Bayesian Rank Selection in Multivariate Regressions

Bin Jiang, Anastasios Panagiotelis,
George Athanasopoulos, Rob J Hyndman, and Farshid Vahid

Department of Econometrics and Business Statistics, Monash University,
Clayton, Victoria 3800, Australia

29 January 2016

Abstract

Estimating the rank of the coefficient matrix is a major challenge in multivariate regression, including vector autoregression (VAR). In this paper, we develop a novel fully Bayesian approach that allows for rank estimation. The key to our approach is reparameterizing the coefficient matrix using its singular value decomposition and conducting Bayesian inference on the decomposed parameters. By implementing a stochastic search variable selection on the singular values of the coefficient matrix, the ultimate selected rank can be identified as the number of nonzero singular values. Our approach is appropriate for small multivariate regressions as well as for higher dimensional models with up to about 40 predictors. In macroeconomic forecasting using VARs, the advantages of shrinkage through proper Bayesian priors is well documented. Consquently, the shrinkage approach proposed here that selects or average over low rank coefficient matrices is evaluated in a forecasting environment. We show in both simulations and empirical studies that our Bayesian approach provides forecasts that are highly competitive compared to two of most promising benchmarks methods, dynamic factor models and factor augmented VARs.

1 Introduction

There has been a growing literature on the Bayesian vector autoregression (VAR) approach to macroeconomic forecasting, beginning with Sims (1980) who encouraged the rejection of the "incredible" assumptions usually employed for identifying structural models and advocated the use of Bayesian techniques. The Bayesian VAR approach can take advantage of VAR models that require only a small set of plausible identification restrictions and can provide an entire posterior predictive distribution on possible outcomes of the economy in the future which is more useful than approaches producing point forecasts only, as pointed out by Littleman (1986).

Since the 1990s, VAR analyses have been largely used in macroeconomic studies and applications (e.g., Bernanke and Blinder, 1992; Sims, 1992; Leeper et al., 1996; Sims and Zha, 1998; Robertson and Tallman, 1999). However, VAR models are often restricted to include only a small

number of variables in practice (typically less than ten) because of the easily exhausted degrees of freedom, and forecasts based on such low dimensional VARs are likely to be contaminated as stressed by Bernanke et al. (2005). The problem of estimating higher dimensional VARs is extremely challenging and only a few contributions have been made to this literature. In general, these contributions can be categorized roughly into two strands.

One strand utilizes Bayesian shrinkage to exploit the sparse structure of the coefficient matrix which means that most of the VAR coefficients are so small that their effects are negligible. For instance, Bańbura et al. (2010) implement Bayesian shrinkage on VAR coefficients through the utilization of a natural conjugate extension of the Minnesota prior (e.g. Doan et al., 1984; Littleman, 1986) and show that Bayesian VARs can perform better in forecasting high dimensional time series than factor models. Koop (2013) imposes the stochastic search variable selection (SSVS) prior of George et al. (2008) on each element of the coefficient matrix to conduct Bayesian shrinkage in an automatic fashion, and implements a combination of the SVSS prior and the Minnesota prior to improve the forecast compared to using each method alone.

In fact, because macroeconomic time series are highly correlated, the VAR coefficient matrix is not only sparse but also low-rank; thus using a hybrid method of shrinkage and rank reduction leads to additional gains on forecasting accuracy. This motivates the other strand which considers reduced rank multivariate regressions. A typical work is Carriero et al. (2011) who investigated a reduced rank approximation of the posterior estimate obtained using the Minnesota prior in a large VAR and demonstrated that forecasts taking into account both sparsity and rank deficiency can outperform those with either shrinkage or rank reduction (e.g. dynamic factor models) only. However, since the rank is generally unknown, rank selection is a major challenge in forecasting methods involving rank reduction. In practice, researchers have to rely on the sensitivity analysis of the results obtained with alternative ranks. For example, Bernanke et al. (2005) determined how many factors should be included in the factor augmented vector autoregression (FAVAR) approach by checking the sensitivity of the results to different numbers of factors. Carriero et al. (2011) searched over all possible rank values and identified the 'true' rank of their estimate as the one minimizing the forecast error. It is worth mentioning that rank selection is also a challenging issue even in the classical literature on small-scale reduced-rank regression models, since many tests or selecting procedures (e.g. Anderson, 1951; Geweke, 1996; Kleibergen and Paap, 2002; Strachan, 2003) are developed but techniques estimating the rank are rarely provided (e.g. Strachan and Inder, 2004). A common weakness of these procedures and the aforementioned sensitivity analysis is that rank selection requires the evaluation of forecasting performance over alternative rank values which can provide only point estimates of the rank and could be cumbersome, especially in high dimensional cases.

In this paper we focus on forecasting using reduced rank models which could be either VARs or standard multivariate regressions. This paper has two major contributions. The first is that we propose a novel fully Bayesian approach which allows for rank estimation. The novelty of this approach is that we reparameterize the coefficient matrix using its singular value decomposition and work on the posterior inference for the reparameterized model with decomposed parameters. By imposing a mixture of

a point mass and proper density as a prior on singular values (see Mitchell and Beauchamp, 1988; Smith and Kohn, 1996, for examples in the variable selection context) we are able to estimate the rank, noticing that the rank of the coefficient matrix is equivalent to the number of its non-zero singular values. It is worth noting that in addition to rank selection, such Bayesian SSVS shrinks nonzero singular values as well so that our approach can exploit both rank reduction and shrinkage on the coefficient matrix simultaneously and automatically. Therefore, our approach can possibly add value to forecasting accuracy in comparison with existing methods such as dynamic factor models and FAVAR approach which emphasize rank reduction. Some numerical evidence for this claim can be found in our simulation and empirical studies. Hoff (2007) also considers Bayesian model averaging and dimension selection with the singular value decomposition, but he is interested in modelling the multivariate data matrix as a reduced-rank mean matrix plus i.i.d. Gaussian noise which is completely different from multivariate reduced-rank regressions that we examine. Koop (2013) uses the SSVS prior for variable selection on the coefficient matrix, but our approach differs from his work in that we conduct variable selection on the singular values of the coefficient matrix rather than on each VAR coefficient. An additional advantage of applying this novel SVD reparameterization is that in this case the problem of estimating the coefficient matrix can be broken up into a set of conditional problems of estimating decomposed parameters for which proper non-informative priors can be determined and by doing so we can avoid the imposition of priors that are too subjective and informative on the coefficient matrix (cf. Yang and Berger, 1994). Although the singular value decomposition can ease the difficulty of rank selection, it raises the challenging problem of simulating the left and right singular vectors that are orthonormal and the posteriors of which are not recognizable as known distributions.

The second major contribution of this paper is that we devise a random walk Metropolis-Hastings algorithm for sampling from such irregular posteriors subject to nonlinear constraints. The principal feature of this algorithm is that the proposal for the next sample of the matrix of either left or right singular vectors is made by slightly permuting its current value in a random direction characterized by the product of a set of randomly chosen Givens rotation matrices. An advantage of this approach is computational efficiency since the Givens matrices are sparse, however, it should be noted that any random walk proposal in high dimensions will lead to a highly correlated Markov chain and low Monte Carlo efficiency. In spite of this, we demonstrate that our approach can handle the so-called 'medium' and 'medium-large' scale VARs with 20 to 40 variables (e.g. Bańbura et al., 2010; Koop, 2013), while for large scale VARs with over 50 predictors we might require an extremely large number of MCMC iterations for convergence which could be computationally prohibitive.

The rest of the paper is organized as follows. Section 2 describes the reduced rank regression model as well as the reparameterization using singular value decomposition and discusses the priors imposed on the reparameterized model. In Section 3, we provide the posterior odds ratio and the conditional posteriors required for the Bayesian SSVS and posterior inference via Gibbs sampler. Section 4 presents our simulation study and Section 5 covers the empirical application. Section 6 concludes. All the technical details are documented in the appendix.

2 Model and Preliminaries

Let $Y_t = (y_{1t}, y_{2t}, \ldots, y_{Nt})$ be an N-dimensional time series in row vector form where N can be large. Then a multivariate VAR(p) model can be expressed as

$$Y_t = Y_{t-1}A_1 + Y_{t-2}A_2 + \cdots + Y_{t-p}A_{t-p} + e_t \tag{2.1}$$

where e_t is an error vector distributed independently $N(0, \Sigma)$ across the time period t with Σ being a $N \times N$ positive definite matrix and $\{A_i; i = 1, 2, \ldots, p\}$ are $N \times N$ autoregressive coefficient matrices. It is assumed that means and trends in this model have been properly eliminated. Moreover, letting $X_t = (Y_{t-1}, Y_{t-2}, \ldots, Y_{t-p})$ and $\Theta = (A_1', A_2', \ldots, A_p')'$, one can write (2.1) alternatively as

$$Y_t = X_t\Theta + e_t$$

or in a more convenient matrix notation

$$\underset{T\times N}{Y} = \underset{T\times Np}{X}\ \underset{Np\times N}{\Theta} + \underset{T\times N}{E} \tag{2.2}$$

where $Y = (Y_1', Y_2', \ldots, Y_T')'$, $X = (X_1', X_2', \ldots, X_T')'$ and $E = (e_1', e_2', \ldots, e_T')'$ such that vec(E) follows the multivariate normal distribution $N(0, \Sigma \otimes I_T)$. Throughout this paper, we assume that rank $(\Theta) = r \ll N \leq Np$ and r is unknown. It is worth noting that equation (2.2) that we shall frequently work with should not necessarily be a VAR(p) model but it can also be a more general multivariate regression model.

In the classical reduced rank regression literature, the key to conduct statistical inference is to reparameterize the rank-deficient coefficient matrix as a product of two low-rank matrices, i.e.,

$$\underset{T\times N}{Y} = \underset{T\times Np}{X}\ \underset{Np\times N}{\Theta} + \underset{T\times N}{E} = \underset{T\times Np}{X}\ \underset{Np\times r}{\Phi} \cdot \underset{r\times N}{\Psi} + \underset{T\times N}{E}$$

(e.g., Geweke, 1996). However, in this paper we do not follow this standard procedure but make use of a novel parametrization due to singular value decomposition $\Theta = U\Lambda V'$ which gives

$$\underset{T\times N}{Y} = \underset{T\times Np}{X}\ \underset{Np\times N}{U} \cdot \underset{N\times N}{\Lambda} \cdot \underset{N\times N}{V'} + \underset{T\times N}{E} \tag{2.3}$$

in which U is a semi-orthogonal matrix (i.e. $U'U = I_N$ where I_N is a $N \times N$ identity matrix), Λ is a diagonal matrix containing nonnegative singular values and V is an orthogonal matrix such that $V'V = VV' = I_N$. It should be noted that Λ contains only r positive diagonal elements under the low-rank assumption that we impose. One can easily see that our parametrization in (2.3) explicitly allows for rank selection since the rank of Θ is equivalent to the number of nonzero diagonal entities of Λ. Therefore, we are able to implement stochastic search variable selection (e.g., George and McCulloch, 1993) on the diagonal of Λ via Gibbs sampling to obtain the posterior distribution on rank r. Specifically, denote that $\Lambda = \text{diag}\{\lambda_1, \ldots, \lambda_N\}$ where $\lambda_j \geq 0$ for each j and in

order to facilitate variable selection we introduce the latent vector $\gamma = (\gamma_1, \ldots, \gamma_N)$ such that $\lambda_j > 0$ iff $\gamma_j = 1$ and $\lambda_j = 0$ iff $\gamma_j = 0$ following Mitchell and Beauchamp (1988); Smith and Kohn (1996); Clyde et al. (1996); George and McCulloch (1997); Brown et al. (2002); Panagiotelis and Smith (2008) and others. Furthermore, we specify the prior for each λ_i conditioning on γ_i as

$$p(\lambda_j \mid \gamma_j = 1) = \sqrt{\frac{2\tau^2}{\pi}} \exp\left(-\frac{\tau^2 \lambda_j^2}{2}\right) \cdot 1\{\lambda_j > 0\}, \tag{2.4}$$

$$p(\lambda_j \mid \gamma_j = 0) = 1\{\lambda_j = 0\}, \tag{2.5}$$

so that given $\gamma_j = 1$ the conditional prior distribution of λ_j is half-normal with hyper-parameter τ^2 for each j. Basically, the half-normal prior works in approximately the same way as the normal prior but restricts λ_j to be positive real numbers.

Notice that parameters U, Λ and V in (2.3) are unidentified. To see this, one can change the ordering of the singular values in Λ and the ordering of the corresponding columns of U and V without altering the product $U\Lambda V'$. Moreover, given that Θ is rank deficient, Λ contains many zeros in the diagonal so that arbitrarily changing the ordering of the columns of either U or V corresponding to these zero singular values would leave $U\Lambda V'$ unchanged. Furthermore, one can also rotate the subspace spanned by those columns corresponding to zero singular values of either U or V without changing the value of $U\Lambda V'$. In addition, one cannot identify the column signs of U and V either. These imply that the posterior inference on either U, Λ or V would be senseless, but posterior inference on the coefficient Θ and its rank r is still valid because they are invariant to rotations, permutations and changes of sign in the matrices in the singular value decomposition and are thus identified. A great advantage of working with the partially identified model (2.3) is that it allows for forecasting and rank selection without requiring additional identifying restrictions and also gives us a large amount of computational flexibility.

An empirical justification for model (2.3) is that it is an alternative for factor models since the product XU can be regarded as certain "common factors" since the elements of this product are also linear combinations of the predictors and the remaining product $\Lambda V'$ can represent the associated "factor loadings". It is in this sense that model (2.3) is directly comparable to dynamic factor models. If the model (2.3) were augmented with an additional sparse coefficient matrix, then it would be comparable to the factor augmented VARs. We leave this extension to future research but still compare the forecasting performance of our model with both dynamic factor models and factor augmented VARs in the simulation and empirical studies later on, noting that the former is a more appropriate benchmark.

2.1 Priors for Rank Selection and Smoothing

First of all, we make use of the notation $q_\gamma = \sum_j \gamma_j$ to represent the model complexity which in our case means the selected rank for the coefficient Θ. As revealed by Kohn et al. (2001), Fernandez et al. (2001), Wolfe et al. (2004) and Scott and Berger (2010) in the context of linear regression, the prior on γ can be extremely informative when the number of available candidates for selection is large. A typical

example is the flat prior $p(\gamma) = 2^{-N}$ which imposes a strong belief on the latent vector γ with q_γ being approximately $N/2$ so that this prior heavily discriminates against low-rank models which are in fact better predictive models under the assumption of dynamic factor models in macroeconomic forecasting (e.g., Forni et al., 2000; Stock and Watson, 2011). Therefore, we address this issue by proposing the following prior for γ

$$p(\gamma) = \begin{cases} \frac{1}{q^{\max}+1}\binom{N}{q_\gamma}^{-1}, & 0 \le q_\gamma \le q^{\max} \\ 0 & q_\gamma > q^{\max}, \end{cases}$$

where $q^{\max}$ is the highest rank we allow for. Notice that this prior is a slightly modified version of the one implemented by George and McCulloch (1993), Cripps et al. (2005) and Panagiotelis and Smith (2008) and it actually implies a uniform prior on q_γ, i.e.,

$$p(q_\gamma) = \begin{cases} \binom{N}{q_\gamma} \cdot p(\gamma) = \frac{1}{q^{\max}+1}, & 0 \le q_\gamma \le q^{\max} \\ 0 & q_\gamma > q^{\max}. \end{cases}$$

Because the coefficient matrix Θ is assumed to have a small rank, it is unnecessary to put prior weight on the entire model space and hence we restrict the selected rank not to exceed $q^{\max}$. Under this prior,

$$\begin{aligned} p(\gamma_j = 1 \mid \gamma_{/j}) &= \frac{q_{\gamma_{/j}} + 1}{N+1}, \qquad 0 \le q_{\gamma_{/j}} \le q^{\max} \\ p(\gamma_j = 0 \mid \gamma_{/j}) &= \frac{N - q_{\gamma_{/j}}}{N+1}, \qquad 0 \le q_{\gamma_{/j}} \le q^{\max} \end{aligned} \tag{2.6}$$

and

$$\begin{aligned} p(\gamma_j = 1 \mid \gamma_{/j}) &= 0, \qquad q_{\gamma_{/j}} > q^{\max} \\ p(\gamma_j = 0 \mid \gamma_{/j}) &= 1, \qquad q_{\gamma_{/j}} > q^{\max} \end{aligned} \tag{2.7}$$

where $\gamma_{/j} = \big(\gamma_1, \ldots, \gamma_{j-1}, 0, \gamma_{j+1}, \ldots, \gamma_N\big)$ and $q_{\gamma_{/j}} = \sum_{i \neq j} \gamma_i$. Note that these conditional priors are derived to facilitate the calculation of posterior odds ratio for the ultimate Bayesian rank selection.

The hyper-parameter τ^2 in (2.4) is important for determining the degree of Bayesian shrinkage on nonzero singular values. In this paper, we employ the prior $\log(\tau^2) \sim N(a, b)$ with the hyper-priors $a \sim N(0, 100)$ and $b \sim IG(101, 10100)$ following Panagiotelis and Smith (2008) and the simulation and empirical studies later on will show that adding this level of hierarchy to the prior for τ^2 does improve the quality of the fits.

2.2 Priors on U and V

Due to the parametrization based on singular value decomposition adopted in (2.3), we require quadratic constraints on U and V such that $U'U = I_N$ and $V'V = VV' = I_N$. As a consequence, imposing some common priors (e.g., the use of normal priors advocated for problems with

linear constraints) on parameters U and V subject to the two constraints would lead to posteriors with nonlinear constraints, but sampling from such irregular posteriors is rather challenging. In fact, these constraints formulate two specific topological spaces in matrix algebra, one of which denoted by $\mathcal{V}_{N,Np} = \{U(Np \times N) \mid U'U = I_N\}$ is the so-called Stiefel manifold containing all the $Np \times N$ semi-orthogonal matrices and the other of which denoted by $O(N) = \{V(N \times N) \mid V'V = VV' = I_N\}$ is the orthogonal group of $N \times N$ orthogonal[1] matrices (e.g., Muirhead, 2005, p.67). In this paper, we will make use of uniform priors for parameters U and V over the spaces specified above which are first introduced in Bayesian analysis by Strachan and Inder (2004) who notice that uniform distributions are proper over these compact spaces. Koop et al. (2009) also implement such priors to simulate the cointegration space in Bayesian analysis of cointegrated models. A brief demonstration on the construction of uniform priors over these spaces are presented as follows. Interested readers are referred to James (1954) and Muirhead (2005) for more conceptual and technical details.

The crucial point of deriving the uniform distribution on $\mathcal{V}_{N,Np}$ is obtaining its volume which requires integration over this space. James (1954) has proved the existence of an invariant measure on the Stiefel manifold under orthogonal transformations. Moreover, it is also unique up to a finite multiplicative constant. In order to define this measure, we introduce its differential form denoted by $(U'dU)$ which in our case can be written as

$$(U'dU) = \bigwedge_{i=1}^{N} \bigwedge_{j=i+1}^{Np} u_j' du_i$$

where $\left(U, U^{\perp}\right) = (u_1, \ldots, u_N \mid u_{N+1}, \ldots, u_{Np})$ forms a $Np \times Np$ orthogonal matrix such that the columns of $U^{\perp}$ are perpendicular to those of U and the symbol '$\bigwedge$' is referred to as the 'exterior product' (e.g., Muirhead, 2005, p.52). We are now able to give the following invariant measure on $\mathcal{V}_{N,Np}$

$$\mu(\mathcal{M}) = \int_{\mathcal{M}} (U'dU), \qquad \mathcal{M} \subseteq \mathcal{V}_{N,Np}$$

and it is an invariant measure in the sense that $\mu(Q\mathcal{M}) = \mu(\mathcal{M}T) = \mu(\mathcal{M})$ for any orthogonal matrices $Q \in O(Np)$ and $T \in O(N)$. It is worth noticing that each element $U \in \mathcal{V}_{N,Np}$ can be regarded as the coordinates of a point lying on the surface of dimension $N^2p - N(N+1)/2$ in the Euclidean space of dimension N^2p, because the constraint $U'U = I_N$ implies $N(N+1)/2$ functionally independent restrictions on U. Therefore, the measure μ defined on the Stiefel manifold $\mathcal{V}_{N,Np}$ is equivalent to the ordinary Lebesgue measure. Due to Theorem 2.1.15 of Muirhead (2005, p.70) the volume of the Stiefel manifold $\mathcal{V}_{N,Np}$ can be evaluated as

$$\mathrm{Vol}\left(\mathcal{V}_{N,Np}\right) = \int_{\mathcal{V}_{N,Np}} (U'dU) = \frac{2^N \pi^{N^2 p/2}}{\Gamma_N(Np/2)}$$

[1]In linear algebra, an orthogonal matrix is a square matrix with real elements the columns and rows of which are orthogonal unit vectors which are usually referred to as orthonormal vectors. Thus, the terminologies 'orthogonal' and 'orthonormal' are sometimes used interchangeably when applied to such a matrix. We make a note here to avoid any potential confusion between these two terminologies.

where $\Gamma_N(\cdot)$ is the multivariate Gamma function (e.g., Definition 2.1.10 of Muirhead, 2005, p.61). Thus, the density of the uniform distribution over $\mathcal{V}_{N,Np}$ can be expressed as

$$p(U) = \frac{1}{\text{Vol}\,(\mathcal{V}_{N,Np})} = \frac{\Gamma_N(Np/2)}{2^N \pi^{N^2 p/2}} \tag{2.8}$$

for $U \in \mathcal{V}_{N,Np}$. Since the orthogonal group $O(N)$ is actually a special case of the Stiefel manifold $\mathcal{V}_{N,Np}$ with $p = 1$, we can immediately obtain the density of the uniform distribution over $O(N)$ which is

$$p(V) = \frac{1}{\text{Vol}\,(O(N))} = \frac{\Gamma_N(N/2)}{2^N \pi^{N^2/2}} \tag{2.9}$$

for $V \in O(N)$ with the differential form

$$(V'dV) = \bigwedge_{i<j}^{N} v_j' dv_i$$

where v_i and v_j represent the ith and jth orthonormal columns of V respectively. So the invariant measure can be then defined by

$$\nu(\mathcal{S}) = \int_{\mathcal{S}} (V'dV), \qquad \mathcal{S} \subseteq O(N),$$

which is the well known Haar measure.

2.3 Bayesian Shrinkage on Θ via SVD Parameterization

Given priors on the decomposed parameters (U, Λ, V), an interesting question arises as to what Bayesian shrinkage these priors would imply on the regression coefficient Θ via SVD parameterization. For the sake of simplicity, we ignore the variable selection on the singular values for rank reduction and just investigate Bayesian shrinkage. Specifically, suppose that Θ has full column rank so that the singular values are all nonzero and shrinkage is imposed upon these singular values in a straightforward manner by assuming independent half normal priors. The resulting prior for the coefficient matrix Θ in this case is formulated in the proposition below.

Proposition 2.1. *Providing the rotation matrices U and V have uniform priors given by* (2.8) *and* (2.9)*, the singular values $\{\lambda_i; i = 1, 2, \ldots, N\}$ are ordered decreasingly and follow i.i.d. $N(0, 1/\tau^2)$ priors, then the SVD parameterization $\Theta = U\Lambda V'$ implies that the prior for the coefficient matrix Θ is*

$$\begin{aligned} p(\Theta)(d\Theta) &\propto \exp\left(-\frac{\tau^2}{2} tr\,\Lambda^2\right)(U'dU)(d\Lambda)(V'dV) \\ &= \exp\left(-\frac{\tau^2}{2} tr\,\Theta'\Theta\right)|\Theta'\Theta|^{-\frac{Np-N}{2}} \left(\prod_{i<j}^{N}(\lambda_i^2 - \lambda_j^2)\right)^{-1}(d\Theta) \end{aligned} \tag{2.10}$$

in which $(d\Theta)$ represents the differential form of Θ and $(d\Lambda) = \bigwedge_{i=1}^{N} d\lambda_i$ denotes the differential form of

Λ.

Notice that in Proposition 2.1 we set the singular values of Θ in the descending order for the ease of presentation whereas the theory and implementation of our Bayesian approach in fact do not require such ordering. The expression (2.10) suggests a prior that shrinks Θ towards the equality of its singular values. This phenomenon is also observed in the work of Yang and Berger (1994) which focuses on the development of the reference prior for high dimensional covariance matrices. They reparameterize the covariance matrix based on orthogonal decomposition and the resulting reference prior (e.g., Yang and Berger, 1994, Eq. (15)) involves the term $\left(\prod_{i<j}(d_i - d_j)\right)^{-1}$ where d_i is the ith largest eigenvalue of the covariance matrix under their notation thus that the reference prior puts much weight near the region of the equality of the eigenvalues of the covariance matrix. Their reparameterization slightly differs from the one we use in that in our case Θ is not necessarily positive definite so that orthogonal decomposition is not appropriate and we apply singular value decomposition instead. Yang and Berger (1994) suggest that such Bayesian shrinkage would produce a better eigenstructure for a large covariance matrix estimator from which the estimation of Θ in our case may also benefit. It is worth noting that the Minnesota prior of Doan et al. (1984) and Littleman (1986) shrinks the coefficient matrix towards a random walk representation which can be regarded as a special case of such Bayesian shrinkage since the coefficient matrix under this situation is actually an identity matrix.

3 Posterior Inference

Following the standard procedure of stochastic search variable selection (e.g., George and McCulloch, 1993; Koop et al., 2007; Panagiotelis and Smith, 2008), we extend the regression setup in (2.3) with the latent variable γ and implement the Gibbs sampler relying on the resulting hierarchical Bayes model which delivers an auxiliary Markov chain

$$U^{(0)}, \Lambda^{(0)}, V^{(0)}, \Sigma^{(0)}, \gamma^{(0)}, U^{(1)}, \Lambda^{(1)}, V^{(1)}, \Sigma^{(1)}, \gamma^{(1)}, \ldots, U^{(k)}, \Lambda^{(k)}, V^{(k)}, \Sigma^{(k)}, \gamma^{(k)}, \ldots \tag{3.1}$$

where the superscript k indicates the kth sweep of the MCMC simulation. The Gibbs sampling scheme guarantees that these values can be regarded as samples drawn from the joint posterior distribution once the Markov chain (3.1) converges. The generation of this Markov chain requires the derivation of conditional posteriors of interesting parameters which will be given in the subsequent parts of this section. For simplifying the notation, we let $\Pi = \{U, \Lambda, V, \Sigma, \gamma\}$ represent the set of parameters.

3.1 Posterior Odds Ratio and Conditional Posteriors

An ultimate goal of this paper is to develop a Bayesian technique that allows for rank selection on the coefficient matrix Θ. In practice, we underpin this problem by tracing out the selected rank q_γ at every single sweep of the MCMC simulation and then estimating the rank using the posterior mode. The selected rank at each sweep is calculated as the number of nonzero singular values λ_i's which is

determined by the value of γ simulated based on the following posterior odds ratio

$$\frac{p(\gamma_k = 0 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X)}{p(\gamma_k = 1 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X)} \tag{3.2}$$

for each integer $1 \leq k \leq N$. Notice that the posterior odds ratio (3.2) indicates a MCMC step that simulates γ_k with λ_k being analytically integrated out so that it avoids a reducible Markov chain.

Before deriving the posterior odds ratio (3.2), we find it more convenient to post-multiply V on both sides of the model (2.3) and consider its a 'canonical SUR' form as follows

$$\widetilde{y} = \text{vec}(\widetilde{Y}) = \widetilde{X}\lambda + \text{vec}(\widetilde{E}) = \begin{bmatrix} \widetilde{x}_1 & 0 & \cdots & 0 \\ 0 & \widetilde{x}_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \widetilde{x}_N \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_N \end{bmatrix} + \widetilde{e} \tag{3.3}$$

in which we let $\widetilde{Y} = YV = (\widetilde{y}_1, \widetilde{y}_2, \ldots, \widetilde{y}_N)$, $\widetilde{X} = XU = (\widetilde{x}_1, \widetilde{x}_2, \ldots, \widetilde{x}_N)$ and $\lambda = (\lambda_1, \lambda_2, \ldots, \lambda_N)'$. Moreover, we have $\widetilde{E} = EV \sim N(0, V'\Sigma V \otimes I_T)$. The following theorem provides the posterior odds ratio for each γ_k in this case.

Theorem 3.1 (Posterior Odds Ratio)**.** *Let $\gamma_{i \neq k}$ represent all the remaining components of γ excluding γ_k and denote that*

$$\widetilde{\Sigma} = V'\Sigma^{-1}V = \begin{bmatrix} \widetilde{\sigma}_{11} & \widetilde{\sigma}_{12} & \cdots & \widetilde{\sigma}_{1N} \\ \widetilde{\sigma}_{21} & \widetilde{\sigma}_{22} & \cdots & \widetilde{\sigma}_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \widetilde{\sigma}_{N1} & \widetilde{\sigma}_{N2} & \cdots & \widetilde{\sigma}_{NN} \end{bmatrix} \tag{3.4}$$

Given priors (2.4) *and* (2.5) *on each singular value λ_k we can obtain the posterior odds ratio for each γ_k based on* (3.3) *as below*

$$\begin{aligned} \frac{p(\gamma_k = 0 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X)}{p(\gamma_k = 1 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X)} = \frac{p(\gamma_k = 0 \mid \gamma_{i \neq k})}{p(\gamma_k = 1 \mid \gamma_{i \neq k})} \cdot \left(\frac{4\tau^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^{-1/2} \\ \times \exp\left(-\frac{1}{2} \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) / \Phi\left(\frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\sqrt{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}} \right) \end{aligned} \tag{3.5}$$

where $H_{/k} = \sum_{j=1}^{k-1} (\widetilde{y}_j - \widetilde{x}_j\lambda_j)\widetilde{\sigma}_{jk} + \sum_{j=k+1}^{N} (\widetilde{y}_j - \widetilde{x}_j\lambda_j)\widetilde{\sigma}_{kj}$.

It is worth mentioning that the use of the general expression

$$\frac{p(\gamma_k = 0 \mid \gamma_{i \neq k})}{p(\gamma_k = 1 \mid \gamma_{i \neq k})}$$

in equation (3.5) gives the flexibility of using different priors on the latent vector γ for the posterior

odds ratio. In our case, the uniform prior on q_γ leads to

$$\frac{p(\gamma_k = 0 \mid \gamma_{i\neq k})}{p(\gamma_k = 1 \mid \gamma_{i\neq k})} = \begin{cases} \frac{N - q_{\gamma_{/j}}}{q_{\gamma_{/j}} + 1}, & 0 \leq q_{\gamma_{/j}} \leq q^{\max} \\ \infty, & q_{\gamma_{/j}} > q^{\max} \end{cases}$$

due to (2.6) and (2.7). This implies that when implementing a Metropolis-Hastings algorithm for rank selection, we always have to accept $\gamma_k = 0$ if the selected rank at the current state reaches the maximum $q^{\max}$.

In addition to the posterior odds ratio for each γ_k given in Theorem 3.1, the conditional posterior for each singular value λ_k is summarized as below.

Theorem 3.2. *When $\gamma_k = 1$, the conditional posterior density of λ_k is*

$$p(\lambda_k \mid \{\Pi \backslash \lambda_k, \gamma_k\}, \gamma_k = 1, Y, X) \propto \exp\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}{2} \left(\lambda_k - \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^2 \right) \cdot 1\{\lambda_k > 0\}$$

so that in this situation it follows a truncated normal distribution. Given that $\gamma_k = 0$, the posterior $p(\lambda_k \mid \{\Pi \backslash \lambda_k, \gamma_k\}, \gamma_k = 0, Y, X)$ only has probability mass at the point $\lambda_k = 0$ so that λ_k is equal to zero with probability one under this situation.

Due to the use of the standard Jeffreys prior on Σ, i.e., $p(\Sigma) \propto |\Sigma|^{-(N+1)/2}$, we can easily show that

$$p(\Sigma \mid U, \Lambda, V, \gamma, Y, X) \propto |\Sigma|^{-\frac{T+N+1}{2}} \exp\left(-\frac{1}{2} tr\left(\left(Y - XU\Lambda V'\right)' \left(Y - XU\Lambda V'\right) \Sigma^{-1} \right) \right)$$

and then it follows that

$$\Sigma \mid U, \Lambda, V, \gamma, Y, X \sim IW\left(\left(Y - XU\Lambda V'\right)' \left(Y - XU\Lambda V'\right),\ T \right);$$

(e.g., Zellner, 1971, pp.225–227).

It should be noted that we derive the posteriors above by treating the hyper-parameter τ^2 as fixed for simplicity. In fact, the posterior inference on τ^2 is quite standard and easily accommodated so that we omit it here.

3.2 The Algorithm of Proposing Rotation Matrices U and V

A major difficulty in implementing the Bayesian approach developed in this paper is how one generates the rotation matrices U and V arising due to SVD parameterization. First notice that the conditional posteriors of U and V cannot be recognized as standard distributions since they are both proportional to the likelihood such that

$$\begin{aligned} p(U \mid \Lambda, V, \Sigma, \gamma, Y, X) &\propto p(V \mid U, \Lambda, \Sigma, \gamma, Y, X) \\ &\propto \pi(Y \mid U, \Lambda, V, \Sigma, \gamma, X) \end{aligned}$$

$$\propto \exp\left(-\frac{1}{2}\mathrm{tr}\left(\left(Y - XU\Lambda V'\right)'\left(Y - XU\Lambda V'\right)\Sigma^{-1}\right)\right),$$

where γ has been included implicitly and $\pi(Y \mid U, \Lambda, V, \Sigma, \gamma, X)$ denotes the likelihood function. Therefore, we require a feasible Metropolis-Hastings algorithm to simulate rotation matrices U and V within the Gibbs sampler.

Without any loss of generality, we only consider the development of the Metropolis-Hastings algorithm of sampling U and the situation of V would be the same with the obvious change of the notation. Given that the parameter U lies in the Stiefel manifold $\mathcal{V}_{N,Np} = \{U(Np \times N) \mid U'U = I_N\}$, its orthonormal column vectors actually span a N-dimensional plane that passes through the origin in the Euclidean space $\mathbb{R}^{Np}$. As this plane moves in any possible directions (i.e. the column vectors forming it point in any directions from the origin) in the coordinate system, U can vary over the entire Stiefel manifold $\mathcal{V}_{N,Np}$. Thus, a natural way of proposing the candidate for the next sample of U (denoted by U^*) is to slightly permute its value $U^{(k)}$ at the current sweep k in a random direction. In fact, the action of 'randomly permuting' $U^{(k)}$ can be characterized by postmultiplying it with a set of Givens rotation matrices which have the following general form (e.g., Golub and Van-Loan, 2012)

$$G_{\ell,m} = \begin{array}{c} \\ \ell^{th} \\ \\ m^{th} \\ \end{array} \begin{array}{c} \begin{array}{ccccc} \ell^{th} & & & m^{th} & \end{array} \\ \begin{bmatrix} I & 0 & 0 & 0 & 0 \\ 0 & \cos\theta & 0 & -\sin\theta & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & \sin\theta & 0 & \cos\theta & 0 \\ 0 & 0 & 0 & 0 & I \end{bmatrix} \end{array} \tag{3.6}$$

where $-\pi/2 < \theta < \pi/2$ denotes the angle of the rotation and $G_{\ell,m}$ is clearly orthogonal. For an arbitrary Givens matrix $G_{\ell,m}$, the product $U^{(k)}G_{\ell,m}$ implies that one rotates the ℓth and the mth columns of $U^{(k)}$ in the two-dimensional plane spanned by these two vectors by θ radians counterclockwise. For instance, suppose that we have

$$U^{(k)} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

the columns of which are clearly orthonormal and span the xy plane in the standard $\mathbb{R}^3$ coordinate system. Now we postmultiply $U^{(k)}$ by

$$\begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}$$

and it follows accordingly that

$$U^* = \begin{bmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \\ 0 & 0 \end{bmatrix}.$$

As can be obviously observed, the column vectors of U^* differ from those of U^k by θ radians in the xy plane. Therefore, by randomly selecting a set of Givens matrices we are able to make a proposal for U. Specifically, let $I_\gamma(k)$ be the index set corresponding to the elements of $\gamma^{(k)} = \left(\gamma_1^{(k)}, \ldots, \gamma_j^{(k)}, \ldots, \gamma_N^{(k)}\right)$ such that $\gamma_j^{(k)} = 1$ at the sweep k and denote by $I_{/\gamma}(k)$ the set of remaining indices. Moreover, suppose that $q^{(k)}$ stands for the rank of the coefficient Θ selected at the sweep k. The random walk Metropolis-Hastings sampling procedure for U can be formulated as follows.

1. Select $q^{\max} - q^{(k)}$ elements from the index set $I_{/\gamma}(k)$ without replacement and denote the set of these elements by $I_+(k)$.

2. Construct the set $I_{\gamma+}(k) = \{I_\gamma(k), I_+(k)\}$.

3. Let $\Delta^{(k)}$ be the set of all possible 2-combinations out of the set $I_{\gamma+}(k)$ and let $\delta^{(k)} = (\ell, m) \in \Delta^{(k)}$ be an arbitrary element in this set that implies a clockwise rotation in the plane spanned by the ℓth and the mth columns of $U^{(k)}$. The Givens matrix corresponding to the rotation $\delta^{(k)}$ (which is denoted by $G_{\ell,m}^{(k)}$) is given by (3.6) with θ replaced by $\theta^{(k)}$ that we simulate from $N\left(0, \upsilon^2\right)$.

4. Randomly permute $U^{(k)}$ by postmultiplying the permutation matrix $P^{(k)}$ which is represented by the product of a sequence of Givens matrices $G_{\ell,m}^{(k)}$, i.e.,

$$P^{(k)} = \prod_{(\ell,m)\in\Delta^{(k)}} G_{\ell,m}^{(k)}$$

   in which the order of $G_{\ell,m}^{(k)}$ that forms the product $P^{(k)}$ is randomly chosen.

5. Given that the proposed random matrix $U^* = U^{(k)}P^{(k)}$, accept this proposal with probability

$$\min\left(1, \pi(Y \mid U^*, \Lambda, V, \Sigma, \gamma^{(k)}, X)/\pi(Y \mid U^{(k)}, \Lambda, V, \Sigma, \gamma^{(k)}, X)\right)$$

   where $\pi(Y \mid U^*, \Lambda, V, \Sigma, \gamma^{(k)}, X)$ and $\pi(Y \mid U^{(k)}, \Lambda, V, \Sigma, \gamma^{(k)}, X)$ are the likelihood functions for U^* and $U^{(k)}$ respectively and here we use $\gamma^{(k)}$ to emphasize that the simulated rank does not change when a proposal for U is made.

6. Set $k = k + 1$ and go back to step 1.

In practice, this procedure implies $N \times q^{(k)}$ variables that one has to simulate at the kth sweep of the MCMC scheme. If $N = 20$ and $q^{(k)} \leq q^{\max} = 10$ for example, then we need to estimate up to 200 variables at each sweep which will certainly slow down the convergence of acceptance and rejection samplers like the Metropolis-Hastings algorithm we propose here. Thus, an appropriate initial value can largely increase the efficiency of this algorithm, especially when we are interested in Bayesian

estimation for high dimensional problems. Fortunately, in the frequentist world, there is a long history of considering maximum likelihood (ML henceforth) estimation of the multivariate linear regression with low-rank assumption (e.g., Anderson, 1951; Davies and Tso, 1982; Izenman, 1975; Robinson, 1973, 1974; Tso, 1981) and such reduced-rank ML estimator can be easily implemented with its closed-form expression being documented by Reinsel (2006, p. 4) in detail. Since the reduced-rank estimation conventionally assumes that the rank is known, it is also necessary to determine a decent starting value for the rank. A feasible approximation is the number of nonzero canonical correlations between data matrices Y and X of the model (2.2) which can be obtained by the standard Wilk's Lambda F test for the significance of canonical dimensions.

4 Simulation

In this simulation study, we mainly investigate two issues, one is whether a Bayesian approach based on singular value decomposition that averages over different ranks (henceforth BRA-SVD) can correctly select the rank of the coefficient matrix and the other is to compare the one-step ahead out-of-sample forecasting performance of our BRA-SVD method with existing benchmarks, the dynamic factor model (DFM henceforth), the factor augmented vector autoregressive approach (FAVAR henceforth) and ordinary least squares (OLS henceforth).

4.1 DFM and FAVAR

The DFM and FAVAR approaches are two of the most promising methods in the recent literature on macroeconomic forecasting. Dynamic factor models generally assume that the information contained in a large set of stationary macroeconomic variables can be summarized with a small number of unobserved factors. In this paper, we use the dynamic factor model of Stock and Watson (2012) which is expressed as below

$$Y_{it} = \delta_i' F_{t-1} + u_{it}, \qquad i = 1, \ldots, N \qquad t = 1, \ldots, T \tag{4.1}$$

$$X_t = \Lambda F_t + \epsilon_t. \tag{4.2}$$

where X_t denotes the informational time series at time t (in our simulation we have simply $X_t = Y_t$), Y_t is a $n \times 1$ vector of target variables to forecast and F_t is the $K \times 1$ vector of latent factors where K is small relative to n. Notice that the unobserved factors F_t in (4.2) are often estimated by the principal components of the predictors. Due to (4.1) we formulate the one-step ahead forecast utilizing the first lag of the first K principal components as regressors with the coefficient δ estimated using OLS. In the simulation study, we consider the one-step ahead DFM forecast with 3 factors which is denoted by DFM3.

The FAVAR framework of Bernanke et al. (2005) also takes advantage of the factor structure

$$X_t = \Lambda F_t + \epsilon_t$$

but augments the unobserved factors into a standard VAR model

$$\begin{bmatrix} F_t \\ Y_{it} \end{bmatrix} = \Phi_i(L) \begin{bmatrix} F_{t-1} \\ Y_{it-1} \end{bmatrix} + v_{it}, \qquad i = 1, 2, \ldots, N \qquad t = 1, 2, \ldots, T, \tag{4.3}$$

in which $\Phi_i(L)$ is a lag polynomial operator of appropriate finite order. Similar to DFM forecasting, we use the first K principal components of X_t as the estimate of the latent factors at time period t to derive one-step ahead forecast. In the following numerical study, we utilize a simple version of the FAVAR approach by assuming $\Phi_i(L) = \Phi_i$ where Φ_i is a constant matrix in (4.3) and the resulting model then coincides with a dynamic factor model given by (4.1) plus an AR(1) term. In order to stress this point, we let DFM3-AR1 to represent the simplified FAVAR approach that we use with 3 factors.

4.2 Monte Carlo Designs and Results

We use two Monte Carlo designs to evaluate the forecasting performance of our Bayesian approach. Specifically, the first Monte Carlo design is based on a standard VAR(1) model presented as below

$$\underset{1\times N}{Y_t} = \underset{1\times N}{Y_{t-1}} \cdot \underset{N\times N}{\Theta} + \underset{1\times N}{E_t} = Y_{t-1} \cdot \underset{N\times N}{U} \cdot \underset{N\times N}{\Lambda} \cdot \underset{N\times N}{V'} + E_t$$

where the error term $E_t \sim N(0, \Sigma)$. The coefficient matrix Θ is assumed to be rank-deficient with $rank(\Theta) = 3 \ll N = 20$. The number of variables is set to $N = 20$ to resemble what Bańbura et al. (2010) call a 'medium' VAR. As found by Bańbura et al. (2010) and Koop (2013), adding more variables to such a 'medium' VAR in practice only leads to minor gains in forecasting accuracy generally and sometimes even cause deterioration in forecasting performance.

As demonstrated in Table 1, in this design we consider numerical experiments in different combinations of autocorrelation and noise when the number of observations are $T = 51$, 101 and 201 respectively. In cases where autocorrelation is high, we mean that the first three nonzero eigenvalues of the coefficient matrix Θ are large and set $\Lambda^H = \Lambda = \text{diag}\{0.9, 0.75, 0.5, 0, \ldots, 0\}$ in the simulation. Alternatively, in situations of a low autocorrelation we set $\Lambda^L = 0.5\Lambda^H$. Moreover, when we refer to a large noise, we set the covariance matrix of the error term $\Sigma^L = \Sigma \sim W_N(I_N/N,\ N)$. Otherwise, in cases where the noise is small we assume that $\Sigma^S = \Sigma \sim W_N(0.01 \times I_N/N,\ N)$. Overall, we have 12 scenarios and in each scenario we replicate 100 datasets for each of which T observations of N series are simulated underlying this data generating process. For dataset i, we compute the posterior mode of the selected rank of the coefficient Θ denoted by $\widehat{r}_i$ and the posterior estimate $\widehat{\Theta}_i$ using BRA-SVD method based on the first $T-1$ observations. Furthermore, we formulate the one-step ahead forecast $\widehat{Y}_T^i = Y_{T-1}^i \widehat{\Theta}_i$ where Y_{T-1}^i represents the $(T-1)$th observation of dataset i. Thus, the average selected rank for each scenario due to BRA-SVD can be interpreted as

$$\widehat{r}_{\text{BRA-SVD}} = \frac{1}{100} \cdot \sum_{i=1}^{100} \widehat{r}_i$$

Table 1: *Monte Carlo designs.*

T	Autocorrelation (Λ)	Noise (Σ)
Monte Carlo Design 1*: VAR(1)*		
51	High	Large
101	High	Large
201	High	Large
51	High	Small
101	High	Small
201	High	Small
51	Low	Large
101	Low	Large
201	Low	Large
51	Low	Small
101	Low	Small
201	Low	Small
Monte Carlo Design 2*: Regression*		
51	n.a.	Large
101	n.a.	Large
201	n.a.	Large
51	n.a.	Small
101	n.a.	Small
201	n.a.	Small

and the root mean squared error (henceforth RMSE) for the one-step ahead forecast compared with the Tth observation Y_T^i based on BRA-SVD is

$$\text{RMSE}_{\text{BRA-SVD}} = \sqrt{\frac{1}{100} \cdot \sum_{i=1}^{100} \left\| Y_T^i - \widehat{Y}_T^i \right\|^2}$$

by noting that $\|\cdot\|$ denotes the Euclidean norm of a vector. In order to evaluate whether BRA-SVD method can correctly select the rank of the coefficient matrix and compare its forecasting performance with existing benchmarks, we report the average selected rank $\widehat{r}_{\text{BRA-SVD}}$ and the relative RMSEs of the OLS, DFM3, DFM3-AR1 and BRA-SVD against the oracle estimator $\text{E}[Y_T^i \mid Y_{T-1}^i] = Y_{T-1}^i\Theta$ which should produce the best forecast but is infeasible in practice.

We consider the following regression model in the second Monte Carlo design

$$\underset{T\times N}{Y} = \underset{T\times N}{X} \cdot \underset{N\times N}{\Theta} + \underset{T\times N}{E} = X \cdot \underset{N\times N}{U} \cdot \underset{N\times N}{\Lambda} \cdot \underset{N\times N}{V'} + E$$

where each row of E follows the multivariate normal distribution $N(0, \Sigma)$ and we still assume that $\text{rank}(\Theta) = 3 \ll N = 20$. It should be noted that we generate exactly the same large and small noise as those in VAR(1) model. Moreover, the coefficient matrix Θ is generated as the same as the ones in cases of the high autocorrelation in Monte Carlo design 1 with $\Lambda = \text{diag}\{0.9, 0.75, 0.5, 0, \ldots, 0\}$. The 6 different scenarios of this design have also been illustrated in Table 1. Elements of the data matrix X for the regressors are generated i.i.d. from the standard normal distribution. It should be noted

that the one-step ahead out-of-sample forecast of the regression model differs slightly from that of the VAR(1) model. We take the BRA-SVD forecasting as an example. For dataset i, the posterior estimate $\widehat{\Theta}_i$ is first computed based on the first $T-1$ observations of the dependent variable Y and regressors X, and then the one-step ahead forecast for Y_T^i is formulated with $\widehat{Y}_T^i = X_T^i \widehat{\Theta}_i$ where X_T^i is the Tth observation of the regressors X in dataset i. Since the second design corresponds to the classical reduced rank regression, we only report the average selected rank and the relative RMSEs of one-step ahead forecasts of the OLS, DFM3[2] and BRA-SVD against the oracle estimator $E[Y_T^i \mid X_T^i] = X_T^i \Theta$ based on 100 datasets randomly simulated from this model for each scenario. Finally, one should notice that we set $q^{\max} = 10$ when implementing the BRA-SVD approach so that our sampler would not be allowed to visit a rank higher than 10 given that the value of true rank in both designs is 3.

Table 2 summarizes the average and median selected ranks in different scenarios for the two Monte Carlo designs. Overall, the BRA-SVD can deliver posterior estimates which are more or less 3 when the numbers of available observations are 100 and 200. In cases with 50 observations, the BRA-SVD can either overestimate or underestimate the rank depending on the strength of noise. For the VAR(1) model, the values of average and median selected ranks using BRA-SVD approach are quite stable across different levels of autocorrelation in respect to the same level of noise. Moreover, Figures 1 and 2 illustrate the distribution of the selected rank using BRA-SVD regarding the 100 replicated data sets for every scenario of each Monte Carlo design in detail. As can be seen clearly from these figures, when the sample size is 50, the posterior mode of the rank is often not equal to the true value across 100 replications, especially for higher levels of noise. As sample size increases and noise decreases the posterior mode of rank is nearly always equal to the true rank across 100 replications. Table 3 shows that in terms of the RMSE measure the performance of BRA-SVD forecasting is closest to that of the oracle estimator for both the VAR(1) and regression models. This is actually what we expect since our Bayesian approach not only selects the correct rank but also shrinks the selected singular values so that it can gain additional forecasting accuracy compared to methods like DFM and FAVAR which only apply dimension reduction.

[2]In the context of static regression models, the terminology 'DFM3' is inappropriate. Here we just use it to refer to a linear model where the dependent variables are regressed on the first 3 classical factors (e.g., Anderson, 1984) of the regressors for the ease of presenting Table 3.

Table 2: *Average and median (in brackets) selected ranks.*

$T-1$	High/Large	High/Small	Low/Large	Low/Small
		VAR(1)		
50	3.63(4)	2.27(2)	3.65(4)	1.90(2)
100	3.47(3)	2.98(3)	2.74(3)	2.40(2)
200	3.22(3)	3.05(3)	2.89(3)	2.84(3)
		Regression		
50	2.34(2)	3.03(3)	n.a.	n.a.
100	3.02(3)	3.04(3)	n.a.	n.a.
200	3.12(3)	3.05(3)	n.a.	n.a.

Note: The term '—/—' refers to the combination of the levels of autocorrelation and noise which has been listed in Table 1. For example, 'High/Large' means the case of both a high autocorrelation and a large noise for the VAR(1) model. In the simulation study on the regression model, we do not take into account autocorrelation, and use the terms 'High/Large' and 'High/Small' just to represent the two cases of large and small noise respectively for simplicity..

Table 3: *Relative RMSEs of OLS, DFM3, DFM3-AR1, BRA-SVD, Oracle forecasts.*

T-1	AutoCorr./Noise	OLS	DFM3	DFM3-AR1	BRA-SVD	Oracle
			VAR(1)			
50	High/Large	1.326	1.060	1.074	1.024	1.000
100	High/Large	1.124	1.045	1.050	1.016	1.000
200	High/Large	1.044	1.027	1.029	1.004	1.000
50	High/Small	1.272	1.051	1.061	1.020	1.000
100	High/Small	1.151	1.058	1.067	1.041	1.000
200	High/Small	1.064	1.035	1.036	1.011	1.000
50	Low/Large	1.266	1.024	1.038	1.010	1.000
100	Low/Large	1.124	1.026	1.031	1.009	1.000
200	Low/Large	1.024	1.011	1.010	1.002	1.000
50	Low/Small	1.322	1.037	1.060	1.028	1.000
100	Low/Small	1.171	1.028	1.033	1.017	1.000
200	Low/Small	1.039	1.004	1.007	1.005	1.000
			Regression			
50	n.a./Large	1.283	1.059	n.a.	1.026	1.000
100	n.a./Large	1.104	1.053	n.a.	1.017	1.000
200	n.a./Large	1.053	1.043	n.a.	1.011	1.000
50	n.a./Small	1.356	2.645	n.a.	1.039	1.000
100	n.a./Small	1.131	2.891	n.a.	1.030	1.000
200	n.a./Small	1.078	2.702	n.a.	1.014	1.000

Note: Relative RMSE is computed using the RMSE of each method divided by that of the benchmark, the oracle estimator. All relative RMSE values reported in this paper are rounded up to three decimal places..

Figure 1: *Distributions of the selected ranks in different scenarios:* VAR(1).

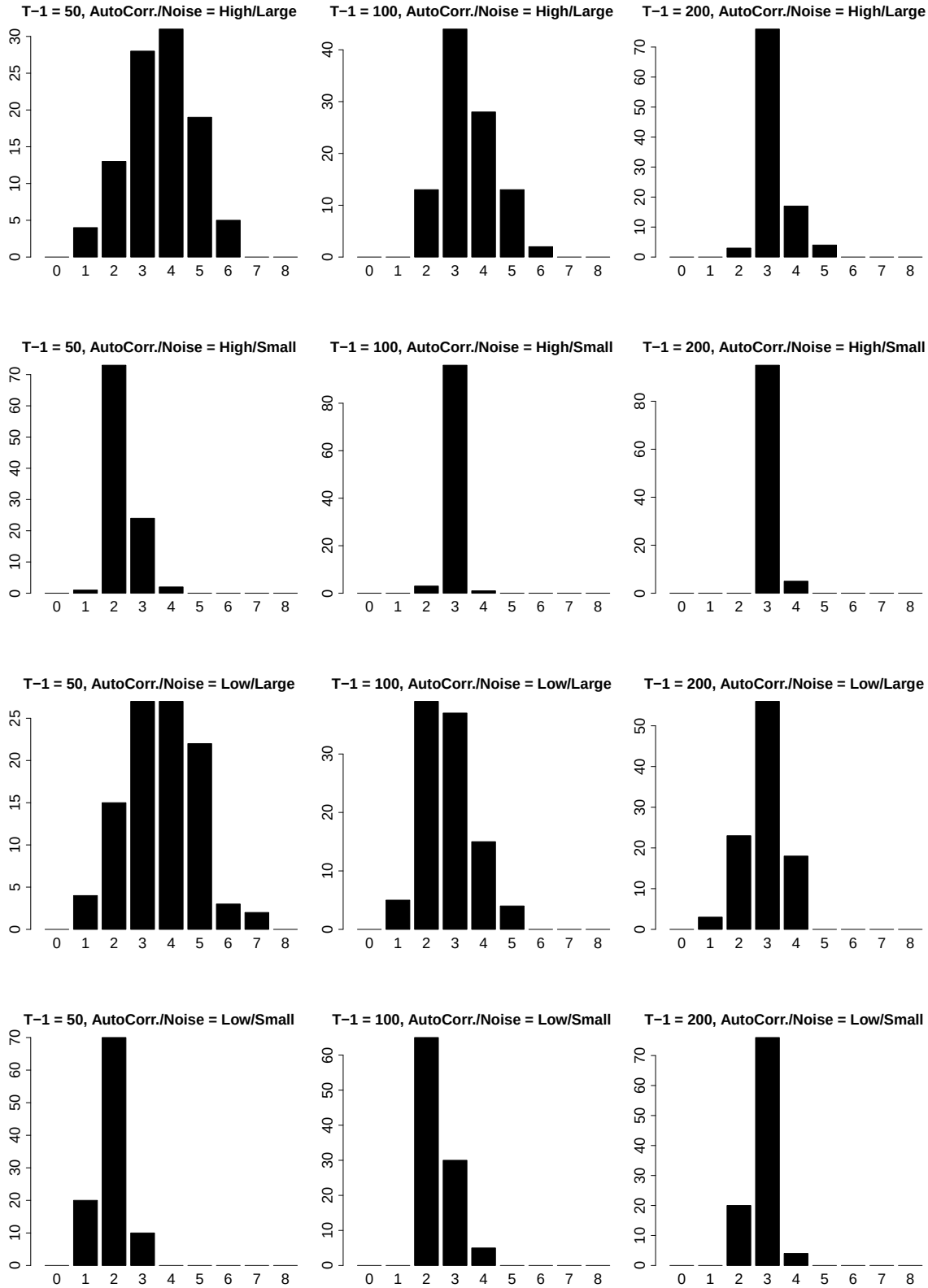
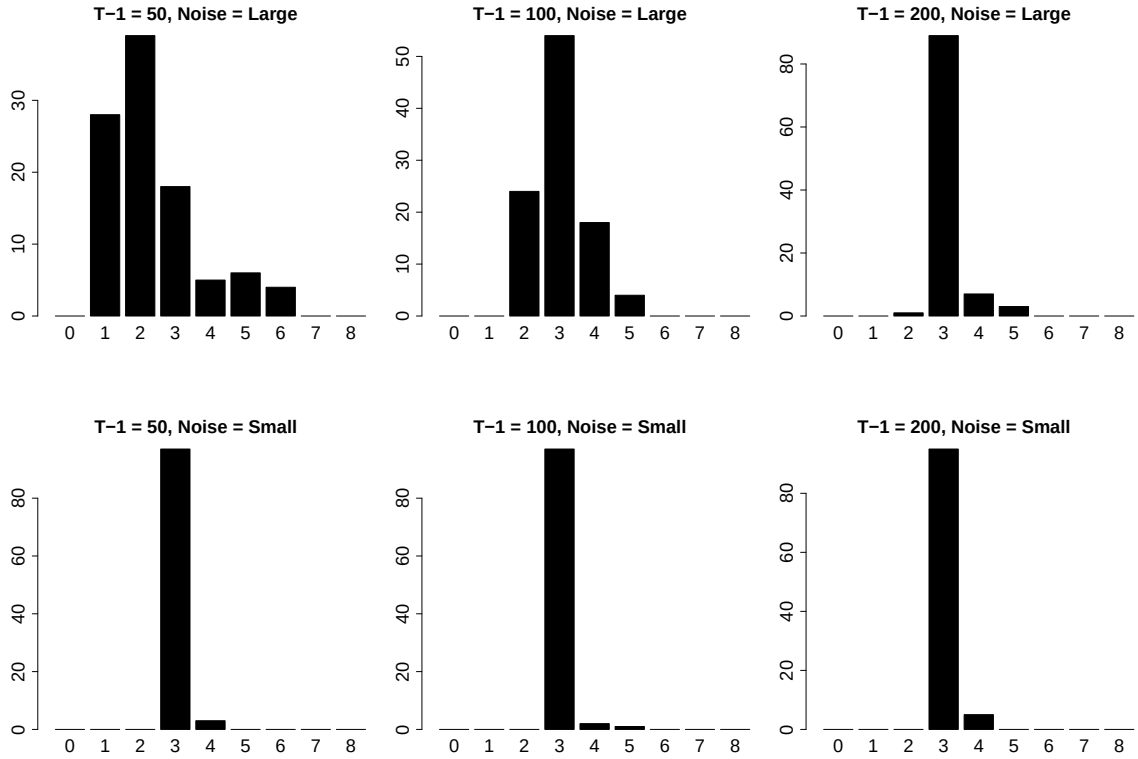


Figure 2: *Distributions of the selected ranks in different scenarios:* Regression.



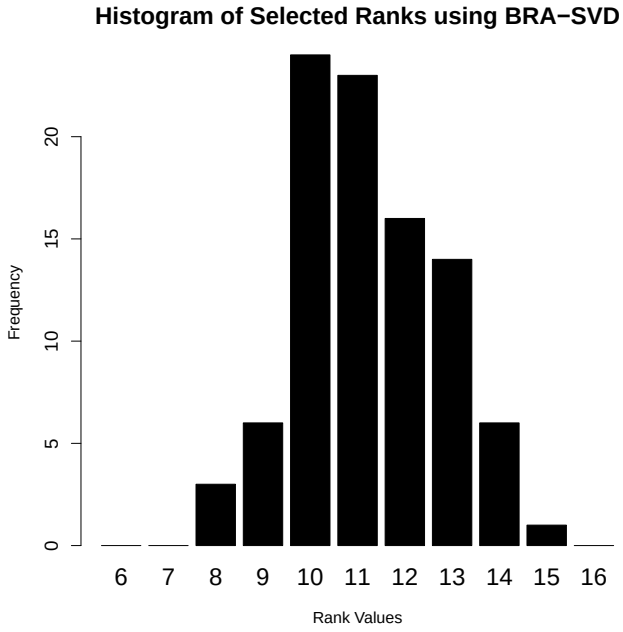
5 Empirical Application

The empirical data we use are originally from Stock and Watson (2012) and are quarterly observations from Q3 1960 until Q3 2008 (with earlier observations used for the lags of regressors as necessary) on 20 macroeconomic aggregate time series selected for the 'medium' VAR model studied by Koop (2013). All the variables are differenced or transformed to ensure stationarity, the details of which can be found in an earlier manuscript version of Stock and Watson's paper (Stock and Watson, 2009). We consider h-step-ahead forecasts for $h = 1, 2, 3$ and 4. First, one-step-ahead forecasts in Q3 1985 are produced using the 100 observations of all variables from Q3 1960 to Q2 1985 as the dependent variables, then the sample is rolled forwards so that an one-step-ahead forecast is produced for Q4 1985 using observations from Q4 1960 to Q3 1985 as a training sample and so on. Other multiple steps ahead forecasts are generated in a similar way. Overall, $94 - h$ rolling pseudo out-of-sample forecasts are obtained for each h as a result. The relative RMSE with the DFM5 being the benchmark is used for evaluating the performance of our BRA-SVD method in comparison with that of the OLS, DFM5 and DFM5-AR1. Notice here that DFM5 and DFM5-AR1 refer to the DFM and simple FAVAR forecasts with 5 factors respectively. All approaches make use of only one lag of the dependent variables ($p = 1$). It should be noted that BRA-SVD forecasts are Bayesian forecasts $E\left[Y_{t+h}|\mathcal{I}_t\right]$ where $Y_{t+1}, ..., Y_{t+h-1}$ as well as the unknown parameters are integrated out since they are simulated in an MCMC scheme. When applying the BRA-SVD method, we simulate 150,000 sweeps of the Gibbs sampling scheme to obtain the approximate samples from the posterior distribution and discard the first 100,000 sweeps as burn-in for each rolling window. Moreover, we set $q^{\max} = 19$ the maximal possible rank since the

Table 4: *h-step-ahead rolling forecasting performance.*

		OLS	DFM5	DFM5-AR1	BRA-SVD
$h = 1$	avg. Relative RMSE	1.026	1.000	0.888	0.917
	Best Performance	0(0%)	2(10%)	13(65%)	5(25%)
	Estimated Rank	n.a.	n.a.	n.a.	11.23
$h = 2$	avg. Relative RMSE	1.060	1.000	0.917	0.919
	Best Performance	0(0%)	3(15%)	9(45%)	8(40%)
	Estimated Rank	n.a.	n.a.	n.a.	11.21
$h = 3$	avg. Relative RMSE	0.986	1.000	0.927	0.947
	Best Performance	3(15%)	1(5%)	10(50%)	6(30%)
	Estimated Rank	n.a.	n.a.	n.a.	11.20
$h = 4$	avg. Relative RMSE	0.986	1.000	0.938	0.959
	Best Performance	3(15%)	2(10%)	11(55%)	4(20%)
	Estimated Rank	n.a.	n.a.	n.a.	11.20

Figure 3: *Distribution of the selected ranks based on BRA-SVD method.*



Note: Here we only consider the distribution of the selected ranks for the case of one-step-ahead forecast ($h = 1$) since the rank distributions in other cases are almost the same.

data are centered.

Table 4 summarises the average relative RMSEs of the h-step-ahead ($h = 1, 2, 3$ and 4) forecasts based on the OLS, DFM5, DFM5-AR1 and BRA-SVD methods, the mean selected rank obtained from BRA-SVD and the number of variables for which each forecasting technique performs the best. The details of the relative RMSEs of individual forecasts generated by these predictive techniques can be found in Tables 5–8. The average relative RMSE of the h-step-ahead forecasting for each method is simply the average of the relative RMSEs of the individual h-step-ahead forecasts derived by that method. For $h = 1, 2, 3$ and 4, BRA-SVD always has the second lowest average relative RMSE with its overall

Table 5: *Relative RMSEs for OLS, DFM5, DFM5-AR1, BRA-SVD h-step-ahead forecasts,* $h = 1$.

Variables	Description	OLS	DFM5	DFM5-AR1	BRA-SVD
GDP251	Real GDP, quantity index	1.169	1.000	1.069	0.989
GDP252	Real Personal Cons. Exp., Index	1.195	1.000	1.009	0.984
IPS10	Industrial production index: total	1.236	1.000	0.922	1.107
UTL11	Capacity utilization: manufacturing	0.256	1.000	0.233	0.233
CES002	Employees, nonfarm: total private	0.982	1.000	0.709	0.882
LHUR	Unemp. rate: All workers, 16 and over	1.213	1.000	0.967	1.048
HSFR	Housing starts: Total	0.325	1.000	0.301	0.320
GDP273A	Personal Cons Exp., price index	0.992	1.000	0.929	0.962
CPIAUCSL	CPI all items	1.008	1.000	0.910	0.908
PSCCOMR	Real spot market price index	1.224	1.000	0.980	0.996
CES275R	Real avg hrly earnings: non-farm	0.966	1.000	0.819	0.887
FYFF	Interest rate: federal funds	1.051	1.000	1.000	0.881
FYGT10	US treasury const. mat., 10-yr	1.072	1.000	0.955	1.041
FM1	Money stock: M1	1.255	1.000	1.017	1.057
FM2	Money stock: M2	1.017	1.000	1.028	0.986
FMRRA	Depository inst reserves: total	1.193	1.000	0.911	1.002
EXRUS	US effective exchange rate	1.082	1.000	0.987	0.998
FSPIN	S&P stock price index: industrials	1.135	1.000	0.999	1.027
FMRNBA	Depository inst reserves: nonborrowed	1.092	1.000	1.038	1.018
PWFSA	Producer price index: finished goods	1.053	1.000	0.968	1.024

Table 6: *Relative RMSEs for OLS, DFM5, DFM5-AR1, BRA-SVD h-step-ahead forecasts,* $h = 2$.

Variables	Description	OLS	DFM5	DFM5-AR1	BRA-SVD
GDP251	Real GDP, quantity index	1.188	1.000	0.933	0.950
GDP252	Real Personal Cons. Exp., Index	1.135	1.000	1.012	1.058
IPS10	Industrial production index: total	1.201	1.000	0.892	1.010
UTL11	Capacity utilization: manufacturing	0.515	1.000	0.423	0.431
CES002	Employees, nonfarm: total private	1.123	1.000	0.780	1.032
LHUR	Unemp. rate: All workers, 16 and over	1.153	1.000	0.909	1.019
HSFR	Housing starts: Total	0.439	1.000	0.420	0.425
GDP273A	Personal Cons Exp., price index	0.999	1.000	0.994	0.943
CPIAUCSL	CPI all items	0.999	1.000	0.986	0.944
PSCCOMR	Real spot market price index	1.415	1.000	0.912	0.906
CES275R	Real avg hrly earnings: non-farm	1.050	1.000	0.924	1.001
FYFF	Interest rate: federal funds	1.018	1.000	0.907	0.812
FYGT10	US treasury const. mat., 10-yr	1.010	1.000	0.966	0.936
FM1	Money stock: M1	0.987	1.000	0.961	0.961
FM2	Money stock: M2	1.130	1.000	1.040	1.035
FMRRA	Depository inst reserves: total	1.151	1.000	1.020	1.031
EXRUS	US effective exchange rate	0.978	1.000	0.956	0.936
FSPIN	S&P stock price index: industrials	1.068	1.000	1.000	1.003
FMRNBA	Depository inst reserves: nonborrowed	1.620	1.000	1.287	0.971
PWFSA	Producer price index: finished goods	1.018	1.000	1.008	0.969

Table 7: *Relative RMSEs for OLS, DFM5, DFM5-AR1, BRA-SVD h-step-ahead forecasts,* $h = 3$.

Variables	Description	OLS	DFM5	DFM5-AR1	BRA-SVD
GDP251	Real GDP, quantity index	1.113	1.000	0.953	1.032
GDP252	Real Personal Cons. Exp., Index	1.021	1.000	0.992	1.015
IPS10	Industrial production index: total	1.127	1.000	0.916	1.003
UTL11	Capacity utilization: manufacturing	0.744	1.000	0.603	0.623
CES002	Employees, nonfarm: total private	1.171	1.000	0.857	1.057
LHUR	Unemp. rate: All workers, 16 and over	1.122	1.000	0.949	1.004
HSFR	Housing starts: Total	0.569	1.000	0.557	0.574
GDP273A	Personal Cons Exp., price index	1.007	1.000	0.992	0.975
CPIAUCSL	CPI all items	0.988	1.000	0.993	0.982
PSCCOMR	Real spot market price index	0.970	1.000	0.978	0.988
CES275R	Real avg hrly earnings: non-farm	1.026	1.000	0.956	1.010
FYFF	Interest rate: federal funds	0.965	1.000	0.923	0.864
FYGT10	US treasury const. mat., 10-yr	0.989	1.000	1.014	0.998
FM1	Money stock: M1	0.947	1.000	0.941	0.936
FM2	Money stock: M2	1.043	1.000	0.966	0.983
FMRRA	Depository inst reserves: total	0.999	1.000	0.996	0.978
EXRUS	US effective exchange rate	0.932	1.000	0.964	0.950
FSPIN	S&P stock price index: industrials	0.987	1.000	0.983	0.962
FMRNBA	Depository inst reserves: nonborrowed	0.992	1.000	0.991	0.994
PWFSA	Producer price index: finished goods	1.003	1.000	1.015	1.003

Table 8: *Relative RMSEs for OLS, DFM5, DFM5-AR1, BRA-SVD h-step-ahead forecasts,* $h = 4$.

Variables	Description	OLS	DFM5	DFM5-AR1	BRA-SVD
GDP251	Real GDP, quantity index	1.019	1.000	0.918	0.963
GDP252	Real Personal Cons. Exp., Index	1.006	1.000	0.938	0.987
IPS10	Industrial production index: total	1.082	1.000	0.960	1.029
UTL11	Capacity utilization: manufacturing	0.847	1.000	0.689	0.726
CES002	Employees, nonfarm: total private	1.122	1.000	0.865	1.032
LHUR	Unemp. rate: All workers, 16 and over	1.114	1.000	0.990	1.015
HSFR	Housing starts: Total	0.687	1.000	0.667	0.692
GDP273A	Personal Cons Exp., price index	1.007	1.000	0.986	0.979
CPIAUCSL	CPI all items	0.979	1.000	0.985	0.980
PSCCOMR	Real spot market price index	0.975	1.000	0.973	0.973
CES275R	Real avg hrly earnings: non-farm	0.977	1.000	0.930	0.962
FYFF	Interest rate: federal funds	0.987	1.000	0.940	0.949
FYGT10	US treasury const. mat., 10-yr	0.946	1.000	0.951	0.955
FM1	Money stock: M1	1.003	1.000	0.984	0.980
FM2	Money stock: M2	0.989	1.000	0.978	0.977
FMRRA	Depository inst reserves: total	1.024	1.000	1.015	1.024
EXRUS	US effective exchange rate	0.950	1.000	0.994	0.969
FSPIN	S&P stock price index: industrials	1.003	1.000	0.990	0.988
FMRNBA	Depository inst reserves: nonborrowed	1.005	1.000	1.005	1.003
PWFSA	Producer price index: finished goods	1.002	1.000	0.997	0.998

performance slightly inferior to that of DFM5-AR1. In terms of individual forecasts, DFM5-AR1 has the lowest RMSEs amongst all models for the most individual series while BRA-SVD comes to the second place. It is worth noting that the BRA-SVD method provides the most accurate one-step-ahead forecasts for three variables: GDP251, CPIAUCSL and FYFF, which respectively represent the Real GDP, CPI and US Federal funds that are often the indicators of primary interest in macroeconomic forecasting with VAR models (e.g., Christiano et al., 1999; An and Schorfheide, 2007). In general, these results support the use of BRA-SVD as an alternative to factor models for forecasting. In fact, there is no consensus on which method possesses the optimal forecasting performance. As one may notice, when the forecast horizon increases to $h = 3$ and 4, even the OLS is competitive for the DFM5 and also it can outperform the other models in forecasting three specific variables. Therefore, the selection of forecasting methodologies really depends on the variable of interest and the forecast horizon. Finally, Figure 3 shows the distribution of the ranks selected by BRA-SVD for the coefficient matrix in the VAR(1) model investigated here. It has a clear mode around 10 with ranks outside the range between 8 and 15 being quite unlikely.

6 Conclusion

In this paper, we propose a novel fully Bayesian approach that can address the important issue of rank selection in multivariate regressions. We assess the performance of our Bayesian approach for the 'medium' VAR in both simulation and empirical studies, and the results show that this approach can correctly select the rank, and provides forecasting accuracy that is highly competitive in comparison with dynamic factor models and factor augmented VARs. Since the 'medium' VAR can generate forecasts that remain rather robust when more predictors are included, as noticed by Bańbura et al. (2010) and Koop (2013), our approach which can handle the 'medium' and 'medium-large' VARs with 20 to 40 predictors could be a competing candidate in forecasting toolbox for macroeconometricans.

Finally, a future research topic is worth mentioning here. In this paper, we adopt the SVD reparameterization for the coefficient matrix to achieve rank selection, because we are also interested in the canonical correlations between the independent variable and the predictors, and the left and right singular vectors actually reveal such correlations. However, if one only focuses on rank estimation, the reparameterization based on the QR decomposition may deserve further attention although the appropriate sampling scheme for such a model could be challenging. In fact, working with the reparameterized model due to different matrix decompositions might be a promising avenue to estimate invariants (e.g. determinant, rank, trace and so on) under these transformations, and this is quite a novel research direction compared to classical methodologies.

A Appendix

Proof of Proposition 2.1. According to James (1954, pp. 70-71), we can obtain the Jacobian of the SVD parameterization shown as below

$$(d\Theta) = \left(\prod_{i=1}^{N} \lambda_i\right)^{Np-N} \prod_{i<j}^{N} (\lambda_i^2 - \lambda_j^2)(U'dU)(d\Lambda)(V'dV) \tag{A.1}$$

in which we let $(d\Theta)$ be the differential form of Θ and denote that $(d\Lambda)$ is the differential form of Λ which is equivalent to $\bigwedge_{i=1}^{N} d\lambda_i$, the exterior product of the N diagonal elements of $d\Lambda$. The i.i.d. normal priors for λ_is imply that

$$p(\Lambda)(d\Lambda) \propto \prod_{i=1}^{N} \exp\left(-\frac{\tau^2\lambda_i^2}{2}\right)(d\Lambda) = \exp\left(-\frac{\tau^2}{2} tr\Lambda^2\right)(d\Lambda).$$

Thus, noting the uniform priors on U and V given in (2.8) and (2.9) and the Jacobian in (A.1) we have

$$\begin{aligned} p(\Theta)(d\Theta) &\propto \exp\left(-\frac{\tau^2}{2} tr\Lambda^2\right)(U'dU)(d\Lambda)(V'dV) \\ &= \exp\left(-\frac{\tau^2}{2} tr\Theta'\Theta\right)|\Theta'\Theta|^{-\frac{Np-N}{2}} \left(\prod_{i<j}^{N}(\lambda_i^2 - \lambda_j^2)\right)^{-1}(d\Theta). \end{aligned}$$

The proof is complete. □

Proof of Theorem 3.1. First of all, notice that the likelihood functions for the model (2.3) and its SUR form (3.3) are equivalent

$$\begin{aligned} p(Y \mid U, \Lambda, V, \Sigma, \gamma, X) &\propto |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2} tr\left(\left(Y - XU\Lambda V'\right)'\left(Y - XU\Lambda V'\right)\Sigma^{-1}\right)\right) \\ &= |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}(\widetilde{y} - \widetilde{X}\lambda)'(V'\Sigma^{-1}V \otimes I_T)(\widetilde{y} - \widetilde{X}\lambda)\right) \\ &= |\Sigma|^{-\frac{T}{2}} \exp\left(-\frac{1}{2}(\widetilde{y} - \widetilde{X}\lambda)'(\widetilde{\Sigma} \otimes I_T)(\widetilde{y} - \widetilde{X}\lambda)\right) \end{aligned}$$

by noting (3.4) so that the posterior odd ratio obtained in this theorem can apply directly to our original model.

The derivation of the posterior odds ratio involves two scenarios where $\gamma_k = 0$ and $\gamma_k = 1$ and they are treated separately in this proof. Throughout this proof, we make use of the notation $\{\Pi \backslash \lambda_k, \gamma_k\}$ to stand for all the interest parameters exclusive of $\{\lambda_k, \gamma_k\}$.

Initially, we examine the scenario where $\gamma_k = 1$. Note that by combining the likelihood and priors

$$p(\lambda_k, \gamma_k = 1 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X)$$

$$
\begin{aligned}
&\propto \exp\left(-\frac{1}{2}(\widetilde{y}-\widetilde{X}\lambda)'(\widetilde{\Sigma}\otimes I_T)(\widetilde{y}-\widetilde{X}\lambda)\right)\sqrt{\frac{2\tau^2}{\pi}}\exp\left(-\frac{\tau^2\lambda_k^2}{2}\right)\cdot 1\{\lambda_k>0\}\cdot p(\gamma_k=1\mid\gamma_{i\neq k})\\
&=\exp\left(\frac{1}{2}\begin{bmatrix}\widetilde{y}_1-\widetilde{x}_1\lambda_1\\ \widetilde{y}_2-\widetilde{x}_2\lambda_2\\ \vdots\\ \widetilde{y}_N-\widetilde{x}_N\lambda_N\end{bmatrix}'\begin{bmatrix}\widetilde{\sigma}_{11}I_T & \widetilde{\sigma}_{12}I_T & \cdots & \widetilde{\sigma}_{1N}I_T\\ \widetilde{\sigma}_{21}I_T & \widetilde{\sigma}_{22}I_T & \cdots & \widetilde{\sigma}_{2N}I_T\\ \vdots & \vdots & \ddots & \vdots\\ \widetilde{\sigma}_{N1}I_T & \widetilde{\sigma}_{N2}I_T & \cdots & \widetilde{\sigma}_{NN}I_T\end{bmatrix}\begin{bmatrix}\widetilde{y}_1-\widetilde{x}_1\lambda_1\\ \widetilde{y}_2-\widetilde{x}_2\lambda_2\\ \vdots\\ \widetilde{y}_N-\widetilde{x}_N\lambda_N\end{bmatrix}\right)\\
&\quad\times\sqrt{\frac{2\tau^2}{\pi}}\exp\left(-\frac{\tau^2\lambda_k^2}{2}\right)\cdot 1\{\lambda_k>0\}\cdot p(\gamma_k=1\mid\gamma_{i\neq k})
\end{aligned}
$$

where $\widetilde{y}_i-\widetilde{X}_i\lambda_i$ is a $T\times 1$ vector for any $i=1,2,\ldots,N$. Thus, we have

$$
\begin{aligned}
&p(\lambda_k,\gamma_k=1\mid\{\Pi\backslash\lambda_k,\gamma_k\},Y,X)\\
&\quad\propto\exp\left(-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\widetilde{y}_i-\widetilde{x}_i\lambda_i)'(\widetilde{y}_j-\widetilde{x}_j\lambda_j)\widetilde{\sigma}_{ij}\right)\sqrt{\frac{2\tau^2}{\pi}}\exp\left(-\frac{\tau^2\lambda_k^2}{2}\right)\cdot 1\{\lambda_k>0\}\cdot p(\gamma_k=1\mid\gamma_{i\neq k})\\
&\quad\propto\exp\left(-\frac{1}{2}\big(2\sum_{i<j}(\widetilde{y}_i-\widetilde{x}_i\lambda_i)'(\widetilde{y}_j-\widetilde{x}_j\lambda_j)\widetilde{\sigma}_{ij}+\sum_{i=1}^{N}(\widetilde{y}_i-\widetilde{x}_i\lambda_i)'(\widetilde{y}_i-\widetilde{x}_i\lambda_i)\widetilde{\sigma}_{ii}\big)\right)\\
&\qquad\times\sqrt{\frac{2\tau^2}{\pi}}\exp\left(-\frac{\tau^2\lambda_k^2}{2}\right)\cdot 1\{\lambda_k>0\}\cdot p(\gamma_k=1\mid\gamma_{i\neq k})\\
&\quad\propto\exp\left(-(\widetilde{y}_k-\widetilde{x}_k\lambda_k)'\big(\sum_{j=1}^{k-1}(\widetilde{y}_j-\widetilde{x}_j\lambda_j)\widetilde{\sigma}_{jk}+\sum_{j=k+1}^{N}(\widetilde{y}_j-\widetilde{x}_j\lambda_j)\widetilde{\sigma}_{kj}\big)-\frac{1}{2}(\widetilde{y}_k-\widetilde{x}_k\lambda_k)'(\widetilde{y}_k-\widetilde{x}_k\lambda_k)\widetilde{\sigma}_{kk}\right)\\
&\qquad\times\sqrt{\frac{2\tau^2}{\pi}}\exp\left(-\frac{\tau^2\lambda_k^2}{2}\right)\cdot 1\{\lambda_k>0\}\cdot p(\gamma_k=1\mid\gamma_{i\neq k})\\
&\quad\propto\exp\left(-(\widetilde{y}_k-\widetilde{x}_k\lambda_k)'H-\frac{1}{2}(\widetilde{y}_k-\widetilde{x}_k\lambda_k)'(\widetilde{y}_k-\widetilde{x}_k\lambda_k)\widetilde{\sigma}_{kk}-\frac{\tau^2\lambda_k^2}{2}\right)\\
&\qquad\times\sqrt{\frac{2\tau^2}{\pi}}\cdot 1\{\lambda_k>0\}\cdot p(\gamma_k=1\mid\gamma_{i\neq k})
\end{aligned}
$$

where $H_{/k}=\sum_{j=1}^{k-1}(\widetilde{y}_j-\widetilde{x}_j\lambda_j)\widetilde{\sigma}_{jk}+\sum_{j=k+1}^{N}(\widetilde{y}_j-\widetilde{x}_j\lambda_j)\widetilde{\sigma}_{kj}$. Moreover, it holds that

$$
\begin{aligned}
&p(\lambda_k,\gamma_k=1\mid\{\Pi\backslash\lambda_k,\gamma_k\},Y,X)\\
&\qquad\propto\exp\left(-\widetilde{y}_k'H_{/k}+\widetilde{x}_k'H_{/k}\lambda_k-\frac{\widetilde{\sigma}_{kk}}{2}\left(\widetilde{y}_k'\widetilde{y}_k-2\widetilde{y}_k'\widetilde{x}_k\lambda_k+\widetilde{x}_k'\widetilde{x}_k\lambda_k^2\right)-\frac{\tau^2\lambda_k^2}{2}\right)\\
&\qquad\qquad\times\sqrt{\frac{2\tau^2}{\pi}}1\{\lambda_k>0\}p(\gamma_k=1\mid\gamma_{i\neq k})\\
&\qquad\propto\exp\left(-\widetilde{y}_k'H_{/k}+\widetilde{x}_k'H_{/k}\lambda_k-\frac{\widetilde{\sigma}_{kk}}{2}\widetilde{y}_k'\widetilde{y}_k+\widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{x}_k\lambda_k-\frac{\widetilde{\sigma}_{kk}}{2}\widetilde{x}_k'\widetilde{x}_k\lambda_k^2-\frac{\tau^2\lambda_k^2}{2}\right)\\
&\qquad\qquad\times\sqrt{\frac{2\tau^2}{\pi}}1\{\lambda_k>0\}p(\gamma_k=1\mid\gamma_{i\neq k})\\
&\qquad\propto\exp\left(-\frac{1}{2}\left((\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k+\tau^2)\lambda_k^2-2\widetilde{x}_k'(H_{/k}+\widetilde{\sigma}_{kk}\widetilde{y}_k)\lambda_k+2\widetilde{y}_k'H_{/k}+\widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k\right)\right)
\end{aligned}
$$

$$
\begin{aligned}
&\times \sqrt{\frac{2\tau^2}{\pi}} 1\{\lambda_k > 0\} p(\gamma_k = 1 \mid \gamma_{i\neq k}) \\
\propto \exp&\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}{2} \left(\lambda_k^2 - 2 \cdot \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \lambda_k + \frac{2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) \right) \\
&\times \sqrt{\frac{2\tau^2}{\pi}} 1\{\lambda_k > 0\} p(\gamma_k = 1 \mid \gamma_{i\neq k}) \\
\propto \exp&\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}{2} \left(\left(\lambda_k - \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^2 + \frac{2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right.\right. \\
&\left.\left. - \left(\frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^2 \right) \right) \times \sqrt{\frac{2\tau^2}{\pi}} p(\gamma_k = 1 \mid \gamma_{i\neq k}) \cdot 1\{\lambda_k > 0\} \\
\propto \exp&\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}{2} \left(\lambda_k - \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^2 \right) \cdot 1\{\lambda_k > 0\} \cdot p(\gamma_k = 1 \mid \gamma_{i\neq k}) \\
&\times \sqrt{\frac{2\tau^2}{\pi}} \exp\left(-\frac{1}{2} \left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) \right). \qquad \text{(A.2)}
\end{aligned}
$$

By integrating out λ_k, we can obtain

$$
\begin{aligned}
&p(\gamma_k = 1 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X) \\
&\quad \propto p(\gamma_k = 1 \mid \gamma_{i\neq k}) \cdot \int_{\lambda_k > 0} \exp\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}{2} \left(\lambda_k - \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^2 \right) d\lambda_k \\
&\qquad \times \sqrt{\frac{2\tau^2}{\pi}} \exp\left(-\frac{1}{2} \left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) \right) \\
&\quad \propto p(\gamma_k = 1 \mid \gamma_{i\neq k}) \cdot \left(\frac{2\pi}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^{1/2} \left(1 - \Phi\left(-\frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\sqrt{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}} \right) \right) \\
&\qquad \times \sqrt{\frac{2\tau^2}{\pi}} \exp\left(-\frac{1}{2} \left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) \right) \\
&\quad \propto p(\gamma_k = 1 \mid \gamma_{i\neq k}) \cdot \left(\frac{2\pi}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^{1/2} \Phi\left(\frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\sqrt{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}} \right) \\
&\qquad \times \sqrt{\frac{2\tau^2}{\pi}} \exp\left(-\frac{1}{2} \left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) \right) \\
&\quad = p(\gamma_k = 1 \mid \gamma_{i\neq k}) \cdot \left(\frac{4\tau^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right)^{1/2} \Phi\left(\frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\sqrt{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}} \right) \\
&\qquad \times \exp\left(-\frac{1}{2} \left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2} \right) \right)
\end{aligned}
$$

When $\gamma_k = 0$, we can similarly get

$$
\begin{aligned}
&p(\lambda_k, \gamma_k = 0 \mid \{\Pi \backslash \lambda_k, \gamma_k\}, Y, X) \\
&\qquad \propto \exp\left(-\frac{1}{2} (\widetilde{y} - \widetilde{X}\lambda)' (\widetilde{\Sigma} \otimes I_T) (\widetilde{y} - \widetilde{X}\lambda) \right) \cdot 1\{\lambda_k = 0\} \cdot p(\gamma_k = 0 \mid \gamma_{i\neq k})
\end{aligned}
$$

$$\propto p(\gamma_k = 0 \mid \gamma_{i\neq k}) \cdot \exp\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}{2}\left(\lambda_k - \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}\right)^2\right) \cdot 1\{\lambda_k = 0\}$$
$$\times \exp\left(-\frac{1}{2}\left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}\right)\right). \tag{A.3}$$

As a consequence,

$$\begin{aligned}
&p(\gamma_k = 0 \mid \{\Pi\backslash\lambda_k, \gamma_k\}, Y, X) \\
&\quad\propto p(\gamma_k = 0 \mid \gamma_{i\neq k}) \cdot \int \exp\left(-\frac{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}{2}\left(\lambda_k - \frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}\right)^2\right) \cdot 1\{\lambda_k = 0\} d\lambda_k \\
&\qquad\times \exp\left(-\frac{1}{2}\left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}\right)\right) \\
&\quad\propto p(\gamma_k = 0 \mid \gamma_{i\neq k}) \cdot \exp\left(-\frac{1}{2}\frac{\left(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)\right)^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}\right) \\
&\qquad\times \exp\left(-\frac{1}{2}\big(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k - \frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k}\big)\right) \\
&\quad\propto p(\gamma_k = 0 \mid \gamma_{i\neq k}) \cdot \exp\left(-\frac{1}{2}\left(2\widetilde{y}_k' H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k'\widetilde{y}_k\right)\right).
\end{aligned}$$

The posterior odds ratio is then given by

$$\begin{aligned}
\frac{p(\gamma_k = 0 \mid \{\Pi\backslash\lambda_k, \gamma_k\}, Y, X)}{p(\gamma_k = 1 \mid \{\Pi\backslash\lambda_k, \gamma_k\}, Y, X)} &= \frac{p(\gamma_k = 0 \mid \gamma_{i\neq k})}{p(\gamma_k = 1 \mid \gamma_{i\neq k})} \cdot \left(\frac{4\tau^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}\right)^{-1/2} \\
&\quad\times \exp\left(-\frac{1}{2}\frac{(\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k))^2}{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}\right) / \Phi\left(\frac{\widetilde{x}_k'(H_{/k} + \widetilde{\sigma}_{kk}\widetilde{y}_k)}{\sqrt{\widetilde{\sigma}_{kk}\widetilde{x}_k'\widetilde{x}_k + \tau^2}}\right).
\end{aligned}$$

The proof is complete. □

Proof of Theorem 3.2. The results follow immediately from (A.2) and (A.3). □

References

An, S. and Schorfheide, F. (2007). Bayesian analysis of dsge models. *Econometric Reviews*, 26:113–172.

Anderson, T. W. (1951). Estimating linear restrictions on regression coefficients for multivariate normal distributions. *The Annals of Mathematical Statistics*, pages 327–351.

Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: Wiley.

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

Bernanke, B. S. and Blinder, A. S. (1992). The federal funds rate and the channels of monetary transmission. *The American Economic Review*, pages 901–921.

Bernanke, B. S., Boivin, J., and Eliasz, P. (2005). Measuring the effects of monetary policy: A factor-augmented vector autoregressive (favar) approach. *The Quarterly Journal of Economics*, pages 387–422.

Brown, P., Vannucci, M., and Fearn, T. (2002). Bayesian model averaging with selection of regressor. *Journal of the Royal Statistical Society Series B*, 64:519–536.

Carriero, A., Kapetanios, G., and Marcellino, M. (2011). Forecasting large datasets with Bayesian reduced rank multivariate models. *Journal of Applied Econometrics*, 26(5):735–761.

Christiano, L., Eichenbaum, M., and Evans, C. (1999). Monetary policy shocks: what have we learned and to what end? In Taylor, J. and Woodford, M., editors, *Handbook of Macroeconomics*, volume 1, pages 65–148, Amsterdam. Elsevier.

Clyde, M., DeSimone, H., and Parmigiani, G. (1996). Prediction via orthogonalized model mixing. *Journal of the American Statistical Association*, 91:1197–1208.

Cripps, E., Carter, C., and Kohn, R. (2005). Variable selection and covariance selection in multivariate regression models. In Dey, D. K. and Rao, C. R., editors, *Handbook of Statistics. Bayesian Thinking: Modeling and Computation*, volume 25, pages 519–552. Elsevier, North-Holland, Amsterdam.

Davies, P. T. and Tso, M. K. S. (1982). Procedures for reduced-rank regression. *Applied Statistics*, pages 244–255.

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

Fernandez, C., Ley, E., and Steel, M. (2001). Benchmark priors for Bayesian model averaging. *Journal of Econometrics*, 100:381–427.

Forni, M., Hallin, M., Lippi, M., and Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 82(4):540–554.

George, E. I. and McCulloch, R. E. (1993). Variable selection via Gibbs sampling. *Journal of the American Statistical Association*, 88(423):881–889.

George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian variable selection. *Statistica Sinica*, 7:339–373.

George, E. I., Sun, D., and Ni, S. (2008). Bayesian stochastic search for VAR model restrictions. *Journal of Econometrics*, 142(1):553–580.

Geweke, J. (1996). Bayesian reduced rank regression in econometrics. *Journal of Econometrics*, 75(1):121–146.

Golub, G. H. and Van-Loan, C. F. (2012). *Matrix Computations*, volume 3. The Johns Hopkins University Press.

Hoff, P. D. (2007). Model averaging and dimension selection for the singular value decomposition. *Journal of the American Statistical Association*, 102(478):674–685.

Izenman, A. J. (1975). Reduced-rank regression for the multivariate linear model. *Journal of Multivariate Analysis*, 5(2):248–264.

James, A. T. (1954). Normal multivariate analysis and the orthogonal group. *The Annals of Mathematical Statistics*, pages 40–75.

Kleibergen, F. and Paap, R. (2002). Priors, posteriors and bayes factors for a Bayesian analysis of cointegration. *Journal of Econometrics*, 111(2):223–249.

Kohn, R., Smith, M., and Chan, D. (2001). Nonparametric regression using linear combinations of basis functions. *Statistics and Computing*, 11:313–332.

Koop, G., Len-Gonzlez, R., and Strachan, R. W. (2009). Efficient posterior simulation for cointegrated models with priors on the cointegration space. *Econometric Reviews*, 29(2):224–242.

Koop, G., Poirier, D. J., and Tobias, J. L. (2007). *Bayesian econometric methods*. Cambridge University Press.

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

Leeper, E., Sims, C., and Zha, T. (1996). What does monetary policy do? *Brookings Papers on Economic Activity*, 2:1–63.

Littleman, R. B. (1986). Forecasting with Bayesian vector autoregressions-five years of experience. *Brookings Papers on Economic Activity*, 4(1):25–38.

Mitchell, T. and Beauchamp, J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83:1023–1036.

Muirhead, R. J. (2005). *Aspects of Multivariate Statistical Theory*. New Jersey: Wiley.

Panagiotelis, A. and Smith, M. (2008). Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *Journal of Econometrics*, 143(2):291–316.

Reinsel, G. (2006). Reduced-rank regression. *Encyclopedia of Statistical Sciences*, 11.

Robertson, J. C. and Tallman, E. W. (1999). Vector autoregressions: forecasting and reality. *Economic Review*, pages 4–18.

Robinson, P. M. (1973). Generalized canonical analysis for time series. *Journal of Multivariate Analysis*, 3(2):141–160.

Robinson, P. M. (1974). Identification, estimation and large-sample theory for regressions containing unobservable variables. *International Economic Review*, pages 680–692.

Scott, J. G. and Berger, J. O. (2010). Bayes and empical-bayes multiplicity adjustment in the variable-selection problem. *Annals of Statistics*, pages 2587–2619.

Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48.

Sims, C. A. (1992). Interpreting the macroeconomic time series facts: The effects of monetary policy. *European Economic Review*, 36:975–1000.

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

Smith, M. and Kohn, R. (1996). Nonparametric regression using Bayesian variable selection. *Journal of Econometrics*, 75:317–344.

Stock, J. and Watson, M. (2009). Generalized shrinkage methods for forecasting using many predictors. *Manuscript, Harvard University.*

Stock, J. and Watson, M. (2012). Generalized shrinkage methods for forecasting using many predictors. *Journal of Business and Economic Statistics*, 30(4):481–493.

Stock, J. H. and Watson, M. W. (2011). Dynamic factor models. *Oxford Handbook of Economic Forecasting*, 1:35–59.

Strachan, R. W. (2003). Valid Bayesian estimation of the cointegrating error correction model. *Journal of Business and Economic Statistics*, 21(1):185–195.

Strachan, R. W. and Inder, B. (2004). Bayesian analysis of the error correction model. *Journal of Econometrics*, 123(2):307–325.

Tso, M. K. S. (1981). Reduced-rank regression and canonical analysis. *Journal of the Royal Statistical Society*, Series B:183–189.

Wolfe, P., Godsill, S., and Ng, W.-J. (2004). Bayesian variable selection and regularization for time–frequency surface estimation. *Journal of the Royal Statistical Society Series B*, 66:575–589.

Yang, R. and Berger, J. (1994). Estimation of a covariance matrix using the reference prior. *The Annals of Statistics*, 22:1195–1211.

Zellner, A. (1971). *An introduction to Bayesian inference in econometrics.* Wiley, New York.