Essex Finance Centre

Working Paper Series

Working Paper No14: 12-2016

"Adaptive Minnesota Prior for High-Dimensional Vector Autoregressions"

"Dimitris Korobilis and Davide Pettenuzzo"

Essex Business School, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ
Web site: http://www.essex.ac.uk/ebs/

Adaptive Minnesota Prior for High-Dimensional Vector Autoregressions

Dimitris Korobilis
University of Essex*

Davide Pettenuzzo
Brandeis University†

December 13, 2016

Abstract

We develop a novel, highly scalable estimation method for large Bayesian Vector Autoregressive models (BVARs) and employ it to introduce an "*adaptive*" version of the Minnesota prior. This flexible prior structure allows each coefficient of the VAR to have its own shrinkage intensity, which is treated as an additional parameter and estimated from the data. Most importantly, our estimation procedure does not rely on computationally intensive Markov Chain Monte Carlo (MCMC) methods, making it suitable for high-dimensional VARs with more predictors that observations. We use a Monte Carlo study to demonstrate the accuracy and computational gains of our approach. We further illustrate the forecasting performance of our new approach by applying it to a quarterly macroeconomic dataset, and find that it forecasts better than both factor models and other existing BVAR methods.

Keywords: Bayesian VARs, Minnesota prior, Large datasets, Macroeconomic forecasting

JEL Classifications: C11, C13, C32, C53

*Essex Business School, Wivenhoe Park, Colchester, CO4 3SQ, United Kingdom. d.korobilis@essex.ac.uk
†Brandeis University, Sachar International Center, 415 South St, Waltham, MA. dpettenu@brandeis.edu

1 Introduction

A number of recent contributions have highlighted the empirical success of large Bayesian Vector Autoregressive models (BVARs) in forecasting macroeconomic variables. These include, among others, Banbura et al. (2010), Koop (2011), Carriero et al. (2009, 2012), Koop and Korobilis (2013), Giannone et al. (2015), and Koop et al. (2016). The popularity of BVARs dates back to the early work of Litterman (1979) and Doan et al. (1984) and the so-called Minnesota prior, which imposes the belief that the dynamics of most macroeconomic variables can be described accurately using a univariate random walk process. In this paper, we are concerned with a specific aspect of these models, namely the choice of informativeness of their priors.[1]

We focus our attention on the very popular and successful Minnesota prior, which relies on a small number of hyperparameters - often just one - to control the overall level of informativeness of the prior. There are two general strategies to eliciting the values of these hyperparameters, which differ mainly in the way they trade-off flexibility in the specification of the prior and computational tractability. The first approach relies on the original formulation of Litterman (1979), who suggested shrinking the lags of the dependent variable and the lags of the other variables with a different intensity. This type of prior is sometimes referred to as an "asymmetric" Minnesota prior (see for example Carriero et al., 2016), and is implemented by imposing a heavier shrinkage on the lags of the other variables to reinforce the random walk nature of the prior. The flexibility of this prior translates into costly and time-intensive computations, as the posterior distribution of the VAR coefficients can only be obtained through the use of Markov Chain Monte Carlo (MCMC) methods.[2] This, in turn, makes the asymmetric Minnesota prior hard to employ in large-dimensional VARs.[3]

The second strategy focuses on a restricted version of the original Minnesota prior, where all

[1] As noted by Giannone et al. (2015), the choice of informativeness of the prior distribution bears important consequences for the performance of the BVARs, both in terms of in-sample fit as well as out-of-sample forecasting.

[2] In the pre-MCMC era, researchers used to fix the VAR covariance matrix to its OLS estimate (or even impose a diagonal structure), and then draw inference about intercepts and coefficients of lagged variables using analytical results. This is also the approach followed by Litterman (1986). See also Kadiyala and Karlsson (1997), and Koop and Korobilis (2010) for excellent reviews on BVARs.

[3] In a recent paper, Carriero et al. (2016) propose a transformation of the original n-dimensional BVAR that breaks its estimation into n univariate regressions. Their approach can handle asymmetric priors (and time-varying volatility) and yields significant computational benefits, reducing the computational complexity by a factor of n^2 relative to the existing algorithms. However, this approach also relies on MCMC methods and can still be cumbersome in high dimensions.

the coefficients are shrunk with the same intensity. The symmetry restriction paves the way to the use of the natural conjugate prior and the availability of analytical results for the parameter posterior distributions and predictive densities. On the one hand, the symmetry of the priors makes this approach particularly suited to handle large-dimensional VARs. In fact, this version of the Minnesota prior has been used successfully by both Banbura et al. (2010) and Carriero et al. (2012) to estimate BVARs with more than 120 variables. On the other hand, the symmetry of the prior brings a number of undesirable consequences. Most importantly, it imposes that the prior covariance of the coefficients in any two equations must be proportional to one another, which is a very restrictive feature. Suppose for example that the researcher wishing to employ the BVAR believes in money neutrality. This belief could be implemented by setting a very tight prior around zero for the lagged money growth coefficients in the GDP equation, and a much looser prior for the same coefficients in the other equations of the VAR. Unfortunately, the natural conjugate prior cannot be adapted to fit this situation, and may ultimately lead to an inferior model.[4]

In this paper, we propose a novel estimation procedure that combines the flexibility of the original asymmetric Minnesota prior of Litterman (1979) with the computational tractability of the natural conjugate prior. Following van den Boom et al. (2015a,b) we introduce a rotation of the original data where, one at a time, we break the dependence between one of the BVAR coefficients and all the remaining ones, which are effectively treated as nuisance parameters. This rotation allows us to reduce a high-dimensional posterior derivation into a number of independent scalar problems, each one yielding an approximated marginal posterior distribution that can be derived analytically. We exploit this framework to introduce a new and highly flexible version of the Minnesota prior, where each coefficient is allowed its own shrinkage hyperparameter. These, in turn, are estimated from the data by maximizing the marginal likelihood of the rotated regressions. We label this new prior "*adaptive Minnesota prior*", to explicitly highlight the fact that the shrinkage intensity is allowed to vary with the data, in a highly flexible way. Most importantly, thanks to the independence between the approximated marginal posteriors and the availability of analytical results, our proposed approach is highly scalable and can be easily

[4]In larger dimensional VARs, were one would expect more complex patterns of sparsity to occur, this problem will likely be exacerbated.

parallelized to fully exploit the power of modern computer clusters.[5]

We implement a thorough Monte Carlo analysis to investigate the performance of our proposed approach, and compare it against some of the existing methods including a VAR with flat priors (estimated by OLS) and the popular BVAR approach of Giannone et al. (2015). We simulate VARs of different sizes, ranging from small ($n = 3$) to large ($n = 20$) systems, and overall find that our adaptive Minnesota prior delivers the most accurate estimates of the true underlying BVAR coefficients. This is particularly true in the large VAR case, where the importance of shrinkage is higher. We also find that these improvements are highly related to the ability of our approach to implement individualized shrinkage on the various VAR coefficients. Specifically, we find that our approach generally imposes a heavier shrinkage on the VAR coefficients when their true underlying values are in fact zero. We then carry out a substantial macroeconomic forecasting exercise involving VARs with up to 40 dependent variables and five lags. We investigate the forecasting performance of seven key macroeconomic variables, comparing the adaptive Minnesota prior to various popular alternatives including, in addition to the VAR with flat priors and the BVAR approach of Giannone et al. (2015), the Dynamic factor model of Stock and Watson (2002), the Factor-Augmented VAR of Bernanke et al. (2005), the homoskedastic BVAR with asymmetric priors of Carriero et al. (2016), and the BVAR of Banbura et al. (2010). Our results are quite encouraging for the adaptive Minnesota prior, showing superior forecast improvements in many cases and comparable forecast performance in the remainder.

Our adaptive Minnesota prior is related to a number of existing approaches that use the data to infer the prior informativeness of the VAR coefficients. These include Litterman (1979), Doan et al. (1984), Banbura et al. (2010), Bloor and Matheson (2011), Giannone et al. (2015), and Huber and Feldkircher (2016). The last two papers, in particular, are the most closely related to ours. Giannone et al. (2015) integrate the choice of the prior informativeness in the estimation of a BVAR by adding an extra layer to the prior structure, and placing a separate prior on the hyperparameters, in a hierarchical fashion. As with our approach, the prior hyperparameters are estimated from the data. However, their method relies on the very restrictive natural conjugate

[5]This work made use of the High Performance Computing Cluster (HPC64) at Brandeis University.

prior, and imposes that all coefficients are shrunk with the same intensity. In contrast, our approach works with the less restrictive asymmetric Minnesota prior, and most importantly yields individualized shrinkage factors on the BVAR coefficients. Huber and Feldkircher (2016) consider a BVAR with stochastic volatility and do allow for idiosyncratic shrinkage on the BVAR coefficients and covariance terms. The idiosyncratic shrinkage factors are estimated from the data, using a multi-layered hierarchical setup and a Metropolis-within-Gibbs sampler. Relative to their approach, our adaptive Minnesota prior is significantly simpler to implement, with essentially no tuning required on the prior hyperparameters, and posterior distributions that are available analytically, making it particularly suited for high dimensional VARs. To give an idea of the computational burden involved with implementing our approach, estimation of the largest VAR considered in this paper, a 40 variable VAR with five lags, takes about 40 seconds on a 64-bit Windows PC with a 3.4 Ghz Quad-Core Intel i7-3770 processor and MatLab (R2015b).

The rest of the paper is organized as follows. Section 2 provides a general description of Bayesian VARs and introduces our adaptive Minnesota prior. Next, Section 3 goes over the estimation procedure we rely on to implement the adaptive Minnesota prior and automatically select, one coefficient at a time, the optimal level of informativeness of the prior. Section 4 describes in details our Monte Carlo exercise, while Section 5 is devoted to the macroeconomic forecasting application. Finally, Section 6 offers some concluding remarks.

2 Adaptive Minnesota Prior

Our starting point is the following n-dimensional VAR(p) model

$$\boldsymbol{y}_t = \boldsymbol{c} + \boldsymbol{A}_1\boldsymbol{y}_{t-1} + \ldots + \boldsymbol{A}_p\boldsymbol{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad t = 1, ..., T, \tag{1}$$

where $\boldsymbol{y}_t$ is an $n \times 1$ vector of time series of interest, $\boldsymbol{c}$ is an $n \times 1$ vector of intercepts, $\boldsymbol{A}_1, ..., \boldsymbol{A}_p$ are $n \times n$ matrices of coefficients on the lagged dependent variables, and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Omega})$, with $\boldsymbol{\Omega}$ an $n \times n$ covariance matrix.

We are particularly interested in the situation when the dimension of the VAR, n, is large, in which case the system in (1) suffers from the well-known curse of dimensionality. To deal with this problem, we estimate the model using the Bayesian VAR (BVAR) approach, which

deals with the over-parametrization of the VAR by imposing some informative prior beliefs on the parameters $\boldsymbol{c}$, $\boldsymbol{a}$, and $\boldsymbol{\Omega}$, where $\boldsymbol{a} = vec\left([\boldsymbol{A}_1, ..., \boldsymbol{A}_p]\right)'$ denotes the $n^2p \times 1$ vector that groups together all VAR coefficients associated with the lags of $\boldsymbol{y}_t$. We consider the following prior for the VAR coefficients:

$$\begin{aligned} c_i &\sim \mathcal{N}\left(\underline{c}_i, \underline{V}_{c_i}\right), \qquad i = 1, ..., n, \\ a_{i,j} &\sim \mathcal{N}\left(\underline{a}_{i,j}, \underline{V}_{a_{i,j}}\right), \; i = 1, ..., n, \quad j = 1, ..., np, \\ \boldsymbol{\Omega} &\sim \mathcal{IW}\left(\underline{\boldsymbol{\Psi}}, \underline{d}\right), \end{aligned} \tag{2}$$

where c_i and $a_{i,j}$ denote, respectively, the intercept and j-th coefficient in the i-th VAR equation. Note that we have implicitly assumed prior independence among the elements of $\boldsymbol{c}$ and $\boldsymbol{a}$. This assumption is frequent in Bayesian analysis, as we do not have any reason to believe that the coefficients $a_{i,j}$ should be a-priori correlated (this assumption is equivalent to assuming that the coefficient vector $\boldsymbol{a}$ has a multivariate Normal prior with diagonal covariance matrix). Despite the independence assumption for c_i and $a_{i,j}$, the prior in (2) is quite flexible and can accommodate a number of prior choices for VARs, including the very popular and successful Minnesota prior.[6] This is also our starting point. However, we introduce an important twist to the way we specify the Minnesota prior. Generally, this type of prior relies on a small number of hyperparameters - often just one - to control the overall level of informativeness of the prior for the parameter vector $\boldsymbol{a}$. In our approach, we propose instead to modify the standard Minnesota prior and allow the shrinkage intensity on each VAR coefficient to be potentially different. We thus consider the following prior for the generic $a_{i,j}$ VAR coefficient:

$$\underline{a}_{i,j} = \begin{cases} \delta_i & \text{if own first lag} \\ 0 & \text{otherwise} \end{cases}, \qquad \underline{V}_{a_{i,j}} = \lambda_{i,j}^2 \times \begin{cases} \frac{1}{l_{i,j}^2} & \text{if own lags} \\ \frac{\psi}{l_{i,j}^2} \frac{\sigma_i^2}{\sigma_k^2} & \text{otherwise} \end{cases}. \tag{3}$$

In what follows, we will refer to the prior in (3) as an "*adaptive Minnesota prior*". In fact, as it can be easily seen from (3), this prior closely resembles a Minnesota prior and shares with it a number of important features. First, the own first lag of each variable (i.e. $a_{i,i}$) is centered around δ_i, which in turn is generally set to either zero or one depending on the degree of mean reversion on the i-th variable in the VAR.[7] On the other hand, all other coefficients, including

[6] We should point out that there are some exceptions to this rule, and in fact some of the priors considered in the BVAR literature do require prior correlation among the coefficients of the same equation. See for example the sum of coefficients and unit root priors proposed by Sims (1993) and Sims and Zha (1998).

[7] With highly persistent variables entering the VAR in levels or log-levels (e.g. interest rates or log-GDP), it is typical to set $\delta_i = 1$. In all other cases, it is customary to work with $\delta_i = 0$.

the more distant own lags (i.e., $a_{i,j}$, $i \neq j$), are centered around zero. Second, the prior variance of the VAR coefficients embodies the notion that more distant lags become increasingly less important. This belief is reflected in the term $1/l_{i,j}^2$, which controls the rate at which the prior variance decreases as the lag length increases (we use $l_{i,j}$ to denote the lag length associated with the $a_{i,j}$ coefficient, which can be backed out by $l = \lfloor j/i \rfloor$).[8] The parameter ψ, usually set to be lower than one, implements additional shrinkage on the cross-equation coefficients, which are deemed to be a-priori less important in describing the dynamics of a given variable. Finally, the prior variance is rescaled by σ_i^2/σ_k^2, the ratio between the variance of variables i and k, which controls for the different scale and variability in the data (k denotes the variable that the $a_{i,j}$ coefficient belongs to, i.e. $k = j - n\,(l_{i,j} - 1)$).[9]

As we mentioned above, the key novelty of the prior specification in (3) is the presence of the prior hyperparameter $\lambda_{i,j}$, which controls the tightness of the prior distribution and is allowed to differ across the VAR coefficients. As a result, this simple modification gives rise to a highly flexible variant of the original Minnesota prior. In contrast, all the existing Minnesota prior formulations restrict the shrinkage parameter to be equal for all VAR coefficients (i.e., $\lambda_{i,j} = \lambda$). As one may expect, plugging our flexible prior into equation (1) will lead to a complex parameter posterior that can only be approximated with computationally intensive MCMC methods. In the next section we develop an alternative algorithm that allows to compute the marginal posterior of each VAR coefficient and its optimal level of prior informativeness $\lambda_{i,j}$ analytically.

3 A New Bayesian VAR methodology

In this section we introduce a novel and highly scalable algorithm for BVAR inference that relies on the adaptive Minnesota prior in (3) and at the same time automatically selects the hyperparameters $\lambda_{i,j}$ ($i = 1, ..., n$, $j = 1, ..., np$), controlling the optimal level of informativeness of the prior. Our algorithm operates by breaking the evaluation of the multivariate posterior distribution of the BVAR parameters $\{\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\Omega}\}$ into a number of one-dimensional problems, each one focusing on a single element of the parameter set. The approach proceeds in three steps. In the first step, we rewrite the original VAR model in (1) in a form that allows to estimate

[8] We denote with $\lfloor x \rfloor$ the floor of x, i.e. the smallest integer greater or equal to x.

[9] We follow standard practice and set σ_i, $i = 1, ..., n$, equal to the standard deviation of the residuals from n univariate AR(1) models estimated using OLS. See Litterman (1986) for more details.

the VAR coefficients $\{\boldsymbol{c}, \boldsymbol{a}\}$ and the elements of the covariance matrix $\boldsymbol{\Omega}$ one equation at a time. Next, in the second step, we introduce a rotation of the original data which breaks the dependence between a given coefficient of the VAR and all the remaining ones within the same equation, which are effectively treated as nuisance parameters. This rotation allows to focus on the marginal posterior distributions of the individual VAR coefficients, one at a time. In the last step of our procedure, we derive analytical formulae for the moments of the marginal posterior distribution of each VAR coefficient. Furthermore, we show how to integrate the adaptive Minnesota prior in (3) within this last step of the algorithm, and how to automatically select the prior hyperparameters $\lambda_{i,j}$ from the data.

3.1 Triangularization of the VAR

To specify a prior distribution with implied moments as described in (3), we need first to rewrite the VAR in (1) in a more convenient form. To that effect, we follow Carriero et al. (2016) and Koop et al. (2016) and decompose the VAR covariance matrix $\boldsymbol{\Omega}$ in (1) as $\boldsymbol{\Omega} = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}\left(\boldsymbol{\Sigma}\boldsymbol{\Gamma}^{-1}\right)'$, where

$$\boldsymbol{\Gamma}^{-1} = \begin{bmatrix} 1 & 0 & ... & 0 & 0 \\ \gamma_{2,1} & 1 & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & 0 \\ \gamma_{n-1,1} & ... & \gamma_{n-1,n-2} & 1 & 0 \\ \gamma_{n,1} & ... & \gamma_{n,n-2} & \gamma_{n,n-1} & 1 \end{bmatrix}, \tag{4}$$

and $\boldsymbol{\Sigma} = diag\left(\sigma_1, ..., \sigma_n\right)$. Under this decomposition the residuals of the original VAR(p) in (1) can be written using the identity $\boldsymbol{\varepsilon}_t = \boldsymbol{\Gamma}^{-1}\boldsymbol{\Sigma}\boldsymbol{u}_t$, with $\boldsymbol{u}_t \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{I}_n\right)$, which implies that the i-th row of this identity is

$$\varepsilon_{i,t} = \gamma_{i,1}\sigma_1 u_{1,t} + ... + \gamma_{i,i-1}\sigma_{i-1}u_{i-1,t} + \sigma_i u_{i,t}. \tag{5}$$

As a result, the VAR(p) in equation (1) admits the following triangular structure,

$$\begin{aligned} y_{1,t} &= c_1 + \boldsymbol{a}_{1,\cdot}\boldsymbol{X}_t + \sigma_1 u_{1t}, \\ y_{2,t} &= c_2 + \boldsymbol{a}_{2,\cdot}\boldsymbol{X}_t + \gamma_{2,1}\sigma_1 u_{1,t} + \sigma_2 u_{2,t}, \\ &\vdots \\ y_{n,t} &= c_n + \boldsymbol{a}_{n,\cdot}\boldsymbol{X}_t + \gamma_{n,1}\sigma_1 u_{1,t} + ... + \gamma_{n,n-1}\sigma_{n-1}u_{n-1,t} + \sigma_n u_{n,t}, \end{aligned} \tag{6}$$

where $\boldsymbol{a}_{i,\cdot} = [a_{i,1}, ..., a_{i,np}]$ denotes the vector of coefficients in the i-th VAR equation, and $\boldsymbol{X}_t = \left[\boldsymbol{y}'_{t-1}, ..., \boldsymbol{y}'_{t-p}\right]'$. Next, we modify the prior in (2) to exploit the triangularization of the VAR, breaking the $\mathcal{IW}$ prior for $\boldsymbol{\Omega}$ into $p(\boldsymbol{\Omega}) = p\left(\boldsymbol{\Gamma}^{-1}\right) p(\boldsymbol{\Sigma})$, and write

$$\begin{aligned} c_i &\sim \mathcal{N}\left(\underline{c}_i, \underline{V}_{c_i}\right), \qquad i = 1, ..., n, \\ a_{i,j} &\sim \mathcal{N}\left(\underline{a}_{i,j}, \underline{V}_{a_{i,j}}\right), \quad i = 1, ..., n, \quad j = 1, ..., np, \\ \gamma_{i,i'} &\sim \mathcal{N}\left(\underline{\gamma}_{i,i'}, \underline{V}_{\gamma_{i,i'}}\right), \quad i, i' = 1, ..., n, \quad i > i', \\ \sigma_i^2 &\sim \mathcal{IG}\left(\underline{\Psi}_{i,i}, \underline{d}\right), \qquad i = 1, ..., n. \end{aligned} \tag{7}$$

As noted by Carriero et al. (2016), the re-parametrization of the VAR(p) in (6)-(7) allows for estimation of the system recursively, equation-by-equation.[10] For example, consider the generic equation i, which we rewrite as

$$y_{i,t} = c_i + \boldsymbol{a}_{i,\cdot}\boldsymbol{X}_t + \gamma_{i,1}\sigma_1 u_{1,t} + ... + \gamma_{i,i-1}\sigma_{i-1}u_{i-1,t} + \sigma_i u_{i,t}, \tag{8}$$

or, more compactly,

$$y_{i,t} = \boldsymbol{Z}_{i,t}\boldsymbol{\beta}_i + \sigma_i u_{i,t}, \tag{9}$$

with $\boldsymbol{Z}_{i,t} = (\boldsymbol{X}'_t, \widetilde{u}_{1,t}, ..., \widetilde{u}_{i-1,t})$, $\boldsymbol{\beta}_i = \left(c_i, \boldsymbol{a}_{i,\cdot}, \boldsymbol{\gamma}_{i,\cdot}\right)'$, $\boldsymbol{\gamma}_{i,\cdot} = (\gamma_{i,1}, ..., \gamma_{i,i-1})$, and $\widetilde{u}_{l,t} = \sigma_l u_{l,t}$, $l = 1, ..., i-1$. Furthermore, using (7) rewrite the prior for $\left(\boldsymbol{\beta}_i, \sigma_i^2\right)$ as

$$\begin{aligned} \boldsymbol{\beta}_i &\sim \mathcal{N}\left(\underline{\boldsymbol{\beta}}_i, \underline{\boldsymbol{V}}_{\boldsymbol{\beta}_i}\right), \\ \sigma_i^2 &\sim \mathcal{IG}\left(\underline{\Psi}_{i,i}, \underline{d}\right), \end{aligned} \tag{10}$$

where

$$\underline{\boldsymbol{\beta}}_i = \begin{bmatrix} \underline{c}_i \\ \underline{a}_{i,1} \\ \vdots \\ \underline{a}_{i,np} \\ \underline{\boldsymbol{\gamma}}'_{i,\cdot} \end{bmatrix}, \qquad \underline{\boldsymbol{V}}_{\boldsymbol{\beta}_i} = \begin{bmatrix} \underline{V}_{c_i} & 0 & ... & 0 & 0 \\ 0 & \underline{V}_{a_{i,1}} & \ddots & \vdots & \vdots \\ \vdots & \ddots & \ddots & 0 & \vdots \\ 0 & ... & 0 & \underline{V}_{a_{i,np}} & 0 \\ 0 & ... & ... & 0 & \underline{\boldsymbol{V}}_{\boldsymbol{\gamma}_{i,\cdot}} \end{bmatrix}, \tag{11}$$

[10] It is worth pointing out an important feature that affects all models that rely on the triangularization in (6). Since in (7) the priors for $\boldsymbol{\Gamma}^{-1}$ and $\boldsymbol{\Sigma}$ are elicited separately, the implied prior for $\boldsymbol{\Omega}$ will change if one changes the equation ordering. As a result, different orderings of the variables in the VAR will lead to different prior specifications for $\boldsymbol{\Omega}$ and potentially different joint posteriors of the BVAR parameters $\{\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\Omega}\}$. As noted by Primiceri (2005), this problem will likely be less severe in the case as it is here in which the elements of the covariance matrix in $\boldsymbol{\Gamma}^{-1}$ do not vary with time, because the likelihood will quickly dominate the prior as the sample size increases. On this point, see also the estimation algorithms of Smith and Kohn (2002) and George et al. (2008) and discussions therein.

$\underline{\boldsymbol{\gamma}}_{i,\cdot} = \left(\underline{\gamma}_{i,1}, ..., \underline{\gamma}_{i,i-1}\right)$, and $\underline{\boldsymbol{V}}_{\boldsymbol{\gamma}_{i,\cdot}} = diag\left(\underline{V}_{\gamma_{i,1}}, ..., \underline{V}_{\gamma_{i,i-1}}\right)$.[11] Provided that all previous $i-1$ equations have been already estimated, all terms on the right hand side of (9) involving the previous equation error terms can be replaced by their estimated counterparts. Hence, the full posterior for the VAR parameters $\left\{\boldsymbol{c}, \boldsymbol{a}, \boldsymbol{\Gamma}^{-1}, \boldsymbol{\Sigma}\right\}$ can now be obtained recursively in separate blocks, one equation at a time.

To conclude this section, we note that one could exploit the triangularization in (6)-(7) to allow for a separate Minnesota shrinkage hyperparameter in each VAR equation (i.e., λ_i, $i = 1, ..., n$). Nevertheless, implementing equation-specific adaptive shrinkage within this setting would require computationally intensive MCMC methods, needed to obtain posterior predictive distributions and marginal likelihoods. This would render the approach prohibitively expensive even for medium size VARs. In the next subsection we show how to further transform the triangular VAR in (6) to obtain suitable analytical expressions that are appropriate for adaptive shrinkage in large dimensional settings.

3.2 A useful rotation of the original VAR

Reconsider now (9), the i-th equation of the triangular VAR in (6). We now describe how we further transform equation (9) in order to obtain suitable expressions for the posterior distributions of the VAR coefficients under the adaptive Minnesota prior. In order to achieve that, we rotate each VAR equation so that we can isolate the marginal effect of the j-th element of $\boldsymbol{\beta}_i$, $j = 1, ..., k_i$, and derive an analytical expression for its marginal posterior.[12] First, let us stack all time series observations of the i-th VAR equation (9) as follows,

$$\boldsymbol{y}_i = \boldsymbol{Z}_i \boldsymbol{\beta}_i + \boldsymbol{v}_i, \tag{12}$$

where $\boldsymbol{y}_i = (y_{i,1}, ..., y_{i,T})'$ is of dimension $T \times 1$, $\boldsymbol{Z}_i = \left(\boldsymbol{Z}'_{i,1}, ..., \boldsymbol{Z}'_{i,T}\right)'$ is a $T \times k_i$ matrix containing all right-hand-side variables of equation i, and $\boldsymbol{\beta}_i$ is the corresponding $k_i \times 1$ vector of VAR coefficients. Finally, $\boldsymbol{v}_i = (\sigma_i u_{i,1}, ..., \sigma_i u_{i,T})' \sim \mathcal{N}\left(\boldsymbol{0}, \sigma_i^2 \boldsymbol{I}_T\right)$.

[11] Following the Bayesian VAR tradition, only the elements within the vector $\boldsymbol{\beta}_i$ that correspond to the group of coefficients $\boldsymbol{a}_{i,\cdot}$ will feature the adaptive Minnesota moments described in the previous section, while the intercept c_i and the covariance terms $\boldsymbol{\gamma}_{i,\cdot}$ will have non-informative priors. However, it is worth noting that our algorithm can be easily extended to allow for Bayesian shrinkage on those elements as well. In high dimensional VARs, where the number of covariance terms can be extremely large, this added flexibility can be very beneficial.

[12] We use the subscript i in k_i to denote the fact that after the triangularization of the VAR the number of regressors within each VAR equation will now be different.

Consider now the generic j-th coefficient $\beta_{i,j}$ $(j = 1, ..., k_i)$. Starting with (12), we follow van den Boom et al. (2015a,b) and define the following rotation,

$$y_i^* = \boldsymbol{q}_1' \boldsymbol{y}_i, \qquad \widetilde{\boldsymbol{y}}_i = \boldsymbol{Q}_2' \boldsymbol{y}_i, \tag{13}$$

where $\boldsymbol{q}_1 = \boldsymbol{Z}_{i,j} / \|\boldsymbol{Z}_{i,j}\|$ is a $T \times 1$ unit vector in the direction of j-th column of $\boldsymbol{Z}_i$ and $\boldsymbol{Q}_2$ is an arbitrarily chosen $T \times T-1$ matrix, subject to the constraint $\boldsymbol{Q}_2 \boldsymbol{Q}_2' = \boldsymbol{I}_T - \boldsymbol{q}_1 \boldsymbol{q}_1'$. Note that since the $T \times T$ matrix $\boldsymbol{Q} = [\boldsymbol{q}_1 | \boldsymbol{Q}_2]$ is of full rank, the suggested rotation provides a one-to-one mapping between the original data $\boldsymbol{y}_i$ and the rotated data $(y_i^*, \widetilde{\boldsymbol{y}}_i)$.

Combining (12) and (13), it is possible to show that the distribution of the rotated data $(y_i^*, \widetilde{\boldsymbol{y}}_i)$ is given by[13]

$$\left[\begin{array}{c} y_i^* \\ \widetilde{\boldsymbol{y}}_i \end{array} \right] \Bigg| \boldsymbol{\beta}_i, \sigma_i^2 \sim \mathcal{N} \left(\left[\begin{array}{c} \|\boldsymbol{Z}_{i,j}\| \, \beta_{i,j} \\ 0 \end{array} \right] + \left[\begin{array}{c} \boldsymbol{q}_1' \boldsymbol{Z}_{i,(-j)} \boldsymbol{\beta}_{i,(-j)} \\ \boldsymbol{Q}_2' \boldsymbol{Z}_{i,(-j)} \boldsymbol{\beta}_{i,(-j)} \end{array} \right], \sigma_i^2 \boldsymbol{I}_T \right), \tag{14}$$

where $\boldsymbol{\beta}_{i,(-j)} = \boldsymbol{\beta_i} \setminus \beta_{i,j}$ is a $(k_i - 1) \times 1$ vector of regression coefficients and, similarly, $\boldsymbol{Z}_{i,(-j)} = \boldsymbol{Z}_i \setminus \boldsymbol{Z}_{i,j}$ is a $T \times (k_i - 1)$ matrix. The rescaled regression in (14) separates a single observation for which the scalar y_i^* depends on $\beta_{i,j}$ from the remaining $T-1$ observations, for which the vector $\widetilde{\boldsymbol{y}}_i$ is conditionally independent of the effect of $\beta_{i,j}$. Most importantly, the rescaled regression in (14) shows that the $(T-1) \times 1$ vector $\widetilde{\boldsymbol{y}}_i$ is conditionally independent of $\beta_{i,j}$. This result leads to the following expression for $p(\beta_{i,j} | \boldsymbol{y}_i)$, the marginal posterior distribution of $\beta_{i,j}$:

$$p(\beta_{i,j} | \boldsymbol{y}_i) \propto p(y_i^* | \beta_{i,j}, \widetilde{\boldsymbol{y}}_i) \, p(\beta_{i,j}). \tag{15}$$

In other words, the marginal posterior distribution of $\beta_{i,j}$ is proportional to the rotated likelihood $p(y_i^* | \beta_{i,j}, \widetilde{\boldsymbol{y}}_i)$ and the prior $p(\beta_{i,j})$.[14] Most importantly, equation (15) reveals that thank to the rotation in (13), it is now possible to conduct posterior inference on the whole vector of VAR coefficients $\boldsymbol{\beta}_i$ one element at a time. Empirically, the usefulness of this result hinges on two important conditions. First, the elements of $p(\boldsymbol{\beta}_i)$ should be a priori independent, that is, $p(\boldsymbol{\beta}_i) = \prod_{j=1}^{k_i} p(\beta_{i,j})$. Second, the rotated likelihood $p(y_i^* | \beta_{i,j}, \widetilde{\boldsymbol{y}}_i)$ should to be available in closed form. As for the first condition, we note that the prior distribution we introduced in

[13] We provide a proof of this result in the Appendix A.

[14] The result in (15) is obtained by first noting that the one-to-one mapping between $\boldsymbol{y}_i$ and $(y_i^*, \widetilde{\boldsymbol{y}}_i)$ implies that $p(\beta_{i,j} | \boldsymbol{y}_i) = p(\beta_{i,j} | y_i^*, \widetilde{\boldsymbol{y}}_i)$. Next, rewrite $p(\beta_{i,j} | y^*, \widetilde{\boldsymbol{y}}_i) = p(\beta_{i,j}, y_i^* | \widetilde{\boldsymbol{y}}_i) / p(y_i^* | \widetilde{\boldsymbol{y}}_i)$. (15) then follows from applying Bayes theorem to $p(\beta_{i,j}, y_i^* | \widetilde{\boldsymbol{y}}_i)$ and noting that: (i) $p(y_i^* | \widetilde{\boldsymbol{y}}_i)$ does not not convey any information about $\beta_{i,j}$, so can effectively be treated as a normalizing constant; (ii) $\widetilde{\boldsymbol{y}}_i$ and $\beta_{i,j}$ are conditionally independent.

equations (10)-(11) satisfies this requirement. The main challenge in using (15) for posterior inference is therefore to evaluate the rotated likelihood, and this is where we turn our attention next.

We show in Appendix A that under a natural conjugate prior for $\left(\boldsymbol{\beta}_{i,(-j)}, \sigma_i\right)$ we have that

$$\begin{aligned} p\left(y_i^* | \beta_{i,j}, \widetilde{\boldsymbol{y}}_i\right) &= \|\boldsymbol{Z}_{i,j}\| \beta_{i,j} + t_{2\overline{d}}\left(\overline{\mu}_{i,j}, \overline{\tau}_{i,j}^2\right) \\ &\approx \|\boldsymbol{Z}_{i,j}\| \beta_{i,j} + \mathcal{N}\left(\overline{\mu}_{i,j}, \overline{\tau}_{i,j}^2\right), \end{aligned} \tag{16}$$

where

$$\overline{\mu}_{i,j} = \boldsymbol{q}_1' \boldsymbol{Z}_{i,(-j)} \overline{\boldsymbol{\beta}}_{i,(-j)}, \tag{17}$$

and

$$\overline{\tau}_{i,j}^2 = \frac{\overline{\Psi}_{i,i}}{\overline{d}} \left(1 + \boldsymbol{q}_1' \boldsymbol{Z}_{i,(-j)} \overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}} \boldsymbol{Z}_{i,(-j)}' \boldsymbol{q}_1\right). \tag{18}$$

The exact formulas for the posterior moments $\overline{\boldsymbol{\beta}}_{i,(-j)}$, $\overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}}$, $\overline{\Psi}_{i,i}$, and $\overline{d}$ are standard to derive, and are also provided in Appendix A. The key concept in equations (16) is that we have chosen to approximate the Student-t predictive distribution using a Normal distribution. An immediate question is how good an approximation this will be. If σ_i^2 is known, then the formulas are exact. In other words, the rotated likelihood $p\left(y_i^* | \beta_{i,j}, \widetilde{\boldsymbol{y}}_i\right)$ is indeed normal with the moments specified above. When σ_i^2 is unknown then the approximation can still be quite accurate, and the accuracy will increases with the sample size.[15]

3.3 Implementing the adaptive Minnesota prior

Armed with an analytical expression for the rotated likelihood in (15), we are now ready to implement the adaptive Minnesota prior set forth in (3), and allow for a different shrinkage parameter $\lambda_{i,j}$ for each coefficient of the VAR. We accomplish this by specifying a Gaussian prior of the form

$$p\left(\beta_{i,j} | \lambda_{i,j}\right) \sim \mathcal{N}\left(\underline{\beta}_{i,j}, \underline{V}_{\beta_{i,j}}\right), \qquad j = 1, ..., k_i, \tag{19}$$

where we remind that $\beta_{i,j}$ can be any of the three types of elements in the vector $\boldsymbol{\beta}_i$, namely intercept c_i, covariances $\boldsymbol{\gamma}_{i,i'}$, and coefficients on lags $\boldsymbol{a}_{i,\cdot}$. Therefore, the moments $\underline{\beta}_{i,j}$ and $\underline{V}_{\beta_{i,j}}$ are adjusted accordingly depending on the type of parameter they refer to. First note that

[15] This is related to the fact that a Student-t distribution with a sufficient number of degrees of freedom - typically 100 or more - converges to a Normal distribution.

because of the approximation in (16), we can derive the marginal likelihood for y_i^* analytically. This takes the form

$$\begin{aligned} p\left(y_i^* \mid \lambda_{i,j}, \widetilde{\boldsymbol{y}}_i\right) &= \int p\left(y_i^* | \beta_{i,j}, \lambda_{i,j}, \widetilde{\boldsymbol{y}}_i\right) p\left(\beta_{i,j} | \lambda_{i,j}\right) d\beta_{i,j} \\ &= \mathcal{N}\left(y_i^* | \|\boldsymbol{Z}_{i,j}\| \underline{\beta}_{i,j} + \overline{\mu}_{i,j}, \|\boldsymbol{Z}_{i,j}\|^2 \underline{V}_{\beta_{i,j}} + \overline{\tau}_{i,j}^2\right). \end{aligned} \tag{20}$$

Next, we can choose the shrinkage parameter $\lambda_{i,j}$ that maximizes the Marginal Likelihood of the model, i.e.

$$\widehat{\lambda}_{i,j} = \arg\max_{\lambda_{i,j}} \; p\left(y_i^* \mid \lambda_{i,j}, \widetilde{\boldsymbol{y}}_i\right). \tag{21}$$

The key takeaway from (20) and (21) is that we have now derived a principled way to choose the shrinkage parameter $\lambda_{i,j}$ that maximizes the Marginal Likelihood of the model for $\beta_{i,j}$, and we can do so separately for each of the k_i VAR coefficients within equation i, and similarly for each of the n VAR equations. Finally, conditional on the optimal shrinkage intensity $\widehat{\lambda}_{i,j}$, it is straightforward to compute the marginal posterior of the VAR coefficient $\beta_{i,j}$, which is available analytically and is of the form

$$p\left(\beta_{i,j} | \widehat{\lambda}_{i,j}, \boldsymbol{y}_i\right) \sim \mathcal{N}\left(\overline{\beta}_{i,j}, \overline{V}_{\beta_{i,j}}\right), \tag{22}$$

where both $\overline{\beta}_{i,j}$ and $\overline{V}_{\beta_{i,j}}$ depend on $\widehat{\lambda}_{i,j}$ (indirectly, through $\underline{V}_{\beta_{i,j}}$), and are given by

$$\overline{V}_{\beta_{i,j}} = \frac{\overline{\tau}_{i,j}^2 \underline{V}_{\beta_{i,j}}}{\|\boldsymbol{Z}_{i,j}\|^2 \underline{V}_{\beta_{i,j}} + \overline{\tau}_{i,j}^2}, \qquad \overline{\beta}_{i,j} = \frac{\|\boldsymbol{Z}_{i,j}\| \underline{V}_{\beta_{i,j}} \left(y_i^* - \overline{\mu}_{i,j}\right)}{\|\boldsymbol{Z}_{i,j}\|^2 \underline{V}_{\beta_{i,j}} + \overline{\tau}_{i,j}^2} + \frac{\overline{\tau}_{i,j}^2 \underline{\beta}_{i,j}}{\|\boldsymbol{Z}_{i,j}\|^2 \underline{V}_{\beta_{i,j}} + \overline{\tau}_{i,j}^2}. \tag{23}$$

In practice we need to sequentially optimize $\lambda_{i,j}$ in (21) and compute the posterior mean and variance of $\beta_{i,j}$ in (23) for all $j = 1, ..., k_i$ and all VAR equations $i = 1, ..., n$. This procedure will require multiple "for loops" (i.e. as many as the number of VAR coefficients). Doing so might sound quite cumbersome, and at first the benefits relative to MCMC methods (which also involve expensive "for loops") might not be readily clear to the reader. However, it is important to clarify at this point that all elements in in (21) and (23) are scalars, which means that all formulas are extremely fast to evaluate. In addition, all the steps described in this section can be parallelized, so the algorithm can be easily adapted to fully exploit the power of modern high-performance computer clusters. These features guarantee that the proposed algorithm can be implemented very efficiently and with very low computational costs, even for large-dimensional VARs.

4 Monte Carlo Analysis

In this section we use three simulated examples to illustrate the performance of the adaptive Minnesota prior approach introduced in Section 2, and compare it against some of the existing methods, including a VAR with flat prior and no shrinkage estimated by OLS (henceforth, VAR-OLS), and a Bayesian VAR where the optimal shrinkage parameter is selected as in Giannone et al. (2015) (henceforth, BVAR-GLP). To investigate the importance of shrinkage as a function of the VAR size, we consider three different cases, each one differing in the number of variables included in the VAR. More specifically, we investigate small ($n = 3$), medium ($n = 7$), and large-scale ($n = 20$) VAR models. In all three cases, we set the number of lags to $p = 2$. We model the persistence of each variable in the VAR by setting the first own lag coefficient to be in the range $[0.4, 0.6]$, i.e.

$$\boldsymbol{A}_1 = diag\left(\rho_1, \rho_2, ..., \rho_n\right), \tag{24}$$

where $\rho_i \sim \mathcal{U}(0.4, 0.6)$, $i = 1, ..., n$. The coefficients on the subsequent own lags, $(A_l)_{i,i}$ are then generated according to the rule that $(A_l)_{i,i} = (A_1)_{i,i} / l^2$ $(l = 2, ..., p)$, implying a geometric decay in their magnitudes, with the more distant lags having a lesser impact, in the the spirit of the Minnesota prior.[16] As for the coefficients on the other lags, we set them according to the following rule:

$$(A_l)_{i,j} = \begin{cases} \mathcal{N}\left(0, \sigma_\xi^2\right) & \text{with prob } \xi \\ 0 & \text{with prob } (1-\xi) \end{cases} \qquad l = 1, ..., p, \quad i \neq j, \tag{25}$$

where $\xi \in (0, 1)$ is the probability of obtaining a non-zero coefficient. We set $\sigma_\xi^2 = 0.1$ and calibrate the inclusion probability according the the VAR size by setting $\xi = 1/(n-1)$. This implies that the number of non-zero coefficients decreases with the VAR size, a feature that is consistent with actual data.[17] Next, we decompose the covariance matrix $\boldsymbol{\Omega}$ as $\boldsymbol{\Omega} = \boldsymbol{\Pi}\boldsymbol{\Pi}'$ where $\boldsymbol{\Pi}$ is lower triangular, i.e.

$$\boldsymbol{\Pi} = \begin{bmatrix} 1 & 0 & ... & 0 \\ \pi_{2,1} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ \pi_{n,1} & ... & \pi_{n,n-1} & 1 \end{bmatrix}, \tag{26}$$

[16] When $p = 2$, these settings imply a total persistence for the variables in the VAR in the range $[0.5, 0.75]$.

[17] We restrict our attention to covariance-stationary VARs and discard all simulated DGPs producing non-stationary variables.

and set its generic element $\pi_{i,j}$ according to the following rule:

$$\pi_{i,j} = \begin{cases} \mathcal{U}(0,1) & \text{with prob } \xi_\pi \\ 0 & \text{with prob } 1-\xi_\pi \end{cases} \qquad i > j, \tag{27}$$

where we fix $\xi_\pi = 0.5$. Finally, as mentioned above, we only implement the adaptive shrinkage prior on the elements of the vector $\boldsymbol{a}$ while we specify an uninformative prior on the intercept and covariance terms, with mean equal to 0 and variance equal to 10. As for the remaining prior hyperparameters in (3), we set $\delta_i = 0.9$ $(i = 1, ..., n)$, while we set the additional shrinkage parameter ψ according to the VAR size, with $\psi = 0.1$ in the small and medium VARs, and $\psi = 0.05$ in the large VAR.

Next, for each VAR size we generate 1000 VAR(p) models of size $T = 150$ each, and evaluate the performance of the competing methods in two ways. First, we compare the optimal degree of shrinkage estimated by the BVAR methods. Second, we look at the effectiveness of the various methods in recovering the parameters of the true data generating process. To this end, for each of the approaches considered in this section, we compute the Mean Absolute Deviation (MAD), defined as

$$MAD^{(r,s)} = \frac{1}{n^2 p} \sum_{l=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \left(A_l^{(r)}\right)_{i,j} - \left(\widehat{A}_l^{(r,s)}\right)_{i,j} \right|, \tag{28}$$

where s denotes the method used, i.e. $s \in$ (VAR-OLS, BVAR-GLP, adaptive Minnesota prior), $r = 1, ..., 1000$ keeps track of the MC simulations, $n^2 p$ denotes the total number of lag coefficients in the VAR, $\left(A_l^{(r)}\right)_{i,j}$ is the true VAR coefficient from the r-th simulation, and $\left(\widehat{A}_l^{(r,s)}\right)_{i,j}$ denotes the (posterior mean of the) corresponding estimate according to method s.

Figure 1 displays the optimal degree of shrinkage we obtain by using either the BVAR-GLP procedure of Giannone et al. (2015) or our adaptive Minnesota prior approach.[18] The top three panels of the figure show the empirical distribution, computed over the 1,000 Monte Carlo simulations, of the posterior mean of λ obtained using the BVAR-GLP procedure. Moving from left to right, the mean of the empirical distribution goes from 0.424 in the small VAR case to 0.373 in the large VAR. This is consistent with Giannone et al. (2015, see Figure 1),

[18] In practice, instead of finding the maximum of the marginal likelihood function, we follow Banbura et al. (2010) and define a very fine grid. We then select the value of $\lambda_{i,j}$ that maximizes the marginal likelihood, as in (21).

who also find that the optimal degree of shrinkage decreases with the VAR size.[19] Next, the middle panels of Figure 1 display the empirical distributions of the *average* optimal shrinkage parameter $\widehat{\lambda}_{i,j}$ obtained with our Adaptive shrinkage procedure. Overall, as it was the case with the BVAR-GLP approach, we find that more shrinkage is required when the VAR size increases. The average shrinkage intensity goes from 0.462 in the small VAR case to 0.345 in the large VAR.

A notable feature of our procedure is that it yields individualized shrinkage coefficients for the different VAR coefficients. While we find that on average our approach and the BVAR-GLP method yield very similar shrinkage, we also uncover a large degree of heterogeneity among the values of $\lambda_{i,j}$ obtained using our algorithm. The bottom three panels of Figure 1 provide additional details on this regard. While it would be impossible to plot each individual hyperparameter, we show in the three bottom panels of the figure the average degree of shrinkage broken down according to whether the corresponding true VAR coefficients are either equal to zero (i.e., $E\left[\lambda_{i,j}|a_{i,j}=0\right]$) or not (i.e., $E\left[\lambda_{i,j}|a_{i,j}\neq 0\right]$). It is interesting to note how across all VAR sizes the two empirical distributions are markedly different. That is, whenever the coefficients in the VAR are equal to zero, the estimated $\widehat{\lambda}_{i,j}$ tend to be substantially smaller than the estimated prior hyperparameters associated with the non-zero VAR coefficients. In other words, our adaptive procedure generally imposes a heavier shrinkage when the underlying VAR coefficient is in fact zero.[20] In contrast, traditional Minnesota approaches rely on an "average" shrinkage parameter applied to all coefficients - even if the VAR has hundreds of thousands of coefficients.

We next look into whether the additional flexibility that our approach brings is useful in obtaining improved accuracy in the estimation phase. Figure 2 provides evidence on this regard, summarizing the accuracy, as measured by the MAD metrics, of the VAR-OLS and the two BVAR approaches in recovering the true underlying VAR coefficients. While we find no major

[19] There are a number of differences between the exercise we implement here and the simulation performed by Giannone et al. (2015). In particular, the latter work with actual macro data while we simulate our series. Also, their sample is a bit larger than ours, as is the lag length of their VAR models. In their case, they set $T = 200$ and $p = 5$. The stronger persistence in their data along with the longer lag length leads in their case to a much stronger shrinkage than what we find, especially in the case of the large VAR.

[20] Figure C.1 in Appendix C further breaks down the shrinkage intensity by own-lag and other-lag coefficients, i.e. $a_{i,i}$ vs. $a_{i,j}$ ($i = 1, ..., n$, $j = 1, ..., np$ and $i \neq j$). The latter group is further divided according to whether the true other-lag coefficient is equal to zero or not.

differences between the three methods when focusing on small VARs, the second and third panels of the figure clearly reveal the increasing beneficial effect of parameter shrinkage. In addition, it appears that for both the medium and the large VARs, the adaptive Minnesota approach is outperforming both alternative methods. In the medium VAR case, our procedure yields an average MAD of 5.51% (with a standard deviation, computed over the $1,000$ simulations, of 2.09%), while the corresponding figures for the VAR-OLS and BVAR-GLP are 11.2% and 7.59% (standard deviations are 4.07% and 2.29%, respectively). In the large VAR case, our approach yields an average MAAD of 5.66% with a standard deviation of 1.36%. In contrast, the VAR-OLS delivers an average MAD of 19.14% (standard deviation of 10.65%), while the BVAR-GLP average MAD is equal to 6.35% (standard deviation of 2.43%).

5 Macroeconomic Forecasting

5.1 Data and Models

We use 124 quarterly variables for the US spanning the period 1959Q1 to 2015Q4.[21] The data, which are obtained from the Federal Reserve Economic Data (FRED) and are available at https://fred.stlouisfed.org, cover a wide range of key macroeconomic variables that applied economists monitor regularly, such as different measures of output, prices, interest and exchange rates, and stock market performance. We provide the full list of data and their transformations in order to achieve stationarity, in Appendix B. Out of the 124 series, we further distinguish seven "variables of interest", that is, key variables of interest which we will inspect very closely in order to evaluate how well the different models perform. Consistent with previous studies (Banbura et al., 2010; Giannone et al., 2015), the first three key variables are: real gross domestic product (GDP), GDP deflator (GDPDEFL), and federal funds rate (FEDFUNDS). We also evaluate the various models in terms of an additional set of four series: total employment (PAYEMS), unemployment rate (UNRATE), consumer prices (CPIAUCSL), and the 10-year rate on government securities (GS10).

We estimate several BVAR models, and for the sake of comparability, whenever possible, we try to use the same exact prior settings. In particular, in all BVAR models we set $\delta_i = 0.9$

[21] For the variables which are originally observed at the monthly frequency, we transform them into quarterly series by computing the average of their monthly values within each quarter.

$(i = 1, ..., n)$ and calibrate the additional shrinkage parameter ψ according to the VAR size. We start by estimating a BVAR using our adaptive Minnesota prior in (3), with the prior hyperparameters $\lambda_{i,j}$ tuned optimally using the algorithm described in Section 2. We label this model with the mnemonic BVAR-KP. We next consider three variants of the BVAR with Minnesota prior, as suggested by Banbura et al. (2010) (denoted with the mnemonic BVAR-BGR), Carriero et al. (2016) (denoted BVAR-CCM), and Giannone et al. (2015) (denoted BVAR-GLP). For the BVAR-BGR approach, we estimate the common shrinkage hyperparameter λ by focusing on the same grid as in our BVAR-KP approach. As for the BVAR-CCM approach, which require setting a-priori the overall prior tightness, we follow the recommendation of Sims and Zha (1998) and set $\lambda = 0.2$. Finally, the BVAR-GLP method is fully automatic and requires no further tuning of the prior hyperparameters.[22] Along with the various BVAR models, we also consider a dynamic factor model (DFM) as in Stock and Watson (2002), estimated with OLS and with factors extracted using PCA, as well as a factor-augmented VAR (FAVAR) as in Bernanke et al. (2005), also estimated using OLS and PCA.

We estimate all competing methods on three VAR sizes, medium (seven variables), large (20 variables) and x-large (40 variables), where in all models the seven variables of interest, as detailed above, are in common. We do this to gradually assess the role of shrinkage in each competing method. Finally, we set the maximum number of lags to $p = 5$. For all models/methods that rely on the Minnesota prior (i.e., our KP approach plus the three BVAR variants) we focus on a VAR(p) specification and let the prior shrink to zero the irrelevant coefficients. For the DFM and the FAVAR, we use the Bayesian information criterion (BIC) to select the optimal number of factors (minimum is 1 and maximum is $\lfloor\sqrt{n}\rfloor$, with n the VAR size) and the optimal number of lags (ranging from one to five).

5.2 Measuring Predictive Accuracy

We use the first twenty five years of data, 1959:Q3–1984:Q4, to obtain initial parameter estimates for all models, which are then used to predict outcomes from 1985:Q1 ($h = 1$) to 1985:Q4 ($h = 4$). The next period, we include data for 1985:Q1 in the estimation sample, and use the resulting

[22] Note that the BVAR-GLP approach allows other prior variants, such as the sum-of-coefficients prior. We have estimated a number of these variants and, with the exception of the sum-of-coefficients prior, by and large the results do not seem to change significantly. As expected, with stationary data as it our case, the sum-of-coefficients prior does not work particularly well.

estimates to predict the outcomes from 1985:Q2 to 1986:Q1. We proceed recursively in this fashion until 2015:Q4, thus generating a time series of point and density forecasts for each forecast horizon h, with $h = 1, ..., 4$.[23]

Next, for each of the seven key variables listed above we summarize the precision of the h-step-ahead point forecasts for model i, relative to that from a benchmark VAR(p^*), by means of the ratio of MSFEs:

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e_{i,j,\tau+h}^2}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e_{bcmk,j,\tau+h}^2}, \tag{29}$$

where the benchmark VAR(p^*) has flat prior and is estimated using OLS, p^* denotes the largest lag length that can be estimated in a VAR with OLS and the data at hand, $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, and $e_{i,j,\tau+h}^2$ and $e_{bcmk,j,\tau+h}^2$ are the squared forecast errors of variable j at time τ and forecast horizon h associated with model i ($i \in$ {DFM,FAVAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}) and the VAR(p^*) model, respectively.[24] The point forecasts used to compute the forecast errors are obtained by averaging over the draws from the various models' h-step-ahead predictive densities. Values of $MSFE_{ijh}$ below one suggest that model i produces more accurate point forecasts than the VAR(p^*) benchmark for variable j and forecast horizon h.

We also assess the accuracy of the point forecasts of the various methods using the multivariate loss function of Christoffersen and Diebold (1998). Specifically, we compute the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model i and the WMSFE of the benchmark VAR(p^*) model as follows:

$$WMSFE_{ih} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{i,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{bcmk,\tau+h}}, \tag{30}$$

where $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$ are time $\tau + h$ weighted forecast errors of model i and the benchmark model, $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are either the (3×1) or the (7×1) vector of forecast errors for the key series we focus on, and W is either a (3×3) or a (7×7) matrix of weights. Following Carriero et al. (2011), we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the

[23] Note that when $h > 1$, point forecasts are iterated and predictive simulation is used to produce the predictive densities.

[24] That is, $p = 5$ for the medium VAR, $p = 2$ for the large VAR, and $p = 1$ for the x-large VAR.

series to be forecast.

As for the quality of the density forecasts, we follow Geweke and Amisano (2010) and compute the average log predictive likelihood differential between model i and the VAR(p^*) benchmark,

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h} \right), \tag{31}$$

where $LPL_{i,j,\tau+h}$ ($LPL_{bcmk,j,\tau+h}$) denotes model i's (benchmark's) log predictive score of variable j, computed at time $\tau + h$, i.e., the log of the h-step-ahead predictive density evaluated at the outcome. Positive values of $ALPL_{ijh}$ indicate that for variable j and forecast horizon h on average model i produces more accurate density forecasts than the benchmark model.

In order to test the statistical significance of differences in point and density forecasts, we consider pairwise tests of equal predictive accuracy (henceforth, EPA; Diebold and Mariano, 1995; West, 1996) in terms of MSFE, WMSFE, and ALPL. All EPA tests we conduct are based on a two sided test with the null hypothesis being the VAR(p^*) benchmark. We use standard normal critical values. Based on simulation evidence in Clark and McCracken (2013), when computing the variance estimator which enters the test statistic we rely on Newey and West (1987) standard errors, with truncation at lag $h-1$, and incorporate the finite sample correction due to Harvey et al. (1997). In the tables, we use ***, ** and * to denote results which are significant at the 1%, 5% and 10% levels, respectively, in favor of the model listed at the top of each column.

5.3 Forecasting Results

We now present results on the short-term forecasting performance of the various methods describesd above, based on the model sizes and forecast metrics outlined in the previous subsections. Starting with Table 1, we report the $WMSFE$ computed using all seven series (right panels), as well as the smaller subset comprising the three key variables (left panels). In particular, we compute the $WMSFE$ metric from all six competing models, namely DFM, FAVAR, BVAR-BGR, BVAR-CCM, BVAR-GLP and BVAR-KP, and for all three VAR sizes, medium, large and x-large.

A quick look at Table 1 reveals that in 20 of the 24 cases considered (that is, 3 VAR sizes, 4 forecast horizons, and 2 sets of variables) our proposed prior and estimation method generates the most accurate point forecasts. This is true whether we combine the three or the seven variables of interest in the calculation of the multivariate $WMSFE$ metric. Moreover, in the four cases where our model does not come out on top, we find its performance to be still quite strong, securing it the second place not too far behind the best model, the BVAR-CCM. Interestingly, it also appears that both our BVAR-KP and the BVAR-CCM are overall doing substantially better than the two other BVAR methods, BVAR-BGR and BVAR-GLP, which rely on the more restrictive natural conjugate prior. As it can be seen moving from the top to the bottom of the table, the wedge between the former and the latter sets of models tends to widen with the VAR size, implying that as the number of coefficients to estimate grows, the symmetry restrictions imposed by the natural conjugate prior become more and more detrimental to the overall performance of the VAR. All in all, these results help establish the benefits of using the independent Normal priors and the asymmetric version of the Minnesota prior originally proposed by Litterman (1979). They also showcase the usefulness and forecast accuracy of our adaptive Minnesota prior approach, which further improves over the BVAR-CCM method.[25]

Next, Figure 3 to Figure 5 show the cumulative sum of weighted forecast errors generated by each of the competing models we considered, over the whole out-of-sample period. The purpose of these graphs is to investigate the evolution of the forecast performance of the various models over time. That is, while the $WMSFE$ metric gives a picture of the average forecast performance over the whole evaluation period, the cumulative sums are meant to reveal if there are undesirable volatile behaviors during sub-samples of the evaluation period by any of the competing methods.[26] These weighted forecast errors are computed by multiplying the simple forecast errors of all three (or seven) variables of interest by the weighting matrix W, as defined

[25] It is worth reminding the reader at this point that the BVAR-CCM is the only BVAR method we considered that requires manual intervention in the tuning of the overall shrinkage intensity parameter λ. While in this particular setting the recommendation of Sims and Zha (1998) of setting $\lambda = 0.2$ appears to work quite well, we have also found in our experimentation that many other (reasonable) values of λ yields considerably worse forecasts for the BVAR-CCM. On the other hand, our flexible and fully automated approach of selecting the Minnesota shrinkage hyperparameter(s) $\lambda_{i,j}$ does not suffer from such shortcomings. In light of the fact that computationally our approach is considerably faster than the BVAR-CCM method, the excellent performance we observe is even more remarkable.

[26] The typical example of this would be a model producing a very low forecast error in one period, followed by an extremely large forecast error in the next period.

previously. As in Table 1, results are presented relative to the benchmark VAR(p^*) model. That is, each line depicted in the figures denotes the cumulative differential between the weighted forecasts of a given model and the benchmark VAR(p^*) model. More specifically, for each model size and for each forecast horizon h, we plot the time series

$$CSWFED_{iht} = \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(we_{bcmk,\tau+h} - we_{i,\tau+h} \right), \tag{32}$$

where i denotes the model considered, h is the forecast horizon, and $t \in \left(\underline{t}, \bar{t}\right)$ denotes the time period. Positive values indicate smaller weighted forecast errors relative to the benchmark VAR(p^*) and, as a consequence, better forecasting performance. All three figures reveal that the excellent performance of the BVAR-KP in Table 1 is not due to any specific and short-lived episodes, but is rather built gradually over the whole evaluation period, as indicated by the (almost always) increasing $CSWFED$ lines. The BVAR-CCM follows closely behind, performing as well or slightly better than the BVAR-KP at the very short forecast horizon (this happens for $h = 1$ in the large and x-large VARs), and trailing at longer forecast horizons. All other factor models and BVARs fall significantly behind.

We next dissect the forecast performance of the various methods by looking at how they fare when focusing on the individual series we are most interested on. In particular, Table 2, Table 3, and Table 4 report individual MSFE ratios for the medium, large and x-large VARs, broken down by each of the seven variables of interest and by forecast horizon. Note that, as in Table 1, these ratios are relative to the benchmark VAR(p^*) model, with values smaller than one implying an improvement over the benchmark model.[27] Overall, all three tables confirm the excellent forecast performance of our proposed BVAR-KP approach. The only case when our method is not performing well is for the one-step ahead ($h = 1$) forecasts of the consumer price index (CPIAUCSL). In this case, forecasts from the BVAR-KP are similar to those of the DFM, while the alternative BVAR methods tend to perform better. Nevertheless, it is worth noting that in the x-large case all approaches fail to improve for the one-step ahead forecasts of the consumer price index (relative to the benchmark), and in the medium and large cases

[27] Note that Table 2 for the medium VAR does not include forecasts for the FAVAR model. This is because the medium VAR only includes the seven variables of interest, which means that the equivalent FAVAR would only be including all the variables of interest and no factors. In other words, in this case the FAVAR would collapse down to a simple VAR.

the improvements are never statistically significant. With regards to all other variables, our approach provides significant improvements over the benchmark VAR(p^*), sometimes reaching improvement in MSFEs of the order of 80%.

We conclude this section by evaluating performance over the whole predictive densities. So-called density forecasts are important part of macroeconomic forecasting, since they allow us to evaluate the degree of uncertainty produced around a point forecast. Table 5, Table 6 and Table 7 report the $ALPL$ statistics for all seven series and three different VAR sizes. The evidence here appears more mixed, with no single method emerging as a clear winner. At $h = 1$ using the medium and large VARs our method is the best performing one in four out of the seven series, while the BVAR-BGR works better in four out of seven cases in the x-large case for the same $h = 1$ horizon. However, these improvements are statistically significant only for the GS10 series. Other than this clear pattern, we can generally conclude that, based on all forecast horizons, the BVAR-KP and BVAR-BGR methods are performing consistently well, with the former doing well in the medium and large VARs and the latter doing well in the large and x-large dimensions. The fact that there is no clear winner when looking at the accuracy of the whole density forecasts should not come as a surprise. All competing methods, including the DFM and FAVAR, are imposing some sort of shrinkage to the VAR coefficients. In large dimensional VARs this shrinkage can introduce a substantial bias in the mean estimates while at the same time significantly reducing their variances. This is true both regarding the variance of the estimated BVAR coefficients as well as the final estimate of the VAR covariance matrix. What the final impact of these different shrinkage procedures is, remains uncertain, as the optimal trade-off between bias and variance will depend on the application and data used. In order to get an idea of how this bias/variance trade-off works with shrinkage estimators, consider again the case of one step-ahead ($h = 1$) forecasting of the consumer price index, and in particular the x-large VAR case. While this was the only case where BVAR-KP failed to generate a point forecast that was anywhere near the unrestricted VAR-OLS benchmark, the corresponding density forecast is not only better than the VAR-OLS benchmark, but it is the best among all shrinkage-based competing methods!

6 Conclusions

We have introduced a novel methodology for estimating BVARs which features a number of desirable properties, including scalability, flexibility, computational efficiency, and forecast accuracy. Our approach works extremely well with BVARs of both small, medium, and high dimensions, delivering analytical approximations to the marginal posterior distributions of the BVAR coefficients that are very accurate. In addition, our proposed algorithm for posterior inference is multiple times faster than the conventional Bayesian VAR methods that rely on simulation methods. We exploit the flexibility of this novel approach to extend the traditional Minnesota prior in an important new direction, where we allow each VAR coefficient to have its own shrinkage intensity. The hyperparameters controlling the tightness of the priors, in turn, are estimated alongside the BVAR coefficients, in a fully automated way and using only information contained in the data. We implement a thorough Monte Carlo analysis to quantify the benefits of our novel approach, and find that it can recover very accurately the underlying VAR coefficients. We also demonstrate, using an extensive macroeconomic application, the benefits of our adaptive shrinkage procedure in preventing over-fitting of large VARs and providing excellent forecasting performance.

References

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

Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *The Quarterly Journal of Economics*, 120, 387–422.

Bloor, C. and T. Matheson (2011): "Real-time conditional forecasts with Bayesian VARs: An application to New Zealand," *The North American Journal of Economics and Finance*, 22, 26 – 42.

Carriero, A., T. Clark, and M. Marcellino (2016): "Large Vector Autoregressions

with stochastic volatility and flexible priors," Working paper 16-17, Federal Reserve Bank of Cleveland.

Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25, 400–417.

——— (2011): "Forecasting large datasets with Bayesian reduced rank multivariate models," *Journal of Applied Econometrics*, 26, 735–761.

——— (2012): "Forecasting government bond yields with large Bayesian vector autoregressions," *Journal of Banking and Finance*, 36, 2026 – 2047.

Christoffersen, P. F. and F. X. Diebold (1998): "Cointegration and Long-Horizon Forecasting," *Journal of Business & Economic Statistics*, 16, 450–458.

Clark, T. and M. McCracken (2013): "Advances in forecast evaluation," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, Elsevier, vol. 2, Part B, 1107–1201.

Diebold, F. X. and R. S. Mariano (1995): "Comparing Predictive Accuracy," *Journal of Business & Economic Statistics*, 13, 253–263.

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

George, E. I., D. Sun, and S. Ni (2008): "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553–580.

Geweke, J. and G. Amisano (2010): "Comparing and evaluating Bayesian predictive distributions of asset returns," *International Journal of Forecasting*, 26, 216 – 230.

Giannone, D., G. E. Primiceri, and M. Lenza (2015): "Prior Selection for Vector Autoregressions," *Review of Economics and Statistics*, 97, 436–451.

Harvey, D., S. Leybourne, and P. Newbold (1997): "Testing the equality of prediction mean squared errors," *International Journal of Forecasting*, 13, 281 – 291.

Huber, F. and M. Feldkircher (2016): "Adaptive shrinkage in Bayesian vector autoregressive models," *Journal of Business & Economic Statistics*, forthcoming.

Kadiyala, K. R. and S. Karlsson (1997): "Numerical methods for estimation and inference in Bayesian VAR-models," *Journal of Applied Econometrics*, 12, 99–132.

Koop, G. (2011): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28, 177–203.

Koop, G. and D. Korobilis (2010): "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends in Econometrics*, 3, 267–358.

——— (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177, 185–198.

Koop, G., D. Korobilis, and D. Pettenuzzo (2016): "Bayesian Compressed Vector Autoregressions," Brandeis University, Department of Economics and International Businesss School Working Paper.

Litterman, R. B. (1979): "Techniques of forecasting using vector autoregressions," Working Papers 115, Federal Reserve Bank of Minneapolis.

——— (1986): "Forecasting with Bayesian Vector Autoregressions: Five Years of Experience," *Journal of Business and Economic Statistics*, 4, pp. 25–38.

Newey, W. K. and K. D. West (1987): "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–708.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *The Review of Economic Studies*, 72, 821–852.

Sims, C. A. (1993): "A Nine-Variable Probabilistic Macroeconomic Forecasting Model," in *Business Cycles, Indicators, and Forecasting*, ed. by J. H. Stock and M. W. Watson, University of Chicago Press, 179–212.

Sims, C. A. and T. Zha (1998): "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39, 949–968.

Smith, M. and R. Kohn (2002): "Parsimonious Covariance Matrix Estimation for Longitudinal Data," *Journal of the American Statistical Association*, 97, 1141–1153.

Stock, J. H. and M. W. Watson (2002): "Forecasting Using Principal Components From a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–1179.

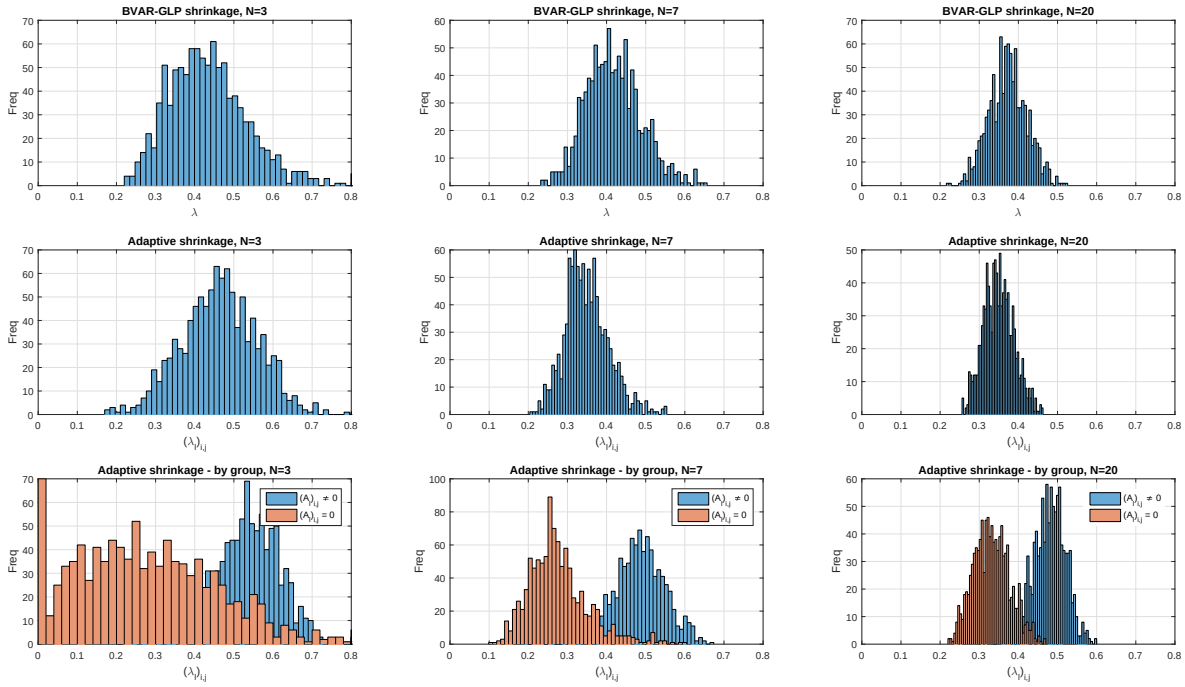
van den Boom, W., G. Reeves, and D. B. Dunson (2015a): "Quantifying Uncertainty in Variable Selection with Arbitrary Matrices," in *Proceedings of the IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing (CAMSAP), Cancun, Mexico*.

——— (2015b): "Scalable Approximations of Marginal Posteriors in Variable Selection," Working paper, Duke University Department of Statistical Science. Available at arXiv:1506.06629.

West, K. D. (1996): "Asymptotic Inference about Predictive Ability," *Econometrica*, 64, pp. 1067–1084.

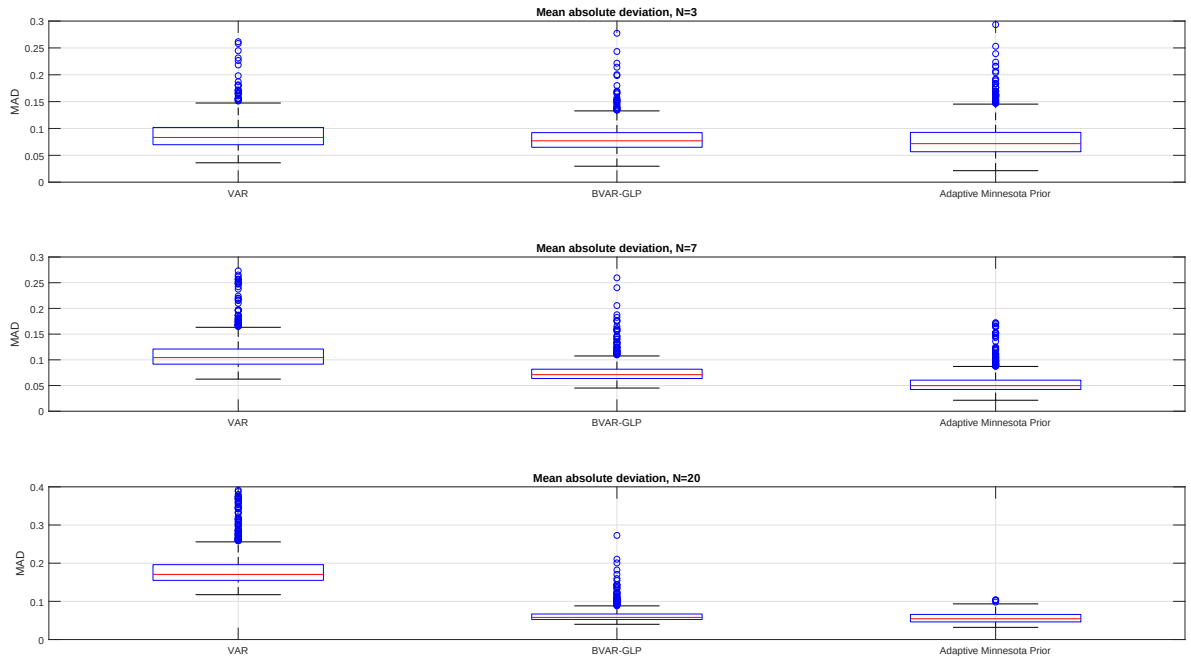
Figures and Tables

Figure 1. Monte Carlo simulation - Shrinkage intensity



The top three panels of this figure plot the empirical distribution of the estimated shrinkage parameter $\widehat{\lambda}$ obtained using the Giannone et al. (2015) approach for a small ($n = 3$), medium ($n = 7$), and large ($n = 20$) VAR(p). The middle three panels plot the empirical distribution of the estimated shrinkage intensity $\widehat{\lambda}_{i,j}$ ($i = 1, ..., n$ and $j = 1, ..., np$) estimated using our adaptive shrinkage procedure, and averaged over all VAR coefficients. The bottom three panels plot the average shrinkage intensity estimated by our adaptive procedure, broken down according to whether the corresponding VAR coefficients in the simulated data are equal to zero (i.e., $E\left[\widehat{\lambda}_{i,j}|a_{i,j} = 0\right]$) or not (i.e., $E\left[\widehat{\lambda}_{i,j}|a_{i,j} \neq 0\right]$). All empirical distributions are obtained by simulating $1,000$ VAR(p) of sample size $T = 150$ and lag length $p = 2$. See Section 4 for additional details on the design of the Monte Carlo simulation.

Figure 2. Monte Carlo simulation - Mean Absolute Deviations

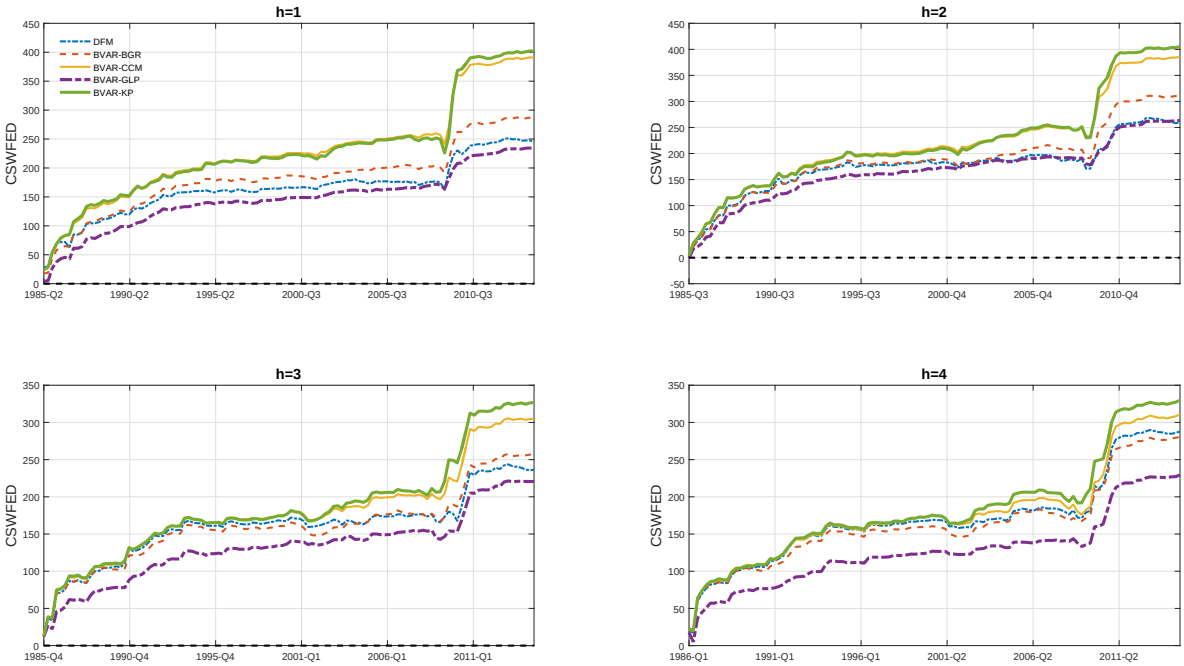


This figure reports box plots for the empirical distributions of the Mean Absolute Deviations (MAD), obtained from estimating a VAR(p) with OLS, a BVAR using the Giannone et al. (2015) (BVAR-GLP), and our adaptive Minnesota prior approach. These empirical distributions are obtained by simulating $1,000$ VAR(p) of sample size $T = 150$ and lag length $p = 2$, and for each simulated dataset computing the optimal degree of shrinkage, either using the Giannone et al. (2015) approach or our procedure. For each of the three approaches and each of the 1000 simulations we next compute the Mean Absolute Deviation (MAD), defined as

$$MAD^{(r,s)} = \frac{1}{n^2 p} \sum_{l=1}^{p} \sum_{i=1}^{n} \sum_{j=1}^{n} \left| \left(A_l^{(r)}\right)_{i,j} - \left(\widehat{A}_l^{(r,s)}\right)_{i,j} \right|$$

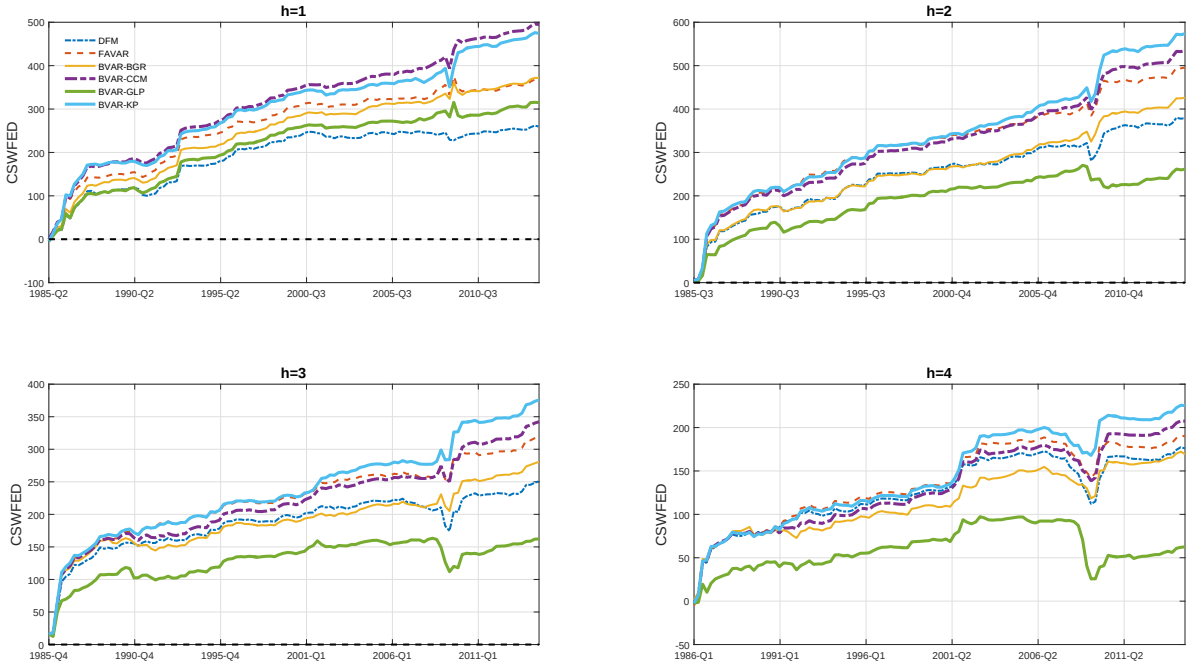
where s denotes the method used, i.e. $s \in$ (VAR-OLS, BVAR-GLP, adaptive Minnesota prior), $r = 1, ..., 1000$ keeps track of the MC simulations, $n^2 p$ denotes the total number of lag coefficients in the VAR, $\left(A_l^{(r)}\right)_{i,j}$ is the true DGP coefficient from the r-th simulation, and $\left(\widehat{A}_l^{(r,s)}\right)_{i,j}$ denotes the (posterior mean of the) corresponding estimate according to method s. Results are reported separately for small ($n = 3$), medium ($n = 7$), and large ($n = 20$) VARs.

Figure 3. Cumulative sum of weighted forecast error differentials, Medium VAR



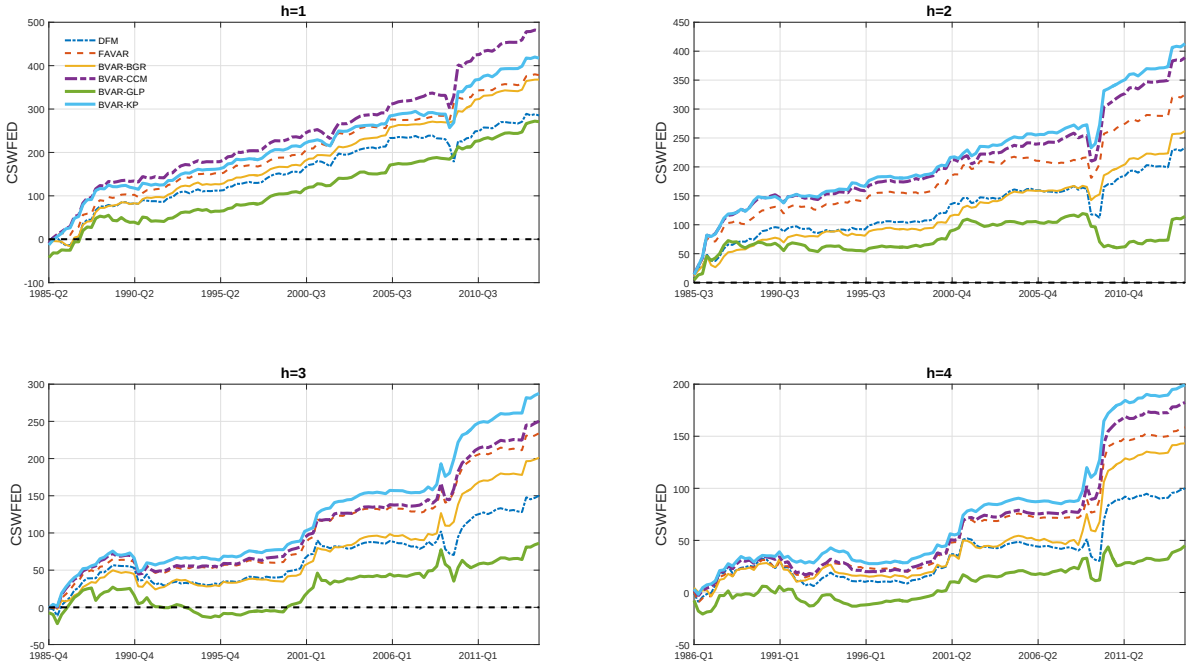
This figure plots the cumulative sum of weighted forecast errors generated by the VAR(p) model minus the cumulative sum of weighted forecast errors generated by model i for a medium size VAR. We define the weighted forecast error of model i and the VAR(p) model at time $\tau + h$ as $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$, where $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. We set $N = 7$, to focus on the following key seven series, {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}. In addition, we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. $\underline{t}$ and $\overline{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Each panel displays results for a different forecast horizon.

Figure 4. Cumulative sum of weighted forecast error differentials, Large VAR



This figure plots the cumulative sum of weighted forecast errors generated by the VAR(p^*) model minus the cumulative sum of weighted forecast errors generated by model i for a large size VAR. p^* denotes the largest lag length that can be estimated in a VAR with flat priors and the data at hand. We define the weighted forecast error of model i and the VAR(p^*) model at time $\tau + h$ as $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$, where $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. We set $N = 7$, to focus on the following key seven series, {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}. In addition, we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {VAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Each panel displays results for a different forecast horizon.

Figure 5. Cumulative sum of weighted forecast error differentials, X-large VAR



This figure plots the cumulative sum of weighted forecast errors generated by the VAR(p) model minus the cumulative sum of weighted forecast errors generated by model i for a x-large VAR. p^* denotes the largest lag length that can be estimated in a VAR with flat priors and the data at hand We define the weighted forecast error of model i and the VAR(p) model at time $\tau + h$ as $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$, where $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. We set $N = 7$, to focus on the following key seven series, {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}. In addition, we set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Each panel displays results for a different forecast horizon.

Table 1. Out-of-sample forecast performance: Multivariate results

	DFM	*FAVAR*†	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*†	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
	Medium VAR											
	3 series						*7 series*					
h=1	0.692***		0.641***	0.522***	0.722***	**0.494*** **	0.755***		0.713***	0.609***	0.765***	**0.598*** **
h=2	0.705***		0.650***	0.580***	0.711***	**0.555*** **	0.769***		0.723***	0.657***	0.766***	**0.640*** **
h=3	0.755***		0.732***	0.686***	0.768***	**0.675*** **	0.790***		0.772***	0.730***	0.805***	**0.711*** **
h=4	0.710***		0.700***	0.675***	0.749***	**0.669*** **	0.759***		0.765***	0.739***	0.808***	**0.724*** **
	Large VAR											
	3 series						*7 series*					
	DFM	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
h=1	0.643***	0.578***	0.584***	0.456***	0.671***	**0.447*** **	0.761***	0.663***	0.659***	**0.546*** **	0.711***	0.564***
h=2	0.657***	0.563***	0.626***	0.534***	0.783***	**0.483*** **	0.707***	0.615***	0.670***	0.587***	0.800***	**0.556*** **
h=3	0.744**	0.680**	0.707**	0.646***	0.800**	**0.620*** **	0.787**	0.728***	0.761***	0.709***	0.863*	**0.680*** **
h=4	0.829	0.807*	0.808*	0.778**	0.883*	**0.761** **	0.837**	0.824**	0.843**	0.808**	0.942	**0.791*** **
	X-large VAR											
	3 series						*7 series*					
	DFM	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
h=1	0.626***	0.582***	0.577***	**0.441*** **	0.732***	0.476***	0.734***	0.648***	0.658***	**0.552*** **	0.747***	0.610***
h=2	0.788**	0.714***	0.787***	0.623***	0.988	**0.580*** **	0.794***	0.711***	0.769***	0.656***	0.899	**0.634*** **
h=3	0.857**	0.793***	0.791***	0.732***	0.921	**0.704*** **	0.861**	0.783***	0.814***	0.768***	0.919	**0.734*** **
h=4	0.934	0.885	0.847**	**0.816** **	0.926	0.817**	0.904	0.848**	0.863**	0.826**	0.955	**0.809*** **

This table reports the ratio between the multivariate weighted mean squared forecast error (WMSFE) of model i and the WMSFE of the benchmark VAR(p^*) model, computed as

$$WMSFE_{ih} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{i,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} we_{bcmk,\tau+h}},$$

where p^* is the largest lag length that can be estimated in a VAR with flat priors and the data at hand, $we_{i,\tau+h} = \left(e'_{i,\tau+h} \times W \times e_{i,\tau+h}\right)$ and $we_{bcmk,\tau+h} = \left(e'_{bcmk,\tau+h} \times W \times e_{bcmk,\tau+h}\right)$ denote the weighted forecast errors of model i and the benchmark model at time $\tau + h$, $e_{i,\tau+h}$ and $e_{bcmk,\tau+h}$ are the $(N \times 1)$ vector of forecast errors, and W is an $(N \times N)$ matrix of weights. The left panels are based on $N = 3$, and focus on the following three series {FEDFUNDS, GDP, GDPDEFL} . The right panels focus on $N = 7$ and the following series {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}. We set the matrix W to be a diagonal matrix featuring on the diagonal the inverse of the variances of the series to be forecast. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,FAVAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest WMSFE and across all models for any given VAR size - forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

† The factor-augmented VAR (FAVAR) of medium size only has the seven variables of interest observed but no additional variables to extract factors from. Therefore, the FAVAR estimated on the medium size is equivalent to the VAR estimated with OLS, and for that reason we do not report its results.

Table 2. Out-of-sample point forecast performance, Medium VAR

Variable	*DFM*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
			$h=1$					$h=2$		
PAYEMS	0.876	0.667**	0.572**	0.746**	**0.566*** **	0.686	0.672**	0.573**	0.730**	**0.549*** **
CPIAUCSL	0.976	0.975	0.966	**0.956**	1.029	1.010	0.956	**0.945**	0.966	0.950
FEDFUNDS	0.575***	0.467***	0.301***	0.655***	**0.230*** **	0.566***	0.443***	0.374***	0.597***	**0.336*** **
GDP	0.846	0.821*	0.743**	0.782***	**0.738** **	0.788**	0.808*	**0.745** **	0.780***	0.745**
UNRATE	0.776	0.734**	**0.670** **	0.842**	0.691*	0.801	0.741**	0.679**	0.812*	**0.669** **
GDPDEFL	**0.821***	0.886	0.846	0.826**	0.904	0.922	0.923	0.845*	0.888	**0.816***
GS10	0.789**	0.800**	0.670***	0.766***	**0.650*** **	0.885	0.867	0.801*	0.833**	**0.801***
			$h=3$					$h=4$		
PAYEMS	0.685	0.693*	0.638**	0.751*	**0.595** **	0.690*	0.702**	0.681*	0.777*	**0.627** **
CPIAUCSL	1.057	0.950	0.962	0.978	**0.946**	1.010	1.014	1.021	**0.999**	1.019
FEDFUNDS	0.588***	0.559***	0.498***	0.666***	**0.490*** **	0.481***	0.500***	0.448***	0.603***	**0.436*** **
GDP	0.823*	0.787**	0.767**	0.800***	**0.764** **	0.920	**0.827***	0.850	0.860**	0.859
UNRATE	0.759*	0.745**	0.696**	0.815	**0.664** **	0.727*	0.746**	0.715*	0.838	**0.685** **
GDPDEFL	0.944	0.952	0.891	0.901	**0.861**	**0.852** **	0.912	0.870**	0.879**	0.854**
GS10	0.850*	0.896	0.845*	0.855*	**0.829***	**0.905**	0.962	0.911	0.923	0.905

This table reports the ratio between the MSFE of model i and the MSFE of the benchmark VAR(p) for the medium size VAR, computed as

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{i,j,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{bcmk,j,\tau+h}},$$

where $p = 5$, $e^2_{i,j,\tau+h}$ and $e^2_{bcmk,j,\tau+h}$ are the squared forecast errors of variable j at time τ and forecast horizon h generated by model i and the VAR(p) model, respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, $j \in$ {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 3. Out-of-sample point forecast performance, Large VAR

Variable	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
	$h=1$						$h=2$					
PAYEMS	1.211	0.582***	0.571**	0.481***	0.641***	**0.471*** **	0.907	0.622***	0.718**	0.578***	0.876	**0.525*** **
CPIAUCSL	1.282	1.095	1.049	1.020	**0.959**	1.245	0.820*	0.777**	0.754**	**0.740** **	0.787***	0.756**
FEDFUNDS	0.474***	0.471***	0.489***	0.295***	0.672**	**0.203*** **	0.418***	0.311***	0.450***	0.356***	0.727**	**0.288*** **
GDP	1.005	0.673	0.674*	0.592*	0.656*	**0.584***	1.105	0.863	0.872	0.748	0.955	**0.699**
UNRATE	0.770**	0.731**	0.607***	**0.563*** **	0.795**	0.571***	0.715***	0.674***	0.686***	0.601***	0.885	**0.595*** **
GDPDEFL	**0.612*** **	0.710**	0.697***	0.663***	0.686***	0.839	0.687***	0.793*	0.743**	0.694***	0.711**	**0.676*** **
GS10	0.776**	0.721**	0.773**	**0.625*** **	0.697***	0.640***	0.693***	**0.639*** **	0.734**	0.656***	0.770**	0.662**
	$h=3$						$h=4$					
PAYEMS	0.943	0.748**	0.845	0.751**	1.021	**0.678*** **	0.841	0.764**	0.818*	0.792*	1.055	**0.733** **
CPIAUCSL	0.814	0.793*	0.791*	0.799**	0.844**	**0.781***	**0.871**	0.881	0.890	0.893	0.883*	0.889
FEDFUNDS	0.508**	0.465**	0.508**	0.438**	0.692*	**0.425** **	0.510**	0.528**	0.553**	0.498**	0.673***	**0.491** **
GDP	1.013	0.888	0.873	0.822	0.962	**0.779**	1.175	1.091	1.011	1.022	1.120	**0.988**
UNRATE	0.818	0.823	0.830	0.779*	1.019	**0.752***	0.788**	0.829*	0.829**	0.774**	1.073	**0.766** **
GDPDEFL	0.778**	0.766**	0.823*	0.768**	0.760**	**0.742** **	0.951	**0.940**	1.022	0.977	0.947	0.970
GS10	0.783**	0.766***	0.824*	0.779**	0.878**	**0.765** **	**0.881***	0.901	0.974	0.898	0.965	0.906

This table reports the ratio between the MSFE of model i and the MSFE of the benchmark VAR(p^*) for the large size VAR, computed as

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{i,j,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{bcmk,j,\tau+h}},$$

where p^* is the largest lag length that can be estimated in a VAR with flat priors and the data at hand, $e^2_{i,j,\tau+h}$ and $e^2_{bcmk,j,\tau+h}$ are the squared forecast errors of variable j at time τ and forecast horizon h generated by model i and the VAR(p^*) model, respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,FAVAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, $j \in$ {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 4. Out-of-sample point forecast performance, X-large VAR

Variable	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
	$h=1$						$h=2$					
PAYEMS	1.084	0.558***	0.569***	0.554***	0.544***	**0.518*** **	0.963	0.601***	0.714**	0.621***	0.721***	**0.536*** **
CPIAUCSL	1.299	1.079	1.013	1.004	**0.942**	1.385	0.961	0.972	0.894	**0.870**	0.933	0.922
FEDFUNDS	0.444***	0.485***	0.501***	0.271***	0.621***	**0.220*** **	0.553***	0.507***	0.716***	0.466***	1.086	**0.391*** **
GDP	1.012	0.792*	0.728**	0.674**	0.756*	**0.649** **	1.227	0.962	0.913	0.784	0.852	**0.750***
UNRATE	0.640**	0.581***	**0.546*** **	0.585***	0.730*	0.567***	0.749*	**0.655*** **	0.687***	0.712**	0.828	0.665***
GDPDEFL	0.622***	0.575***	0.585***	**0.540*** **	0.901	0.769	**0.708*** **	0.799*	0.770**	0.717***	0.967	0.717***
GS10	0.762**	0.701***	0.843	**0.692** **	0.751**	0.706**	0.665***	0.631***	0.717***	**0.606*** **	0.749***	0.643***
	$h=3$						$h=4$					
PAYEMS	0.981	0.725***	0.817*	0.756**	0.872	**0.632*** **	0.918	0.769**	0.849	0.761**	0.990	**0.664*** **
CPIAUCSL	0.930	0.916	0.924	0.908*	1.023	**0.906**	0.984	**0.981**	0.986	0.989	1.002	0.997
FEDFUNDS	0.609***	0.581***	0.636***	0.558***	0.852	**0.532*** **	0.644**	0.611***	0.639***	0.600***	0.717***	**0.589*** **
GDP	1.112	0.959	0.865	0.821*	0.892	**0.795***	1.204	1.116	**0.925**	0.929	1.016	0.934
UNRATE	0.806	**0.699*** **	0.792*	0.790	0.930	0.744**	0.791**	**0.723*** **	0.827*	0.774**	1.038	0.735***
GDPDEFL	0.869*	0.889	0.944	0.891	1.091	**0.857***	1.012	**1.004**	1.107	1.033	1.169	1.045
GS10	0.779**	**0.764*** **	0.819**	0.775**	0.859*	0.771**	0.843	**0.830**	0.872	0.850	0.900	0.866

This table reports the ratio between the MSFE of model i and the MSFE of the benchmark VAR(p^*) for the X-large size VAR, computed as

$$MSFE_{ijh} = \frac{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{i,j,\tau+h}}{\sum_{\tau=\underline{t}}^{\bar{t}-h} e^2_{bcmk,j,\tau+h}},$$

where p^* is the largest lag length that can be estimated in a VAR with flat priors and the data at hand, $e^2_{i,j,\tau+h}$ and $e^2_{bcmk,j,\tau+h}$ are the squared forecast errors of variable j at time τ and forecast horizon h generated by model i and the VAR(p^*) model, respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,FAVAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, $j \in$ {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 5. Out-of-sample density forecast performance, Medium VAR

Variable	*DFM*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
			$h=1$					$h=2$		
PAYEMS	0.535	0.704	0.742	0.472*	**0.745**	0.232	0.322*	0.328*	0.285*	**0.357***
CPIAUCSL	**2.367**	1.856	1.983	1.455	2.096	0.622	0.309	0.616	-0.420	**0.828**
FEDFUNDS	0.055	0.183**	0.191	0.065	**0.211**	0.049	**0.160*****	0.108	0.127***	0.111
GDP	-0.008	0.045	0.020	0.047	**0.059**	0.086	**0.142***	0.081	0.085	0.056
UNRATE	**1.077**	0.894	1.046	0.520	0.944	0.388	0.452	0.456	0.261	**0.478**
GDPDEFL	0.028	-0.005	-0.026	**0.062**	-0.030	-0.032	-0.003	-0.007	0.011	**0.019**
GS10	0.281*	0.266*	0.311**	0.292**	**0.353****	0.077	0.084	0.097	0.116*	**0.116**
			$h=3$					$h=4$		
PAYEMS	0.334	0.282**	0.382*	0.338*	**0.388***	0.068	-0.047	0.132	-0.090	**0.178**
CPIAUCSL	0.891	0.939	0.411	**1.123**	0.697	**1.210**	-0.251	-0.214	-0.217	0.075
FEDFUNDS	-0.023	**0.063****	0.006	0.035*	0.020	0.019	**0.108*****	0.054	0.073***	0.066
GDP	0.136	0.150**	0.100	0.122**	**0.170****	-0.023	-0.001	-0.024	**0.029**	-0.048
UNRATE	**0.352**	0.153**	0.321	0.252	0.196**	**0.260**	-0.007	0.171	-0.010	0.075
GDPDEFL	-0.026	-0.013	-0.009	**0.020**	0.017	0.010	0.017	-0.003	**0.031**	0.017
GS10	0.057	0.036	0.034	**0.071**	0.069	0.023	-0.005	-0.014	**0.024**	0.009

This table reports the average log predictive likelihood (ALPL) differential between model i and the benchmark VAR(p) for the medium VAR, computed as

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h} \right),$$

where $LPL_{i,j,\tau+h}$ and $LPL_{bcmk,j,\tau+h}$ are the log predictive likelihoods of variable j at time τ and forecast horizon h generated by model i and the VAR(p^*), respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, $j \in$ {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 6. Out-of-sample density forecast performance, Large VAR

Variable	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
	$h=1$						$h=2$					
PAYEMS	0.043	0.410	0.491	0.484	0.221***	**0.492**	0.087	0.275**	0.253**	0.320**	-0.106	**0.342****
CPIAUCSL	1.584	1.460	0.964	0.925*	-0.250	**1.682**	**1.453**	1.363	0.060	1.047	-1.164	1.090
FEDFUNDS	0.291	0.289	0.251**	**0.359***	0.004	0.315	0.103*	0.123*	**0.216*****	0.147**	0.043	0.073
GDP	0.149	0.283	0.354	0.310	**0.394***	0.324	-0.080	0.037	0.077	0.114	-0.105	**0.120**
UNRATE	0.488	0.497	0.594	0.650	0.168*	**0.708**	0.182*	0.206*	**0.251*****	0.248***	-0.218	0.249***
GDPDEFL	0.105	0.100	0.125**	0.094	**0.159*****	-0.001	0.088*	0.058	0.099**	0.086*	**0.150*****	0.067
GS10	0.185**	0.219**	0.209**	0.252**	0.218***	**0.255****	0.310	**0.340**	0.299	0.309	0.264	0.309
	$h=3$						$h=4$					
PAYEMS	-0.011	0.068	0.032	0.042	-0.421	**0.141****	0.124	0.163**	-0.033	0.047	-0.912	**0.243***
CPIAUCSL	1.069	**1.709**	-0.239	0.382	-1.421	0.828	0.145	**0.477**	-0.514	-0.095	-0.711	0.474
FEDFUNDS	0.039	0.045	**0.184*****	0.097**	0.134***	-0.009	0.032	0.045	**0.172*****	0.079**	0.138***	-0.018
GDP	0.111	0.182	0.150*	0.253	0.118*	**0.292**	-0.127	-0.039	-0.128	-0.032	-0.188	**0.019**
UNRATE	**0.030**	-0.003	-0.020	-0.007	-0.230	-0.072	0.116	0.071	-0.146	0.016	-0.528	**0.138****
GDPDEFL	0.021	0.026	0.060	0.037	**0.109*****	0.022	-0.002	-0.008	0.018	-0.015	**0.058****	-0.025
GS10	0.098**	**0.104****	0.085*	0.076	0.045*	0.069	0.043	**0.047***	0.032	0.019	0.028	0.006

This table reports the average log predictive likelihood (ALPL) differential between model i and the benchmark VAR(p^*) for the large VAR, computed as

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h} \right),$$

where p^* is the largest lag length that can be estimated in a VAR with flat priors and the data at hand, $LPL_{i,j,\tau+h}$ and $LPL_{bcmk,j,\tau+h}$ are the log predictive likelihoods of variable j at time τ and forecast horizon h generated by model i and the VAR(p^*), respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,FAVAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, $j \in$ {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Table 7. Out-of-sample density forecast performance, X-large VAR

Variable	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*	*DFM*	*FAVAR*	*BVAR-BGR*	*BVAR-CCM*	*BVAR-GLP*	*BVAR-KP*
	$h=1$						$h=2$					
PAYEMS	-0.045	0.300***	**0.324*** **	0.237**	0.323***	0.234**	0.179	**0.410***	0.221**	0.335	0.133	0.356*
CPIAUCSL	0.536	0.520	0.331	1.095**	-1.996	**1.452**	0.278	0.014	-0.633	0.020	-1.955	**1.034**
FEDFUNDS	**0.369**	0.362	0.368***	0.338	0.310***	0.326	0.035	0.037	**0.176*** **	-0.011	-0.247	-0.086
GDP	0.059	0.126	**0.244** **	0.155	0.156*	0.202*	-0.137	-0.028	-0.030	-0.028	-0.037	**0.059**
UNRATE	0.575*	0.695**	**0.755** **	0.605*	0.475	0.630*	0.179*	0.301*	0.047	0.225*	0.039	**0.301** **
GDPDEFL	0.185*	0.244***	**0.267*** **	0.206**	0.100	0.094	0.062	0.032	**0.098*** **	0.040	-0.098	-0.014
GS10	0.213**	**0.260** **	0.193**	0.223*	0.209**	0.226*	0.217***	**0.233*** **	0.193***	0.216**	0.129**	0.179*
	$h=3$						$h=4$					
PAYEMS	0.083	**0.261**	0.074	0.209	-0.470	0.249	0.001	0.161	-0.352	0.105	-0.910	**0.211**
CPIAUCSL	-0.058	0.075	-1.184	-0.451	-2.532	**0.346**	0.030	**0.126**	-0.807	-0.508	-2.157	-0.020
FEDFUNDS	0.034	0.037	**0.213*** **	0.003	0.129***	-0.103	0.044*	0.051**	**0.224*** **	0.022	0.171***	-0.099
GDP	0.072	0.129	0.061	**0.181**	0.105	0.160	-0.076	-0.057	-0.260	-0.010	-0.250	**0.019**
UNRATE	**0.072**	0.045	-0.131	0.056	-0.478	0.032	0.057	0.137	-0.324	**0.180**	-0.657	0.057
GDPDEFL	0.014	0.007	**0.042**	-0.013	-0.190	-0.064	-0.016	-0.016	**0.019**	-0.035	-0.228	-0.100
GS10	0.091*	**0.108** **	0.099***	0.065	0.054	0.038	0.040	0.043	**0.057**	0.018	0.056	-0.031

This table reports the average log predictive likelihood (ALPL) differential between model i and the benchmark VAR(p^*) for the X-large VAR, computed as

$$ALPL_{ijh} = \frac{1}{\bar{t} - \underline{t} - h + 1} \sum_{\tau=\underline{t}}^{\bar{t}-h} \left(LPL_{i,j,\tau+h} - LPL_{bcmk,j,\tau+h} \right),$$

where p^* is the largest lag length that can be estimated in a VAR with flat priors and the data at hand, $LPL_{i,j,\tau+h}$ and $LPL_{bcmk,j,\tau+h}$ are the log predictive likelihoods of variable j at time τ and forecast horizon h generated by model i and the VAR(p^*), respectively. $\underline{t}$ and $\bar{t}$ denote the start and end of the out-of-sample period, $i \in$ {DFM,FAVAR,BVAR-BGR,BVAR-CCM,BVAR-GLP,BVAR-KP}, $j \in$ {PAYEMS, CPIAUCSL,FEDFUNDS, GDP, UNRATE, GDPDEFL, GS10}, and $h = 1, ..., 4$. All forecasts are generated out-of-sample using recursive estimates of the models, with the out of sample period starting in 1985:Q1 and ending in 2015:Q4. Bold numbers indicate the lowest MSFE across all models for a given variable-forecast horizon pair. * significance at the 10% level; ** significance at the 5% level; *** significance at the 1% level.

Appendix A Proofs

In this section, we provide detailed derivations and proofs for all the main results in the paper.

A.1 Proof of rescaled regression

In this subsection, we provide details on the derivation of the rescaled regression in equation (14). Start from the original regression model for the i-th equation of the VAR(p) in (1),

$$\boldsymbol{y}_i = \boldsymbol{Z}_i\boldsymbol{\beta}_i + \boldsymbol{v}_i, \tag{A.1}$$

Next, as in the text, introduce the rotation

$$y_i^* = \boldsymbol{q}_1'\boldsymbol{y}_i, \qquad \widetilde{\boldsymbol{y}}_i = \boldsymbol{Q}_2'\boldsymbol{y}_i, \tag{A.2}$$

where $\boldsymbol{q}_1 = \boldsymbol{Z}_{i,j}/\left\|\boldsymbol{Z}_{i,j}\right\|$ is a $T \times 1$ unit vector in the direction of j-th column of $\boldsymbol{Z}_i$, $\boldsymbol{Q}_2$ is an arbitrarily chosen $T \times T-1$ matrix subject to the constraint $\boldsymbol{Q}_2\boldsymbol{Q}_2' = \boldsymbol{I}_T - \boldsymbol{q}_1\boldsymbol{q}_1'$, and $\boldsymbol{Q} = [\boldsymbol{q}_1|\boldsymbol{Q}_2]$ is the full-rank $T \times T$ matrix providing a one-to-one mapping between the original data $\boldsymbol{y}_i$ and the rotated data $(y_i^*, \widetilde{\boldsymbol{y}}_i)$. Start by rewriting (A.1) as

$$\boldsymbol{y}_i = \boldsymbol{Z}_{i,j}\beta_{i,j} + \boldsymbol{Z}_{i,(-j)}\boldsymbol{\beta}_{i,(-j)} + \boldsymbol{v}_i \tag{A.3}$$

Next, pre-multiply both LHS and RHS of (A.3) by $\boldsymbol{Q}$, to obtain

$$\boldsymbol{Q}\boldsymbol{y}_i = \boldsymbol{Q}\boldsymbol{Z}_{i,j}\beta_{i,j} + \boldsymbol{Q}\boldsymbol{Z}_{i,(-j)}\boldsymbol{\beta}_{i,(-j)} + \boldsymbol{Q}\boldsymbol{v}_i, \tag{A.4}$$

or, using the fact that $\boldsymbol{Q} = [\boldsymbol{q}_1|\boldsymbol{Q}_2]$,

$$\begin{bmatrix} \boldsymbol{q}_1' \\ \boldsymbol{Q}_2' \end{bmatrix} \boldsymbol{y}_i = \begin{bmatrix} \boldsymbol{q}_1' \\ \boldsymbol{Q}_2' \end{bmatrix} \boldsymbol{Z}_{i,j}\beta_{i,j} + \begin{bmatrix} \boldsymbol{q}_1' \\ \boldsymbol{Q}_2' \end{bmatrix} \boldsymbol{Z}_{i,(-j)}\boldsymbol{\beta}_{i,(-j)} + \boldsymbol{Q}\boldsymbol{v}_i. \tag{A.5}$$

Now, combining (A.5) with (A.2) and using the definition of $\boldsymbol{q}_1$, we have that

$$\begin{bmatrix} y_i^* \\ \widetilde{\boldsymbol{y}}_i \end{bmatrix} = \begin{bmatrix} \left(\boldsymbol{Z}_{i,j}'\boldsymbol{Z}_{i,j}/\left\|\boldsymbol{Z}_{i,j}\right\|\right) \\ \boldsymbol{Q}_2'\boldsymbol{q}_1\left\|\boldsymbol{Z}_{i,j}\right\| \end{bmatrix} \beta_{i,j} + \begin{bmatrix} \boldsymbol{q}_1'\boldsymbol{Z}_{i,(-j)} \\ \boldsymbol{Q}_2'\boldsymbol{Z}_{i,(-j)} \end{bmatrix} \boldsymbol{\beta}_{i,(-j)} + \boldsymbol{Q}\boldsymbol{v}_i, \tag{A.6}$$

Further simplifications lead to

$$\begin{bmatrix} y_i^* \\ \widetilde{\boldsymbol{y}}_i \end{bmatrix} = \begin{bmatrix} \left\|\boldsymbol{Z}_{i,j}\right\|\beta_{i,j} \\ \boldsymbol{0} \end{bmatrix} + \begin{bmatrix} \boldsymbol{q}_1'\boldsymbol{Z}_{i,(-j)}\boldsymbol{\beta}_{i,(-j)} \\ \boldsymbol{Q}_2'\boldsymbol{Z}_{i,(-j)}\boldsymbol{\beta}_{i,(-j)} \end{bmatrix} + \boldsymbol{Q}\boldsymbol{v}_i, \tag{A.7}$$

where we have exploited the following two results:

1. $\left(\boldsymbol{Z}_{i,j}'\boldsymbol{Z}_{i,j}/\left\|\boldsymbol{Z}_{i,j}\right\|\right)=\left\|\boldsymbol{Z}_{i,j}\right\|$. This is due to the fact that $\boldsymbol{Z}_{i,j}'\boldsymbol{Z}_{i,j}=\left\|\boldsymbol{Z}_{i,j}\right\|^2$;

2. By definition, $\boldsymbol{Q}_2'$ and $\boldsymbol{q}_1$ are orthogonal. They all are columns of the orthogonal matrix $\boldsymbol{Q}=[\boldsymbol{q}_1|\boldsymbol{Q}_2]$), so by construction $\boldsymbol{Q}_2'\boldsymbol{q}_1\left\|\boldsymbol{Z}_{i,j}\right\|=\boldsymbol{0}$.

Finally, note that $E\left(\boldsymbol{Q}\boldsymbol{v}_i\right)=\boldsymbol{0}$ while $var\left(\boldsymbol{Q}\boldsymbol{v}_i\right)=\sigma_i^2\boldsymbol{Q}\boldsymbol{Q}'=\sigma_i^2\boldsymbol{I}_T$, which gives equation (14). ■

A.2 Approximating the rotated likelihood

In this subsection, we provide details on the results in equations (16), (17), and (18). Start by focusing on the top row of (14), and note that the conditional density $p\left(y_i^*|\boldsymbol{\beta}_i,\sigma_i^2\right)$ can be decomposed as follows

$$y_i^*=\left\|\boldsymbol{Z}_{i,j}\right\|\beta_{i,j}+y_i^+ \tag{A.8}$$

where

$$y_i^+|\boldsymbol{\beta}_{i,(-j)},\sigma_i^2\sim\mathcal{N}\left(\boldsymbol{q}_1'\boldsymbol{Z}_{i,(-j)}\boldsymbol{\beta}_{i,(-j)},\sigma_i^2\right) \tag{A.9}$$

Notice that the newly defined $p\left(y_i^+|\boldsymbol{\beta}_{i,(-j)},\sigma_i^2\right)$ can be interpreted as essentially the predictive distribution associated with the auxiliary regression that is defined in the second row of (14). This leads to the following result,

$$\begin{aligned}p\left(y_i^*|\beta_{i,j},\widetilde{\boldsymbol{y}}_i\right)&=\left\|\boldsymbol{Z}_{i,j}\right\|\beta_{i,j}+p\left(y_i^+|\widetilde{\boldsymbol{y}}_i\right)\\&=\left\|\boldsymbol{Z}_{i,j}\right\|\beta_{i,j}+\int\int p\left(y_i^+|\boldsymbol{\beta}_{i,(-j)},\sigma_i^2,\widetilde{\boldsymbol{y}}_i\right)p\left(\boldsymbol{\beta}_{i,(-j)},\sigma_i^2|\widetilde{\boldsymbol{y}}_i\right)d\boldsymbol{\beta}_{i,(-j)}d\sigma_i^2\end{aligned} \tag{A.10}$$

The key to solving (A.10) it to compute the integral in the second row of the equation, which in turn will depend on the prior distribution adopted for $p\left(\boldsymbol{\beta}_{i,(-j)},\sigma_i^2\right)$. There are many alternatives available for this. To stay consistent with the rest of our approach, we have chosen to rely on a natural conjugate Minnesota-type prior, which can cope well with the potentially high dimension of the $(k_i-1)\times 1$ vector of VAR coefficients $\beta_{i,(-j)}$, and at the same time yields a closed-form expression for the predictive density in (A.10). In particular, we specify

$$\begin{aligned}\boldsymbol{\beta}_{i,(-j)}|\lambda_{i,(-j)},\sigma_i^2&\sim\mathcal{N}\left(\underline{\boldsymbol{\beta}}_{i,(-j)},\sigma_i^2\underline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}}\right)\\\sigma_i^2&\sim\mathcal{IG}\left(\underline{\Psi}_{(i,i)},\underline{d}\right)\end{aligned} \tag{A.11}$$

where we have made explicit that the prior for $\boldsymbol{\beta}_{i,(-j)}$ depends also on the shrinkage parameter $\lambda_{i,(-j)}$, which enter through the prior variance $\underline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}}$. The key difference between the prior

in (A.11) and the prior for the individual VAR coefficients $\beta_{i,j}$ that we specified in (19) is the reliance in (A.11) on a common shrinkage coefficient $\lambda_{i,(-j)}$.[28] Our motivation for this choice is that the sole purpose of the auxiliary regression is to compute the predictive density in (A.10), which is only the first step in implementing the adaptive shrinkage prior we discuss in the main body of the paper.

Continuing on, it is easy to show that the posterior distribution $p\left(\boldsymbol{\beta}_{i,(-j)}, \sigma_i^2 | \lambda_{i,(-j)}, \widetilde{\boldsymbol{y}}_i\right)$ also belongs to the normal-inverse-gamma family, and is given by

$$\begin{aligned} \boldsymbol{\beta}_{i,(-j)} | \lambda_{i,(-j)}, \sigma_i^2, \widetilde{\boldsymbol{y}}_i &\sim \mathcal{N}\left(\overline{\boldsymbol{\beta}}_{i,(-j)}, \sigma_i^2 \overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}}\right) \\ \sigma_i^2 | \widetilde{\boldsymbol{y}}_i &\sim \mathcal{IG}\left(\overline{\Psi}_{i,i}, \overline{d}\right) \end{aligned} \tag{A.12}$$

where $\overline{d} = \underline{d} + (T-1)/2$,

$$\overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}} = \left(\underline{\boldsymbol{V}}^{-1}_{\boldsymbol{\beta}_{i,(-j)}} + \boldsymbol{Z}'_{i,(-j)} \boldsymbol{Q}_2 \boldsymbol{Q}'_2 \boldsymbol{Z}_{i,(-j)}\right)^{-1}, \tag{A.13}$$

$$\overline{\boldsymbol{\beta}}_{i,(-j)} = \overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}} \left(\underline{\boldsymbol{V}}^{-1}_{\boldsymbol{\beta}_{i,(-j)}} \underline{\boldsymbol{\beta}}_{i,(-j)} + \boldsymbol{Z}'_{i,(-j)} \boldsymbol{Q}_2 \widetilde{\boldsymbol{y}}_i\right), \tag{A.14}$$

and

$$\overline{\Psi}_{i,i} = \underline{\Psi}_{i,i} + \frac{1}{2}\left(\widetilde{\boldsymbol{y}}'_i \widetilde{\boldsymbol{y}}_i + \underline{\boldsymbol{\beta}}'_{i,(-j)} \underline{\boldsymbol{V}}^{-1}_{\boldsymbol{\beta}_{i,(-j)}} \underline{\boldsymbol{\beta}}_{i,(-j)} - \overline{\boldsymbol{\beta}}'_{i,(-j)} \overline{\boldsymbol{V}}^{-1}_{\boldsymbol{\beta}_{i,(-j)}} \overline{\boldsymbol{\beta}}_{(-j)}\right). \tag{A.15}$$

The marginal likelihood for $\widetilde{\boldsymbol{y}}_i$ is also easy to compute, and is given by:

$$\begin{aligned} p\left(\widetilde{\boldsymbol{y}}_i | \lambda_{i,(-j)}\right) &= \int\int p\left(\widetilde{\boldsymbol{y}}_i | \boldsymbol{\beta}_{i,(-j)}, \sigma_i^2\right) p\left(\boldsymbol{\beta}_{i,(-j)}, \sigma_i^2 | \lambda_{i,(-j)}\right) d\boldsymbol{\beta}_{i,(-j)} d\sigma_i^2 \\ &= \int\int \mathcal{N}\left(\boldsymbol{Q}'_2 \boldsymbol{Z}_{i,(-j)} \boldsymbol{\beta}_{i,(-j)}, \sigma_i^2\right) \mathcal{N}\left(\underline{\boldsymbol{\beta}}_{i,(-j)}, \sigma_i^2 \underline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}}\right) \mathcal{IG}\left(\underline{\Psi}_{i,i}, \underline{d}\right) d\boldsymbol{\beta}_{i,(-j)} d\sigma_i^2 \\ &= MVSt_{2\underline{d}}\left(\widetilde{\boldsymbol{y}}_i \mid \boldsymbol{Q}'_2 \boldsymbol{Z}_{i,(-j)} \underline{\boldsymbol{\beta}}_{i,(-j)}, \frac{\overline{\Psi}_{i,i}}{\underline{d}}\left(I_{T-1} + \boldsymbol{Q}'_2 \boldsymbol{Z}_{i,(-j)} \underline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}} \boldsymbol{Z}'_{i,(-j)} \boldsymbol{Q}_2\right)\right). \end{aligned} \tag{A.16}$$

The last step to conclude the derivation of equation (16) is to choose the shrinkage parameter $\lambda_{i,(-j)}$. Two approaches present themselves. The first option is to fix $\lambda_{i,(-j)}$, calibrating it to

[28] Again, in the spirit of the Bayesian VAR tradition, we only shrink the elements of the vector $\boldsymbol{\beta}_{i,(-j)}$ that correspond to the group of coefficients $\boldsymbol{a}_{i,\cdot}$. The intercept c_i and the covariance terms $\boldsymbol{\gamma}_{i,\cdot}$ will have non-informative priors. We have experimented with alternative approaches, including one where we apply the shrinkage coefficient $\lambda_{i,(-j)}$ also to the covariance terms in $\boldsymbol{\gamma}_{i,\cdot}$.

the desired overall shrinkage level. The second approach is to choose $\lambda_{i,(-j)}$ by maximizing the marginal likelihood in (A.16), i.e. mirroring the approach we implemented for $\lambda_{i,j}$. That is,

$$\widehat{\lambda}_{i,(-j)} = \arg \max_{\lambda_{i,(-j)}} \; p\left(\widetilde{\boldsymbol{y}}_i | \lambda_{i,(-j)}\right). \tag{A.17}$$

Regardless of the approach chosen, i.e. whether we fix $\lambda_{i,(-j)}$ or compute $\widehat{\lambda}_{i,(-j)}$ as in (A.17), we are now ready to derive the predictive density of the auxiliary regression (to ease the notation, we drop the dependence on the shrinkage parameter):

$$\begin{aligned}
p\left(y_i^* | \beta_{i,j}, \widetilde{\boldsymbol{y}}_i\right) &= \|\boldsymbol{Z}_{i,j}\| \, \beta_{i,j} + \int \int p\left(y_i^+ | \boldsymbol{\beta}_{i,(-j)}, \sigma_i^2, \widetilde{\boldsymbol{y}}_i\right) p\left(\boldsymbol{\beta}_{i,(-j)}, \sigma_i^2 | \widetilde{\boldsymbol{y}}_i\right) d\boldsymbol{\beta}_{i,(-j)} d\sigma_i^2 \\
&= \|\boldsymbol{Z}_{i,j}\| \, \beta_{i,j} + \int \int \mathcal{N}\left(\boldsymbol{q}_1' \boldsymbol{Z}_{i,(-j)} \boldsymbol{\beta}_{i,(-j)}, \sigma_i^2\right) \times \\
&\times \mathcal{N}\left(\overline{\boldsymbol{\beta}}_{i,(-j)}, \sigma_i^2 \lambda_{i,(-j)}^2 \overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}}\right) \mathcal{IG}\left(\overline{\Psi}_{i,i}, \overline{d}\right) d\boldsymbol{\beta}_{i,(-j)} d\sigma_i^2 \\
&= \|\boldsymbol{Z}_{i,j}\| \, \beta_{i,j} + t_{2\overline{d}}\left(\overline{\mu}_{i,j}, \overline{\tau}_{i,j}^2\right) \\
&\approx \|\boldsymbol{Z}_{i,j}\| \, \beta_{i,j} + \mathcal{N}\left(\overline{\mu}_{i,j}, \overline{\tau}_{i,j}^2\right)
\end{aligned} \tag{A.18}$$

where

$$\overline{\mu}_{i,j} = \boldsymbol{q}_1' \boldsymbol{Z}_{i,(-j)} \overline{\boldsymbol{\beta}}_{i,(-j)} \tag{A.19}$$

and

$$\overline{\tau}_{i,j}^2 = \frac{\overline{\Psi}_{i,i}}{\overline{d}} \left(1 + \boldsymbol{q}_1' \boldsymbol{Z}_{i,(-j)} \overline{\boldsymbol{V}}_{\boldsymbol{\beta}_{i,(-j)}} \boldsymbol{Z}_{i,(-j)}' \boldsymbol{q}_1\right). \tag{A.20}$$

This concludes the derivations of equations (16), (17), and (18). ■

Appendix B Data and transformations

Table B.1. List of series

Series id	*Tcode*	*Medium*	*Large*	*X-large*	*FRED*	*Description*
1	5		X	X	RPI	Real Personal Income
2	5		X	X	W875RX1	RPI ex. Transfers
3	5		X	X	DPCERA3M086SBEA	Real PCE
4	5		X	X	CMRMTSPLx	Real M&T Sales
5	5		X	X	RETAILx	Retail and Food Services Sales
6	5			X	INDPRO	IP Index
7	5				IPFPNSS	IP: Final Products and Supplies
8	5				IPFINAL	IP: Final Products
9	5				IPCONGD	IP: Consumer Goods
10	5				IPDCONGD	IP: Durable Consumer Goods
11	5				IPNCONGD	IP: Nondurable Consumer Goods
12	5				IPBUSEQ	IP: Business Equipment
13	5				IPMAT	IP: Materials
14	5				IPDMAT	IP: Durable Materials
15	5				IPNMAT	IP: Nondurable Materials
16	5				IPMANSICS	IP: Manufacturing
17	5				IPB51222S	IP: Residential Utilities
18	5				IPFUELS	IP: Fuels
19	2				CUMFNS	Capacity Utilization: Manufacturing
20	2			X	HWI	Help-Wanted Index for US
21	2			X	HWIURATIO	Help Wanted to Unemployed ratio
22	5			X	CLF16OV	Civilian Labor Force
23	5				CE16OV	Civilian Employment
24	2	X	X	X	UNRATE	Civilian Unemployment Rate
25	2				UEMPMEAN	Average Duration of Unemployment
26	5				UEMPLT5	Civilians Unemployed $\leq$ 5 Weeks
27	5				UEMP5TO14	Civilians Unemployed 5-14 Weeks
28	5				UEMP15OV	Civilians Unemployed $>$ 15 Weeks
29	5				UEMP15T26	Civilians Unemployed 15-26 Weeks
30	5				UEMP27OV	Civilians Unemployed $>$ 27 Weeks
31	5				CLAIMSx	Initial Claims
32	5	X	X	X	PAYEMS	All Employees: Total nonfarm
33	5				USGOOD	All Employees: Goods-Producing
34	5				CES1021000001	All Employees: Mining and Logging
35	5				USCONS	All Employees: Construction
36	5				MANEMP	All Employees: Manufacturing
37	5				DMANEMP	All Employees: Durable goods
38	5				NDMANEMP	All Employees: Nondurable goods
39	5				SRVPRD	All Employees: Service Industries
40	5				USTPU	All Employees: TT&U
41	5				USWTRADE	All Employees: Wholesale Trade
42	5				USTRADE	All Employees: Retail Trade
43	5				USFIRE	All Employees: Financial Activities
44	5				USGOVT	All Employees: Government
45	5			X	CES0600000007	Hours: Goods-Producing

Table B.1 continued

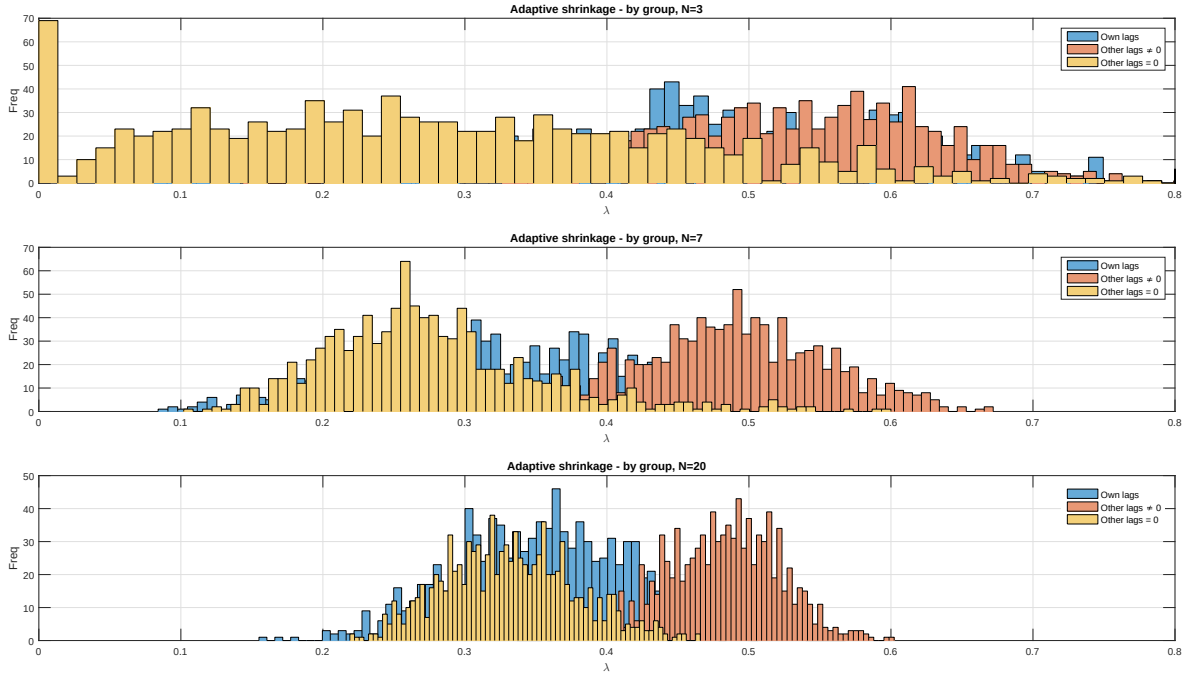
Series id	*Tcode*	*Medium*	*Large*	*X-large*	*FRED*	*Description*
46	2				AWOTMAN	Overtime Hours: Manufacturing
47	5				AWHMAN	Hours: Manufacturing
48	5				HOUST	Starts: Total
49	5				HOUSTNE	Starts: Northeast
50	5				HOUSTMW	Starts: Midwest
51	5				HOUSTS	Starts: South
52	5				HOUSTW	Starts: West
53	5				AMDMNOx	Orders: Durable Goods
54	5				AMDMUOx	Unfilled Orders: Durable Goods
55	5				BUSINVx	Total Business Inventories
56	2				ISRATIOx	Inventories to Sales Ratio
57	5			X	M1SL	M1 Money Stock
58	5			X	M2SL	M2 Money Stock
59	5			X	M2REAL	Real M2 Money Stock
60	5		X	X	BUSLOANS	Commercial and Industrial Loans
61	5				REALLN	Real Estate Loans
62	5		X	X	NONREVSL	Total Nonrevolving Credit
63	2		X	X	CONSPI	Credit to PI ratio
64	5			X	S&P 500	S&P 500
65	5			X	S&P: indust	S&P Industrial
66	2			X	S&P div yield	S&P Divident yield
67	5			X	S&P PE ratio	S&P Price/Earnings ratio
68	2	X	X	X	FEDFUNDS	Effective Federal Funds Rate
69	2		X	X	CP3M	3-Month AA Comm. Paper Rate
70	2			X	TB3MS	3-Month T-bill
71	2			X	TB6MS	6-Month T-bill
72	2			X	GS1	1-Year T-bond
73	2			X	GS5	5-Year T-bond
74	2	X	X	X	GS10	10-Year T-bond
75	2			X	AAA	Aaa Corporate Bond Yield
76	2			X	BAA	Baa Corporate Bond Yield
77	1				COMPAPFF	CP - FFR spread
78	1				TB3SMFFM	3 Mo. - FFR spread
79	1				TB6SMFFM	6 Mo. - FFR spread
80	1				T1YFFM	1 yr. - FFR spread
81	1				T5YFFM	5 yr. - FFR spread
82	1				T10YFFM	10 yr. - FFR spread
83	1				AAAFFM	Aaa - FFR spread
84	1				BAAFFM	Baa - FFR spread
85	5		X	X	EXSZUS	Switzerland / U.S. FX Rate
86	5		X	X	EXJPUS	Japan / U.S. FX Rate
87	5		X	X	EXUSUK	U.S. / U.K. FX Rate
88	5		X	X	EXCAUS	Canada / U.S. FX Rate
89	5				WPSFD49107	PPI: Final demand less energy
90	5				WPSFD49501	PPI: Personal cons

Table B.1 continued

Series id	*Tcode*	*Medium*	*Large*	*X-large*	*FRED*	*Description*
91	5				WPSID61	PPI: Processed goods
92	5				WPSID62	PPI: Unprocessed goods
93	5			X	OILPRICEx	Crude Oil Prices: WTI
94	5				PPICMM	PPI: Commodities
95	6	X	X	X	CPIAUCSL	CPI: All Items
96	5				CPIAPPSL	CPI: Apparel
97	5				CPITRNSL	CPI: Transportation
98	5				CPIMEDSL	CPI: Medical Care
99	5				CUSR0000SAC	CPI: Commodities
100	5				CUUR0000SAD	CPI: Durables
101	5				CUSR0000SAS	CPI: Services
102	5				CPIULFSL	CPI: All Items Less Food
103	5				CUUR0000SA0L2	CPI: All items less shelter
104	5				CUSR0000SA0L5	CPI: All items less medical care
105	5				PCEPI	PCE: Chain-type Price Index
106	5				DDURRG3M086SBEA	PCE: Durable goods
107	5				DNDGRG3M086SBEA	PCE: Nondurable goods
108	5				DSERRG3M086SBEA	PCE: Services
109	5				CES0600000008	Ave. Hourly Earnings: Goods
110	5				CES2000000008	Ave. Hourly Earnings: Construction
111	5				CES3000000008	Ave. Hourly Earnings: Manufacturing
112	5				MZMSL	MZM Money Stock
113	5				DTCOLNVHFNM	Consumer Motor Vehicle Loans
114	5				DTCTHFNM	Total Consumer Loans and Leases
115	5			X	INVEST	Securities in Bank Credit
116	5	X	X	X	GDP	Real Gross Domestic Product
117	5				PCDG	PCE: Durable Goods
118	5				PCESV	PCE: Services
119	5				PCND	PCE: Nondurable Goods
120	5				FPI	Fixed Private Investment
121	5				PRFI	Private Residential Fixed Investment
122	5				GCEC1	Government Cons Expenditures & Gross Inv
123	6	X	X	X	GDPDEFL	GDP deflator
124	5				PCEDEFL	PCE deflator

Appendix C Additional results

Figure C.1. Monte Carlo simulation - Shrinkage intensity by coefficient group



This figure plot the empirical distribution of the shrinkage parameter $\widehat{\lambda}_{i,j}$ estimated using our adaptive shrinkage procedure ($i = 1, ..., n$ and $j = 1, ..., np$). The empirical distributions are obtained by simulating $1,000$ VAR(p) of sample size $T = 150$ and lag length $p = 2$. See Section 4 for additional details on the design of the Monte Carlo simulation. Next, for each simulated dataset we compute the optimal degree of shrinkage using our adaptive shrinkage procedure. We construct empirical distributions of the estimated shrinkage parameters $\widehat{\lambda}_{i,j}$ according to whether the corresponding VAR coefficients belong to the own-lags category ($E\left(\widehat{\lambda}_{i,i}\right)$, $i = 1, ..., n$) or the other-lags category ($E\left(\widehat{\lambda}_{i,j}\right)$, $i = 1, ..., n$ $j = 1, ..., np$, and $i \neq j$). We further break down the latter group according to whether the corresponding coefficients in the simulated data are equal to zero (i.e., $E\left[\widehat{\lambda}_{i,j}|a_{i,j} = 0\right]$) or not (i.e., $E\left[\widehat{\lambda}_{i,j}|a_{i,j} \neq 0\right]$).