \cdots & 0 & \dots & \kappa_{1,N-1,K} & 0 & \dots & \kappa_{1,N,K} \\
\kappa_{1,2,1} & \dots & 0 & z_{1,2} & \dots & 0 & & \kappa_{2,N-1,1} & \dots & 0 & \kappa_{2,N,1} & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & \kappa_{1,2,K} & 0 & \dots & z_{K,2} & & 0 & \dots & \kappa_{2,N-1,K} & 0 & \dots & \kappa_{2,N,K} \\
 & & & \vdots & & & \ddots & & & \vdots & & & \\
\kappa_{1,N-1,1} & \dots & 0 & \kappa_{2,N-1,1} & \dots & 0 & & z_{1,N-1} & \dots & 0 & \kappa_{N-1,N,1} & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & \kappa_{1,N-1,K} & 0 & \dots & \kappa_{2,N-1,K} & \cdots & 0 & \dots & z_{K,N-1} & 0 & \dots & \kappa_{N-1,N,K} \\
\kappa_{1,N,1} & \dots & 0 & \kappa_{2,N,1} & \dots & 0 & & \kappa_{N-1,N,1} & \dots & 0 & z_{1,N} & \dots & 0 \\
\vdots & \ddots & \vdots & \vdots & \ddots & \vdots & & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\
0 & \dots & \kappa_{1,N,K} & 0 & \dots & \kappa_{2,N,K} & & 0 & \dots & \kappa_{N-1,N,K} & 0 & \dots & z_{K,N}
\end{pmatrix}.
$$

Hence, $z_{j,n}$ are the diagonal elements of $\boldsymbol{V}_{\beta}^{-1}$ and $\kappa_{i,m,j}$ are the non-zero off diagonal elements.

2.3 Estimation of hyperparameter

The choice of hyperparameters is not straightforward. While overly high values lead to a rather flat prior which does not offer protection against overfitting, overly low values will suppress important signals from the data. The estimation of the hyperparameters of our prior may lead to a balanced solution between the two extremes. We follow Prüser (2022) and assume half Cauchy prior for the hyperparameters of the Minnesota prior

$$\kappa_{1,n} \sim C^{+}(0,1), \tag{11}$$

$$\kappa_{2,n} \sim C^{+}(0,1). \tag{12}$$

In order to obtain conditional posterior distributions for each of the hyperparameters, we follow Makalic and Schmidt (2016) and exploit the scale mixture representation of the half-Cauchy distribution. The scalar mixture representation stems from the fact that, if X and w are random variables such that $X^2|w \sim IG(\frac{1}{2}, \frac{1}{w})$ and $w \sim IG(\frac{1}{2}, 1)$, then $X \sim C^{+}(0,1)$. Since the Gaussian and inverse gamma distributions are conjugate distributions, it is straightforward to derive the posteriors of the hyperparameters. The conditional posterior distributions are given by:

$$\kappa_{1,n}^2 \sim IG\left(\frac{PG+1}{2}, \frac{1}{v_{\kappa_{1,n}^2}} + 0.5\sum_{p=1}^{P}\sum_{j\in \boldsymbol{I}_o}\frac{\overline{\gamma}_{nj}^2}{p^2}\right), \tag{13}$$

$$\kappa_{2,n}^2 \sim IG\left(\frac{PG(G-1)+1}{2}, \frac{1}{v_{\kappa_{2,n}^2}} + 0.5\sum_{p=1}^{P}\sum_{j\in \boldsymbol{I}_c}\frac{\overline{\gamma}_{nj}^2}{p^2}\right), \tag{14}$$

$$v_{\kappa_{1,n}^2} \sim IG\left(1, 1+\frac{1}{\kappa_{1,n}^2}\right), \tag{15}$$

$$v_{\kappa_{2,n}^2} \sim IG\left(1, 1+\frac{1}{\kappa_{2,n}^2}\right), \tag{16}$$

where $\boldsymbol{I}_o$ denotes the set of own lags and $\boldsymbol{I}_c$ of cross-variable lags. Gelman (2006) and Polson and Scott (2012) provide strong theoretical arguments for using the half-Cauchy distribution over an inverse-Gamma distribution for the scale parameters. Therefore, we also consider half-Cauchy prior for the hyperparameters of the pooling part of our prior

$$\tau_{i,m} \sim C^+(0,1), \tag{17}$$

$$\lambda_{i,m,j} \sim C^+(0,1). \tag{18}$$

The hyperparameter $\tau_{i,m}$ is country specific and shrinks all VAR coefficients of the country pair to each other. The local hyperparameter $\lambda_{i,m,j}$ is variable specific and can prevent individual VAR coefficients from being pooled together. We borrow this idea from the horseshoe prior, proposed by Carvalho et al. (2010), which provides shrinkage towards zero. The horseshoe prior is free of tuning parameters and has many appealing frequentist properties, see, e.g., Ghosh et al. (2016), Armagan et al. (2013) and van der Pas et al. (2014). We again follow Makalic and Schmidt (2016) and exploit the scale mixture representation of the half-Cauchy distribution. The conditional posterior distributions turn out to be

$$\tau_{i,m}^2 \sim IG\left(\frac{K+1}{2}, \frac{1}{v_{\tau_{i,m}^2}} + 0.5\sum_{j=1}^{K}\frac{(\beta_{j+K(i-1)} - \beta_{j+Km})^2}{\lambda_{i,m,j}^2}\right), \tag{19}$$

$$\lambda^2_{i,m,j} \sim IG\left(1, \frac{1}{v_{\lambda^2_{i,m,j}}} + 0.5\frac{(\beta_{j+K(i-1)} - \beta_{j+Km})^2}{\tau^2_{i,m}}\right), \tag{20}$$

$$v_{\tau^2_{i,m}} \sim IG\left(1, 1 + \frac{1}{\tau^2_{i,m}}\right), \tag{21}$$

$$v_{\lambda^2_{i,m,j}} \sim IG\left(1, 1 + \frac{1}{\lambda^2_{i,m,j}}\right). \tag{22}$$

$$\tag{23}$$

We have derived the conditional posterior of all model parameters.[2] Hence, we use a Gibbs sampler to draw all parameters from their joint posterior distribution. With one exception, we use fixed values for $\tau_{i,m}$ and $\lambda_{i,m,j}$. The reason is that the pooling hyperparameters are highly positively correlated through the VAR coefficients and with increasing N all the hyperparameters will be estimated to be close to zero. Such an informative prior is clearly undesirable. We propose a simple practical solution; we estimate them independently of each other by sampling from their conditional posterior distribution using the draws of the unrestricted posterior (i.e. draws from the posterior from the individual country VARs without pooling prior). This has the advantage that these hyperparameters will no longer be correlated. We obtain point estimates by using the median of these draws for $\tau_{i,m}\lambda_{i,m,j}$. In a second step, we insert these point estimates into $\boldsymbol{V}_{\beta}^{-1}$ and employ the Gibbs sampler to draw the remaining parameter from their joint posterior distribution. In the next section, we carefully investigate the frequentist properties of our estimation strategy.

3 Simulation Study

In this section, we perform a Monte Carlo simulation to display the pooling properties of the priors across different scenarios. Suppose we want to generate a dataset for three countries and three variables per country, such that $N = 3$ and $G = 3$. For ease of future reference let the countries be named $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ and their corresponding variables y^i, π^i, r^i where $i \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$.

[2] We assume an inverse Wishhart prior for the covariance matrix of the country VAR models and due to conjugacy its conditional posterior follows an inverse Wishhart distribution.

Suppose that for a country i we have a stable VAR with two lags of the form:

$$\begin{bmatrix} y_t^i \\ \pi_t^i \\ r_t^i \end{bmatrix} = \begin{bmatrix} a_{11}^i & a_{12}^i & a_{13}^i \\ a_{21}^i & a_{22}^i & a_{23}^i \\ a_{31}^i & a_{32}^i & a_{33}^i \end{bmatrix} \begin{bmatrix} y_{t-1}^i \\ \pi_{t-1}^i \\ r_{t-1}^i \end{bmatrix} + \begin{bmatrix} b_{11}^i & b_{13}^i & b_{13}^i \\ b_{21}^i & b_{22}^i & b_{23}^i \\ b_{31}^i & b_{32}^i & b_{33}^i \end{bmatrix} \begin{bmatrix} y_{t-2}^i \\ \pi_{t-2}^i \\ r_{t-2}^i \end{bmatrix} + \mathbf{e}_t^i, \quad \mathbf{e}_t^i \sim N(0, \Sigma^i). \quad (24)$$

With $N = 3$ equations per DGP we have a total of nine such equations. A concise way to present all the parameters in the panel VAR is to stack the VAR coefficients for country i in a vector β^i, such that the first $GP \times 1$ block corresponds to the parameters of the first equation, the next block the coefficients of the second equation and so on. We then collect the country vectors in a matrix $\mathbb{B} = [\beta^1, \ldots, \beta^N]$ of dimensions $GPG \times N$.

We will explore the following cases: (1) when the three countries are dissimilar to each other; (2) when two countries share similar dynamics across all variables; (3) where only individual variables or relationships are similar across countries. The corresponding $\mathbb{B}$ matrices for the three DGPs are in Table 1. In DGP 1, we set all parameters across countries $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$ to be distinct from one another. In the DGP 2, the parameters of country $\mathcal{B}$ and $\mathcal{C}$ are chosen to be identical, indicated by the shaded areas in the table. In DGP 3, we set the dynamics of the r_t variable to be identical across the three countries, such that the parameters are equal. Moreover, we set some of the relationships among the other variables to be similar, for example r_{t-1} and r_{t-2} w.r.t. y_t. Finally, we add some cases where the parameters across only two of the three countries are identical. Note that the parameter values themselves have been chosen with the sole purpose of producing stable time series and have no other meaning. The covariance matrices Σ^i are chosen to be diagonal for simplicity.

First we will examine the convergence properties of our model. To do so we generate K=200 datasets from each GDP and do so for varying sample sizes, namely T=50, T=120, and T=500. We calculate the mean absolute error (MAE) between the true coefficients β, given in Table 1, and the estimated parameters $\hat{\beta}$ as $\beta^{MAE} = \frac{1}{K}\sum_{k=1}^{K}\left(|\hat{\beta}_k - \beta|\right)$ for each DGP and sample length. The vector β^{MAE} has $G * G * P = 54$ entries per DGP. For ease of exposition we calculate the average of β^{MAE} and present it in Table 2. This is done once for our pooling VAR model with Minnesota prior (pVAR) and for a traditional

Equation		DGP 1 $\mathcal{A}$	DGP 1 $\mathcal{B}$	DGP 1 $\mathcal{C}$	DGP 2 $\mathcal{A}$	DGP 2 $\mathcal{B}$	DGP 2 $\mathcal{C}$	DGP 3 $\mathcal{A}$	DGP 3 $\mathcal{B}$	DGP 3 $\mathcal{C}$
y_t equation	y_{t-1}	0.50	0.80	0.20	0.50	0.80	0.80	0.20	0.70	0.90
	π_{t-1}	0.10	-0.20	0.20	0.10	-0.20	-0.20	0.10	-0.20	0.20
	r_{t-1}	-0.40	-0.10	0.00	-0.40	-0.10	-0.10	0.00	0.00	0.00
	y_{t-2}	-0.25	0.00	-0.30	-0.25	0.00	0.00	-0.05	-0.10	-0.30
	π_{t-2}	-0.20	0.20	0.00	-0.20	0.00	0.00	-0.20	0.00	0.00
	r_{t-2}	-0.40	0.10	0.40	-0.40	0.10	0.10	0.00	0.00	0.00
π_t equation	y_{t-1}	-0.20	0.00	0.30	-0.20	0.00	0.00	-0.20	0.10	0.00
	π_{t-1}	0.90	0.40	0.00	0.90	0.40	0.40	0.30	0.00	0.30
	r_{t-1}	-0.10	0.10	0.30	0.10	0.10	0.10	0.00	0.00	0.00
	y_{t-2}	0.00	-0.40	0.30	0.00	-0.40	-0.40	0.00	0.20	0.00
	π_{t-2}	0.00	-0.30	0.50	0.00	-0.30	-0.30	-0.40	-0.30	0.10
	r_{t-2}	-0.20	0.50	-0.50	-0.20	0.50	0.50	0.00	0.00	0.00
r_t equation	y_{t-1}	0.20	0.00	0.50	0.20	0.00	0.00	0.00	0.00	0.00
	π_{t-1}	0.20	0.00	0.50	0.10	0.00	0.00	0.00	0.00	0.00
	r_{t-1}	0.75	0.25	-0.20	0.75	0.25	0.25	0.95	0.95	0.95
	y_{t-2}	0.00	0.30	-0.20	0.00	0.30	0.30	0.00	0.00	0.00
	π_{t-2}	-0.30	0.20	0.00	-0.30	0.20	0.20	0.00	0.00	0.00
	r_{t-2}	-0.25	-0.10	0.20	-0.24	-0.10	-0.10	-0.20	-0.20	-0.20

Table 1: VAR coefficients for the three data generating processes (DGPs). $\mathcal{A}$, $\mathcal{B}$, $\mathcal{C}$ denote the countries. Shaded areas highlight the coefficients that are identical across countries.

Bayesian VAR with Minnesota prior but without pooling (BVAR).

Table 2 shows that the model converges, with the mean absolute errors falling as the sample size increases. This exercise highlights that the combination of cross-sectional pooling and Minnesota prior does not produce over-shrinkage. The potential issue is that as Minnesota prior pulls the model parameters towards zero their similarity across countries increases and the pooling prior could, by design, further reinforce this notion. If that were the case, the model would not converge to the true parameters.

Furthermore, we see that even in the case of DGP 1, where all parameters are different and a single-country BVAR would be more appropriate, our model performs almost as good in terms of identifying the true parameters. Conversely it performs slightly better for the more complicated cases, DGP 2 and DGP 3, respectively. In relative terms, in the most complex scenario, DGP 3, the pooling prior shines in small sample sizes. Naturally, as the sample length increases, the difference between the models disappears.

One of the main interests of the simulation study is to examine the estimated shrinkage

	DGP 1			DGP 2			DGP 3		
	T=50	T=120	T=500	T=50	T=120	T=500	T=50	T=120	T=500
pVAR	0.126	0.090	0.064	0.114	0.075	0.046	0.067	0.048	0.029
BVAR	0.117	0.085	0.064	0.113	0.075	0.048	0.073	0.051	0.030
relative	1.054	1.064	1.017	0.985	0.993	0.958	0.749	0.832	0.935

Table 2: Mean absolute errors (MAE), average over all coefficients. Estimated on 200 datasets with varying sample length T. BVAR: Bayesian VAR with Minnesota prior. pVAR: pooling VAR with Minnesota shrinkage. We calculate the relative MAE for each individual coefficient and then take the average.

parameters in relation to the true coefficients. We want to see whether they can extract the country similarities given in the DGPs. With three countries there are three possible country pairs - $(\mathcal{A},\mathcal{B}),(\mathcal{A},\mathcal{C})$, and $(\mathcal{B},\mathcal{C})$, yielding $N = 3$ hyperparameters τ, as well as $KN = GPGN = 18 * 3$ parameters λ that shrink the coefficients across equation pairs. Let $\Lambda^{(n,m)}$ denote the product of the two sets of hyperparameters for each country pair (n, m) such that $\Lambda^{(n,m)} = [\tau_{n,m}\lambda_{n,m,1}, \ldots, \tau_{n,m}\lambda_{n,m,K}]'$.

We present the estimated Λ pairs in the last three columns of Table 3 along with the true parameters for DGP 1. The fourth column denotes the estimated shrinkage of the parameters in columns $\mathcal{A}$ and $\mathcal{B}$ (say 0.5 and 0.8 for the first row), the second column $\mathcal{A}$ and $\mathcal{C}$ (say 0.5 and 0.2 for the first row) and similarly the last column captures the estimated shrinkage of $\mathcal{B}$ and $\mathcal{C}$ (say 0.8 and 0.2 for the first row).

In a sense, these numbers indicate the degree of similarity found in the data across the variable pairs. Values closer to 0 indicate high degree of shrinkage, i.e. high degree of similarity, while values away from 0 suggest the opposite. Similarly to the shrinkage parameters in the Minnesota prior, the magnitudes alone are difficult to interpret, i.e. $\Lambda_j^{n,m} = 0.25$ does not have a specific meaning. However, the values may be examined relative to one another as lower magnitudes indicate higher degree of similarity.

The estimated shrinkage pairs in Table 3 appear colorful. We see that most values lie between 0.10 and 0.30 with few cases of much higher values. The lowest instance of the shrinkage pairs is 0.03 in the second own lag of the y_t equation. The true parameter pair in that case $(y^{\mathcal{A}}_{t-2}, y^{\mathcal{C}}_{t-2}) = (-0.25, -0.30)$ is indeed close. The largest $\Lambda^{\mathcal{A},\mathcal{C}}$ is 0.83 in the first own lag of the r_t equation, associated with $r^{\mathcal{A}}_{t-1} = 0.75$ and $r^{\mathcal{C}}_{t-1} = -0.20$. Overall there is a clear trend in Table 3 as the true parameters are further apart, the estimated

		DGP 1			Shrinkage pairs		
		$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\hat{\Lambda}^{(\mathcal{A},\mathcal{B})}$	$\hat{\Lambda}^{(\mathcal{A},\mathcal{C})}$	$\hat{\Lambda}^{(\mathcal{B},\mathcal{C})}$
y_t equation	y_{t-1}	0.50	0.80	0.20	0.15 (0.04)	0.16 (0.03)	0.41 (0.06)
	y_{t-1}	0.10	-0.20	0.20	0.15 (0.04)	0.14 (0.06)	0.38 (0.08)
	r_{t-1}	-0.40	-0.10	0.00	0.12 (0.03)	0.16 (0.04)	0.04 (0.02)
	y_{t-2}	-0.25	0.00	-0.30	0.12 (0.03)	0.03 (0.01)	0.15 (0.04)
	π_{t-2}	-0.20	0.20	0.00	0.26 (0.07)	0.14 (0.06)	0.08 (0.04)
	r_{t-2}	-0.40	0.10	0.40	0.22 (0.04)	0.67 (0.09)	0.27 (0.06)
π_t equation	y_{t-1}	-0.20	0.00	0.30	0.06 (0.03)	0.14 (0.03)	0.08 (0.03)
	π_{t-1}	0.90	0.40	0.00	0.30 (0.05)	0.76 (0.08)	0.24 (0.04)
	r_{t-1}	-0.10	0.10	0.30	0.06 (0.02)	0.13 (0.03)	0.06 (0.02)
	y_{t-2}	0.00	-0.40	0.30	0.27 (0.05)	0.07 (0.02)	0.42 (0.07)
	π_{t-2}	0.00	-0.30	0.50	0.15 (0.04)	0.30 (0.05)	0.62 (0.06)
	r_{t-2}	-0.20	0.50	-0.50	0.31 (0.04)	0.11 (0.03)	0.55 (0.06)
r_t equation	y_{t-1}	0.20	0.00	0.50	0.12 (0.05)	0.06 (0.02)	0.20 (0.06)
	π_{t-1}	0.20	0.00	0.50	0.16 (0.07)	0.25 (0.09)	0.56 (0.08)
	r_{t-1}	0.75	0.25	-0.20	0.30 (0.05)	0.83 (0.09)	0.28 (0.05)
	y_{t-2}	0.00	0.30	-0.20	0.23 (0.06)	0.06 (0.03)	0.36 (0.07)
	π_{t-2}	-0.30	0.20	0.00	0.54 (0.13)	0.31 (0.10)	0.12 (0.04)
	r_{t-2}	-0.25	-0.10	0.20	0.07 (0.03)	0.26 (0.05)	0.16 (0.04)

Table 3: True VAR coefficients and estimated shrinkage hyperparameters for each country pair for DGP1. Standard deviation of the estimated hyperparameters in brackets.

shrinkage pairs are correspondingly high.

When it comes to DGP 1 the usefulness of our approach is not immediately obvious. To provide better inference the model exploits the cross-sectional dimension. This is beneficial when there are enough similarities across the units. Given the stark differences in DGP 1, estimating three single country VARs would probably suffice. However, DGP 2 provides a different view on the performance of the proposed priors.

Table 4 shows the estimated shrinkage pairs from the case where two countries are exactly identical to each other. The last column shows that the model correctly identifies the similarities across $\mathcal{B}$ and $\mathcal{C}$. The parameter for r_{t-1} in the second equation is also identical across the first three countries and this is reflected in the shrinkage parameters.

The final $\Lambda^{n,m}$ entries in Table 4 which are of similar small magnitude are the $\pi^{\mathcal{A}}_{t-1}$ values in the last equation. The identical pairs $(\pi^{\mathcal{A}}_{t-1}, \pi^{\mathcal{B}}_{t-1}) = (0.10, 0, 00)$ and $(\pi^{\mathcal{A}}_{t-1}, \pi^{\mathcal{C}}_{t-1}) = (0.10, 0, 00)$ have equal estimated shrinkage of $\Lambda^{\mathcal{A},\mathcal{B}} = \Lambda^{\mathcal{A},\mathcal{C}} = 0.05$, respectively.

		DGP 2			Shrinkage pairs		
		$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\hat{\Lambda}^{(\mathcal{A},\mathcal{B})}$	$\hat{\Lambda}^{(\mathcal{A},\mathcal{C})}$	$\hat{\Lambda}^{(\mathcal{B},\mathcal{C})}$
y_t equation	y_{t-1}	0.50	0.80	0.80	0.14 (0.04)	0.14 (0.04)	0.02 (0.01)
	π_{t-1}	0.10	-0.20	-0.20	0.12 (0.03)	0.12 (0.04)	0.01 (0.00)
	r_{t-1}	-0.40	-0.10	-0.10	0.13 (0.03)	0.13 (0.03)	0.01 (0.00)
	y_{t-2}	-0.25	0.00	0.00	0.12 (0.03)	0.12 (0.03)	0.02 (0.01)
	π_{t-2}	-0.20	0.00	0.00	0.09 (0.03)	0.10 (0.03)	0.01 (0.00)
	r_{t-2}	-0.40	0.10	0.10	0.21 (0.04)	0.21 (0.04)	0.01 (0.01)
π_t equation	y_{t-1}	-0.20	0.00	0.00	0.07 (0.03)	0.07 (0.03)	0.03 (0.01)
	π_{t-1}	0.90	0.40	0.40	0.29 (0.05)	0.29 (0.05)	0.02 (0.01)
	r_{t-1}	0.10	0.10	0.10	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
	y_{t-2}	0.00	-0.40	-0.40	0.28 (0.06)	0.29 (0.06)	0.03 (0.02)
	π_{t-2}	0.00	-0.30	-0.30	0.16 (0.04)	0.15 (0.04)	0.02 (0.01)
	r_{t-2}	-0.20	0.50	0.50	0.33 (0.05)	0.34 (0.05)	0.02 (0.01)
r_t equation	y_{t-1}	0.20	0.00	0.00	0.11 (0.04)	0.10 (0.05)	0.03 (0.02)
	π_{t-1}	0.10	0.00	0.00	0.05 (0.03)	0.05 (0.03)	0.02 (0.01)
	r_{t-1}	0.75	0.25	0.25	0.30 (0.05)	0.30 (0.05)	0.02 (0.01)
	y_{t-2}	0.00	0.30	0.30	0.26 (0.06)	0.25 (0.06)	0.03 (0.02)
	π_{t-2}	-0.30	0.20	0.20	0.41 (0.08)	0.41 (0.08)	0.02 (0.01)
	r_{t-2}	-0.24	-0.10	-0.10	0.06 (0.03)	0.06 (0.03)	0.02 (0.01)

Table 4: True VAR coefficients and estimated shrinkage hyperparameters for each country pair for DGP 2. Standard deviation of the estimated hyperparameters in brackets.

Table 5 presents the case where the focus is on similar dynamics across countries for specific variables only, notably the r_t equation but also some random lags in the other equations. Without exception, the model shrunk the identical coefficients correctly and also did particularly well in the equations with a mix of similar and different dynamics. Again, in instances where the true values are close to one another (e.g. y_t equation, $(\pi^{\mathcal{A}}_{t-1}, \pi^{\mathcal{C}}_{t-1}) = (0.10, 0.20)$) the estimated shrinkage hyperparameters are low $\Lambda^{(\mathcal{A},\mathcal{C})} = 0.02$.

In reality, actual data is probably a mix of all three DGPs. While a case where two

Equation		DGP 3 $\mathcal{A}$	DGP 3 $\mathcal{B}$	DGP 3 $\mathcal{C}$	Shrinkage pairs $\hat{\Lambda}^{(\mathcal{A},\mathcal{B})}$	Shrinkage pairs $\hat{\Lambda}^{(\mathcal{A},\mathcal{C})}$	Shrinkage pairs $\hat{\Lambda}^{(\mathcal{B},\mathcal{C})}$
y_t equation	y_{t-1}	0.20	0.70	0.90	0.27 (0.05)	0.44 (0.07)	0.07 (0.03)
	π_{t-1}	0.10	-0.20	0.20	0.10 (0.03)	0.02 (0.01)	0.10 (0.02)
	r_{t-1}	0.00	0.00	0.00	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
	y_{t-2}	-0.05	-0.10	-0.30	0.03 (0.01)	0.08 (0.03)	0.07 (0.02)
	π_{t-2}	-0.20	0.00	0.00	0.05 (0.02)	0.05 (0.02)	0.01 (0.00)
	r_{t-2}	0.00	0.00	0.00	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
π_t equation	y_{t-1}	-0.20	0.10	0.00	0.14 (0.03)	0.06 (0.02)	0.05 (0.02)
	π_{t-1}	0.30	0.00	0.30	0.11 (0.03)	0.02 (0.01)	0.11 (0.04)
	r_{t-1}	0.00	0.00	0.00	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
	y_{t-2}	0.00	0.20	0.00	0.06 (0.02)	0.01 (0.00)	0.05 (0.02)
	π_{t-2}	-0.40	-0.30	0.10	0.04 (0.02)	0.25 (0.05)	0.17 (0.04)
	r_{t-2}	0.00	0.00	0.00	0.02 (0.01)	0.02 (0.00)	0.02 (0.00)
r_t equation	y_{t-1}	0.00	0.00	0.00	0.01 (0.00)	0.01 (0.00)	0.01 (0.01)
	π_{t-1}	0.00	0.00	0.00	0.01 (0.01)	0.01 (0.00)	0.01 (0.00)
	r_{t-1}	0.95	0.95	0.95	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
	y_{t-2}	0.00	0.00	0.00	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
	π_{t-2}	0.00	0.00	0.00	0.01 (0.00)	0.01 (0.00)	0.01 (0.00)
	r_{t-2}	-0.20	-0.20	-0.20	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)

Table 5: True VAR coefficients and estimated shrinkage hyperparameters for each country pair for DGP 3. Standard deviation of the estimated hyperparameters in brackets.

countries are exactly identical, as in DGP 2, is hard to imagine, the extreme case of DGP 1, where every possible variable combination is completely different, is also unrealistic. It is realistic to expect that data is a mix of all of the proposed processes - similar, yet not identical dynamics across some variables and different among others.

Finally, these tables show the hyperparameter estimates for the case of $T = 500$ as our goal is to focus on the similarity showcase abstracting from asymptotics. We relegate the tables for $T = 120$ to the Appendix. Those results show that smaller sample sizes lead to the same conclusions. The shrinkage parameter values from the smaller sample are simply larger in absolute value, however, the relative structure among them remains unchanged.

4 Empirical Investigation

The model lends itself best to applications where different units might share common dynamics across some characteristics but not among others. The euro area member states follow a common monetary policy driven by the aggregate price dynamics and the single market allows free movement of goods, capital, services, and persons. At the same time, the member states have individual fiscal and legal systems which lead to different rigidities across the different markets, for example the labour market. Moreover the countries specialize in different industries, for example, the automotive industry is particularly important in some member states while completely non-existent in others. Hence, shocks to such industries would have disproportionate effects on production. This also applies to aggregate shocks, such as the sovereign debt crisis, which affected southern European nations disproportionately to northern member states.[3]

These features make the euro area particularly suitable for our model which allows for a mix of homogeneity and heterogeneity across different macroeconomic variables. First, we perform a forecasting exercise using euro area data and evaluate the forecasting performance and then take a close look at the estimated hyperparameters to recover potential similarities and dissimilarities across the main macroeconomic variables. Finally, we look at the responses of a contractionary monetary policy shock on the member states to evaluate the application of our framework for structural analysis.

4.1 Data

We collect a large dataset on 10 European countries: Germany, France, Italy, Spain, Netherlands, Belgium, Austria, Portugal, Finland, and Greece. Throughout the paper we will refer to the countries using their 2-digit ISO codes. A table with the abbreviations is available in the Appendix, Table B.1. For each country we have chosen 7 variables: real GDP (RGDP), the harmonised index of consumer prices (HICP), EURIBOR, Industrial production (IP), Unemployment (UNEMP), the Economic Sentiment Index (ESI) and the price of crude brent (OIL). The brent price and the interest rates are identical across

[3]The heterogeneity of the euro area member states and their short time series has been a motivation for the development of the Bayesian panel VAR framework of Jarocinski (2010).

countries.

The choice of the data is primarily based on our needs for a balanced panel with identical units. The shrinkage across variables implies that each variable should be present for each country. The chosen countries represent close to 95% of the euro area output, and that these variables are frequently used for structural analysis the coverage should suffice. Notably, we omit Ireland, where the way R&D expenditure was treated in the estimation of GDP was altered in 2015. This change introduced a structural break and sharply increased the volatility of the series as profits from licensing of intellectual property of multinationals now entered investments. We therefore omit Ireland as to avoid issues with outliers and volatility.[4]

Apart from the interest rates and the unemployment rate, the data is transformed to be stationary by taking percentage growth rates. In case the seasonally adjusted data were not available, seasonal adjustment using the X-13 ARIMA procedure has been performed. All data has been standardized so that the VAR may be estimated without a constant. However, we undo this transformation for the calculation of the forecasts and impulse response functions (IRFs). The data is quarterly and spans a total of 79 observations per country from 2000Q2 to 2019Q4. We refrain from including the COVID-19 pandemic as to avoid estimation errors due to the large outliers. Since our model is based on a standard BVAR with a Minnesota prior, any methods that can deal with the pandemic outliers in the standard model (e.g Lenza and Primiceri (2020), Carriero et al. (2021) or Prüser (2021)) are appropriate. The data is from Eurostat and has been obtained through the Macrobond software.

4.2 Forecasting setup

In the forecasting exercise, we compare the performance of a range of different models. First we look at three main specifications - a benchmark case and two versions of our proposed pooling VAR. The benchmark specification is the traditional BVAR with a Minnesota prior (BVAR). Our pooling VARs have also a Minnesota component along

[4]Alternatively one could take the growth of GDP from the production approach - gross value added. However, we wanted to have a balanced panel of identical variables across units.

with shrinkage across both the country and variable dimensions, i.e. with $\lambda_{n,m,j}$ and $\tau_{n,m}$ (pVAR$^{\lambda,\tau}$), as well as a version with country cross pooling only, namely $\tau_{n,m}$ (pVAR$^{\tau}$).

We then include a thorough comparison with a multitude of different models, as listed in Table B.4 in the Appendix. We include a state-of-the art version of the single-country VAR with a more flexible Minnesota prior as suggested by Chan et al. (2021) (BVAR^{+}). In addition, we test two specifications that utilize the cross-sectional dimension. A standard Bayesian Panel VAR that restricts the VAR coefficients to be identical across countries, as well as the Bayesian Panel VAR model of Jarocinski (2010), which allows for deviations of the VAR parameters and is arguably one of the most flexible Panel VAR frameworks.[5] To highlight the importance of using informative prior distributions in VAR models, we consider a Bayesian VAR implementation with a flat prior (BVARf). Furthermore, the combination of the pairwise pooling with a zero shrinkage has, to our knowledge, so far not been proposed in a cross-sectional context. To evaluate the importance of this addition we explore the pooling specification without zero shrinkage, i.e. the models less the Minnesota part (pVARlm$^{\tau}$ and pVARlm$^{\lambda,\tau}$). Finally, we look at a robustness specification that uses an inverse Gamma prior for the hyperparameters instead of the half-Cauchy distributions, which we discuss in the robustness section and in more detail in the Online Appendix.

All models are estimated with four lags, in order to capture potentially different dynamics across the countries and consistent with the quarterly frequency of the data.[6] The models are estimated with a Gibbs sampler chain of 11000 draws from the posterior distribution of which the first 1000 draws are discarded.

Apart from the competing models, we also consider two different variable combinations. A smaller specification with all countries and three variables ($N = 10, G = 3$), where we take the commonly used variables Output, Prices and Interest rates, and a larger specification with all the seven variables ($N = 10, G = 7$). We mirror this setup in our application for the G-7 economies (not to be confused with G, the number of variables in the models), where we test a three and a seven variable specification, $N = 7, G = 3$

[5] The model is implemented and described in the BEAR toolbox, see Dieppe et al. (2016).

[6] Furthermore, we find it to be important to allow for a higher lag length in order to obtain economic plausible impulse response functions.

and $N = 7, G = 7$, respectively.

Our main metric for comparing the models is the root mean squared forecast error (RMSE) calculated as the squared difference between the forecast and the actual data. We also look at the density forecasts using continuous rank probability score (CRPS) as presented in Gneiting and Raftery (2007). We will primarily look at relative RMSE to the benchmark cases - for the $N10G3$ models the relative benchmark is the $N10G3$ BVAR and for the $N10G7$, the seven variable one, respectively.

The forecasting setup is the following. With each model and for each of the two variable specifications we carry out a pseudo-out-of-sample forecasting exercise with an expanding window over 40 periods. That is, we start with the first 39 observations (up to $T = 2011Q4$) to estimate the models and make a one-step ahead forecast for the next quarter ($T + 1 = 2012Q1$). Iteratively, we then create recursive forecasts for up to 8 horizons. The dataset is expanded by one quarter to $2011Q1$ and we repeat the same steps. This is iterated on until the end of the sample. We then take the mean of the respective $T + 1, \ldots, T + 8$ forecasts across the horizon and calculate the total average RMSFEs and CRPSs.

4.3 Forecasting Results

Figure 1 plots selected models, namely the pooling priors for the $N10G3$ specification. The forecasting performance of all model specifications is available in Table 6. The column plots of Figure 1 show the variables RGDP, HICP, and EURIBOR, while the three rows are associated with the one-step ahead forecast error ($T+1$), the one-year ahead forecast error ($T + 5$), and the average over the eight forecast horizons, respectively. The countries are on the x-axis. Relative RMSE calculated to the BVAR baseline. We showcase the performance of the pooling prior with shrinkage both across variables and countries ($\text{BVAR}^{\lambda,\tau}$, triangles), as well as only the country-specific case (BVAR^{τ}, circles).

We find that incorporating shrinkage across unit pairs improves the forecasts considerably for most countries, especially in France, Italy, Spain, and Belgium and to a lesser extent Finland. When it comes to GDP growth, Germany, the Netherlands, Portugal and Greece do not seem to benefit from the pooling priors as much. The findings are similar

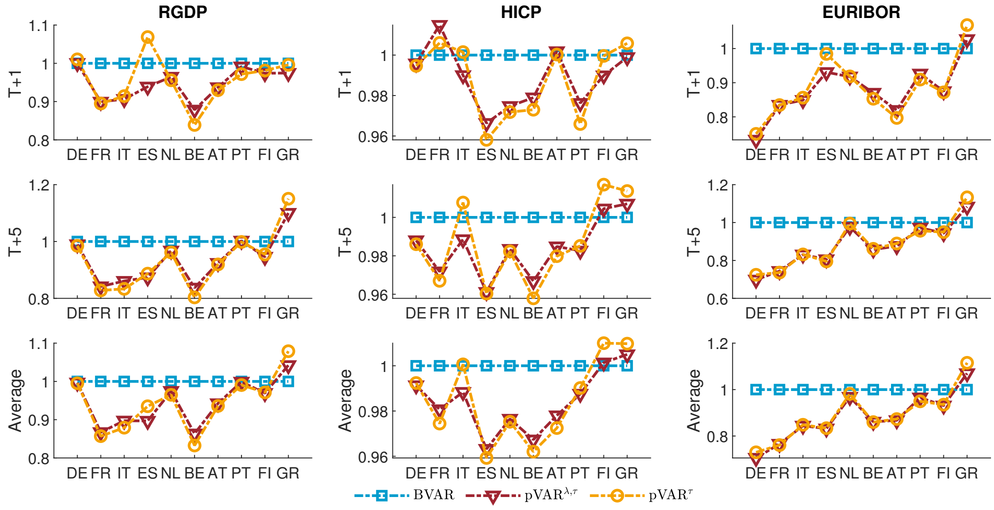

Figure 1: **N=10, G=3 setup:** RMSFEs relative to the BVAR baseline (y-axis) versus countries on the (x-axis). Less is better.

with consumer prices. Notably, the Euribor forecasts improve across the board (bar from Greece). The three month interbank reference rate is identical across countries and due to the common monetary policy we may expect that their dynamics should be comparable within the individual countries' VARs.

Figure 1 shows that the pooling priors improve the forecasts also for horizons beyond the shortest term. Furthermore, the BVAR$^{\lambda,\tau}$ model appears to be slightly better than the BVAR$^{\tau}$ case due to some particular outliers, such as GDP and Euribor forecasts in Spain amd Greece, respectively. When we turn to an analysis of the estimated hyperparmeters we will see that the dynamics of GDP in Spain and Greece are different from the other countries, hence the importance of the local component $\lambda_{i,m,j}$ that allows variable-specific for deviations from country pair shrinkage parameter $\tau_{i,m}$.

Table 6 offers a more detailed look of the performance of the other competing models. We show only the one-step ahead forecast errors in the interest of space. The results for longer horizons are similar. For a sense of scale, the first row of each table reports the absolute forecast error, while the rows below report relative values to those numbers. For each country we highlight in bold the model that yields the lowest relative RMSFE. Surprisingly, the Bayesian VAR extension of Chan et al. (2021) does not seem to outperform the standard version when it comes to GDP or HICP of the European countries as it roughly delivers similar results. It does appear to improve the EURIBOR forecast. This could be due to the fact that the number of variables in this application is small, so not that much shrinkage is needed.

The models that can benefit from the cross-sectional dimensions deliver mixed results for the different variables. As expected in such a large system, the flat prior (BVARf) yields the worst results. However, the standard Bayesian panel VAR (BPVAR) and the approach of Jarocinski (2010) (PVARj) work well in some cases and perform worse in others. The panel VAR works particularly well for many inflation forecasts, while delivering worse GDP forecasts against a single country specification in all situations. Similarly, PVARj delivers arguably the best EURIBOR forecast if one takes the austrian case but does not outperform any model in much of the rest.

In that respect the combination of pooling with shrinkage appears extremely reliable as its performance is consistent. In most situations it delivers some of the best forecasts in terms of RMSFE and in others it performs at least as good as the baseline. Both the pVAR$^{\lambda,\tau}$ and pVAR$^{\lambda}$ specifications show some of the best forecasting performances when it comes to production or the interest rates.

The final two rows of Table 6 emphasize the importance of the combination of zero shrinkage and pooling. Without it both single country specifications as well as panel VARs perform much better when it comes to forecasting as pVARlm$^{\lambda,\tau}$ and pVARlm$^{\tau}$ outperform only the flat prior model.

Next, we turn our attention to the larger model with 7 variables per country. With four lags the number of parameters in each single country VAR grows tremendously. In such an environment shrinkage priors should excel. This is what we observe both in Figure 2 which plots the results for three of the seven variables as well as in Table 7 that details the one-step ahead forecast per country and variable. Moreover, we find that the three-variable BVAR performs in some instances worse than the seven-variable BVAR model (i.e. the absolute RMSFE of the larger model is lower for some variables).

Note that while Figure 2 is the counterpart to the three-variable model in Figure 1, the lines are not perfectly comparable because we always use the standard BVAR model as a reference point. The conclusions are in general similar - we again see gains of the pooling specifications with respect to the standard BVAR. These gains are smaller than

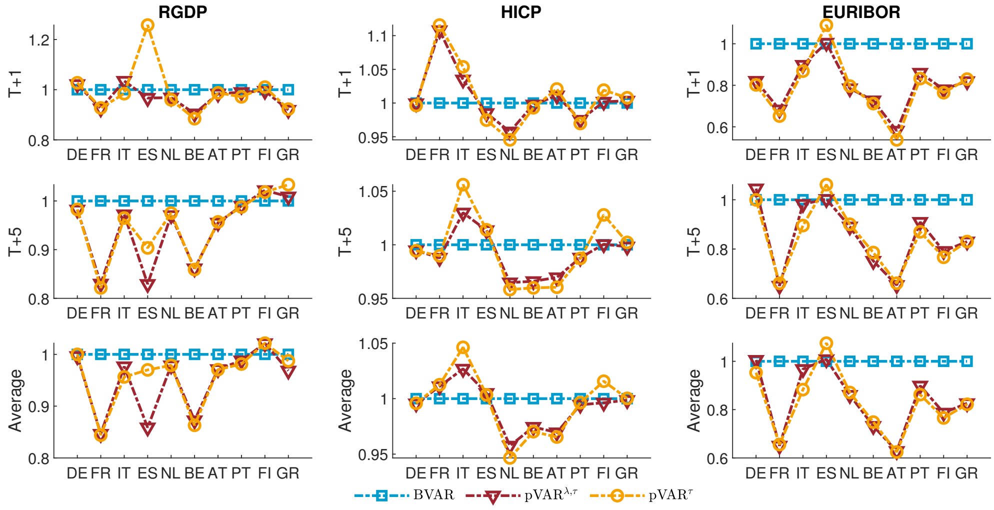

Figure 2: **N=10, G=7 setup:** RMSFEs relative to the BVAR baseline (y-axis) versus countries on the (x-axis). Less is better.

in Figure 1 for the one-step ahead forecast (T+1, first row of figures) and larger over the longer and all forecast horizons.

Table 7 shows that in the larger model the more flexible Minnesota specification of Chan et al. (2021), BVAR$^+$, starts to outperform the three variable case, especially for HICP and for two countries when it comes to output. The first row shows the absolute magnitude of the errors and for all variables and, except for Greece, for all countries the larger model delivers a lower forecast error. While in most cases it is only one to two decimal points, there are cases such as Spain, where the GDP absolute RMSE falls from 0.34 to 0.28. Naturally, the flat prior works even worse for larger models. The pooling priors, especially the BVAR$^{\lambda,\tau}$, again perform best in most specifications: Specifically when it comes to the interest rates they often deliver even better forecasts than in the $N10G3$ case.

These findings are similar across the remaining variables from the larger model, which are plotted on Figure A.2 in the Appendix. The pooling model delivers better forecasting performance than the single country BVAR on average and excels for variables that share dynamics across countries, such as unemployment and the oil prices. Overall, including both λ and τ shrinkage in addition to zero shrinkage works best, while country-wide shrinkage alone (pVAR$^\tau$) produces some outliers such as industrial production for Spain and Greece.

Interestingly, the panel specifications also perform worse than in the three variable case, especially the PVARj approach. These findings suggest that these models are not flexible enough to capture a mix of homo- and heterogeneity given their worsening performance with the larger dataset.

Finally we examine the density forecast performance using the continuous rank probability score as a metric. Tables B.2 and B.3 found in the Appendix are for the three-variable and the seven-variable models, respectively. They follow the structure of the point forecast tables and present the CRPS for all model and variable combinations. Our conclusions closely match the narrative of the point forecasts - combining pair-wise pooling with Minnesota improves the forecasting density in a similar fashion as does so for the point forecasts.

4.4 Analysis of the hyperparameters

Naturally, the model allows for studying the degrees of similarity across countries by looking at the estimated hyperparameters across the country pairs, i.e. $\Lambda^{n,m} = \lambda_{n,m,j}\tau_{n,m}$, where $n, m \in \{DE, FR, ES, IT, NL, BE, AT, PT, FI, GR\}$. In the $N10G7$ case there are 28 parameters for each of the seven equations and $GGP = 196$ coefficients in total per country and correspondingly $N(N-1)/2 = \sum_i^N (i-1) = 45$ country pairs. However, many of these are shrunk to zero due to the Minnesota prior. Therefore we choose to analyse the country pairs of the own lags of GDP, HICP, and EURIBOR as these are usually most informative regarding the variable dynamics.

Figure 3 displays the estimated hyperparameters for the country pairs. The first row represents the hyperparameters for the coefficients of the four own lags of GDP in the GDP equation, the second of the own lags of HICP in the HICP equation, and the third for EURIBOR, respectively. Each box shows the estimated hyperparameter pairs for country i and m, i.e. $\Lambda^{i,m} = \lambda_{i,m,j}\tau_{i,m}$ as coloured squares with darker shades associated with higher values, i.e. dissimilarity between the coefficients, and lighter shades indicating resemblance. By design the boxes are symmetric, since $\Lambda^{i,m} = \Lambda^{m,i}$.

The first box shows the country pairs' hyperparameters for the first lag of GDP. It is immediately obvious that the production dynamics in Spain are different than in most

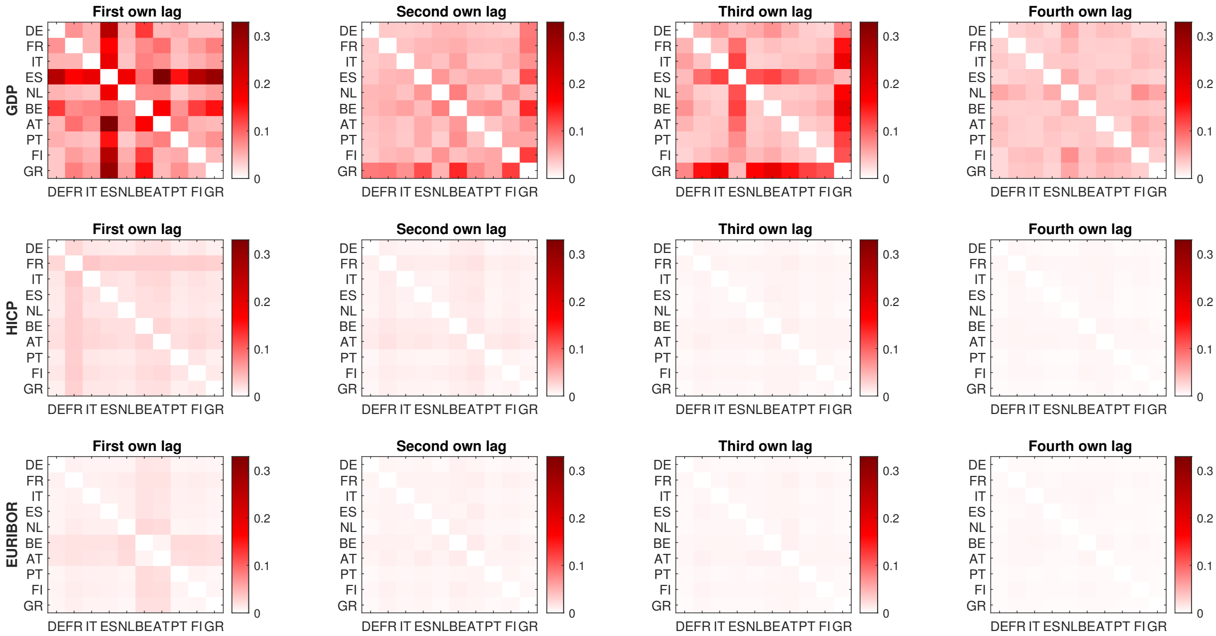

Figure 3: Estimated hyperparameters for each country pair for GDP (first row), HICP (second row), and EURIBOR (third row). Hyperparameters for the coefficients of the four own lags in each respective equation (columns one to four, respectively). Darker shades indicate dissimilarity between the respective parameters, lighter tones point to the contrary.

countries. Interestingly this is true even for Italy, Portugal, and Greece, a group with which Spain is often bundled together. However, the rest of the country pairs also appear rather colorful. The inertia of GDP in the Netherlands and Portugal is more similar to France and Italy, than it is to their direct neighbors Germany and Spain.

The remaining lags of GDP also deliver insight. The dynamics in Greece w.r.t. the second and third lag are particularly different than the other countries, as it is the third lag for Spain. These findings highlight the need of more lags to capture the heterogeneity properly which in turn emphasizes the importance of needing zero shrinkage to achieve parsimony.

Turning to HICP and EURIBOR, the second and third row, respectively, we find much better integration across the member states. The common monetary policy is driving the interest rates and the price dynamics, albeit there seems to be more heterogeneity there. Past lags, especially the third and the fourth do not seem to be important and have most likely been shrunk to zero, hence the low estimates for the hyperparameters.

Finally, Figure 3 explains the worse forecasting performance of $pVAR^{\tau}$ model, compared to the more richly specified $pVAR^{\lambda,\tau}$ when it comes to GDP for Spain. When it comes to prices and interest rates, the Spanish dynamics are much more aligned to their European counterparts, however the differences in the GDP process require the λ component to be captured properly.

4.5 Impulse Response Functions

Next, we evaluate the inference properties of our model when it comes to structural identification. We calculate the impulse responses (IRs) following a one standard deviation shock to the interest rates for each country using the BVAR and the pooling BVAR$^{\lambda,\tau}$. For identification we rely on the conventional Cholesky decomposition and we use the $N10G3$ specification.

We plot the IRs for Germany, France and Italy on Figure 4 and relegate the other seven countries to the Appendix. On average the median impulse responses are highly similar. However, the pooling prior appears to provide sharper inference leading to much narrower probability intervals. This could prove pivotal as it helps differentiate significant responses. In this example we find that using a small BVAR would imply that RGDP, while on average declining within one to two quarters, does not do so in a statistically meaningful way. However, using the pooling prior highlights clearer the effects of contractionary monetary policy and in all instances the zero line lies outside of the probability bands.

While the simple structural analysis is aimed at highlighting the estimation uncertainty reduction and not at addressing an economic question, it is interesting to note that the price puzzle is not a feature of estimation uncertainty, as even the sharper inference does not change the shape of the impulse responses. The euro area data suggests that prices either decline in median or do not respond at all to tightening monetary policy shocks.

4.6 Robustness and alternative specification

We look to an alternative prior specification, an alternative empirical application and alternative identification strategy to test the limitations of our model and the robustness of the findings. We relegate these tests and a more in-depth discussion to the Online

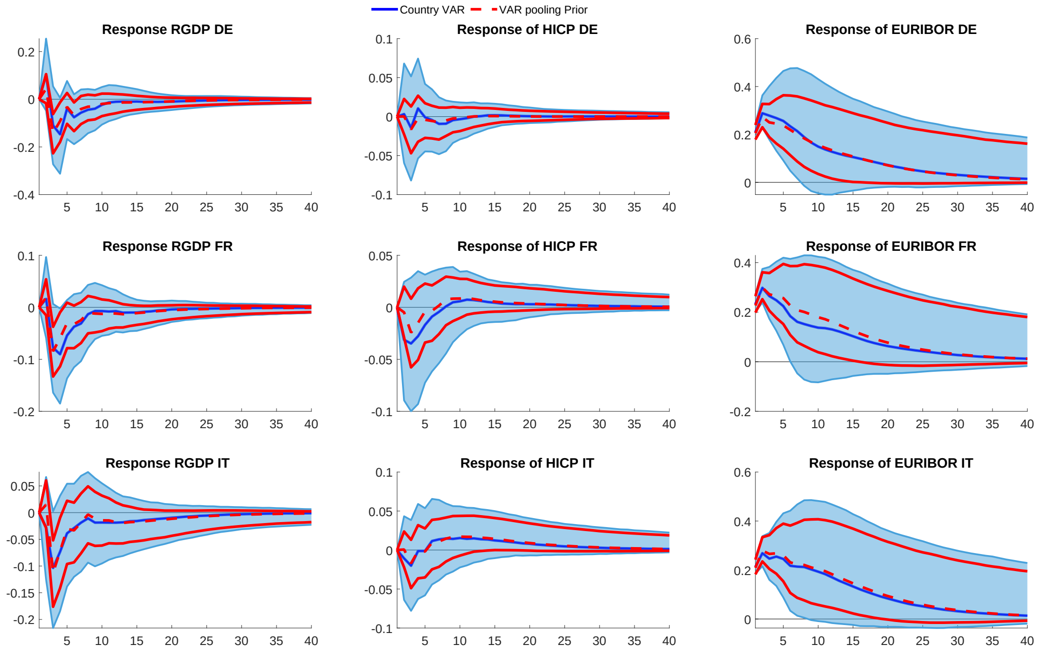

Figure 4: Impulse responses following a one standard deviation shock to the interest rates. Shaded areas represent one standard deviation probability intervals.

Appendix and summarize our results here.

We have made a choice of the hyperprior on the hyperparameters, namely the half-Cauchy prior. We explore additional specifications with an inverse Gamma for λ and τ, which proves to be also extremely effective if a bit more volatile. For many instances where our main specification ($pVAR^{\lambda,\tau}$) outperforms the competition, the inverse Gamma prior does so even better. At the same time, in the cases where the opposite is true and the horseshoe is equal or slightly worse than the benchmark, inverse Gamma exhibits a worse performance than that. Therefore we think that the horseshoe prior is a safer choice that yields the most optimal gains at no apparent cost.

We also explore a different dataset, namely the G-7 group of advanced economies. Again we find that our model is as equal or better than the benchmark (BVAR with Minesota prior), again with the best performance when it relates to the interest rates. These series are, in contrast to the euro area case, different from one another, especially for the UK, Japan, and the US. However, overall the differences between the competing models are lower for the G-7 application. For example, the performance when it comes to forecasting inflation is almost identical across the board (apart from a flat prior, which is not a good choice). Notably, the forecast errors for Germany, France and Italy are worse when using the G-7 dataset in comparison to the euro area data. This goes to show that our approach benefits most from using countries that share common characteristics but can hold on its own even if that is not the case.

Finally, we also test the structural identification results by adopting a different identification strategy, namely sign restrictions instead of a recursive scheme and again report that the error bands of our model are narrower than the respective single country counterparts.

5 Conclusion

In this paper we introduce a new prior for VAR models which can exploit the panel structure of the data to deliver more efficient estimates as well as provides shrinkage towards zero to alleviate overparameterization concerns. We illustrate our approach by means of several applications. The first application uses synthetic data to investigate the properties of the model across different data-generating processes. It turns out that our approach is highly flexible and can adapt to different scenarios. This frees the researcher from the need to make assumptions about the structure or homogeneity between economies prior to taking our model to the data. The second application analyzes the predictive gains from our prior in a forecasting exercise for actual data. We test our model on a dataset of ten of the largest euro area member states since they share a lot of characteristics while also hiding large heterogeneities. Additionally we look at a dataset with the G-7 advanced economies and find that even in smaller panels there are gains from the cross-sectional information to be made. We show that pooling across countries alone is not enough to help the estimation, but also forecasts from single country specifications can be improved upon. It is the combination of cross-country pooling with Minnesota that achieves the excellent performance. Finally, we show that our approach provides narrower probability intervals for structural analysis stemming from the lower estimation uncertainty.

RGDP	**DE**	**FR**	**IT**	**ES**	**NL**	**BE**	**AT**	**PT**	**FI**	**GR**
BVAR (abs)	0.60	0.35	0.37	0.34	0.42	0.32	0.54	0.62	0.78	1.34
BVAR	**1.00**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
BVAR$^+$	1.06	1.07	1.08	1.09	1.02	1.00	0.99	1.03	1.01	1.02
BVARf	1.24	1.27	1.49	1.47	1.41	1.24	1.20	1.09	1.36	1.33
pVAR$^{\lambda,\tau}$	1.00	0.90	**0.91**	**0.94**	0.96	0.88	0.94	0.99	**0.97**	0.97
pVAR$^{\tau}$	1.01	**0.89**	0.91	1.07	**0.96**	**0.84**	**0.93**	**0.97**	0.98	1.00
BPVAR	1.01	1.07	1.10	1.14	1.09	1.06	0.96	1.00	1.01	1.01
PVARj	1.08	1.15	1.03	1.17	1.13	1.35	0.96	1.01	1.10	**0.97**
pVARlm$^{\lambda,\tau}$	1.19	1.23	1.40	1.31	1.28	1.16	1.10	1.01	1.21	1.13
pVARlm$^{\tau}$	1.19	1.24	1.41	1.30	1.28	1.15	1.11	1.01	1.22	1.13

HICP	**DE**	**FR**	**IT**	**ES**	**NL**	**BE**	**AT**	**PT**	**FI**	**GR**
BVAR (abs)	0.33	0.27	0.30	0.49	0.38	0.36	0.29	0.46	0.28	0.53
BVAR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
BVAR$^+$	**0.99**	**1.00**	1.02	1.02	1.02	1.01	1.00	1.01	1.08	1.00
BVARf	1.14	1.06	1.13	1.19	1.15	1.16	1.14	1.15	1.32	1.08
pVAR$^{\lambda,\tau}$	1.00	1.01	0.99	0.97	0.97	0.98	1.00	0.98	0.99	1.00
pVAR$^{\tau}$	0.99	1.01	1.00	0.96	**0.97**	**0.97**	1.00	**0.97**	1.00	1.01
BPVAR	1.04	1.03	**0.96**	**0.92**	0.97	1.01	**0.96**	0.98	**0.95**	**0.94**
PVARj	1.08	1.10	1.07	1.00	0.99	1.03	1.02	1.03	1.07	1.05
pVARlm$^{\lambda,\tau}$	1.07	1.05	1.07	1.10	1.07	1.09	1.06	1.14	1.18	1.04
pVARlm$^{\tau}$	1.07	1.06	1.07	1.10	1.06	1.09	1.07	1.14	1.11	1.04

EURIBOR	**DE**	**FR**	**IT**	**ES**	**NL**	**BE**	**AT**	**PT**	**FI**	**GR**
BVAR (abs)	0.19	0.18	0.20	0.17	0.20	0.19	0.21	0.23	0.21	0.17
BVAR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**
BVAR$^+$	1.01	0.97	0.93	1.04	0.98	0.94	0.94	0.94	0.97	1.16
BVARf	1.46	1.26	1.61	1.57	1.77	1.06	1.19	1.33	1.31	1.67
pVAR$^{\lambda,\tau}$	**0.73**	0.83	0.85	**0.93**	0.92	0.87	0.82	0.93	0.87	1.03
pVAR$^{\tau}$	0.75	**0.83**	0.86	0.98	0.92	**0.85**	0.80	0.91	**0.87**	1.07
BPVAR	0.96	0.93	**0.76**	0.97	**0.88**	1.04	0.86	**0.76**	0.89	1.14
PVARj	1.04	0.99	0.94	1.20	0.92	1.09	**0.73**	1.00	1.00	2.05
pVARlm$^{\lambda,\tau}$	1.35	1.24	1.46	1.42	1.50	1.02	1.03	1.33	1.20	1.62
pVARlm$^{\tau}$	1.37	1.25	1.49	1.48	1.52	1.02	1.05	1.38	1.27	1.62

Table 6: **N=10, G=3 setup:** One step ahead root mean squared forecast errors for different models, relative to the BVAR forecast. The first row is the absolute error of the BVAR benchmark. Values in bold denote the lowest relative RMSFE for that country.

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
RGDP	**BVAR (abs)**	0.59	0.33	0.38	0.28	0.41	0.29	0.48	0.63	0.75	1.44
RGDP	**BVAR**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
RGDP	**BVAR$^+$**	**0.95**	1.05	1.02	1.15	1.07	1.06	1.02	1.03	**0.99**	0.95
RGDP	**BVARf**	2.49	1.91	1.94	2.21	2.77	2.12	2.19	1.79	2.30	2.11
RGDP	**pVAR$^{\lambda,\tau}$**	1.02	**0.92**	1.03	**0.97**	0.97	0.90	**0.98**	0.99	1.00	0.92
RGDP	**pVAR$^{\tau}$**	1.03	0.93	**0.98**	1.26	**0.96**	**0.88**	0.99	**0.97**	1.01	0.92
RGDP	**BPVAR**	0.99	1.35	1.05	1.27	1.08	1.36	1.10	0.98	1.07	**0.89**
RGDP	**PVARj**	1.11	1.56	1.05	1.77	1.42	1.64	1.17	1.09	1.13	0.93
RGDP	**pVARlm$^{\lambda,\tau}$**	1.54	1.65	1.53	1.65	1.71	1.82	1.49	1.16	1.59	1.23
RGDP	**pVARlm$^{\tau}$**	1.23	1.54	1.46	1.53	1.55	1.70	1.38	1.15	1.56	1.17
		DE	**FR**	**IT**	**ES**	**NL**	**BE**	**AT**	**PT**	**FI**	**GR**
HICP	**BVAR (abs)**	0.33	0.24	0.28	0.47	0.39	0.35	0.29	0.46	0.27	0.51
HICP	**BVAR**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
HICP	**BVAR$^+$**	**0.98**	**0.90**	**0.98**	1.04	1.00	1.02	1.02	1.03	1.07	**0.96**
HICP	**BVARf**	2.27	1.36	1.46	1.62	1.74	1.60	1.65	1.70	2.16	1.36
HICP	**pVAR$^{\lambda,\tau}$**	1.00	1.11	1.03	0.98	0.96	1.00	1.01	0.97	1.00	1.00
HICP	**pVAR$^{\tau}$**	1.00	1.12	1.05	0.97	**0.95**	**0.99**	1.02	**0.97**	1.02	1.01
HICP	**BPVAR**	1.06	1.14	1.01	**0.94**	0.98	1.06	**0.98**	1.00	**0.98**	0.99
HICP	**PVARj**	1.18	1.18	1.12	1.06	1.09	1.13	1.17	1.06	1.23	1.02
HICP	**pVARlm$^{\lambda,\tau}$**	1.20	1.13	1.11	1.27	1.38	1.25	1.30	1.34	1.53	1.05
HICP	**pVARlm$^{\tau}$**	1.16	1.11	1.07	1.25	1.13	1.23	1.21	1.28	1.26	1.05
		DE	**FR**	**IT**	**ES**	**NL**	**BE**	**AT**	**PT**	**FI**	**GR**
EURIBOR	**BVAR (abs)**	0.15	0.21	0.16	0.15	0.16	0.19	0.24	0.20	0.19	0.20
EURIBOR	**BVAR**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
EURIBOR	**BVAR$^+$**	0.91	0.90	0.97	**0.96**	1.11	0.87	0.83	1.12	0.93	0.99
EURIBOR	**BVARf**	3.91	2.03	2.73	2.45	3.50	2.32	1.61	2.78	2.12	2.11
EURIBOR	**pVAR$^{\lambda,\tau}$**	0.82	0.68	0.90	1.00	**0.78**	0.73	0.58	0.86	0.78	**0.82**
EURIBOR	**pVAR$^{\tau}$**	**0.80**	**0.65**	**0.87**	1.09	0.80	**0.71**	**0.54**	**0.83**	**0.76**	0.83
EURIBOR	**BPVAR**	1.26	0.93	0.99	1.28	0.97	1.09	0.72	0.84	0.98	1.04
EURIBOR	**PVARj**	1.61	1.30	1.30	1.47	1.58	1.29	0.87	1.19	1.26	1.52
EURIBOR	**pVARlm$^{\lambda,\tau}$**	2.58	1.70	2.63	2.08	2.84	1.57	1.42	2.06	1.86	1.75
EURIBOR	**pVARlm$^{\tau}$**	3.34	1.78	2.63	1.99	2.58	1.63	1.50	2.13	1.93	1.83

Table 7: **N=10, G=7 setup:** One step ahead root mean squared forecast errors for different models, relative to the BVAR forecast. The first row is the absolute error of the BVAR benchmark. Values in bold denote the lowest relative RMSFE for that country.

References

Armagan, A. Dunson, D., Bajwa, W., Lee, J., and Strawn, N. (2013). Posterior consistency in linear models under shrinkage priors. *Biometrika*, 100(4):1011–1018.

Bańbura, M., Giannone, D., and Reichlin, L. (2010). Large Bayesian vector auto regressions. *Journal of Applied Econometrics*, 25(1):71–92.

Canova, F. and Ciccarelli, M. (2009). Estimating multicountry VAR models. *International Economic Review*, 50(3):929–959.

Carriero, A., Clark, T., and Marcellino, M. (2016). Common drifting volatility in large Bayesian VARs. *Journal of Business and Economic Statistics*, 34(3):375–390.

Carriero, A., Clark, T., Marcellino, M., and Mertens, E. (2021). Addressing Covid-19 outliers in BVARs with stochastic volatility. *Federal Reserve Bank of Cleveland, Working Paper No. 21-02*.

Carvalho, C., Polson, N., and Scott, J. (2010). The horseshoe estimator for sparse signals. *Biometrika*, 97(2):465–480.

Chan, J. (2020). Large Bayesian VARs: A flexible Kronecker error covariance structure. *Journal of Business and Economic Statistics*, 38(1):68–79.

Chan, J. (2021). Minnesota-type adaptive hierarchical priors for large Bayesian VARs. *International Journal of Forecasting*, 37(3):1212–1226.

Chan, J., Koop, G., and Yu, X. (2021). Large order-invariant Bayesian VARs with stochastic volatility. *Working Paper*.

Cross, J., Hou, C., and Poon, A. (2020). Macroeconomic forecasting with large Bayesian VARs: Global-local priors and the illusion of sparsity. *International Journal of Forecasting*, 36(3):899–915.

Dieppe, A., Roye van, B., and Legrand, R. (2016). The BEAR toolbox. *Working Paper Series European Central Bank*, (1934).

Doan, T., Litterman, R., and Sims, C. (1984). Forecasting and conditional projection using realistic prior distributions. *Econometric Reviews*, 3:1–100.

Feldkircher, M., Huber, F., Koop, G., and Pfarrhofer, M. (2022). Approximate bayesian inference and forecasting in huge-dimensional multicountry VARs. *International Economic Review*.

Gelman, A. (2006). Prior distributions for variance parameters in hierarchical models. *Bayesian Analysis*, 1(3):515–534.

Ghosh, P., Tang, X., Ghosh, M., and Chakrabarti, A. (2016). Asymptotic properties of Bayes risk of a general class of shrinkage priors in multiple hypothesis testing under sparsity. *Bayesian Analysis*, 11(3):753–796.

Giannone, D., Lenza, M., and Primiceri, G. (2015). Prior selection for vector autoregressions. *Review of Economics and Statistics*, 97(2):436–451.

Gneiting, T. and Raftery, A. (2007). Strictly proper scoring rules, prediction, and estimation,. *Journal of the American Statistical Association*, 102(477):359–378.

Huber, F. and Feldkircher, M. (2019). Adaptive shrinkage in Bayesian vector autoregressive models. *Journal of Business and Economic Statistics*, 37(1):27–39.

Jarocinski, 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.

Koop, G. (2013). Forecasting with medium and large Bayesian VARs. *Journal of Applied Econometrics*, 28(2):177–203.

Koop, G. and Korobilis, D. (2015). Model uncertainty in panel vector autoregressions. *European Economic Review*, 81:115–131.

Korobilis, D. (2016). Prior selection for panel vector autoregressions. *Computational Statistics and Data Analysis*, 101:110–120.

Korobilis, D. and Pettenuzzo, D. (2019). Adaptive hierarchical priors for high-dimensional Vector Autoregressions. *Journal of Econometrics*, 212(1):241–271.

Lenza, M. and Primiceri, G. (2020). How to estimate a vector autoregression after march 2020. *Journal of Applied Econometrics*.

Litterman, R. (1986). Forecasting with Bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics*, 4(1):25–38.

Makalic, E. and Schmidt, D. (2016). A simple sampler for the horseshoe estimator. *IEEE Signal Processing Letters*, 23(1):179–182.

Pesaran, H., Shin, Y., and Smith, R. (1999). Pooled mean group estimation of dynamic heterogeneous panels. *Journal of the American Statistical Association*, 94(446):621–634.

Polson, N. and Scott, J. (2012). On the half-Cauchy prior for a global scale parameter. *Bayesian Analysis*, 7(4):887–902.

Prüser, J. (2021). The horseshoe prior for time-varying parameter VARs and monetary policy. *Journal of Economic Dynamics and Control*, 129:103–188.

Prüser, J. (2022). Data-based priors for vector error correction models. *International Journal of Forecasting*.

Sims, C. and Zha, T. (1998). Bayesian methods for dynamic multivariate models. *International Economic Review*, 39(4):949–968.

van der Pas, S., Kleijn, B., and van der Vaart, A. (2014). The horseshoe estimator: Posterior concentration around nearly black vectors. *Electronic Journal of Statistics*, 8(2):2585–2618.

Zellner, A. and Hong, C. (1989). Forecasting international growth rates using bayesian shrinkage and other procedures. *Journal of Econometrics*, 40(1):183–202.

Appendix A Additional figures

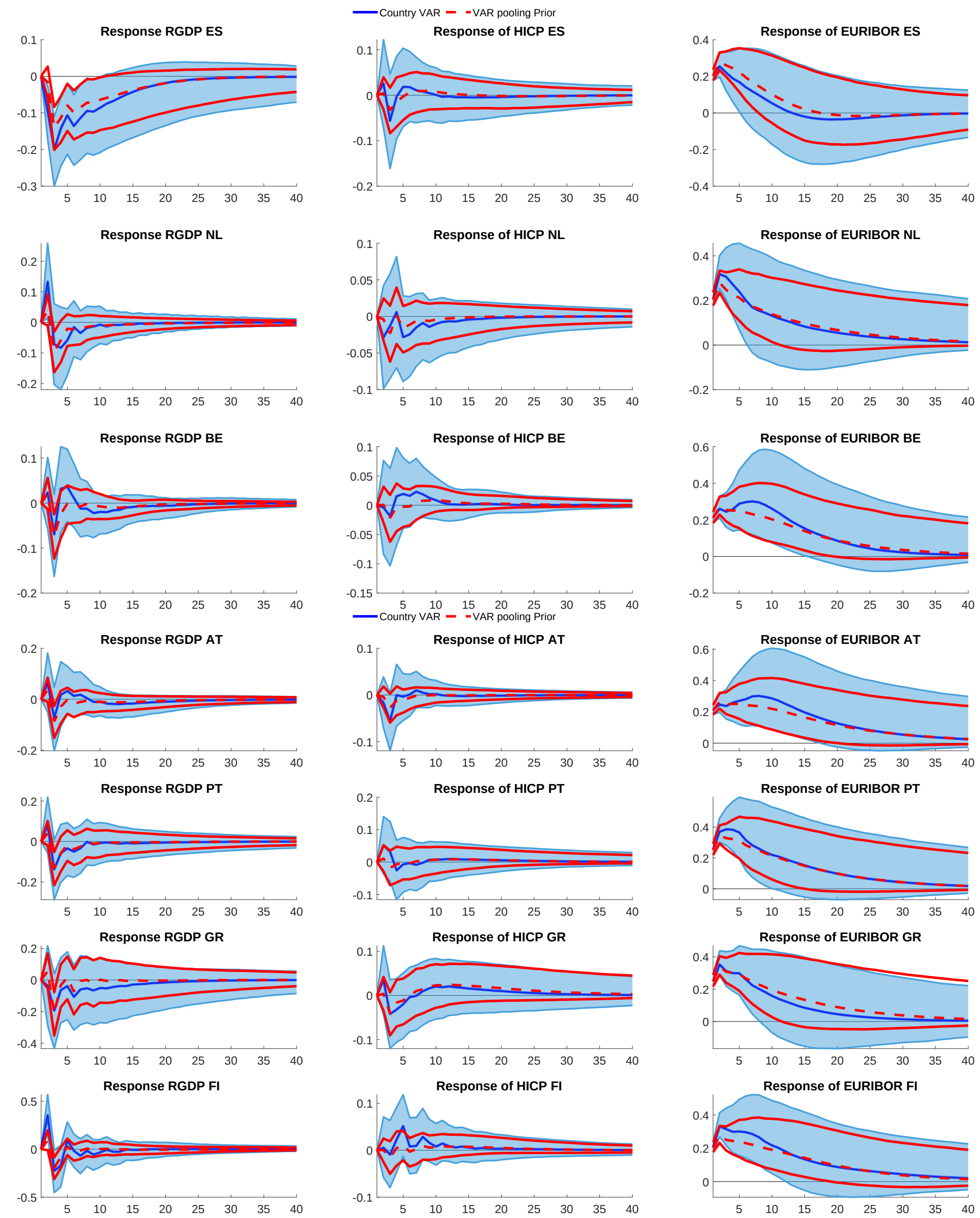

Figure A.1: Impulse responses following a one standard deviation shock to the interest rates.

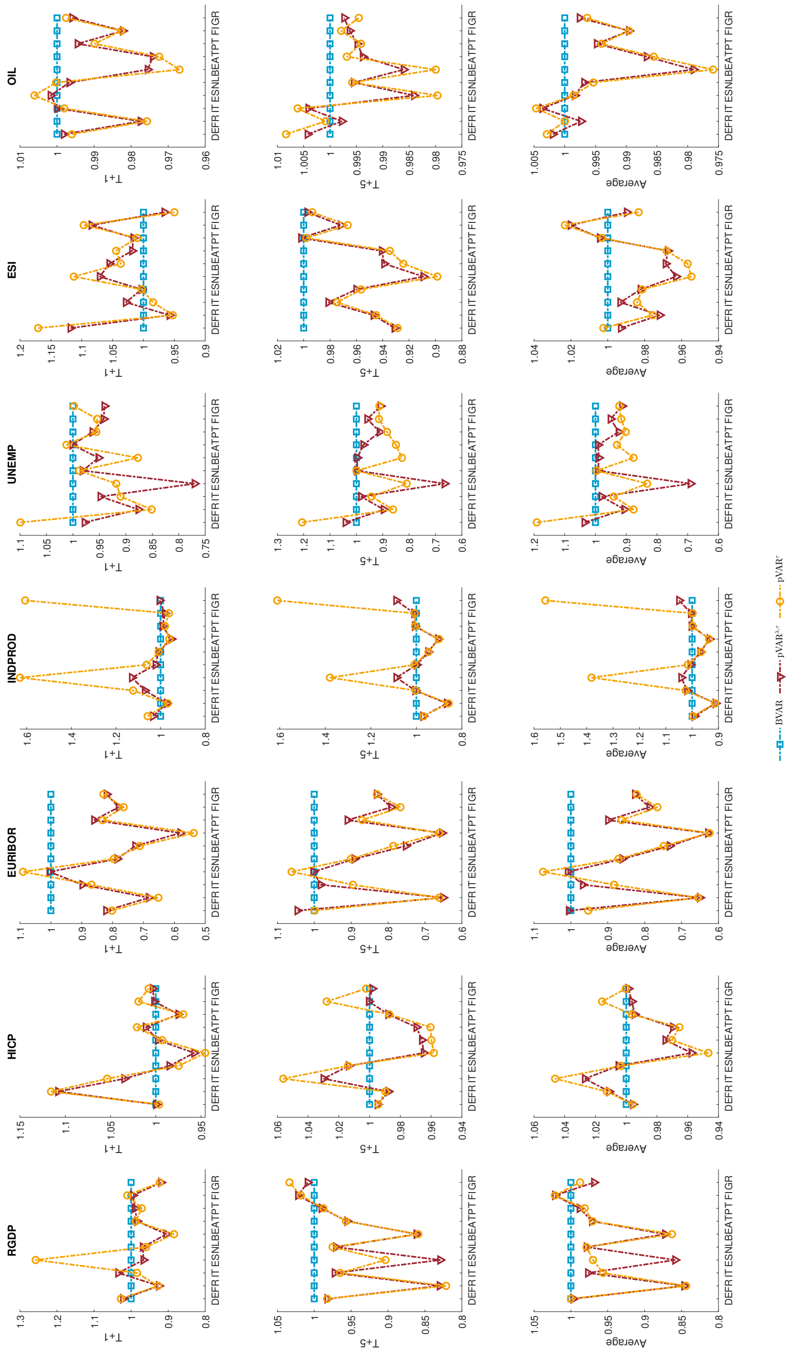

Figure A.2: **N=10, G=7 setup:** RMSFEs relative to the BVAR baseline (y-axis) versus countries on the (x-axis). Less is better.

Appendix B Additional tables

Country name	2-digit ISO code
Germany	DE
France	FR
Italy	IT
Spain	ES
The Netherlands	NL
Belgium	BE
Austria	AT
Portugal	PT
Finland	FI
Greece	GR

Table B.1: Country abbreviations used in the text.

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
RGDP	**BVAR (abs)**	0.36	0.20	0.22	0.20	0.26	0.19	0.31	0.35	0.48	0.71
	BVAR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**
	BVAR$^+$	1.05	1.06	1.07	1.08	1.01	0.99	1.00	1.03	1.00	1.02
	BVARf	1.19	1.26	1.41	1.33	1.28	1.19	1.17	1.07	1.17	1.29
	pVAR$^{\lambda,\tau}$	**0.99**	0.92	**0.94**	**0.96**	0.97	0.91	0.95	1.00	0.98	1.01
	pVAR$^{\tau}$	1.00	**0.91**	0.95	1.06	**0.96**	**0.90**	**0.94**	**0.98**	0.98	1.04
	BPVAR	0.99	1.30	1.22	1.34	1.11	1.31	1.02	1.01	**0.91**	1.06
	PVARj	1.04	1.14	1.02	1.16	1.07	1.26	0.97	1.00	1.04	1.02
	pVARlm$^{\lambda,\tau}$	1.13	1.22	1.33	1.26	1.18	1.11	1.09	1.00	1.11	1.13
	pVARlm$^{\tau}$	1.13	1.23	1.34	1.27	1.19	1.11	1.10	1.00	1.13	1.13

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
HICP	**BVAR (abs)**	0.19	0.15	0.17	0.28	0.21	0.21	0.16	0.26	0.16	0.30
	BVAR	1.00	**1.00**	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	BVAR$^+$	**0.99**	1.00	1.03	1.02	1.02	1.02	1.00	1.02	1.06	0.99
	BVARf	1.14	1.06	1.13	1.19	1.14	1.15	1.14	1.17	1.27	1.08
	pVAR$^{\lambda,\tau}$	0.99	1.02	1.00	0.96	0.97	0.98	1.00	0.98	0.99	1.00
	pVAR$^{\tau}$	0.99	1.01	1.01	0.96	**0.97**	**0.97**	**0.99**	**0.97**	0.99	1.00
	BPVAR	1.03	1.05	**0.96**	**0.92**	0.97	1.00	0.99	0.97	**0.98**	**0.92**
	PVARj	1.09	1.10	1.08	1.01	1.00	1.02	1.02	1.02	1.07	1.06
	pVARlm$^{\lambda,\tau}$	1.08	1.04	1.07	1.11	1.09	1.08	1.06	1.16	1.15	1.05
	pVARlm$^{\tau}$	1.07	1.05	1.07	1.12	1.07	1.08	1.06	1.15	1.09	1.04

Table B.2: **N=10, G=3 setup:** Continuous rank probability scores (CRPS) for the one step ahead forecast errors for different models, relative to the BVAR. Values in bold denote the lowest relative CRPS for that country among all models considered.

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
EURIBOR	**BVAR (abs)**	0.12	0.11	0.12	0.11	0.12	0.11	0.12	0.14	0.13	0.12
	BVAR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	BVAR^{+}	1.02	0.98	0.96	1.03	1.00	0.95	0.96	0.95	0.99	1.06
	BVARf	1.34	1.15	1.39	1.41	1.47	1.05	1.15	1.23	1.21	1.35
	pVAR$^{\lambda,\tau}$	**0.85**	0.90	0.90	**0.94**	0.94	0.90	0.86	0.95	0.92	1.00
	pVAR$^{\tau}$	0.86	**0.90**	0.91	0.98	0.95	**0.89**	0.85	0.93	0.92	1.03
	BPVAR	0.96	0.93	**0.85**	0.95	**0.88**	1.05	0.91	**0.81**	**0.89**	**0.97**
	PVARj	0.98	0.94	0.94	1.07	0.92	1.04	**0.79**	0.98	0.95	1.56
	pVARlm$^{\lambda,\tau}$	1.24	1.12	1.31	1.27	1.33	1.00	1.02	1.25	1.13	1.32
	pVARlm$^{\tau}$	1.26	1.13	1.33	1.33	1.35	1.01	1.04	1.28	1.19	1.33

Continued: **N=10, G=3 setup:** Continuous rank probability scores (CRPS) for the one step ahead forecast errors for different models, relative to the BVAR. Values in bold denote the lowest relative CRPS for that country among all models considered.

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
RGDP	**BVAR (abs)**	0.35	0.19	0.23	0.17	0.25	0.18	0.29	0.36	0.47	0.74
	BVAR	1.00	1.00	**1.00**	**1.00**	1.00	1.00	1.00	1.00	1.00	1.00
	BVAR$^+$	**0.94**	1.04	1.01	1.08	1.03	1.04	1.01	1.03	0.99	0.98
	BVARf	2.29	1.91	1.96	1.83	2.36	2.00	2.08	1.81	1.84	1.92
	pVAR$^{\lambda,\tau}$	1.03	**0.96**	1.05	1.01	0.99	**0.98**	0.99	0.99	1.01	**0.96**
	pVAR$^{\tau}$	1.03	0.96	1.01	1.21	**0.98**	0.99	**0.99**	0.98	1.01	0.98
	BPVAR	0.97	1.46	1.13	1.46	1.09	1.45	1.11	**0.96**	**0.93**	0.97
	PVARj	1.05	1.52	1.01	1.42	1.22	1.46	1.13	1.05	1.05	0.99
	pVARlm$^{\lambda,\tau}$	1.24	1.60	1.43	1.42	1.41	1.58	1.44	1.13	1.36	1.25
	pVARlm$^{\tau}$	1.11	1.50	1.37	1.38	1.35	1.52	1.35	1.13	1.37	1.23

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
HICP	**BVAR (abs)**	0.18	0.14	0.16	0.28	0.22	0.20	0.16	0.25	0.15	0.30
	BVAR	1.00	1.00	1.00	1.00	1.00	1.00	**1.00**	1.00	1.00	1.00
	BVAR$^+$	**0.99**	**0.90**	**0.98**	1.04	1.01	1.02	1.02	1.03	1.06	**0.95**
	BVARf	1.98	1.46	1.50	1.68	1.78	1.75	1.84	1.72	1.94	1.18
	pVAR$^{\lambda,\tau}$	1.00	1.12	1.03	0.98	0.96	1.00	1.01	0.98	**0.99**	1.00
	pVAR$^{\tau}$	1.00	1.13	1.05	0.97	**0.95**	**1.00**	1.02	**0.97**	1.01	1.00
	BPVAR	1.05	1.16	1.00	**0.95**	0.98	1.04	1.01	0.99	1.00	0.96
	PVARj	1.18	1.16	1.07	1.06	1.09	1.11	1.17	1.05	1.22	1.02
	pVARlm$^{\lambda,\tau}$	1.25	1.06	1.08	1.30	1.35	1.24	1.31	1.35	1.42	1.03
	pVARlm$^{\tau}$	1.18	1.05	1.04	1.26	1.15	1.22	1.21	1.30	1.23	1.03

Table B.3: **N=10, G=7 setup:** Continuous rank probability scores (CRPS) for the one step ahead forecast errors for different models, relative to the BVAR. Values in bold denote the lowest relative CRPS for that country among all models considered

		DE	FR	IT	ES	NL	BE	AT	PT	FI	GR
EURIBOR	**BVAR (abs)**	0.11	0.12	0.11	0.10	0.11	0.11	0.14	0.12	0.12	0.13
	BVAR	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	BVAR^{+}	0.96	0.91	0.98	0.99	1.05	0.93	0.86	1.07	0.96	1.00
	BVARf	2.69	1.95	2.44	2.36	2.80	1.89	1.64	2.24	1.97	2.00
	pVAR$^{\lambda,\tau}$	0.95	0.78	0.94	**0.98**	0.91	0.86	0.69	0.91	0.88	**0.89**
	pVAR$^{\tau}$	**0.94**	**0.77**	0.93	1.02	0.91	**0.86**	**0.68**	0.89	**0.87**	0.89
	BPVAR	1.03	0.91	**0.91**	1.06	**0.89**	1.03	0.76	**0.85**	0.91	0.94
	PVARj	1.24	1.19	1.10	1.17	1.28	1.20	0.87	1.09	1.12	1.25
	pVARlm$^{\lambda,\tau}$	1.68	1.55	1.90	1.61	1.99	1.45	1.41	1.79	1.66	1.53
	pVARlm$^{\tau}$	2.05	1.64	1.96	1.61	1.98	1.50	1.49	1.87	1.72	1.58

Continued: **N=10, G=7 setup:** Continuous rank probability scores (CRPS) for the one step ahead forecast errors for different models, relative to the BVAR. Values in bold denote the lowest relative CRPS for that country among all models considered

BVAR	Bayesian VAR with Minnesota prior
BVARf	Bayesian VAR with flat prior
$\mathbf{BVAR}^{+}$	BVAR as in Chan et al. (2021)
$\mathbf{pVAR}^{\tau}$	Pooling across country pairs with Minnesota
$\mathbf{pVAR}^{\lambda,\tau}$	Pooling across country and variable pairs with Minnesota
BPVAR	Traditional Bayesian Panel VAR
PVARj	Bayesian Panel VAR as in Jarocinski (2010)
$\mathbf{pVARlm}^{\tau}$	Pooling (τ only) without Minnesota
$\mathbf{pVARlm}^{\lambda\tau}$	Pooling without Minnesota

Table B.4: Model abbreviations.