A combined shrinkage and pooling prior for VARs

Jan Prüser and Boris Blagov

June 27, 2022

Motivation I

- VAR models are frequently used by applied macroeconomists for forecasting and structural analysis
- They are richly parametrized models which can fit the data well
- However, macroeconomic time series are rather short leading to the risk of over fitting 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))

Motivation II

- Assume we have data for a set of countries
- Countries could be homogeneous or heterogeneous
- If the countries are heterogeneous, we would estimate individual BVARs with each dataset
- 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
- It could be that groups of countries share similar characteristics in some way, while other do not
- This clustering does not have to be across all variables for all countries

Contribution

- We propose a flexible pooling prior which allows for parameter pooling across both a country dimension (two countries are completely alike) and/or variable dimension (dynamics of two or more variables across countries are alike) and at the same time provides shrinkage coefficients towards zero using the Minesota prior to achieve parsimony
- From a frequentist view point our prior penalizes both derivations from zero as well as deviations between the coefficients across countries
- How much such deviations are penalized is determined by a set of hyperparameters
- We provide a two step estimation strategy for the hyperparameter

Literature

- Zellner and Hong (1989), Jarocinski (2010), Canova and Ciccarelli (2009) and Pesaran et al. (1999) assume common dynamics
- Koop and Korobilis (2015), Korobilis (2016) cluster different group of countries
- Our prior is more flexible by allowing for homogeneous and heterogeneous dynamics simultaneously and for both pooling and shrinkage at the same time.

Results Monte Carlo

- We show that our approach is highly flexible and can adapt to different scenarios
- Thus in contrast to panel specifications we do not need to make an assumptions about the structure or homogeneity between economies prior to taking our model to the data
- Examining the convergence properties of our model we want to show that the parameter estimates are perfectly reasonable in normal sample sizes and the parameters are more precisely estimated as the sample size grows

Results Empirical Work

- We apply our model to euro area dataset of seventy variables across ten countries
- A forecasting exercise reveals that using both shrinkage and pooling leads to improvements in forecasting performance both in terms of point and density forecasting for many countries
- In contrast approaches which provide only pooling perform relatively poorly compared to single country VARs with the Minnesota prior.
- A analysis of impulse responses shows that the pooling prior appears to provide sharper inference leading to much narrower probability intervals

VAR

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

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

and in more compact form

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

where $\boldsymbol{Y}_n = (\boldsymbol{y}_{nt}, \ldots, \boldsymbol{y}_{nT})'$ and $\boldsymbol{\Gamma}_n = (\gamma_{n1}, \ldots, \gamma_{nP})'$

VAR

By vectorizing equation (2) we get

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

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 $\tilde{\boldsymbol{y}} = \text{vec}(\overline{\boldsymbol{y}_1}, \dots, \overline{\boldsymbol{y}_N})$ and $\tilde{\boldsymbol{x}} = (\boldsymbol{I}_G \otimes \boldsymbol{x}_1, \dots, \boldsymbol{I}_G \otimes \boldsymbol{x}_N)$ we can write all country VARs as one large VAR

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

We assume a conditional normal prior for $\beta \sim N(\boldsymbol{0}, \boldsymbol{V}_\beta)$. Now we can use standard linear regression results to get the conditional posterior distribution of β.

Conditional Posterior distribution of β

The conditional posterior of β is $N(\hat{\beta}, \boldsymbol{K}_{\beta}^{-1})$ with

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

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

Would the prior inverse covariance matrix $\boldsymbol{V}_{\beta}^{-1}$ be a diagonal matrix 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.

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

Minnesota Prior

The Minnesota prior is given by

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

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

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

The hyperparameters $\kappa_{1,n}$ and $\kappa_{2,n}$ control the informativeness of the prior which we both estimate.

Our Prior

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

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

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

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 prior is normal $\beta \sim N(\mathbf{0}, \boldsymbol{V}_{\beta})$ and $\boldsymbol{V}_{\beta}^{-1}$ has the following structure

$\boldsymbol{V}_{\beta}^{-1}$

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

Estimation of hyperparameter

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

$$\tau_{i,m} \sim C^{+}(0,1), \tag{11}$$

$$\lambda_{i,m,j} \sim C^{+}(0,1). \tag{12}$$

- The hyperparameter $\tau_{i,m}$ is country specific and shrinks all VAR coefficients of the county 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)

Estimation of hyperparameter

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.

Condition posterior distributions of $\tau_{i,m}$ and $\lambda_{i,m,j}$

$$\tau_{i,m}^2 \sim IG\left(\frac{K+1}{2}, \frac{1}{v_{\tau_{i,m}^2}} + 0.5\sum_{j=1}^{K}\frac{(\beta_{j+K(i-1)} - \beta_{j+Km})^2}{\lambda_{i,m,j}^2}\right), \tag{13}$$

$$\lambda_{i,m,j}^2 \sim IG\left(1, \frac{1}{v_{\lambda_{i,m,j}^2}} + 0.5\frac{(\beta_{j+K(i-1)} - \beta_{j+Km})^2}{\tau_{i,m}^2}\right), \tag{14}$$

$$v_{\tau_{i,m}^2} \sim IG\left(1, 1 + \frac{1}{\tau_{i,m}^2}\right), \tag{15}$$

$$v_{\lambda_{i,m,j}^2} \sim IG\left(1, 1 + \frac{1}{\lambda_{i,m,j}^2}\right). \tag{16}$$

$$\tag{17}$$

Estimation of hyperparameter

- In theory it is straightforward to set up a Gibbssampler and sample $\tau_{i,m}$ and $\lambda_{i,m,j}$ jointly with all other model parameters
- In practice this does not work
- The hyperparameters are highly positively correlated through the VAR coefficients and with increasing N the hyperparameters will be estimated to be close to zero
- 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)
- We obtain point estimates by using the median of these draws for $\tau_{i,m}\lambda_{i,m,j}$
- we insert these point estimates into $\boldsymbol{V}_{\beta}^{-1}$ and employ the Gibbssampler to draw the reaming parameter from their joint posterior distribution

Simulation Setup

- Countries: $\mathcal{A}$, $\mathcal{B}$, and $\mathcal{C}$
- A stable VAR with $Y = [y^i, \pi^i, r^i]$ with 2 lags is the data generating process (DGP):
$$Y_t^i = B_1^i Y_{t-1}^i + B_2^i Y_{t-2}^i + e_t^i, \tag{18}$$

with $\mathbf{e}_t^i \sim N(0, \Sigma^i)$, $i \in \{\mathcal{A}, \mathcal{B}, \mathcal{C}\}$

Consider the following cases

1. The three countries are heterogeneous
2. Two of the three countries, $\mathcal{B}$ and $\mathcal{C}$, have identical dynamics
3. A specific variable shares identical dynamics across countries and other parameters are pairwise identical

		DGP 1			DGP 2			DGP 3		
		$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$
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: Shading reflects identical parameters across pairs

Simulation Results: Hyperparameters

- Does the prior work? Do the estimated hyperparameters capture the degree of similarity?
- Remember, interplay between two hyperparameters
  1. $\lambda_{i,m,j}$ country-pair and parameter specific
  2. $\tau_{i,m}$ country-pair specific

  Let

  $$\Lambda_{i,m,j} = \lambda_{i,m,j}\tau_{i,m}$$

  and

  $$\Lambda^{i,m} = [\Lambda_{i,m,1}, \ldots, \Lambda_{i,m,j}, \ldots, \Lambda_{i,m,18}]'$$

- Low values of $\Lambda_{i,m,j}$ suggest similarity between the respective VAR parameters across i and m
- High values of $\Lambda_{i,m,j}$ point to the contrary

DGP1: Heterogeneous countries

		DGP 1			Shrinkage pairs		
		$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\Lambda^{(\mathcal{A},\mathcal{B})}$	$\Lambda^{(\mathcal{A},\mathcal{C})}$	$\Lambda^{(\mathcal{B},\mathcal{C})}$
y_t equation	y_{t-1}	0.50	0.80	0.20	0.15	0.16	0.41
	π_{t-1}	0.10	-0.20	0.20	0.15	0.15	0.38
	r_{t-1}	-0.40	-0.10	0.00	0.12	0.16	0.04
	y_{t-2}	-0.25	0.00	-0.30	0.12	0.03	0.16
	π_{t-2}	-0.20	0.20	0.00	0.25	0.15	0.07
	r_{t-2}	-0.40	0.10	0.40	0.21	0.65	0.26
π_t equation	y_{t-1}	-0.20	0.00	0.30	0.06	0.15	0.08
	π_{t-1}	0.90	0.40	0.00	0.30	0.76	0.24
	r_{t-1}	-0.10	0.10	0.30	0.06	0.13	0.06
	y_{t-2}	0.00	-0.40	0.30	0.26	0.07	0.41
	π_{t-2}	0.00	-0.30	0.50	0.16	0.29	0.62
	r_{t-2}	-0.20	0.50	-0.50	0.31	0.11	0.55
r_t equation	y_{t-1}	0.20	0.00	0.50	0.12	0.06	0.20
	π_{t-1}	0.20	0.00	0.50	0.16	0.26	0.56
	r_{t-1}	0.75	0.25	-0.20	0.31	0.83	0.27
	y_{t-2}	0.00	0.30	-0.20	0.23	0.06	0.36
	π_{t-2}	-0.30	0.20	0.00	0.53	0.29	0.12
	r_{t-2}	-0.25	-0.10	0.20	0.07	0.26	0.15

DGP2: Identical country pair

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

DGP3: Identical variable across countries

		DGP 3			Shrinkage pairs		
		$\mathcal{A}$	$\mathcal{B}$	$\mathcal{C}$	$\Lambda^{(\mathcal{A},\mathcal{B})}$	$\Lambda^{(\mathcal{A},\mathcal{C})}$	$\Lambda^{(\mathcal{B},\mathcal{C})}$
y_t equation	y_{t-1}	0.20	0.70	0.90	0.26	0.45	0.07
	π_{t-1}	0.10	-0.20	0.20	0.09	0.02	0.10
	r_{t-1}	0.00	0.00	0.00	0.02	0.02	0.02
	y_{t-2}	-0.05	-0.10	-0.30	0.03	0.09	0.07
	π_{t-2}	-0.20	0.00	0.00	0.05	0.05	0.01
	r_{t-2}	0.00	0.00	0.00	0.02	0.01	0.01
π_t equation	y_{t-1}	-0.20	0.10	0.00	0.14	0.06	0.05
	π_{t-1}	0.30	0.00	0.30	0.11	0.02	0.11
	r_{t-1}	0.00	0.00	0.00	0.02	0.02	0.02
	y_{t-2}	0.00	0.20	0.00	0.06	0.01	0.05
	π_{t-2}	-0.40	-0.30	0.10	0.04	0.25	0.17
	r_{t-2}	0.00	0.00	0.00	0.02	0.01	0.02
r_t equation	y_{t-1}	0.00	0.00	0.00	0.01	0.01	0.01
	π_{t-1}	0.00	0.00	0.00	0.01	0.01	0.01
	r_{t-1}	0.95	0.95	0.95	0.02	0.02	0.02
	y_{t-2}	0.00	0.00	0.00	0.01	0.01	0.01
	π_{t-2}	0.00	0.00	0.00	0.01	0.01	0.01
	r_{t-2}	-0.20	-0.20	-0.20	0.02	0.02	0.02

Data

- 10 European countries: Germany, France, Italy, Spain, the Netherlands, Belgium Austria, Portugal, Finland, and Greece
- For each country we have 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 data is quarterly and spans a total of 79 observations per country from 2000Q2 to 2019Q4

Forecasting Setup

- We carry out a pseudo-out-of-sample forecasting exercise with an expanding window over 40 periods
- 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
- We then expand the dataset with one quarter (up to $2011Q1$) and repeat until the end of the sample
- 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).

Forecasting Results I

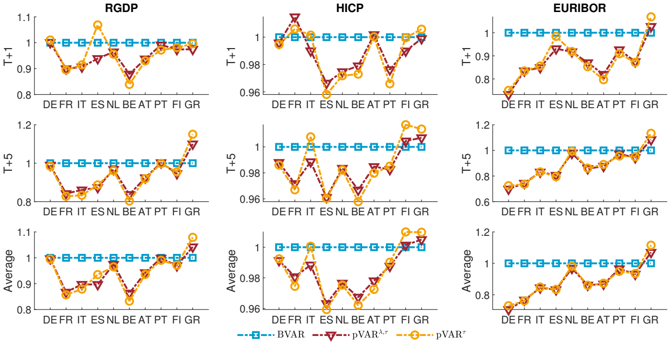

Figure: **N=10, G=3 setup:** RMSFEs relative to the BVAR baseline (y-axis) versus countries on the (x-axis). Less is better.

Forecasting Results II

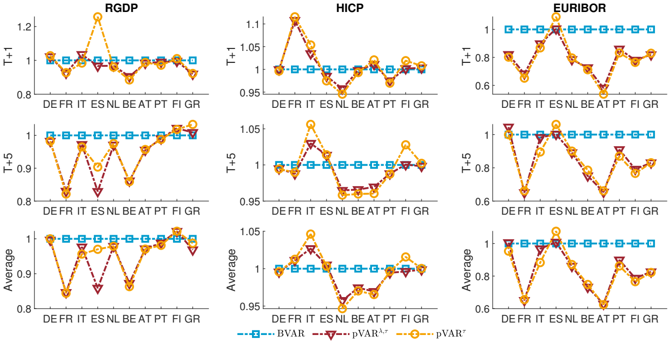

Figure: **N=10, G=7 setup:** RMSFEs relative to the BVAR baseline (y-axis) versus countries on the (x-axis). Less is better.

IRFs I

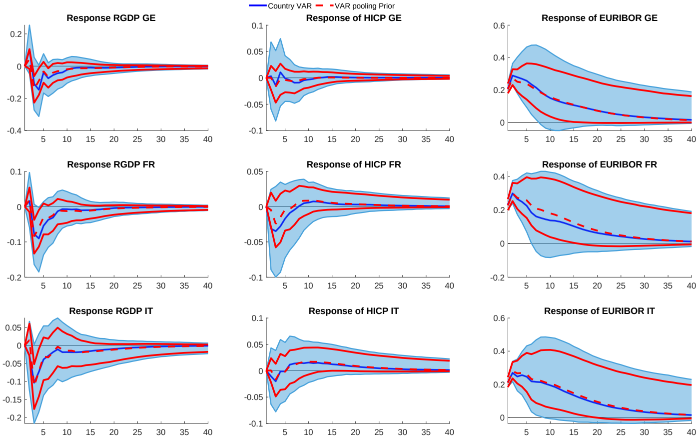

Figure: Impulse responses following a one standard deviation shock to the interest rates. Shaded areas represent one standard deviation probability intervals.

IRFs II

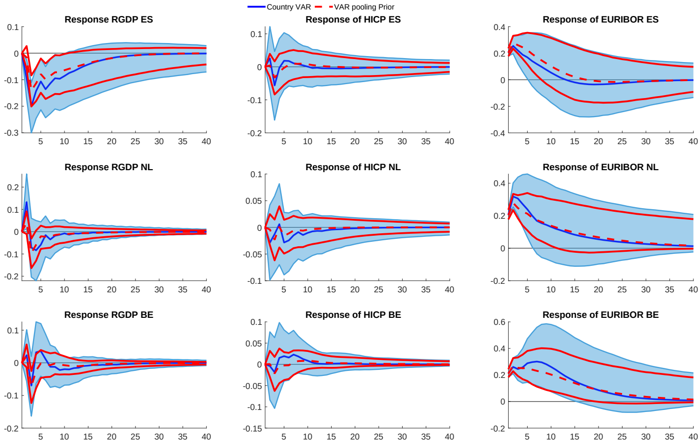

Figure: Impulse responses following a one standard deviation shock to the interest rates. Shaded areas represent one standard deviation probability intervals.

IRFs III

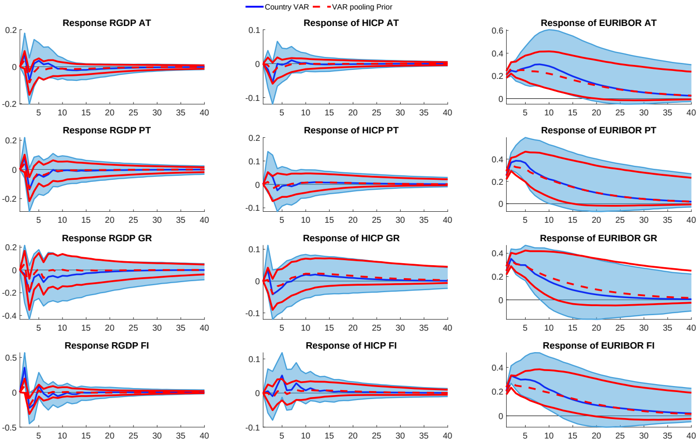

Figure: Impulse responses following a one standard deviation shock to the interest rates. Shaded areas represent one standard deviation probability intervals.

Outlook

- Use Inverse Gamma prior instead of half Cauchy prior
- Apply the model to G7 countries
- Use sign restrictions instead of different recursive identification
- Compare with more other models

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.

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.

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.

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

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.

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.

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.