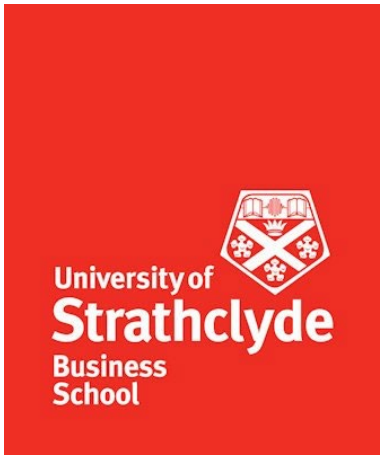
Strathclyde
Discussion Papers
in Economics




# Fast, Order-Invariant Bayesian Inference in VARs using the Eigendecomposition of the Error Covariance Matrix

Ping Wu*

Department of Economics, University of Strathclyde

and

Gary Koop

Department of Economics, University of Strathclyde

No. 23 – 10

Department of Economics
University of Strathclyde, Glasgow

# Fast, Order-Invariant Bayesian Inference in VARs using the Eigendecomposition of the Error Covariance Matrix

Ping Wu*
Department of Economics, University of Strathclyde
and
Gary Koop
Department of Economics, University of Strathclyde

November 11, 2022

**Abstract**

Bayesian inference in Vector Autoregressions (VARs) involves manipulating large matrices which appear in the posterior (or conditional posterior) of the VAR coefficients. For large VARs, the computational time involved with these manipulations becomes so large as to make empirical work impractical. In response to this, many researchers transform their VARs so as to allow for Bayesian estimation to proceed one equation at a time. This leads to a massive reduction in the computational burden. This transformation involves taking the Cholesky decomposition for the error covariance matrix. However, this strategy implies that posterior inference depends on the order the variables enter the VAR. In this paper we develop an alternative transformation, based on the eigendecomposition, which does not lead to order dependence. Beginning with an inverse-Wishart prior on the error covariance matrix, we derive and discuss the properties of the prior it implies on the eigenmatrix and eigenvalues. We then show how an extension of the prior on the eigenmatrix can allow for greater flexibility while maintaining many of the benefits of conjugacy. We exploit this flexibility in order to extend the prior on the eigenvalues to allow for stochastic volatility. The properties of the eigendecomposition approach are investigated in a macroeconomic forecasting exercise involving VARs with 20 variables.

*Keywords:* Eigendecomposition, order invariance, large vector autoregression

*Both authors are research fellows in the Economic Statistics Centre of Excellence (ESCoE). We would like to thank Joshua Chan for many helpful comments.

# 1 Introduction

Vector Autoregressions (VARs) have shown their usefulness in a range of applications in macroeconomics (e.g., Sims, 1980; Cogley and Sargent, 2005; Primiceri, 2005; Koop, 2013; Korobilis, 2013). Beginning with Banbura et al. (2010) macroeconomists have begun working with large VARs, containing tens or even hundreds of dependent variables. Bayesian inference and prediction with such large VARs involves an enormous computational burden. The main computational bottleneck relates to the posterior for the matrix of VAR coefficients $\mathbf{A}$. For VARs involving $n$ variables and $p$ lags this will contain at least $n^2 \times p$ coefficients which is huge for empirically reasonable choices for $n$ and $p$. Manipulation involving features such as the posterior covariance matrix of $\mathbf{A}$ will involve working with matrices of dimension at least $pn^2 \times pn^2$. However, working with the VAR one equation at a time leads to manipulations involving $pn \times pn$. Even though such manipulations must be repeated $n$ times, this equation-by-equation strategy has been shown, e.g. by Carriero et al. (2019), to lead to an $O(n^2)$ reduction in the computational burden. Even for $n = 10$ the benefits of equation by equation estimation can be seen to be substantial and for $n = 100$ they are enormous.

In the conventional (reduced form) VAR, with $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ being the vector of dependent variables and $\Sigma$ being the positive definite error covariance matrix, equation-by-equation estimation is not possible since the off-diagonal elements of $\Sigma$ would be ignored. However, if the VAR is transformed so as to have a diagonal error covariance matrix, equation-by-equation estimation is valid. These considerations motivate most of the recent large VAR literature which uses the Cholesky decomposition to do this transformation. In particular, let $\Sigma^{-1} = \mathbf{L}'\mathbf{C}^{-1}\mathbf{L}$, where $\mathbf{C}$ is a diagonal matrix and $\mathbf{L}$ is a lower triangular matrix with ones on the diagonal, then the Cholesky-transformed VAR has dependent variables $\mathbf{L}\mathbf{y}_t$ and a diagonal error covariance matrix and equation-by-equation estimation

is possible.

Several researchers have pointed out that the use of the Cholesky decomposition leads to order dependence (i.e. posterior inference depends on the way the variables are ordered in the VAR). Arias et al. (2021) demonstrates the importance of the ordering issue in Cholesky-transformed VARs both theoretically and empirically. The authors show that, although point forecasts are not sensitive to the way variables are ordered, predictive standard deviations can be substantially affected. These points are reinforced in Chan et al. (2021) who also show that the problems with order invariance increase in magnitude as the dimension of the VAR increases. In other words, the problem is most acute precisely where the Cholesky transformation is most necessary.

Carriero et al. (2019) theoretically demonstrate that the posterior of the VAR coefficients, conditional on the error covariance matrix $\Sigma$ is invariant to ordering. The lack of order invariance arises due to the fact that the implied prior on $\Sigma$ is not order invariant. Thus, most of the discussion in this literature, including the present paper, focusses on the error covariance matrix.

These considerations have stimulated interest in order invariant approaches and ordering averaging approaches. Order invariant approaches modify the prior to keep the implied prior on $\Sigma$ being order invariant. Given a prior on $\Sigma$ (i.e., the inverse-Wishart distribution), Chan (2021b) derive the implied prior on the impact matrix and the diagonal matrix from the Cholesky decomposition. Since the inverse Wishart prior on $\Sigma$ is order invariant, the implied prior is also independent of the order of the variables. While this prior is order invariant, it is difficult to extend this approach to allow for time variation in the error covariance matrix (which is often empirically important). This difficulty arises since one needs to ensure the implied prior on $\Sigma$ is an inverse-Wishart distribution.

Of the order averaging approaches, Levy and Lopes (2021) is an important recent contribution. After noting that the Cholesky decomposition approach is not order invariant, this

paper proposes a dynamic ordering learning approach to deal with the uncertainty around series ordering. They compute the probability of each ordering, then do order selection or order averaging. However, the limitation of this approach is that it is only practical in relatively small VARs. When $n$ is large, the number of possible orderings becomes enormous, which dramatically increases the computational burden of this approach.

These considerations motivate the present paper. In it, we develop a new approach to Bayesian VAR analysis which is order invariant, computationally practical and allows for extensions such as the addition of stochastic volatillity. We do this by using the eigendecomposition instead of a Cholesky decomposition to transform the VAR and refer to this model as the Eigen-transformed VAR. That is, we replace the Cholesky decomposition by the eigendecomposition in order to transform the VAR in a manner which allows for equation-by-equation estimation. Eigendecompositions have been used in other contexts for grouped covariance estimation in papers such as Hoff (2009b). We extend and adapt these methods for use with VARs, We refer to our model as the Bayesian VAR using eigendecomposition (BVAR-eig).

Note that, by using the eigendecomposition, the ordering issue is solved since the eigenmatrix is orthogonal. We prove order invariance in the appendix and develop an MCMC algorithm that allows for Bayesian inference. We show that its computational burden is similar to that of the MCMC algorithm used with the Cholesky-transformed VAR.

We begin by considering the inverse-Wishart prior for $\Sigma$ which is known to be order invariant. We show that this implies priors for the eigenmatrix (i.e. the matrix of eigenvectors) and eigenvalues which appear in BVAR-eig to have Bingham and inverse-Gamma distributions, respectively. We refer to this as the $\mathcal{BIG}$ prior. In the Bayesian VAR literature, the fact that the natural conjugate Normal-inverted Wishart prior has some restrictive properties is well known, see, (e.g., Chan (2020). In our $\mathcal{BIG}$ prior these properties imply that the eigenvalues control both the scale of the error covariance and the orientation of

the eigenmatrix. We follow a suggestion of Hoff (2009b) and break this dependence by introducting a separate prior parameter to control the orientation of the eigenmatrix. We call this the independent Bingham inverse-gamma or $\mathcal{IBIG}$ prior. The $\mathcal{IBIG}$ prior is potentially interesting in and of itself. However, we introduce it mainly since it allows us to introduce stochastic volatility (SV) into the VAR in a simple way. Since the eigenvalues now only relate to the scale of the variances, we can assume they are SV processes without imposing restrictions on the orientation of the eigenmatrix leading to the model we call BVAR-eigSV. Since SV is empirically important in most applications (e.g., Cogley and Sargent, 2005; Primiceri, 2005; Koop and Korobilis, 2013) we highlight this extension as being of substantive interest.

We carry out a substantial macroeconomic forecasting exercise involving VARs with up to 20 dependent variables. We compare the forecasting performance of our eigendecomposition-based approaches to Cholesky-based approaches (labelled BVAR-chol). Our results are encouraging, particularly for versions of the models with SV. That is, we find substantial forecast improvements in many cases and comparable forecast performance in the remainder.

The rest of the paper is organized as follows. Section 2 provides a description of the theory behind Eigendecomposition, and introduces the homoskedastic version of BVAR-eig. Section 3 presents our empirical work comparing the homoskedastic BVAR-eig to BVAR-chol. Section 4 introduces SV and develops MCMC methods for our BVAR-eigSV model. Section 5 presents our empirical work for models with SV including a comparison of the forecast poerformance of BVAR-eigSV relative to BVAR-cholSV. Section 6 provides some concluding remarks. An appendix proves that the BVAR-eig is order invariant and provides additional details on computation.

# 2 Eigendecomposition and the Bayesian VAR

## 2.1 Transforming the VAR: The BVAR-eig and BVAR-chol

The eigendecomposition of a positive definite covariance matrix $\Sigma$ is given by $\Sigma = \mathbf{U}\mathbf{\Lambda}\mathbf{U}'$. The matrix $\mathbf{U}$ is an $n \times n$ orthogonal matrix of eigenvectors and is called the eigenmatrix. $\mathbf{\Lambda} = \text{diag}(\lambda_1, \ldots, \lambda_n)$ is the matrix of non-negative eigenvalues. Eigendecompositions have been used in fields such as statistics, machine learning, engineering, and signal processing (e.g., Parra et al., 2005). In economics, their most popular use is in principal components analysis (PCA) as used, for instance, in the Factor-Augmented VAR, FAVAR, of Bernanke et al. (2005). PCA is based on the eigendecomposition of the sample covariance matrix of set of time series variables. A common, two step, method for estimating the FAVAR would be to use PCA to select a small number of factors which are then included as dependent variables in a VAR. In contrast, ours is a one step approach which uses the eigendecomposition of the error covariance matrix of the VAR. To our knowledge, the eigendecomposition has not been used in this manner in the Bayesian VAR literature.

To fix the basic ideas, consider a standard homoscedastic VAR of order $p$. Let $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ be an $n \times 1$ vector of variables that is observed over the periods $t = 1, \ldots, T$. Then, the VAR($p$) is given by:

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \Sigma) \tag{1}$$

where $\mathbf{a}$ is an $n \times 1$ vector of intercepts, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $n \times n$ coefficient matrices. Note that there are $n$ equations and each equation has $k = np + 1$ regressors, so there are a total of $nk = n^2p + n$ VAR coefficients. We will refer to this as the reduced form VAR.

The BVAR-eig is produced by multiplying the VAR by $\mathbf{U}'$ leading to a transformed VAR with $\mathbf{U}'\mathbf{y_t}$ on the left hand side and a diagonal error covariance matrix $\mathbf{\Lambda}$. The BVAR-chol

uses the Cholesky-decomposition leading to a transformed VAR with $\mathbf{Ly_t}$ on the left hand side and a diagonal error covariance matrix $\mathbf{C}$. Thus, both have a diagonal error covariance matrix and it is this property which allows for equation-by-equation Bayesian estimation of both of them.

Note, however, that for the BVAR-chol the lower triangularity of $\mathbf{L}$ means that the equation for each variable will have different contemporaneous values of variables appearing in them. For instance, the first variable will appear in all equations, the second will appear in all remaining $(n-1)$ equations, etc.. In contrast, the orthogonality of $\mathbf{U}$ means each equation will have the same variables appearing in it. It is these properties which, through the choices of prior used in these models, that leads to the fact that the BVAR-eig is order-invariant while BVAR-chol is not.

## 2.2 Priors for BVAR-eig and BVAR-chol

For the BVAR-chol, it is standard to use a Normal prior on the elements $\mathbf{L}$ and inverse-gamma priors on diagonal elements in $\mathbf{C}$. As noted above, it is well-established that this leads to a posterior which depends on the way the variables are ordered in the VAR. It is also worth stressing that this prior does not imply an inverse-Wishart prior on $\Sigma$ use of which does imply order invariance in the reduced form VAR.

To develop a prior for BVAR-eig, our strategy will be to begin with the commonly used inverse-Wishart prior on the reduced form VAR error covariance matrix $\Sigma$, then work out what it implies for $\mathbf{U}$ and $\mathbf{\Lambda}$.

Assume the prior for $\Sigma$ to have an inverse-Wishart distribution with shape parameter $v > 0$ and scale matrix $\mathbf{S}_0$:

$$f(\Sigma; v, \mathbf{G}) = \frac{|\mathbf{G}|^{v/2}}{2^{nv/2}\Gamma_n(v/2)}|\Sigma|^{-\frac{v+n+1}{2}}\mathrm{e}^{-\frac{1}{2}\,\mathrm{tr}\left(\mathbf{S}_0\Sigma^{-1}\right)}$$

where $\Gamma_n$ is the multivariate gamma function. We write $\Sigma \sim \mathcal{IW}(v, \mathbf{S}_0)$.

Sub-section 2.4 of Hoff (2009b) shows what the inverse-Wishart prior on $\Sigma$ implies for its eigenmatrix and eigenvalues. It turns out the former has a Bingham distribution and the diagonal elements of the latter have inverse-Gamma distributions. We will not reproduce the full proofs of Hoff (2009b) but provide some details as they allow us to discuss some properties of the Bingham distribution.

A matrix $\mathbf{U}$ is said to have a generalised Bingham distribution if its density function is given by

$$p(\mathbf{U} \mid \mathbf{D}, \mathbf{B}, \mathbf{V}) = c(\mathbf{D}, \mathbf{B})\, \mathrm{e}^{\mathrm{tr}(\mathbf{BU'VDV'U})} \tag{2}$$

where $\mathbf{D}$ and $\mathbf{B}$ are diagonal matrices. It can be shown that, conditional on $\Lambda$, the prior for $\mathbf{U}$ has this form where $\mathbf{B} = -\frac{1}{2}\mathbf{\Lambda}^{-1}$. $\mathbf{V}$ and $\mathbf{D}$ are defined through the eigendecomposition of $\mathbf{S}_0$. That is $\mathbf{S}_0 = \mathbf{VDV}'$.

The Bingham distribution, proposed for a vector by Bingham (1974) and extended to a matrix variate version by Khatri and Mardia (1977) and Gupta and Nagar (2018), is the prior we use on the eigenmatrix. Crucially, the conditional posterior of the eigenmatrix also turns out to be Bingham and a computationally efficient method for taking draws from it is developed in Hoff (2009b).

Conditional on the eigenmatrix $\mathbf{U}$, the inverse-Wishart prior for $\Sigma$ can be shown to imply the prior for the eigenvalues, $\mathbf{\Lambda}$, to be independent inverse-gamma distributions. To be precise, let $u_i$ denote the $i$th column of $\mathbf{U}$, $\lambda_i$ denote the $i$th eigenvalue, then the prior

for $\lambda_i$ is:

$$p(\mathbf{\Lambda} \mid \mathbf{D}, \mathbf{U}, \mathbf{V}) = \prod_{i=1}^{n} p(\lambda_i \mid u_i) \tag{3}$$
$$\propto \prod_{i=1}^{n} \lambda_i^{-\alpha-1} e^{-\frac{\beta}{\lambda_i}}$$

with $\alpha = \frac{v+n-1}{2}$, $\beta = \frac{1}{2} u_i' \left(\mathbf{VDV}'\right) u_i$ denoting the arguments of the inverse-Gamma distribution.

Thus, the inverse-Wishart prior for $\Sigma$ implies a Bingham and inverse-Gamma, or $\mathcal{BIG}$, prior for the eigenmatrix and eigenvalues. We emphasize this equivalence between the $\mathcal{BIG}$ and the inverse Wishart prior which automatically implies order-invariance (and, thus, the proof of order invariance given in the appendix, is not needed for this case). However, as shown in the next sub-section working with the $\mathcal{BIG}$ prior leads to a computationally-attractive MCMC algorithm for this model. A more substantial contribution of this paper lies in our development of the $\mathcal{IBIG}$ prior which allows for the introduction of SV in a straightforward way.

To introduce the $\mathcal{IBIG}$ prior, note first some restrictive properites of the $\mathcal{BIG}$ prior. It assumes that the eigenvalues control both the scale of the error covariance matrix $\Sigma$ and the orientation of the eigenmatrix $\mathbf{U}$. The latter arises is from the fact that the orientation of multivariate density of $\mathbf{Y}$ is described by $\mathbf{U}$ which is controlled by $\mathbf{B}$. The $\mathcal{BIG}$ prior replaces $\mathbf{B}$ with $-\frac{1}{2}\mathbf{\Lambda}^{-1}$ and, thus $-\frac{1}{2}\mathbf{\Lambda}^{-1}$ controls the orientation of $\mathbf{U}$. In this sense the $\mathcal{BIG}$ prior (or equivalently the inverse-Wishart distribution) is restrictive. But we can relax this assumption by following a suggestion of Hoff (2009b). We let $\mathbf{B}$ be a diagonal matrix of parameters which controls the orientation of $\mathbf{U}$, no longer explicitly linked to $\mathbf{\Lambda}$. The role of the latter now solely relates to the scale of the error covariance matrix. Note that $\mathbf{B}$ can be treated as matrix of known prior hyperparameters, but it is also possible to treat it as unknown and estimate it. We will adopt the latter strategy in this paper leading to the

$\mathcal{IBIG}$ prior. More specifically, we use a Normal prior for $\mathbf{B}$ and develop Bayesian methods for its estimation.

The remainder of this sub-section summarizes and provides details of prior hyperparameter choice. We consider three models in this section: the BVAR-chol and the BVAR-eig with $\mathcal{BIG}$ and $\mathcal{IBIG}$ priors. For all models, we make prior hyperparameter choices which, where possible, are the same. They are all inspired by relatively non-informative versions of the Minnesota prior treatment of the error covariance matrix involving a degree of freedom choice $v = n + 2$, and $\mathbf{S}_0 = \text{diag}\left(s_1^2, \ldots, s_n^2\right)$, where $s_i^2$ denotes the sample variance of the residuals from an AR(4) model for the variable $i, i = 1, \ldots, n$.

For BVAR-chol, letting $\sigma_i^2$ denote the diagonal elements of $\mathbf{C}$ and $L_{i,j}$ denote the free elements of the lower triangular matrix $\mathbf{L}$, we use

$$\sigma_i^2 \sim \mathcal{IG}\left(\frac{v+i-n}{2}, \frac{s_i^2}{2}\right), \quad i = 1, \ldots, n$$

$$\left(L_{i,j} \mid \sigma_i^2\right) \sim \mathcal{N}\left(0, \frac{\sigma_i^2}{s_j^2}\right), \quad 1 \leqslant j < i \leqslant n, i = 2, \ldots, n$$

For BVAR-eig, we use the Bingham prior for the eigenmatrix:

$$\mathbf{U} \sim \mathcal{B}\left(\mathbf{D}, \mathbf{B}, \mathbf{V}\right),$$

where $\mathbf{V}$ and $\mathbf{D}$ are obtained as the eigenmatrix and eigenvalues from the eigendcomposition of $\mathbf{S}_0$. The $\mathcal{BIG}$ prior sets $\mathbf{B} = -\frac{1}{2}\mathbf{\Lambda}^{-1}$.

The $\mathcal{IBIG}$ prior treats $\mathbf{D}$, $\mathbf{B}$ and $\mathbf{V}$ as an unknown parameters and estimates them.

For the eigenvalues, $\mathcal{BIG}$ uses the inverse-Gamma priors in Equation (3), that is

$$\lambda_i \sim \mathcal{IG}\left(\frac{v+n-1}{2}, \frac{u_i'\left(\mathbf{VDV}'\right)u_i}{2}\right), \quad i = 1, \ldots, n.$$

Since $\mathcal{IBIG}$ assumes that the prior on $\mathbf{U}$ and $\mathbf{\Lambda}$ are independent, so we use the following inverse-Gamma priors

$$\lambda_i \sim \mathcal{IG}\left(10, \frac{s_i^2}{2}\right), \quad i = 1, \ldots, n.$$

For all models, we use the same prior for the VAR coefficients. Any choice could be made for this without altering the main messages of the paper. We choose the Horseshoe prior. More specifically, we use the inverse-Gamma representation of Horseshoe prior (e.g., Carvalho et al., 2010; Cross et al., 2020):

$$\begin{aligned}
&\mathbf{A}_i \mid \lambda_{i,j}^A, \tau^A \sim \mathcal{N}(0,\ \lambda_{i,j}^A \tau^A),\ i = 1, \ldots,\ j = 1, \ldots, k \\
&\lambda_{i,j}^A \sim \mathcal{IG}(\frac{1}{2},\ \frac{1}{\nu_{i,j}^A}), \quad \tau^A \sim \mathcal{IG}(\frac{1}{2},\ \frac{1}{\xi^A}) \\
&\nu_{i,j}^A \sim \mathcal{IG}(\frac{1}{2},\ 1), \qquad \xi^A \sim \mathcal{IG}(\frac{1}{2},\ 1).
\end{aligned} \tag{4}$$

Please note that the Horseshoe prior is put on the parameters of the reduced form VAR $\mathbf{A}_i$, not the Eigen-transformed VAR.

## 2.3 MCMC Algorithms for BVAR-eig and BVAR-chol

The MCMC algorithm for BVAR-chol we use is the one in Chan et al. (2022) to which the reader is referred for further details. The model of Chan et al. (2022) has SV. In this section, we consider only homoskedastic models and use the inverse-Gamma priors on the error variances as we do for BVAR-eig.

For the BVAR-eig, we will show, drawing on results from Hoff (2009b), that the conditional posteriors of the eigenmatrix and eigenvalues are Bingham and inverse-Gamma and draws from them can easily be taken. But before providing details, it is worth mentioning two issues which arise in this algorithm which are specific to our use of the eigendecomposition. The first is that it suffers from the label switching problem. One will get the

same likelihood by permuting the columns of eigenmatrix and then the eigenvalues accordingly. This can be overcome using the random permutation sampler of Frühwirth-Schnatter (2001). The second is that the MCMC algorithm proposed by Hoff (2009b) breaks down the full eigenmatrix $\mathbf{U}$ into draws of two columns at a time. We believe this feature to be particularly important for high-dimensional VARs and will show that the computation is as fast as Cholesky decomposition. It is also worth emphasizing that the posterior is order invariant. This can be seen intuitively from the fact that it is equivalent to using an inverse-Wishart prior on the reduced form VAR error covariance. More formally, a proof is provided in Appendix A.

Posterior draws can be obtained by sampling sequentially from:

**Step1** : $p\left(\mathbf{U} \mid \mathbf{Y}, \mathbf{A}, \mathbf{\Lambda}\right)$

**Step2** : $p\left(\mathbf{A} \mid \mathbf{Y}, \mathbf{U}, \mathbf{\Lambda}\right)$

**Step3** : $p\left(\mathbf{\Lambda} \mid \mathbf{Y}, \mathbf{U}, \mathbf{A}\right)$

where $\mathbf{Y}$ denotes the data and $\mathbf{A}$ is the matrix containing all the VAR coefficients. The last two steps are standard and will only briefly be described here. Most of this sub-section will deal with Step 1.

To see why the last two steps are standard consider the reduced form VAR defined in (1). Since $\mathbf{U}$ is orthogonal, $\mathbf{U}^{-1} = \mathbf{U}'$ and we can left multiply the VAR by $\mathbf{U}'$ to get

$$\mathbf{U}'\mathbf{y}_t = \mathbf{U}'\mathbf{a} + \mathbf{U}'\mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{U}'\mathbf{A}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Lambda}\right). \tag{5}$$

This is a VAR with diagonal error covariance matrix which allows for equation-by-equation estimation. Each equation takes the form of a homoskedastic regression and standard results for Bayesian regression using the Horseshoe prior hold (e.g., Carvalho et al., 2010; Cross et al., 2020). The eigenvalues play the role of error variances and standard formulae for their inverse-Gamma conditional posteriors apply.

Next we describe the algorithm for drawing the eigenmatrix for the the homoskedastic case using the $\mathcal{BIG}$ prior.

**Step1** : $p\left(\mathbf{U} \mid \mathbf{Y}, \mathbf{A}, \mathbf{\Lambda}\right)$

To obtain this conditional posterior we need likelihood and prior. The likelihood can be obtained by rewriting (5) as:

$$\left(\mathbf{Y} - \mathbf{XA}\right)\mathbf{U} = \mathbf{E} \tag{6}$$

where $\mathbf{E}$ is a $T \times n$ matrix of innovations in which the $t$-th row is $\boldsymbol{\varepsilon}_t'$. It follows that

$$\text{vec}(\mathbf{E}) \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Lambda} \otimes \mathbf{I}_T\right).$$

The likelihood, thus, can be written as:

$$\begin{aligned} p(\mathbf{Y} \mid \mathbf{A}, \mathbf{U}, \mathbf{\Lambda}) &= (2\pi)^{-\frac{Tn}{2}} |\mathbf{\Lambda}|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\left((\mathbf{Y}-\mathbf{XA})\mathbf{U}\right)\mathbf{\Lambda}^{-1}\left((\mathbf{Y}-\mathbf{XA})\mathbf{U}\right)'\right)} \\ &= (2\pi)^{-\frac{Tn}{2}} |\mathbf{\Lambda}|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{XA})\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{XA})'\right)} \\ &= (2\pi)^{-\frac{Tn}{2}} |\mathbf{\Lambda}|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{XA})'(\mathbf{Y}-\mathbf{XA})\mathbf{U}\right)} \end{aligned} \tag{7}$$

The Bingham part of the $\mathcal{BIG}$ prior on $\mathbf{U}$ is:

$$p(\mathbf{U} \mid \mathbf{D}, \mathbf{\Lambda}, \mathbf{V}) = \frac{1}{2^{nv/2}\Gamma_n(v/2)} |\mathbf{D}|^{v/2} |\mathbf{\Lambda}|^{-\frac{v+n+1}{2}} \mathrm{e}^{\mathrm{tr}\left(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{U}'\mathbf{V}\mathbf{D}\mathbf{V}'\mathbf{U}\right)} \tag{8}$$

Multiplying prior by likelihood produces the conditional posterior distribution of $\mathbf{U}$:

$$\begin{aligned}
p(\mathbf{U} \mid \mathbf{Y}, \mathbf{A}, \mathbf{\Lambda}, \mathbf{D}, \mathbf{V}) &= p(\mathbf{U} \mid \mathbf{D}, \mathbf{\Lambda}, \mathbf{V}) \times p(\mathbf{Y} \mid \mathbf{A}, \mathbf{U}, \mathbf{\Lambda}) \\
&\propto \mathrm{e}^{\mathrm{tr}\left(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{U}'\mathbf{V}\mathbf{D}\mathbf{V}'\mathbf{U}\right)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A})\mathbf{U}\right)} \\
&= \mathrm{e}^{\mathrm{tr}\left(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{U}'\mathbf{V}\mathbf{D}\mathbf{V}'\mathbf{U}\right)+\mathrm{tr}\left(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A})\mathbf{U}\right)} \\
&= \mathrm{e}^{\mathrm{tr}\left(-\frac{1}{2}\mathbf{\Lambda}^{-1}\mathbf{U}'\left(\mathbf{V}\mathbf{D}\mathbf{V}'+(\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A})\right)\mathbf{U}\right)} \\
&= \mathrm{e}^{\sum_{i=1}^{n}-\frac{1}{2}\lambda_i^{-1}u_i'(\mathbf{V}\mathbf{D}\mathbf{V}'+(\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A}))u_i} \\
&= \prod_{i=1}^{n} \mathrm{e}^{-\frac{1}{2}\lambda_i^{-1}u_i'(\mathbf{V}\mathbf{D}\mathbf{V}'+(\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A}))u_i} \qquad (9)
\end{aligned}$$

This seems to suggest that one can simply draw the eigenvectors in the eigenmatrix one at a time from the Bingham distribution. However, Hoff (2009b) shows that such a strategy will not work since the chain of draws will be reducible. This arises from the fact that $\mathbf{U}$ is orthogonal and, thus, the conditional distribution of $\mathbf{u}_i$ given the other columns of $\mathbf{U}$ has restricted support. Fortunately, Hoff (2009a) proposes a solution to this: draw the eigenvectors two columns at a time.

To see how this algorithm works, let us take the first and second columns of the eigenvectors, $\{\mathbf{u}_1, \mathbf{u}_2\}$, as an example. Conditional on the remaining columns, their distribution is equivalent to $\mathbf{RS}$ where $\mathbf{R}$ is an orthonormal basis for the nullspace of the remaining columns and $\mathbf{S}$ is orthogonal with density

$$\begin{aligned}
p(\mathbf{S}) &\propto e^{\mathrm{tr}\left(-\frac{1}{2}\mathbf{\Lambda}_{1,2}^{-1}\mathbf{S}'\mathbf{K}\mathbf{S}\right)} \\
&= \mathrm{e}^{-\frac{1}{2}\left(\lambda_1^{-1}s_1'\mathbf{K}s_1+\lambda_2^{-1}s_2'\mathbf{K}s_2\right)} \qquad (10)
\end{aligned}$$

where $\mathbf{K} = \mathbf{R}'\big(\mathbf{V}\mathbf{D}\mathbf{V}' + (\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A})\big)\mathbf{R}$, $s_i$ is the $i$th column of $\mathbf{S}$.

Since $\mathbf{S}$ is orthogonal, it can be parameterized as

$$\mathbf{S} = \begin{pmatrix} \cos\phi & s\sin\phi \\ \sin\phi & -s\cos\phi \end{pmatrix}$$

for some $\phi \in (0, 2\pi)$ and $s = \pm 1$. The second column $s_2$ of $\mathbf{S}$ is a linear function of the first column $s_1$, and the uniform density on the circle is constant in $\phi$, so the joint density of $(\phi, s)$ is simply $p(\mathbf{S}(\phi, s))$. Sampling from this distribution can be accomplished by first sampling $\phi \in (0, 2\pi)$ from a density proportional to $p(\mathbf{S}(\phi, s))$, and then sampling $s$ uniformly from {-1,1}. The density $p(\mathbf{S}(\phi, s))$ can be obtained from Equation (10), by replacing $s_1$ and $s_2$ with $\begin{pmatrix} \cos\phi \\ \sin\phi \end{pmatrix}$ and $\begin{pmatrix} s\sin\phi \\ -s\cos\phi \end{pmatrix}$, respectively.

$$p(\phi) \propto \exp\left\{-\frac{1}{2}\left((\frac{k_{1,1}}{\lambda_1} + \frac{k_{2,2}}{\lambda_2})\cos^2(\phi) + (\frac{k_{2,2}}{\lambda_1} + \frac{k_{1,1}}{\lambda_2})\sin^2(\phi) + (\frac{k_{1,2}}{\lambda_1} + \frac{k_{2,1}}{\lambda_1} - \frac{k_{1,2}}{\lambda_2} - \frac{k_{2,1}}{\lambda_2})\cos(\phi)\sin(\phi)\right)\right\}$$

where $k_{i,j}$ is the $(i, j)$th element in matrix $\mathbf{K}$, which is $2 \times 2$.

To summarize, the Gibbs sampling scheme for the eigenvectors is as follows: Given $\mathbf{U}^{(j)} = \mathbf{U}$, perform steps $a - e$ for each pair $(n_1, n_2) \subset \{1, \ldots, n\}$ in random order:

*Step a:* let $\mathbf{R}$ be the null space of $\mathbf{U}_{[,-(n_1,n_2)]}$;

*Step b:* compute $\mathbf{K} = \mathbf{R}'\big(\mathbf{V}\mathbf{D}\mathbf{V}' + (\mathbf{Y} - \mathbf{X}\mathbf{A})'(\mathbf{Y} - \mathbf{X}\mathbf{A})\big)\mathbf{R}$;

*Step c:* sample $\phi \in (0, 2\pi)$ from the density proportional to $p(\phi)$;

*Step d:* sample $s$ uniformly from {-1,1};

*Step e:* set $\mathbf{R} = \mathbf{E}(\phi, s)$ and $\mathbf{U}_{[,(n_1,n_2)]} = \mathbf{R}\mathbf{S}$.

Set $\mathbf{U}^{(j+1)} = \mathbf{U}$.

This completes our description of the MCMC algorithm for the homoskedastic BVAR-eig with $\mathcal{BIG}$. This algorithm can also be used for the $\mathcal{IBIG}$ if the restriction that $\mathbf{B} = -\frac{1}{2}\mathbf{\Lambda}^{-1}$ is relaxed and $\mathbf{B}$ set to a specific value. Remember that this parameter can also

be estimated. We will describe how this is done when we add SV to the model.

# 3 Forecasting Illustration: Homoskedastic Models

## 3.1 Overview and Data Description

SV, which is discussed in the next section, is typically necessary to ensure good density forecasts using macroeconomic data. Hence, we label this section, where we consider homoskedastic models, as being illustrative. In it, the data is transformed to stationarity there are four lags in all models.

We use a dataset that consists of 20 US quarterly variables with a sample period from 1960Q1 to 2021Q3. It is constructed from the FRED-QD database at the Federal Reserve Bank of St. Louis as described in McCracken and Ng (2016). The dataset contains a range of standard macroeconomic and financial variables, such as Real GDP, industrial production, inflation rates, labor market variables and interest rates. The complete list of variables and how they are transformed is given in Appendix B. The variables are ordered in the same manner as they are listed in the appendix which is the way they are commonly ordered in Bayesian empirical work. Note that the macroeconomic variables are ordered first and the financial variables ordered last.

The forecasting design adopted is iterative forecasting. We consider an initial estimation period from 1960Q1 to 1987Q4. The remaining observations (1988Q1 to 2021Q3) are used as a hold-out period to evaluate our forecasting methods. After obtaining $h \in \{1, 4\}$-step-ahead predictive distributions for a given period in the hold-out, we include this period in the estimation sample and repeat this procedure until we reach the end of the sample. To assess forecasting accuracy, we use root mean square forecast errors (RMSFEs) for point forecasts and average log predictive likelihoods (ALPLs) for density forecasts. To compare each model $M$ against the benchmark $B$, we therefore consider the percentage gains in

terms of RMSFE, defined as

$$(1 - \text{RMSFE}_{i,h}^{M}/\text{RMSFE}_{i,h}^{B}) \times 100$$

and the percentage gain in terms of ALPL, which is

$$(\text{ALPL}_{i,h}^{M} - \text{ALPL}_{i,h}^{B}) \times 100.$$

The benchmark is always the BVAR-chol.

## 3.2 Forecasting Results

Table 1 reports the percentage gains in RMSFE and ALPL relative to the BVAR-chol. The main point to note is that priors matter. The three models in this forecasting exercise all have the same likelihood function and differ only in the prior. Furthermore, prior hyper-parameters are selected to be as comparable as possible across models. But still there are substantial differences in forecast performance across models. Another finding of earlier papers, Arias et al. (2021) and Chan et al. (2021), was that, when using Cholesky-decomposed approaches, ordering mattered more for density forecasts than for point forecasts. In Table 1, we find the percentage differences in RMSFEs are quite small, but the percentage differences in ALPLs are much larger.

For ALPLs, we find BVAR-eig to forecast better for the 20 variables overall (i.e. the joint ALPL of the former is much higher at both forecast horizons). This holds true for both the $\mathcal{BIG}$ and $\mathcal{IBIG}$ priors. In light of this, it is unsurprising that the BVAR-eig also tends to beat the benchmark for most of the variables. But for some of the variables, including some of the primary macroeconomic variables such as IPFINAL and FEDFUNDS, it is failing to do so. But we find the $\mathcal{IBIG}$ prior can improve the bad forecast performance

(i.e. the ALPL for IPFINAL using $\mathcal{BIG}$ is -7.78%, while the ALPL for IPFINAL using $\mathcal{IBIG}$ is -3.10% at $h = 4$). Clearly, the restriction in the $\mathcal{BIG}$ prior that $\mathbf{B} = -\frac{1}{2}\mathbf{\Lambda}^{-1}$ is leading to a slight deterioration in forecast performance.

Table 1: Forecasting performance in homoskedastic case: percentage gains in RMSFE and ALPL (Eigendecomposition against Cholesky decomposition)

| | RMSFE | | | | ALPL | | | |
|---|---|---|---|---|---|---|---|---|
| | $\mathcal{BIG}$ prior | | $\mathcal{IBIG}$ prior | | $\mathcal{BIG}$ prior | | $\mathcal{IBIG}$ prior | |
| Variables | $h = 1$ | $h = 4$ | $h = 1$ | $h = 4$ | $h = 1$ | $h = 4$ | $h = 1$ | $h = 4$ |
| GDPC1 | 1.05 | 0.10 | 1.17 | 0.32 | 17.51 | 11.55 | 9.30 | 5.00 |
| PCECC96 | 1.57 | 0.10 | 1.63 | 0.28 | 20.94 | 38.34 | 6.72 | 9.02 |
| INDPRO | 1.24 | -0.16 | -0.03 | 0.36 | 8.14 | -2.43 | 3.14 | 0.40 |
| IPFINAL | -1.14 | -0.21 | -0.50 | 0.26 | 4.96 | -7.78 | 0.90 | -3.10 |
| PAYEMS | -1.54 | 0.07 | -0.13 | 0.20 | 307.23 | 408.68 | 96.84 | 130.64 |
| MANEMP | 0.29 | 0.98 | -0.23 | 0.77 | 22.29 | 1.60 | 6.88 | 2.55 |
| CE16OV | -2.15 | 0.06 | -0.05 | 0.15 | 243.52 | 255.83 | 56.40 | 88.00 |
| CIVPART | -1.41 | -1.52 | -0.68 | -0.71 | -54.11 | -22.93 | 22.49 | 41.82 |
| UNRATE | 0.18 | 1.32 | 0.52 | 0.91 | 376.30 | 669.04 | 343.48 | 548.25 |
| HOANBS | 0.37 | 0.69 | -0.02 | 0.34 | 59.95 | 37.08 | 27.52 | 5.94 |
| HOUST | 4.64 | 1.12 | 4.88 | 0.07 | 6.26 | -31.41 | 6.56 | -4.33 |
| PERMIT | -0.69 | 2.12 | 4.52 | 0.05 | 5.04 | -29.05 | 6.83 | -2.54 |
| PCECTPI | -1.44 | -4.06 | 0.85 | 0.99 | 14.46 | 6.50 | 4.86 | 7.27 |
| CPIAUCSL | -2.44 | -4.79 | -0.19 | 0.10 | 9.28 | 0.97 | 4.52 | 3.17 |
| OPHNFB | 1.83 | -0.31 | 1.27 | -0.29 | 2.28 | -0.56 | -5.07 | -5.90 |
| FEDFUNDS | -9.23 | -6.36 | -2.80 | -2.97 | -18.46 | -15.18 | -6.29 | -3.14 |
| TB3MS | -7.35 | -5.86 | -6.16 | -3.75 | -25.46 | -12.33 | -7.14 | -3.59 |
| GS1 | -3.22 | -1.47 | -2.37 | -0.34 | -23.51 | -7.95 | -6.35 | -0.31 |
| GS10 | -0.27 | 0.59 | 0.44 | 2.49 | -38.96 | -15.25 | -7.72 | 2.32 |
| BAA10YM | -4.98 | -2.47 | 0.58 | 0.60 | -48.15 | 33.33 | 2.75 | 53.26 |
| Joint Forecasting | | | | | | | | |
| All | | | | | 695.73 | 4710.52 | 770.59 | 4251.37 |

# 4 Adding Stochastic Volatility: The BVAR-eigSV

## 4.1 Overview

In the preceding sections we have assumed the error covariance matrix to be homoskedastic. We showed how the $\mathcal{BIG}$ prior on the eigenvalues and eigenmatrix was equivalent to the

inverted-Wishart prior on the reduced form VAR error covariance matrix but allowed for equation-by-equation estimation. Thus, we did not introduce a new model, but rather developed much improved computation for an existing model. In this section, we move beyond the homoskedastic reduced form VAR to address an empirically important issue: how to allow for the time variation in volatilities which characterizes most macroeconomic and financial data sets and is so necessary for density forecasting.[1] In this section, we show how the eigendecomposition approach can be extended to allow for stochastic volatility in BVARs, leading to the BVAR-eigSV model which is order invariant and allows for equation-by-equation estimation. We do this by allowing for time variation in the eigenvalues. Remember that the Bingham distribution, which we used as a prior for the eigenvectors, depends on three arguments which we called $\mathbf{D}, \mathbf{B}$ and $\mathbf{V}$, but that the $\mathcal{BIG}$ prior restricted $\mathbf{B} = -\frac{1}{2}\mathbf{\Lambda}^{-1}$ but that the $\mathcal{IBIG}$ left $\mathbf{B}$ unrestricted. When adding SV we will only work with the latter prior since the restriction cannot be imposed while retaining a Bingham posterior.

A second contribution of this section lies in hyperparameter estimation. In the Bayesian VAR literature, it is increasingly common to estimate prior hyperparameters. For instance, Giannone et al. (2015) develops methods for estimating the optimal degree of prior shrinkage in Bayesian VARs. This approach relies on the marginal likelihood, which has a closed form expression for the homoskedastic VAR with inverted-Wishart prior. But once SV has been added, the marginal likelihood no longer has a closed form expression and can be very difficult to calculate. In this section, we develop methods for doing so. In particular, we add blocks to the MCMC algorithm for drawing $\mathbf{D}, \mathbf{B}$ and $\mathbf{V}$.

[1]Note that allowing for time variation in the VAR coefficients is also possible. For instance, it is trivial to extend the methods in the paper to allow for them to evolve according to random walks. We do not do so in order to focus on the error covariance matrix which is where the contributions of this paper lie.

The BVAR-eigSV model is

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{U}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}_t) \tag{11}$$

where $\boldsymbol{\Lambda}_t = \text{diag}\left(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}}\right)$ is diagonal. Notice that the eigenvalues may vary over time. Each of the log-volatilities follows an independent random walk:[2]

$$h_{i,t} = h_{i,t-1} + u_{i,t}^h, \quad u_{i,t}^h \sim \mathcal{N}\left(0, \omega_i^2\right), \tag{12}$$

for $t = 1, \ldots, T$, where the initial states are specified as $h_{i,0} \sim \mathcal{N}(0, 10)$. This BVAR-eigSV model takes the advantage of Cholesky decomposition that it allows for equation-by-equation estimation (e.g., Koop et al., 2019; Carriero et al., 2019).

Relative to our earlier BVAR-eig, the new aspects of the MCMC algorithm relate to $\boldsymbol{\Lambda}_t$ as well as $\mathbf{U}$ and its prior hyperparameters: $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{V}$[3]. Accordingly we will only discuss the MCMC steps for these, with other steps being unchanged. $\boldsymbol{\Lambda}_t$ can also be easily dealt with since we use standard methods, in particular we implement the auxiliary mixture sampler of Kim et al. (1998) in conjunction with the precision sampler of Chan and Jeliazkov (2009).

The prior for $\mathbf{U}$, $p(\mathbf{U} \mid \mathbf{D}, \mathbf{B}, \mathbf{V})$, takes the Bingham form given in (2). The hyperparameters in this prior, $\mathbf{B}$, $\mathbf{D}$ and $\mathbf{V}$, are treated as unknown parameters. For reasons to be explained, we do not directly place a prior on these parameters, but work with transformations of them. The prior on the transformed parameters is given in the algorithm.

Noting that, conditional on $\mathbf{U}$, the data provides no additional information for $\mathbf{B}$, $\mathbf{D}$ and

[2]Allowing the the errors in the volatility processes to be correlated with one another would be a simple extension of our model. Alternatively, a factor structure could be used to obtain parsimony as in factor stochastic volatility models (e.g., Pitt and Shephard, 1999; Chib et al., 2006; Kastner and Huber, 2020; Chan, 2021a).

[3]In the SV case, one can still use the $\mathcal{IBIG}$ prior where $\mathbf{D}$ and $\mathbf{V}$ are set to specific values as done in the homoskedastic case. Here we further treat $\mathbf{D}$ and $\mathbf{V}$ as unknown and estimate them.

$\mathbf{V}$, the full conditional posterior distributions simplify and our MCMC algorithm involves the following distributions:

**Step1** : $p(\mathbf{U} \mid \mathbf{Y}, \mathbf{A}, \mathbf{\Lambda}, \mathbf{D}, \mathbf{B}, \mathbf{V})$

**Step2** : $p(\mathbf{B}, \mathbf{D} \mid \mathbf{U}, \mathbf{V})$

**Step3** : $p(\mathbf{V} \mid \mathbf{U}, \mathbf{D}, \mathbf{B})$

which we discuss in turn in this sub-section.

Step 1 requires only minor modification relative to the homoskedastic case.

**Step1** : $p(\mathbf{U} \mid \mathbf{Y}, \mathbf{A}, \mathbf{\Lambda}, \mathbf{D}, \mathbf{B}, \mathbf{V})$

$$p(\mathbf{U} \mid \mathbf{Y}, \mathbf{A}, \mathbf{\Lambda}, \mathbf{D}, \mathbf{B}, \mathbf{V}) \propto \prod_{i=1}^{n} \mathrm{e}^{u_i' \left( b_i \mathbf{V}\mathbf{D}\mathbf{V}' + \sum_{t=1}^{T} -\frac{1}{2} \lambda_{i,t}^{-1} (\mathbf{y}_t' - \mathbf{x}_t'\mathbf{A})'(\mathbf{y}_t' - \mathbf{x}_t'\mathbf{A}) \right) u_i}$$

which is a Bingham distribution. We use the same strategy of drawing two columns of $\mathbf{U}$ at a time that we used for the homoskedastic case with $\mathcal{IBIG}$ prior.

**Step2** : $p(\mathbf{D}, \mathbf{B} \mid \mathbf{U}, \mathbf{V})$

Since $p(\mathbf{D}, \mathbf{B} \mid \mathbf{U}, \mathbf{V}) \propto p(\mathbf{U} \mid \mathbf{D}, \mathbf{B}, \mathbf{V}) p(\mathbf{D}, \mathbf{B} \mid \mathbf{V})$ the Bingham prior for $\mathbf{U}$ is the key component of this conditional posterior. Note, however, that it involves an integrating constant, $c(\mathbf{D}, \mathbf{B})$ which could be ignored in our derivations of the posterior for $\mathbf{U}$, but cannot be ignored in Step 2. The resulting conditional posterior is no longer of a convenient form. Accordingly, we follow an approximate strategy suggested in Hoff (2009b). One aspect of this strategy is to use an approximation to the integrating constant, $\widetilde{c}(\mathbf{D}, \mathbf{B})$ which is given in equation (8) of Hoff (2009b). But this approximation causes an identification issue since it implies the scales of $\mathbf{D}$ and $\mathbf{B}$ are not separately identifiable. Accordingly, Hoff (2009b) reparameterizes the diagonal matrices $\mathbf{D}$ and $\mathbf{B}$ as having diagonal elements

ordered and bounded between zero and one times a scalar constant:

$$\text{diag}(\mathbf{D}) = (d_1, \ldots, d_n) = \sqrt{w}\,(\theta_1, \ldots, \theta_n)$$
$$\text{diag}(\mathbf{B}) = (b_1, \ldots, b_n) = \sqrt{w}\,(\beta_1, \ldots, \beta_n)$$

where $w > 0, 1 = \theta_1 > \theta_2 > \ldots > \theta_{n-1} > \theta_n = 0$ and $1 = \beta_1 > \beta_2 > \ldots > \beta_{n-1} > \beta_n = 0$. Note that these restrictions remove the label switching problem which was present in the homoskedastic model, but is not present here.

This leads to

$$p(\mathbf{U} \mid \theta, \beta, w, \mathbf{V}) \propto w^{m/2} e^{(-w\theta'(\mathbf{I}-\mathbf{M})\beta)} \prod_{i<j} (\theta_i - \theta_j)^{1/2} (\beta_i - \beta_j)^{1/2} \tag{13}$$

where $m = \begin{pmatrix} n \\ 2 \end{pmatrix} = \frac{n(n-1)(n-2)\cdots(n-2+1)}{2!}$, the matrix $\mathbf{M} = (\mathbf{V}'\mathbf{U}) \circ (\mathbf{V}'\mathbf{U})$, and $\circ$ is the Hadamard product operator denoting elementwise multiplication.

It can be see that we have a Gamma density for $w$ (conditional on other parameters) which, combined with a Gamma prior leads to a conditional posterior which is Gamma. The Gamma prior we use is

$$w \sim \mathcal{G}\left(\eta_0/2, \tau_0^2\right))$$

where $\mathcal{G}$ denotes the Gamma distribution. For the hyperparameters, we follow Hoff (2009b) and set $\eta_0 = 2$, $\tau_0^2 = 10$. This choice willl result in an exponential prior distribution that has its mode at $w = 0$ but is very diffuse. Thus our prior has a mode at eigenvector homogeneity but the diffuse feature means we are allowing for a large range of possible values for eigenvector heterogeneity.

With regard to the prior on $p(\theta, \beta)$, it is taken to be such that $1 > \theta_2 > \ldots > \theta_{n-1} > 0$ and $1 > \beta_2 > \ldots > \beta_{n-1} > 0$ are two independent sets of order statistics of uniform random

variables on $[0,1]$. Multiplying this prior by (13) leads to a conditional posteroir that can easily be sampled from on a grid of $[0,1]$.

To summarize the algorithm to update $w$, $\theta$ and $\beta$ involves:

*Step a:* sample $w$ from a Gamma distribution;

*Step b:* for each $i \in \{2,\ldots,n-1\}$ sample $\theta_i \in (\theta_{i-1},\theta_{i+1})$ from the density proportional to

$$e^{-\theta_i\left(w\beta'\mathbf{M}_{[i,]}\right)} \prod_{j:j\neq i} |\theta_i - \theta_j|^{1/2}$$

*Step c:* for each $j \in \{2,\ldots,n-1\}$ sample $\beta_j \in (\beta_{j-1},\beta_{j+1})$ from the density proportional to

$$e^{-\beta_j\left(w\mathbf{M}'_{[,j]}\theta\right)} \prod_{i:i\neq j} |\beta_j - \beta_i|^{1/2}$$

**Step3** : $p\left(\mathbf{V} \mid \mathbf{U},\mathbf{D},\mathbf{B}\right)$

Since $p\left(\mathbf{V} \mid \mathbf{U} \mid \mathbf{D},\mathbf{B}\right) \propto p\left(\mathbf{U} \mid \mathbf{D},\mathbf{B},\mathbf{V}\right) p\left(\mathbf{V} \mid \mathbf{D},\mathbf{B}\right)$ the Bingham prior for $\mathbf{U}$ again plays a key role. However, since the integrating constant, $c(\mathbf{D},\mathbf{B})$ does not depend on $\mathcal{V}$, the complications of Step 2 do not arise. Assuming a uniform prior for $\mathbf{V}$ we obtain

$$\begin{aligned} p(\mathbf{V} \mid \mathbf{U},\mathbf{D},\mathbf{B}) &\propto \mathrm{e}^{\mathrm{tr}(\mathbf{BU'VDV'U})} \\ &= \mathrm{e}^{\mathrm{tr}(\mathbf{DV'UBU'V})} \end{aligned} \tag{14}$$

which is a Bingham distribution. We use the same strategy of drawing two columns of $\mathbf{V}$ at a time as we do when drawing $\mathbf{U}$.

As before, we compare results from our BVAR-eigSV model to those of a Cholesky-decomposed model. This BVAR-cholSV model, along with the necessary MCMC algorithm, is given in Chan et al. (2022). We stress that, to our knowledge, the latter MCMC algorithm is the fastest available algorithm which, as documented in Chan et al. (2022), has computational complexity $O(n^4)$ as opposed to the $O(n^6)$ of full system estimation of

the reduced form VAR. Hence, we compare our methods to the fastest algorithm available instead of to full system estimation which has already been established as being very slow. To get a sense of how long it takes to obtain posterior draws using the proposed eigen-decomposition method, we report the computational times (in seconds) to obtain 10,000 posterior draws in Table 2. As it is evident from the table, our proposed MCMC algorithm has roughly the same computational burden of Chan et al. (2022) and estimating the hyperparameters does not greatly add to the burden.

Table 2: The computation times (in minutes) to obtain 10,000 posterior draws using the proposed Eigendecomposition method compared to the Cholesky decomposition method. All BVARs have $n = 20$ variables and $p = 4$ lags.

| | | Homoskedastic case | SV case |
|---|---|---|---|
| Cholesky | | 9.6 | 10.9 |
| Eigendecomposition | $\mathcal{BIG}$ | 12.5 | - |
| | $\mathcal{IBIG}$ | 12.5 | 13.9 |

[1] $\mathcal{BIG}$ prior is derived from inverse-Wishart distribution. $\mathcal{IBIG}$ prior treats $\mathbf{B}$, $\mathbf{D}$, and $\mathbf{V}$ as unknown.

# 5 Macroeconomic Forecasting Using the BVAR-eigSV

## 5.1 Summary and Overview

In this section we carry out a forecasting exercise using the data set described in Section 3. The forecast exercise is set up in the same way as in that section except that now all of our models include SV. One of the key advantages of our approach is that it is order invariant whereas BVAR-cholSV is not. To show the importance of this property, we repeat our forecast exercise twice. In the first case, the variables are ordered as in Appendix B. In the second, the ordering is reversed. Of course, the forecasts produced by BVAR-eigSV will be the same for these two cases, but it is possible that they differ for BVAR-cholSV.

## 5.2 Results of the Forecasting Exercise

Table 3 reports the forecasting results using the variables in the original order. Table 4 is of the same format but uses the variables in reverse order.

In our previous forecasting exercise using homoskedastic models, an important finding was that priors matter. This finding is reinforced with in Tables 3 and 4. All the models in this paper have the same likelihood function and differ only in the prior. Yet, the forecast performance is substantially different in many cases. To this story can be added another one: ordering matters for the Cholesky-decomposed VARs — and it matters a lot. This reinforces the findings of Arias et al. (2021) and Chan et al. (2021) and motivates the need for computationally efficient order-variant approaches to Bayesian VAR analysis.

For density forecasts, from Table 3, we find BVAR-eigSV to outperform BVAR-cholSV overall (i.e. the joint ALPL of the former is much higher at both forecast horizons) and for most of the variables individually. From Table 4, where the variables are reversed in order, the same statement holds but to a much larger extent. We also find that the forecast performance for some individual variables, of BVAR-cholSV is very poor relative to the BVAR-eigSV.

With Cholesky-decomposed approaches, it would be possible that there is some systematic property relating to the ordering which affects the results. For instance, it is possible that, for typical macroeconomic data sets, that Cholesky approaches always forecast worse for variables ordered first (or vice versa). We find no evidence of this. In Table 3, we find BVAR-cholSV to forecast worst for the first four variables, whereas in Table 4, the reverse holds and it is the last four variables. BVAR-eigSV forecasts well in almost all cases but if there is any general pattern as to when the forecasts of BVAR-cholSV catch up or improve on them it is will the financial variables. There is a slight tendency for eigendecomposed approaches to work relatively better for the macroeconomic variables than the financial

ones.

Table 3: Forecasting performance: percentage gains in RMSFE and ALPL (BVAR-eigSV relative to BVAR-cholSV)

| | RMSFE | | ALPL | |
|---|---|---|---|---|
| | $h = 1$ | $h = 4$ | $h = 1$ | $h = 4$ |
| GDPC1 | -2.07 | 0.16 | 0.38 | -0.46 |
| PCECC96 | 1.03 | 0.33 | -2.53 | 92.56 |
| INDPRO | -1.04 | -0.29 | 7.31 | -10.32 |
| IPFINAL | -0.69 | -0.21 | 6.14 | -18.33 |
| PAYEMS | 4.07 | 1.01 | 179.00 | 75.02 |
| MANEMP | 1.13 | 1.14 | 15.16 | -7.54 |
| CE16OV | 4.02 | 0.24 | -56.55 | -81.28 |
| CIVPART | 2.09 | 4.80 | -0.08 | 57.56 |
| UNRATE | 5.61 | 6.09 | 261.19 | 179.70 |
| HOANBS | 1.44 | 0.45 | 49.89 | 92.50 |
| HOUST | 0.05 | 0.56 | -0.94 | -2.17 |
| PERMIT | 0.26 | 0.40 | 3.48 | -1.27 |
| PCECTPI | -1.60 | -1.51 | 1.16 | 4.50 |
| CPIAUCSL | -2.91 | -1.35 | 0.90 | 3.05 |
| OPHNFB | -1.05 | -0.06 | 0.24 | 2.46 |
| FEDFUNDS | -3.28 | -0.35 | -21.92 | 30.23 |
| TB3MS | -1.20 | 0.03 | -11.99 | 51.73 |
| GS1 | -0.01 | 1.09 | -6.71 | 34.79 |
| GS10 | -1.12 | -0.28 | -3.26 | 3.88 |
| BAA10YM | -0.12 | -0.00 | 6.08 | 47.68 |
| Joint Forecasting | | | | |
| All | | | 2544.38 | 6105.05 |
| First 4 | | | 568.17 | 556.36 |
| Last 4 | | | 26.41 | 417.33 |

Table 4: Reverse ordering Forecasting performance: percentage gains in RMSFE and ALPL (BVVAR-eigSV against BVAR-cholSV)

| | RMSFE | | ALPL | |
|---|---|---|---|---|
| | $h=1$ | $h=4$ | $h=1$ | $h=4$ |
| BAA10YM | 2.02 | 9.55 | -2.17 | 283.59 |
| GS10 | 2.43 | 6.58 | 3.58 | 171.40 |
| GS1 | 3.96 | 7.81 | 13.54 | 99.36 |
| TB3MS | -1.25 | 3.66 | 7.00 | 20.17 |
| FEDFUNDS | 1.39 | 4.94 | 1.88 | 7.77 |
| OPHNFB | 0.42 | 0.81 | 1.21 | 1.14 |
| CPIAUCSL | 3.95 | -0.42 | 2.55 | 19.25 |
| PCECTPI | 2.19 | -1.30 | 2.82 | 15.89 |
| PERMIT | 0.36 | 0.30 | -0.55 | 1.34 |
| HOUST | 1.24 | 0.62 | -1.54 | -1.19 |
| HOANBS | 3.26 | 0.28 | -26.58 | 38.31 |
| UNRATE | 4.11 | 4.57 | 232.45 | 84.37 |
| CIVPART | 1.08 | 3.08 | -8.06 | 192.31 |
| CE16OV | 1.94 | 0.24 | 71.85 | 115.89 |
| MANEMP | 5.05 | 0.14 | 36.69 | 24.54 |
| PAYEMS | 2.09 | 0.07 | 242.81 | 546.93 |
| IPFINAL | 5.17 | 0.48 | 16.08 | -29.12 |
| INDPRO | 5.12 | 0.68 | 25.86 | -5.13 |
| PCECC96 | 3.40 | 0.63 | 1.98 | 28.57 |
| GDPC1 | 4.03 | 0.56 | 16.42 | 39.87 |
| Joint Forecasting | | | | |
| All | | | 30647.79 | 44858.40 |
| First 4 | | | 577.72 | 1363.66 |
| Last 4 | | | 2304.64 | 2115.66 |

# 6 Summary and Conclusions

Bayesian inference in large VARs is computationally very demanding, particular when MCMC methods are used, if the entire system of $n$ equations is estimated at once (i.e. the VAR is estimated in reduced form). However, if the VAR is transformed to a structural form so as to have a diagonal error covariance matrix, then estimation can proceed one equation at a time. This leads to a large reduction in the computational burden, making Bayesian methods practical even in very large VARs. Traditionally, the Cholesky decomposition has

been used to decompose the reduced form VAR error covariance matrix. This leads to fast computation, but results depend on the way the variables are ordered in the VAR. In this paper, we propose use of the eigendecomposition which leads to similar computational benefits, but is order invariant. We show how use of a Bingham prior on the eigenmatrix and inverse Gamme priors on the eigenvalues of the eigendecomposition is equivalent to the use of inverse-Wishart prior on the reduced form error covariance matrix and leads to an MCMC algorithm of a simple form.

In the latter half of the paper, we focus on extending the model to allow for time-varying volatility. We show how this can be done in the eigen-transformed VAR in a manner which leads to only a small increase in the computational burden. This contrasts with conventional methods of adding SV to the reduced form VAR which can greatly add to the computational burden. The forecasting exercises in this paper show the importance of variable ordering in Cholesky-based approaches and the benefits of using our order-invariant Bingham inverse-Gamma prior.

# References

J. E. Arias, J. F. Rubio-Ramírez, M. Shin, et al. *Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models.* Research Department, Federal Reserve Bank of Philadelphia, 2021.

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

B. S. Bernanke, J. Boivin, and P. Eliasz. Measuring the effects of monetary policy: a factor-augmented vector autoregressive (favar) approach. *The Quarterly journal of economics*, 120(1):387–422, 2005.

C. Bingham. An antipodally symmetric distribution on the sphere. *The Annals of Statistics*, pages 1201–1225, 1974.

A. Carriero, T. E. Clark, and M. Marcellino. Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 212(1):137–154, 2019.

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

J. Chan. Large bayesian vars: A flexible kronecker error covariance structure. *Journal of Business and Economic Statistics*, 38:68–79, 2020.

J. Chan. Comparing stochastic volatility specifications for large bayesian vars, 2021a.

J. Chan, A. Carriero, T. E. Clark, and M. Marcellino. Corregindum to: Large bayesian vector autoregressions with stochastic volatility and non-conjugate priors. *Journal of Econometrics*, 227(2):506–512, 2022.

J. C. Chan. Asymmetric conjugate priors for large bayesian vars. *arXiv preprint arXiv:2111.07170*, 2021b.

J. C. Chan and I. Jeliazkov. Efficient simulation and integrated likelihood estimation in state space models. *International Journal of Mathematical Modelling and Numerical Optimisation*, 1(1-2):101–120, 2009.

J. C. Chan, G. Koop, and X. Yu. Large order-invariant bayesian vars with stochastic volatility. *arXiv preprint arXiv:2111.07225*, 2021.

S. Chib, F. Nardari, and N. Shephard. Analysis of high dimensional multivariate stochastic volatility models. *Journal of Econometrics*, 134(2):341–371, 2006.

T. Cogley and T. J. Sargent. Drifts and volatilities: monetary policies and outcomes in the post wwii us. *Review of Economic Dynamics*, 8(2):262–302, 2005.

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

S. Frühwirth-Schnatter. Markov chain monte carlo estimation of classical and dynamic switching and mixture models. *Journal of the American Statistical Association*, 96(453): 194–209, 2001.

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

A. K. Gupta and D. K. Nagar. *Matrix variate distributions*. Chapman and Hall/CRC, 2018.

P. D. Hoff. Simulation of the matrix bingham-von mises-fisher distribution, with applications to multivariate and relational data. *Journal of Computatiponal Graphics and Statistics*, 18(2):438–456, 2009a.

P. D. Hoff. A hierarchical eigenmodel for pooled covariance estimation. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(5):971–992, 2009b.

G. Kastner and F. Huber. Sparse bayesian vector autoregressions in huge dimensions. *Journal of Forecasting*, 39(7):1142–1165, 2020.

C. Khatri and K. V. Mardia. The von mises–fisher matrix distribution in orientation statistics. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(1): 95–106, 1977.

S. Kim, N. Shephard, and S. Chib. Stochastic volatility: likelihood inference and comparison with arch models. *The review of economic studies*, 65(3):361–393, 1998.

G. Koop and D. Korobilis. Large time-varying parameter vars. *Journal of Econometrics*, 177(2):185–198, 2013.

G. Koop, D. Korobilis, and D. Pettenuzzo. Bayesian compressed vector autoregressions. *Journal of Econometrics*, 210(1):135–154, 2019.

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

D. Korobilis. Var forecasting using bayesian variable selection. *Journal of Applied Econometrics*, 28(2):204–230, 2013.

B. P. Levy and H. F. Lopes. Dynamic ordering learning in multivariate forecasting. *arXiv preprint arXiv:2101.04164*, 2021.

L. C. Parra, C. D. Spence, A. D. Gerson, and P. Sajda. Recipes for the linear analysis of eeg. *Neuroimage*, 28(2):326–341, 2005.

M. K. Pitt and N. Shephard. Filtering via simulation: Auxiliary particle filters. *Journal of the American statistical association*, 94(446):590–599, 1999.

G. E. Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.

C. A. Sims. Macroeconomics and reality. *Econometrica*, pages 1–48, 1980.

# Online Appendix

## A Evidence of Order Invariance

It is well-known that Bayesian inference in the VAR with inverse-Wishart prior on the error covariance matrix is order-invariant. Given the equivalence between our $\mathcal{BIG}$ prior and the inverse-Wishart prior it would have sufficed to simply state this fact to establish order-invariance of our approach. However, in the interest of theoretical rigor, in this appendix we provide a proof of order invariance.

### A.1 Original and New parameters

Consider Equation 1, which is a standard homoscedastic VAR of order $p$:

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{U}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$$

Suppose we permute the order of the dependent variables. More precisely, let $\mathbf{P}$ denote an arbitrary permutation matrix of dimension $n$. By left multiplying matrix $\mathbf{P}$, we get:

$$\mathbf{P}\mathbf{y}_t = \mathbf{P}\mathbf{b} + \mathbf{P}\mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{P}\mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{P}\mathbf{U}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda})$$

we define $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$, $\widetilde{\mathbf{b}} = \mathbf{P}\mathbf{b}$, $\widetilde{\mathbf{A}}_1 = \mathbf{P}\mathbf{A}_1, \ldots, \widetilde{\mathbf{A}}_p = \mathbf{P}\mathbf{A}_p$, then we can get the new VAR($p$) :

$$\widetilde{\mathbf{y}}_t = \widetilde{\mathbf{b}} + \widetilde{\mathbf{A}}_1\mathbf{y}_{t-1} + \cdots + \widetilde{\mathbf{A}}_p\mathbf{y}_{t-p} + \mathbf{P}\mathbf{U}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Lambda}) \tag{15}$$

It is clear that if we permute the order of the dependent variables via $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$, new coefficient matrices ($\widetilde{\mathbf{b}}$ and $\widetilde{\mathbf{A}}_j$) can be obtained by permuting the rows and columns of original coefficient matrices accordingly.

Next, we show evidence that the error covariance matrix is also permuted. The error covariance matrix in Equation (1) is

$$\Sigma = \mathbf{U}\mathbf{\Lambda}\mathbf{U}' = \left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)\left(\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\right)'$$

The error covariance matrix in Equation (15) is

$$\widetilde{\Sigma} = \mathbf{P}\mathbf{U}\mathbf{\Lambda}(\mathbf{P}\mathbf{U})' = \left(\mathbf{P}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}'\right)\left(\mathbf{P}\mathbf{U}\mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}'\right)' = \left(\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}}\right)\left(\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}}\right)'$$

The second equality is because all permutation matrices are orthogonal matrices, we have $\mathbf{P}^{-1} = \mathbf{P}'$. New matrices $\widetilde{\mathbf{U}} = \mathbf{P}\mathbf{U}$, $\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}'$ mean that the old eigenvector and eigenvalue matrices are permuted according to $\mathbf{P}$.

Thus far, we have shown that if we permute the order of the dependent variables via $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$, new parameters can be obtained by permuting the old parameters accordingly. This is applicable to both coefficient matrices ($\widetilde{\mathbf{b}}$ and $\widetilde{\mathbf{A}}_j$) and the error covariance matrix $\widetilde{\Sigma}$. We summarize them in Table 5.

Table 5: Original and New parameters

| | Original | New | Relationship |
|---|---|---|---|
| Intercepts | $\mathbf{b}$ | $\widetilde{\mathbf{b}}$ | $\widetilde{\mathbf{b}} = \mathbf{P}\mathbf{b}$ |
| Coefficients | $\mathbf{A}_j$ | $\widetilde{\mathbf{A}}_j$ | $\widetilde{\mathbf{A}}_j = \mathbf{P}\mathbf{A}_j$ |
| Eigenvectors | $\mathbf{U}$ | $\widetilde{\mathbf{U}}$ | $\widetilde{\mathbf{U}} = \mathbf{P}\mathbf{U}$ |
| Eigenvalues | $\mathbf{\Lambda}^{\frac{1}{2}}$ | $\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}}$ | $\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}'$ |

Next, we prove that the conditional likelihood function implied by this new ordering is the same as that of the original ordering.

### A.2 Original and New conditional likelihoods

We first stack the original dependent variables $\mathbf{y}_t$ and new dependent variables $\widetilde{\mathbf{y}}_t$ over time.

We stack the original dependent variables into a $T \times n$ matrix $\mathbf{Y}$ so that its $t$-th row is $\mathbf{y}_t'$. Now, let $\mathbf{X}$ be a $T \times k$ matrix of regressors, where the $t$-th row is $\mathbf{x}_t' = \left(1, \mathbf{y}_{t-1}', \ldots, \mathbf{y}_{t-p}'\right)$. Next, let $\mathbf{A} = (\mathbf{b}, \mathbf{A}_1, \ldots, \mathbf{A}_p)'$ denote the $k \times n$ matrix of VAR coefficients. Then, we can write the original VAR($p$) as follows:

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{E}\mathbf{U}' \tag{16}$$

where $\mathbf{E}$ is a $T \times n$ matrix of innovations in which the $t$-th row is $\boldsymbol{\varepsilon}_t'$. It follows that

$$\text{vec}(\mathbf{E}\mathbf{U}') \sim \mathcal{N}\left(\mathbf{0}, (\mathbf{U}\boldsymbol{\Lambda}\mathbf{U}') \otimes \mathbf{I}_T\right)$$

Stacking the new dependent variables $\widetilde{\mathbf{y}}_t$ into $\widetilde{\mathbf{Y}} = (\widetilde{\mathbf{y}}_1, \ldots, \widetilde{\mathbf{y}}_T)'$, Next, let $\widetilde{\mathbf{A}} = \left(\widetilde{\mathbf{b}}, \widetilde{\mathbf{A}}_1, \ldots, \widetilde{\mathbf{A}}_p\right)'$ denote the $k \times n$ matrix of VAR coefficients. Then, we can write the new VAR($p$) as follows:

$$\widetilde{\mathbf{Y}} = \mathbf{X}\widetilde{\mathbf{A}} + \mathbf{E}\widetilde{\mathbf{U}}' \tag{17}$$

where $\mathbf{E}$ is a $T \times n$ matrix of innovations in which the $t$-th row is $\boldsymbol{\varepsilon}_t'$. It follows that

$$\text{vec}(\mathbf{E}\widetilde{\mathbf{U}}') \sim \mathcal{N}\left(\mathbf{0}, (\widetilde{\mathbf{U}}\widetilde{\boldsymbol{\Lambda}}\widetilde{\mathbf{U}}') \otimes \mathbf{I}_T\right)$$

Table 6 summarizes their relationships. Next, we prove that $p(\widetilde{\mathbf{Y}} \mid \widetilde{\mathbf{A}}, \widetilde{\mathbf{U}}, \widetilde{\boldsymbol{\Lambda}}) = p(\mathbf{Y} \mid \mathbf{A}, \mathbf{U}, \boldsymbol{\Lambda})$. We first show the conditional likelihood of the original ordering, then we show the conditional likelihood of the new ordering.

Table 6: Original and New matrices

| | Original | New | Relationship |
|---|---|---|---|
| Dependent variables | $\mathbf{Y}$ | $\widetilde{\mathbf{Y}}$ | $\widetilde{\mathbf{Y}} = \mathbf{Y}\mathbf{P}'$ |
| Coefficients | $\mathbf{A}$ | $\widetilde{\mathbf{A}}$ | $\widetilde{\mathbf{A}} = \mathbf{A}\mathbf{P}'$ |
| Eigenvectors | $\mathbf{U}$ | $\widetilde{\mathbf{U}}$ | $\widetilde{\mathbf{U}} = \mathbf{P}\mathbf{U}$ |
| Eigenvalues | $\mathbf{\Lambda}^{\frac{1}{2}}$ | $\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}}$ | $\widetilde{\mathbf{\Lambda}}^{\frac{1}{2}} = \mathbf{\Lambda}^{\frac{1}{2}}\mathbf{P}'$ |
| Conditional Likelihoods | $p(\mathbf{Y} \mid \mathbf{A}, \mathbf{U}, \mathbf{\Lambda})$ | $p(\widetilde{\mathbf{Y}} \mid \widetilde{\mathbf{A}}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{\Lambda}})$ | same |

### A.3 Likelihoods of the original ordering

Likelihoods of the original ordering are given by

$$p(\mathbf{Y} \mid \mathbf{A}, \mathbf{U}, \mathbf{\Lambda}) = (2\pi)^{-\frac{Tn}{2}} |\mathbf{U}\mathbf{\Lambda}\mathbf{U}'|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\mathbf{A})(\mathbf{U}\mathbf{\Lambda}\mathbf{U}')^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{A})'\right)} \tag{18}$$

$$= (2\pi)^{-\frac{Tn}{2}} |\mathbf{\Lambda}|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\mathbf{A})(\mathbf{U}\mathbf{\Lambda}\mathbf{U}')^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{A})'\right)} \tag{19}$$

$$= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\mathbf{A})(\mathbf{U}\mathbf{\Lambda}\mathbf{U}')^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{A})'\right)} \tag{20}$$

$$= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left((\mathbf{Y}-\mathbf{X}\mathbf{A})\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{X}\mathbf{A})'\right)} \tag{21}$$

$$= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{X}\mathbf{A})'(\mathbf{Y}-\mathbf{X}\mathbf{A})\right)} \tag{22}$$

Proof from (18) to (19):

Since $\det(\mathbf{B}\mathbf{C}) = \det(\mathbf{B})\det(\mathbf{C})$, and $\det(\mathbf{B}^{-1}) = \frac{1}{\det(\mathbf{B})}$, we can get $|\mathbf{U}\Lambda\mathbf{U}'|^{-\frac{T}{2}} = (|\mathbf{U}||\Lambda||\mathbf{U}'|)^{-\frac{T}{2}} = (|\mathbf{U}||\Lambda||\mathbf{U}^{-1}|)^{-\frac{T}{2}} = \left(|\mathbf{U}||\Lambda|\frac{1}{|\mathbf{U}|}\right)^{-\frac{T}{2}} = |\Lambda|^{-\frac{T}{2}}$.

Proof from (19) to (20):

Matrix $\Lambda$ is diagonal, and for a diagonal matrix $B$, $\det(B) = b_{11}b_{22}\cdots b_{nn} = \prod_{i=1}^{n} b_{ii}$, so $|\mathbf{\Lambda}|^{-\frac{T}{2}} = \left(\prod_{j=1}^{n} \lambda_j\right)^{-\frac{T}{2}} = \prod_{j=1}^{n} \lambda_j^{-T/2}$.

Proof from (20) to (21):

Matrix $\mathbf{U}$ is orthogonal, which means that $\mathbf{U}^{-1} = \mathbf{U}'$.

Then $(\mathbf{U}\mathbf{\Lambda}\mathbf{U}')^{-1} = (\mathbf{U}')^{-1}(\mathbf{\Lambda})^{-1}(\mathbf{U})^{-1} = \mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}'$.

Proof from (21) to (22): $\text{tr}(\mathbf{BCD}) = \text{tr}(\mathbf{CDB}) = \text{tr}(\mathbf{DBC})$

### A.4 Likelihoods of the new ordering

Likelihoods of the new ordering are given by

$$
\begin{aligned}
p(\widetilde{\mathbf{Y}} \mid \widetilde{\mathbf{A}}, \widetilde{\mathbf{U}}, \widetilde{\mathbf{\Lambda}}) &= (2\pi)^{-\frac{Tn}{2}} |(\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}')|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\text{tr}\left((\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})(\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}')^{-1}(\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})'\right)} && (23)\\
&= (2\pi)^{-\frac{Tn}{2}} |\mathbf{\Lambda}|^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2}\text{tr}\left((\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})(\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}')^{-1}(\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})'\right)} && (24)\\
&= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\text{tr}\left((\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})(\mathbf{PU\Lambda U'P'})^{-1}(\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})'\right)} && (25)\\
&= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\text{tr}\left((\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})\mathbf{PU\Lambda}^{-1}\mathbf{U'P'}(\widetilde{\mathbf{Y}}-\mathbf{X}\widetilde{\mathbf{A}})'\right)} && (26)\\
&= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\text{tr}\left((\mathbf{Y}-\mathbf{XA})\mathbf{U\Lambda}^{-1}\mathbf{U'}(\mathbf{Y}-\mathbf{XA})'\right)} && (27)\\
&= (2\pi)^{-\frac{Tn}{2}} \prod_{j=1}^{n} \lambda_j^{-T/2} \mathrm{e}^{-\frac{1}{2}\text{tr}\left(\mathbf{U\Lambda}^{-1}\mathbf{U'}(\mathbf{Y}-\mathbf{XA})'(\mathbf{Y}-\mathbf{XA})\right)} && (28)\\
&= p(\mathbf{Y} \mid \mathbf{A}, \mathbf{U}, \mathbf{\Lambda})
\end{aligned}
$$

Proof from (23) to (24):

Since $\det(\mathbf{BC}) = \det(\mathbf{B})\det(\mathbf{C})$, and $\det(\mathbf{B}^{-1}) = \frac{1}{\det(\mathbf{B})}$, we can get $|(\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}')|^{-\frac{T}{2}} = |\mathbf{PU\Lambda}^{\frac{1}{2}}\mathbf{P'P}(\mathbf{\Lambda}^{\frac{1}{2}})'\mathbf{U'P'}|^{-\frac{T}{2}} = |\mathbf{PU\Lambda U'P'}|^{-\frac{T}{2}} = (|\mathbf{P}||\mathbf{U}||\mathbf{\Lambda}||\mathbf{U'}||\mathbf{P'}|)^{-\frac{T}{2}} = (|\mathbf{P}||\mathbf{U}||\mathbf{\Lambda}||\mathbf{U}^{-1}||\mathbf{P}^{-1}|)^{-\frac{T}{2}} = \left(|\mathbf{P}||\mathbf{U}||\mathbf{\Lambda}||\frac{1}{|\mathbf{U}|}|\frac{1}{|\mathbf{P}|}\right)^{-\frac{T}{2}} = |\mathbf{\Lambda}|^{-\frac{T}{2}}$.

Proof from (24) to (25): $\widetilde{\mathbf{U}}\widetilde{\mathbf{\Lambda}}\widetilde{\mathbf{U}}' = \mathbf{PU\Lambda}^{\frac{1}{2}}\mathbf{P'P}(\mathbf{\Lambda}^{\frac{1}{2}})'\mathbf{U'P'} = \mathbf{PU\Lambda U'P'}$.

Proof from (25) to (26): Same as Proof from (20) to (21).

Proof from (26) to (27):

From Table 6, $\widetilde{\mathbf{Y}} = \mathbf{YP'}$, $\widetilde{\mathbf{A}} = \mathbf{AP'}$.

So $\widetilde{\mathbf{Y}} - \mathbf{X}\widetilde{\mathbf{A}} = (\mathbf{Y} - \mathbf{XA})\mathbf{P'}$, $(\widetilde{\mathbf{Y}} - \mathbf{X}\widetilde{\mathbf{A}})' = \mathbf{P}(\mathbf{Y} - \mathbf{XA})'$.

Then

$(\widetilde{\mathbf{Y}} - \mathbf{X}\widetilde{\mathbf{A}})\mathbf{PU\Lambda}^{-1}\mathbf{U'P'}(\widetilde{\mathbf{Y}} - \mathbf{X}\widetilde{\mathbf{A}})' = (\mathbf{Y} - \mathbf{XA})\mathbf{P'PU\Lambda}^{-1}\mathbf{U'P'P}(\mathbf{Y} - \mathbf{XA})' = (\mathbf{Y} -$

$\mathbf{XA})\mathbf{U}\mathbf{\Lambda}^{-1}\mathbf{U}'(\mathbf{Y}-\mathbf{XA})'$.

Proof from (27) to (28): Same as Proof from (21) to (22).

Combining Section A.1 and Section A.2, we provide evidence that if we permute the order of the dependent variables via $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$, new parameters can be obtained by permuting the old parameters accordingly. This is applicable to both coefficient matrices ($\widetilde{\mathbf{b}}$ and $\widetilde{\mathbf{A}}_j$) and the error covariance matrix $\widetilde{\Sigma}$. And the conditional likelihood function implied by this new ordering is the same as that of the original ordering.

# B Data

Table 7: Description of variables used in the forecasting application

| Variable | Mnemonic | Transformation |
|---|---|---|
| Real Gross Domestic Product | GDPC1 | 400$\Delta$ log |
| Personal Consumption Expenditures | PCECC96 | 400$\Delta$ log |
| Industrial Production Index | INDPRO | 400$\Delta$ log |
| Industrial Production: Final Products | IPFINAL | 400$\Delta$ log |
| All Employees: Total nonfarm | PAYEMS | 400$\Delta$ log |
| All Employees: Manufacturing | MANEMP | 400$\Delta$ log |
| Civilian Employment | CE16OV | 400$\Delta$ log |
| Civilian Labor Force Participation Rate | CIVPART | no transformation |
| Civilian Unemployment Rate | UNRATE | no transformation |
| Nonfarm Business Section: Hours of All Persons | HOANBS | 400$\Delta$ log |
| Housing Starts: Total | HOUST | 400$\Delta$ log |
| New Private Housing Units Authorized by Building Permits | PERMIT | 400$\Delta$ log |
| Personal Consumption Expenditures: Chain-type Price index | PCECTPI | 400$\Delta$ log |
| Consumer Price Index for All Urban Consumers: All Items | CPIAUCSL | 400$\Delta$ log |
| Nonfarm Business Section: Real Output Per Hour of All Persons | OPHNFB | 400$\Delta$ log |
| Effective Federal Funds Rate | FEDFUNDS | no transformation |
| 3-Month Treasury Bill: Secondary Market Rate | TB3MS | no transformation |
| 1-Year Treasury Constant Maturity Rate | GS1 | no transformation |
| 10-Year Treasury Constant Maturity Rate | GS10 | no transformation |
| Moodys Seasoned Baa Corporate Bond Yield Relative to Yield on 10-Year Treasury Constant Maturity | BAA10YM | no transformation |