Large Bayesian Matrix Autoregressions*

Joshua C. C. Chan
Purdue University

Yaling Qi
Purdue University

Novbember 2024

Abstract

High-dimensional matrix-valued time-series are increasingly common in economics and finance. Prominent examples include large cross-region panels and dynamic economic networks. As the dimensions of the matrix grow, conventional approaches based on vector autoregressions—implemented by vectoring the matrix-valued data—become computationally infeasible. We introduce a class of large Bayesian matrix autoregressions (BMARs) that can accommodate time-varying volatility, non-Gaussian errors and COVID-19 outliers. To tackle parameter proliferation, we propose Minnesota-type shrinkage priors on the MAR coefficients. We develop a unified approach for estimating this class of models, which scales well to high dimensions. The empirical relevance of these new BMARs is illustrated using a US state-level dataset that contains 6 macroeconomic times-series for each of the 50 states, with a total of 300 times-series.

JEL classification: C11, C32, C55

Keywords: heavy-tailed distribution, outlier, shrinkage prior, stochastic volatility, tensor, vector autoregression

*We thank Christiane Baumeister, Todd Clark, Gabriel Ehrlich, Sylvia Früwirth-Schnatter, Gergely Ganics, Eva Janssens, Toru Kitagawa, Malte Knüppel, Gary Koop, John Leahy, Kurt Lunsford, Massimiliano Marcellino, James Mitchell, Giovanni Nicolò, Andriy Norets, Michele Piffer, Luca Rossini, Rodney Strachan, Herman van Dyk and Christian Wolf for their constructive comments and suggestions. We are also grateful for the discussions with seminar participants at Brown University, Federal Research Bank of Cleveland, Bocconi University, University of Michigan, MacroFor Workshop and the 2024 ESOBE.

1 Introduction

Matrix-valued data observed over time are common in economics, finance and related areas. A classic example is a cross-country panel dataset in which a few key macroeconomic indicators for each country are observed over time (Canova and Ciccarelli, 2009, 2013; Koop and Korobilis, 2016; Assaf, Li, Song, and Tsionas, 2019). More recently, larger cross-region panels with more regional units and economic variables, such as state-level or other sub-national level time-series datasets, have become widely available (Baumeister, Leiva-León, and Sims, 2022; Bokun, Jackson, Kliesen, and Owyang, 2023; Koop, McIntyre, Mitchell, Poon, and Wu, 2023). Another fast growing category of large matrix-valued time-series is dynamic economic networks, such as bilateral trade volumes among trading partners (Kharrazi, Rovenskaya, and Fath, 2017; Kapetanios, Serlenga, and Shin, 2021) and bilateral outstanding credits between countries (Billio, Casarin, Iacopini, and Kaufmann, 2023).

The growing availability of these complex datasets presents new opportunities, but it also exposes the limitations of conventional multivariate time-series econometric models. More specifically, a standard approach is to treat the matrix-valued observations over time as time-series vectors, which can then be conveniently modeled using vector autoregressions (VARs). There are, however, two disadvantages of this approach. First, vectoring the matrix-valued observation mixes its columns and rows, and consequently it disregards the correlation structure of the data (e.g., elements in the same column or row are expected to be highly correlated). Second, despite recent advances in modeling large VARs, it remains extremely time-consuming to estimate VARs obtained by vectoring the high-dimensional matrix-valued data.[1]

To tackle these challenges, we take up the matrix autoregression (MAR) introduced in Hoff (2015) and Chen, Xiao, and Yang (2021), which regresses the matrix-valued

[1]The seminal paper by Bańbura, Giannone, and Reichlin (2010) demonstrates the feasibility of fitting homoskedastic Bayesian VARs with over a hundred variables. Carriero, Clark, and Marcellino (2019) later introduce an equation-by-equation estimation approach designed for large BVARs with stochastic volatility. Tsionas, Izzeldin, and Trapani (2022) further develop this equation-by-equation approach by treating an n-dimensional BVAR as a set of n univariate equations, and cross-equation dependence is modeled using a copula. Despite these developments, fitting flexible BVARs with hundreds of variables using exact Markov chain Monte Carlo (MCMC) methods remains practically infeasible. To circumvent these computational issues, Feldkircher, Huber, Koop, and Pfarrhofer (2022) develop a fast, approximate approach for estimating large Panel VARs using integrated rotated Gaussian approximations.

observation on its lagged values using a bilinear form. The left matrix in this bilinear form models the row-wise dependence, whereas the right matrix captures the column-wise interactions. The MAR has an equivalent representation as a parsimonious VAR, where the VAR coefficient matrices are represented as Kronecker products of the left and right matrices in the bilinear form. The matrix structure of the data is thus exploited to construct the VAR coefficient matrices using far fewer free parameters relative to an unrestricted VAR. As such, the MAR modeling framework ameliorates the two drawbacks of the standard VAR approach.

We further extend the MAR framework along two directions. First, instead of assuming a time-invariant Gaussian error distribution, we introduce a class of Bayesian MARs (BMARs) that can accommodate time-varying volatility, non-Gaussian errors and COVID-19 outliers. This is motivated by the increasing recognition of the need to allow for time-varying volatility in modeling most macroeconomic datasets (see, e.g., Cogley and Sargent, 2005; Primiceri, 2005; Sims and Zha, 2006). In fact, there is now a large body of empirical evidence that demonstrates the importance of time-varying volatility for model-fit and forecasting in small VARs (Clark, 2011; D'Agostino, Gambetti, and Giannone, 2013; Clark and Ravazzolo, 2015; Chan and Eisenstat, 2018) as well as in large VARs (Koop and Korobilis, 2013; Carriero, Clark, and Marcellino, 2016, 2019; Chan, 2023a). Clark and Mertens (2023) provide a recent review on the benefits of incorporating stochastic volatility in a wide range of applications using Bayesian VARs. In addition, the unexpected extreme movements in many macroeconomic variables at the onset of the COVID-19 pandemic underline the need to allow for non-Gaussian errors and potential outliers.

Our second contribution is to introduce Bayesian shrinkage priors and efficient estimation methods that can handle large datasets. While earlier works have focused on matrix-valued time-series of moderate sizes,[2] we are interested in high-dimensional settings in which the matrix dimensions are large. For example, in our empirical application we analyze a US state-level dataset that contains 6 macroeconomic times-series for each of the 50 states, with a total of 300 times-series. Even though a MAR has far fewer parameters compared to an unrestricted VAR, it may still have more parameters than observations

[2]For example, Chen, Xiao, and Yang (2021) consider an application with a panel of 5 countries, and each has 4 economic indicators. The model is fitted using the iterative least squares and maximum likelihood estimators. Celani and Pagnottoni (2023) provide a Bayesian treatment of the MAR and consider a panel of 9 countries, each with 6 economic indicators.

over time when the dimensions of the matrix are large. Therefore, we introduce Bayesian shrinkage priors on the MAR coefficients. These new priors are inspired by the Minnesota prior of Doan, Litterman, and Sims (1984) and Litterman (1986), and can be viewed as a generalization of the Minnesota prior to the MAR setting. These priors are conjugate and hence facilitate fast estimation. Additionally, we follow Giannone, Lenza, and Primiceri (2015) and estimate the prior hyperparameters that control the overall shrinkage strength from the data, instead of fixing them at some subjective values.

Building upon the fast sampling methods in Carriero, Clark, and Marcellino (2016) and Chan (2020), we develop a unified approach for estimation—by exploiting a certain Kronecker product structure of the likelihood implied by this family of BMARs—that can drastically speed up the computations. In particular, for the matrix-valued observation $\mathbf{Y}_t$ of size $n \times k$, sampling the MAR coefficients using conventional methods would involve $\mathcal{O}(n^6)$ and $\mathcal{O}(k^6)$ elementary operations. The proposed sampling approach instead can be done in computational complexity of the order $\mathcal{O}(n^3)$ and $\mathcal{O}(k^3)$. For more general BMARs whose likelihoods do not have a Kronecker product structure (such as BMARs with the Cholesky or factor stochastic volatility), estimation can still be done in $\mathcal{O}(n^4)$ and $\mathcal{O}(k^4)$ computational complexity, using a similar equation-by-equation estimation approach developed in Carriero, Clark, and Marcellino (2019) and Carriero, Chan, Clark, and Marcellino (2022). These orders-of-magnitude speed-ups make the proposed estimation methods suitable for fitting large datasets.

The empirical relevance of these new models is illustrated using a US state-level dataset that includes 300 times-series. Even with such a large dataset, the proposed BMARs, together with the Minnesota-type shrinkage priors, can be estimated relatively quickly. The estimation results demonstrate the strong interactions between the variables across states, highlighting the importance of modeling all the state variables jointly. In addition, it is also clear that there is a spike in volatility at the onset of the COVID-19 pandemic—the error standard deviation in 2020Q2 is estimated to be between 5-6 times larger than that of regular periods. These results thus underscore the importance of allowing time-varying volatility and heavy-tailed error distributions.[3] In a recursive forecasting exercise, we show that the proposed BMARs provide better density forecasts relative to dynamic

[3]A few recent papers, such as Schorfheide and Song (2021) and Lenza and Primiceri (2022) using US data and Bobeica and Hartwig (2023) using euro area data, have shown that impulse response functions and forecasts from homoskedastic VARs are heavily distorted by the extreme observations related to the COVID-19 pandemic.

factor models. In addition, forecast performance can often be improved substantially by incorporating stochastic volatility and heavy-tailed error distributions.

In addition to contributing to the development of more flexible BMARs, this paper is also related to two other strands of literature. First, it contributes to the emerging literature on modeling multidimensional arrays or tensors (Leng and Tang, 2012; Lock, 2018), particularly third-order tensors. Most of the existing literature does not explicitly model the dynamics even when one of the tensor dimensions is time. There are a few notable exceptions. Hoff (2015) introduces a multilinear tensor autoregression based on the Tucker product. Billio, Casarin, Iacopini, and Kaufmann (2023) and Wang, Zheng, and Li (2024) develop a general linear autoregressive tensor process, where the tensor coefficients are parameterized using a PARAFAC decomposition in the former and a Tucker decomposition in the latter. In contrast to the typical time-invariant Gaussian error distribution considered in earlier works, here we explicitly model the time-varying dynamics by developing a framework that can accommodate time-varying volatility and non-Gaussian errors.[4]

This paper also contributes to the literature on modeling and forecasting regional data. Hamilton and Owyang (2012) is a classic paper that uses US state-level payroll employment data to infer regional recessions. Koop, McIntyre, and Mitchell (2020) and Koop, McIntyre, Mitchell, and Poon (2020) develop a mixed-frequency framework to nowcast UK regional growths using both regional and national data. More recent papers such as Baumeister, Leiva-León, and Sims (2022) and Bokun, Jackson, Kliesen, and Owyang (2023) have used larger US state-level datasets for nowcasting or monitoring state-level economic conditions. Our paper provides a convenient modeling framework that can handle datasets with a large number of regional units and economic indicators.

There is also a related literature on developing factor models for matrix-valued time-series. The seminal paper by Wang, Liu, and Chen (2019) introduces a novel factor model that exploits the matrix structure of the data using a bilinear form (left and right matrices of factor loadings). The model is estimated based on a Box–Ljung type statistics

[4]There are some recent works that develop estimation procedures that are robust against outliers in the context of matrix or tensor factor models. Barigozzi, He, Li, and Trapani (2023) propose estimators based on minimizing the Huber loss, which assigns an L_1-norm weight on outliers. This is similar to our likelihood approach based on distributions that have heavier tails than Gaussian. Alternatively, Barigozzi, Cho, and Maeng (2024) develop a two-stage procedure with data truncation for estimating the factor loadings.

in matrix, which is constructed from the sample autocovariance matrices of the time series. Yu, He, Kong, and Zhang (2022) propose a projection estimation method for this matrix factor model that is especially suitable for high-dimensional settings. They show that the proposed projected estimators of the factor loadings achieve faster convergence rates than alternatives. In terms of model development, Chen, Tsay, and Chen (2020) consider a general framework for imposing linear restrictions—derived from domain or prior knowledge—on the factor loadings matrices. This achieves a more parsimonious parameterization and facilitates interpretation of the latent matrix factors. Liu and Chen (2022) further propose a threshold matrix factor model in which the factor loadings matrices are regime-dependent and the regimes are determined by an observed threshold variable. In a recent paper, Chen, Chen, Bolivar, and Chen (2024) develop a time-varying matrix factor model in which the factor loadings can change smoothly over time to account for potential structural changes.

The rest of this paper is organized as follows. Section 2 first introduces a general framework for modeling matrix-valued time-series with a flexible error covariance structure. It then offers a few different interpretations of the matrix autoregression and discusses some identification issues. Lastly, the section develops Bayesian shrinkage priors that generalize the Minnesota priors to the MAR setting. Section 3 proposes a unified approach to estimate these flexible BMARs using MCMC methods. Section 4 further extends these BMARs to accommodate more general stochastic volatility specifications, such as the Cholesky and factor stochastic volatility. We then discuss in Section 5 the Bayesian approach of using the marginal likelihood to assess if the restrictions implied by the MAR are appropriate. Section 6 considers an application that involves a US state-level dataset with 300 time-series. Lastly, Section 7 concludes and outlines some future research directions.

2 A Flexible Framework for Matrix Autoregressions

We introduce a general framework for Bayesian matrix autoregressions that aims to strike the right balance between flexibility and tractability in high-dimensional settings. On the one hand, this flexible framework can accommodate a wide variety of empirically relevant features, including heavy-tailed error distributions, time-varying volatility and robustness

to outliers. On the other hand, it also facilitates fast computation and can be used to model large datasets.

2.1 The Modeling Framework

To set the stage, let $\mathbf{Y}_t$ denote an $n \times k$ matrix of endogenous variables at time t for $t = 1, \ldots, T$. To fix ideas, one may think of each column of $\mathbf{Y}_t$ containing the n variables for each of the k regions. In our empirical application that models US state-level data, we have $k = 50$ states and each state has $n = 6$ variables, with a total of 300 variables.

A common approach to model the matrix-valued data $\mathbf{Y}_t$ is to first stack its columns into a vector, i.e., $\text{vec}(\mathbf{Y}_t)$, which is then fitted using the vector autoregression with p lags:

$$\text{vec}(\mathbf{Y}_t) = \mathbf{\Phi}_1 \text{vec}(\mathbf{Y}_{t-1}) + \cdots + \mathbf{\Phi}_p \text{vec}(\mathbf{Y}_{t-p}) + \mathbf{e}_t, \tag{1}$$

where $\mathbf{\Phi}_1, \ldots, \mathbf{\Phi}_p$ are $nk \times nk$ coefficient matrices and $\mathbf{e}_t$ is an $nk \times 1$ vector of errors. There are two main drawbacks of modeling $\mathbf{Y}_t$ using the VAR in (1). First, by vectoring $\mathbf{Y}_t$, the columns and rows of $\mathbf{Y}_t$ are mixed. Consequently, the VAR ignores the matrix structure—e.g., the strong connections between the variables in the same region (column) and those between the same variable (row) across regions. The second drawback is the proliferation of parameters when either n or k is large. For example, for $n = 6, k = 50$ and $p = 2$, there are $180,000$ VAR coefficients, which makes estimation and inference practically infeasible.

To tackle these two issues, we follow Hoff (2015) and Chen, Xiao, and Yang (2021) to directly model the evolution of the matrix $\mathbf{Y}_t$ via the following matrix autoregression (MAR):

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{Y}_{t-1} \mathbf{B}_1' + \cdots + \mathbf{A}_p \mathbf{Y}_{t-p} \mathbf{B}_p' + \mathbf{E}_t, \tag{2}$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ and $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are, respectively, $n \times n$ and $k \times k$ coefficient matrices. For simplicity we exclude the intercepts; a matrix of intercepts or any deterministic term can be easily added to the model. The above bilinear form facilitates model interpretation and estimation. In particular, the matrix autoregression in (2) can be represented in the

form of a VAR:

$$\text{vec}(\mathbf{Y}_t) = (\mathbf{B}_1 \otimes \mathbf{A}_1)\text{vec}(\mathbf{Y}_{t-1}) + \cdots + (\mathbf{B}_p \otimes \mathbf{A}_p)\text{vec}(\mathbf{Y}_{t-p}) + \text{vec}(\mathbf{E}_t),$$

where $\otimes$ denotes the Kronecker product. Hence, the MAR can be viewed as a special case of the VAR, where the VAR coefficient matrix is modeled as the Kronecker product $\mathbf{\Phi}_j = (\mathbf{B}_j \otimes \mathbf{A}_j)$. Consequently, the number of VAR coefficients is reduced from n^2k^2p to $(n^2 + k^2)p$. Subsection 2.2 provides more discussions on the interpretation of the MAR and and its relations to the VAR.[5]

While earlier works consider only homoskedastic MARs where the distribution of the $n \times k$ matrix of errors, $\mathbf{E}_t$, is time-invariant, we propose a more general setting in which $\mathbf{E}_t$ has a conditionally Gaussian distribution given the latent variable ω_t:

$$\text{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}_{nk}, \omega_t \mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r), \tag{3}$$

where $\mathbf{\Sigma}_c$ and $\mathbf{\Sigma}_r$ are, respectively, $k \times k$ and $n \times n$ covariance matrices. The homoskedastic MAR considered in Hoff (2015) and Chen, Xiao, and Yang (2021) can be recovered as a special case with $\omega_1 = \cdots = \omega_T = 1$. By assuming different distributions for the mixing variables $\omega_1, \ldots, \omega_T$, this framework encompasses a wide range of flexible error distributions that are found empirically useful for modeling macroeconomic and financial data. Below we give a few important examples.

1. **Heavy-tailed distributions**. Since many distributions can be represented as a scale mixture of normals, the conditionally Gaussian specification in (3) can accommodate many common heavy-tailed distributions that are useful to capture rare but large changes in volatility. For example, if the mixing variable ω_t follows the inverse-gamma distribution $\omega_t \sim \mathcal{IG}(\nu/2, \nu/2)$, then the marginal distribution of $\text{vec}(\mathbf{E}_t)$ unconditional on ω_t has a multivariate t distribution with zero mean, scale matrix $\mathbf{\Sigma}_c \otimes \mathbf{\Sigma}_r$ and degree of freedom parameter ν. Alternatively, if ω_t has a gamma distribution, then marginally $\text{vec}(\mathbf{E}_t)$ has a multivariate normal-gamma distribution, which includes the multivariate Laplace distribution as a special case. Both of these distributions have heavier tails than normals,

[5]For lag length $p = 1$, Chen, Xiao, and Yang (2021) prove that if the product of the spectral radiuses of $\mathbf{A}_1$ and $\mathbf{B}_1$ is less than 1, then the MAR is stationary. Since the MAR has an equivalent VAR representation, for general lag length p, if the spectral radius of the companion matrix of the associated VAR is less then 1, then the MAR is stationary.

and they generally provide better fit for data with infrequent volatility jumps. Empirical studies that find heavy-tailed errors useful in the context of VARs include Clark and Ravazzolo (2015), Cross and Poon (2016) and Chiu, Mumtaz, and Pinter (2017).

2. **Robustness to outliers**. The conditionally Gaussian specification in (3) can also be used for addressing potential outliers using a tailored mixing distribution. An important example is an explicit outlier component of the type proposed in Stock and Watson (2016). More specifically, let $\omega_t = o_t^2$ and o_t follows a mixture of two distributions: a point mass at 1 and a uniform distribution on the interval $(2, 20)$. The former can be thought of as 'regular' observations with scale normalized to 1, whereas the latter captures 'outliers' that have 2-20 times larger standard deviations than regular observations. As demonstrated in Carriero, Clark, Marcellino, and Mertens (2022), this outlier component is especially useful for modeling observations associated with the COVID-19 pandemic.

3. **Time-varying volatility**. One of the most robust empirical findings in modeling macroeconomic data is the importance of allowing for time-varying volatility (e.g., Sims and Zha, 2006; Clark, 2011; Chan and Eisenstat, 2018). The conditionally Gaussian framework in (3) can accommodate certain types of time-varying volatility processes. An important example is the common stochastic volatility model introduced in Carriero, Clark, and Marcellino (2016). In particular, let $\omega_t = \mathrm{e}^{h_t}$, and assume that the log-volatility h_t follows a stationary AR(1) process with 0 mean:

$$h_t = \phi h_{t-1} + u_t^h, \quad u_t^h \sim \mathcal{N}(0, \sigma_h^2), \tag{4}$$

for $t = 2, \ldots, T$, where $|\phi| < 1$ and the initial condition is specified as $h_1 \sim \mathcal{N}(0, \sigma_h^2/(1 - \phi^2))$. The log-volatility h_t here may be interpreted as the level of economy-wide macroeconomic uncertainty (see also Jurado, Ludvigson, and Ng, 2015). Another example is the volatility model with a deterministic break date considered in Lenza and Primiceri (2022), which is designed to model the drastic increase in volatility at the onset of the COVID-19 pandemic and the subsequent gradual decrease in volatility. Their model can also be parameterized using the conditionally Gaussian framework.

Naturally, any combinations of the above heavy-tailed errors and volatility processes can also be incorporated using the conditionally Gaussian framework. For instance, one may consider a MAR with the common stochastic volatility and the outlier component. In that case, $\omega_t = \mathrm{e}^{h_t} o_t^2$, where h_t follows the AR(1) process in (4) and o_t follows the two-

component mixture described above. Other models such as those in Chan (2020) and Hartwig (2021) can also be considered.

While the modeling framework in (2)–(3) is flexible and includes many empirically useful specifications as special cases, it is crucial to recognize its limitations. In particular, the latent variable ω_t is assumed to scale the entire covariance matrix of $\mathbf{E}_t$, implying that each element of $\mathbf{E}_t$ is impacted equally by ω_t. As such, the proposed framework does not nest, for example, a model in which each row of $\mathbf{Y}_t$ has its specific stochastic volatility factor. Estimation of such a model, however, would be practically infeasible in high-dimensional settings. The proposed framework therefore provides the right balance between modeling flexibility and computational tractability.

2.2 Model Interpretation and Identification

In this section we discuss various interpretations of the coefficient matrices in the MAR and some identification issues. For ease of exposition, throughout this section we consider the case with only one lag:

$$\mathbf{Y}_t = \mathbf{A}_1 \mathbf{Y}_{t-1} \mathbf{B}_1' + \mathbf{E}_t. \tag{5}$$

We first cast the MAR(1) in (5) as a VAR and relate the coefficient matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ to the corresponding VAR coefficients. We then discuss how the MAR coefficient matrices are related to the 3-dimensional coefficient array in a more general MAR framework.

2.2.1 Relation to VARs

The coefficient matrix $\mathbf{A}_1$ in the bilinear form in (5) corresponds to row-wise relationships, whereas $\mathbf{B}_1$ represents column-wise interactions. To tease out the impact of the two matrices, it is useful to consider a few special cases. Recall that the MAR may be viewed as a special VAR in which the VAR coefficient matrix is parameterized as $\mathbf{\Phi}_1 = (\mathbf{B}_1 \otimes \mathbf{A}_1)$. If we assume $\mathbf{B}_1 = \mathbf{I}_k$, then we can express the MAR as:

$$\text{vec}(\mathbf{Y}_t) = (\mathbf{I}_k \otimes \mathbf{A}_1)\text{vec}(\mathbf{Y}_{t-1}) + \text{vec}(\mathbf{E}_t).$$

In other words, each column of $\mathbf{Y}_t$ follows the same VAR with the coefficient matrix $\mathbf{A}_1$, and there are no interactions among the columns (in the conditional mean). Similarly, for the special case with $\mathbf{A}_1 = \mathbf{I}_n$, each row of $\mathbf{Y}_t$ follows a VAR with the same coefficient matrix.

The covariance matrix of $\text{vec}(\mathbf{E}_t)$ has a similar interpretation. For simplicity, set $\omega_t = 1$. Then, the matrix of errors as specified in (3) can be equivalently represented as $\mathbf{E}_t = \mathbf{\Sigma}_r^{\frac{1}{2}} \mathbf{Z}_t \mathbf{\Sigma}_c^{\frac{1}{2}}$, where $\mathbf{Z}_t$ is an $n \times k$ matrix consisting of independent standard normal random variables. It is clear from this representation that $\mathbf{\Sigma}_r$ corresponds to row-wise covariances and $\mathbf{\Sigma}_c$ represents column-wise covariances. In particular, if $\mathbf{\Sigma}_c = \mathbf{I}_k$, then $\mathbf{E}_t = \mathbf{\Sigma}_r^{\frac{1}{2}} \mathbf{Z}_t$, which implies that the columns of $\mathbf{E}_t$ are all mutually independent and each row has the same covariance matrix $\mathbf{\Sigma}_r$. More generally, the covariance between the (i_1, j_1) and (i_2, j_2) elements of $\mathbf{E}_t$ is $\text{cov}(e_{t,i_1,j_1}, e_{t,i_2,j_2}) = \sigma_{r,i_1,i_2} \sigma_{c,j_1,j_2}$.

Another interpretation of the MAR is related to the global VAR (Pesaran, Schuermann, and Weiner, 2004) and the multivariate autoregressive index model (Carriero, Kapetanios, and Marcellino, 2016). More specifically, let $y_{t,i,j}$, $a_{1,i,j}$ and $b_{1,i,j}$ denote the (i, j) elements of $\mathbf{Y}_t$, $\mathbf{A}_1$ and $\mathbf{B}_1$, respectively. Then, $y_{t,i,j}$ can be expressed as

$$y_{t,i,j} = \sum_{l_1=1}^{n} \sum_{l_2=1}^{k} a_{1,i,l_1} b_{1,j,l_2} y_{t-1,l_1,l_2} + e_{t,i,j} = \sum_{l_1=1}^{n} a_{1,i,l_1} z_{t-1,l_1,j} + e_{t,i,j},$$

where $z_{t-1,l_1,j} = \sum_{l_2=1}^{k} b_{1,j,l_2} y_{t-1,l_1,l_2}$ is a linear combination of the l_1-th row of $\mathbf{Y}_{t-1}$ across the columns. Under this representation, one can view the MAR as a multi-equation regression with covariates constructed from linear combinations of the columns of $\mathbf{Y}_{t-1}$. In particular, using our running state-level data example, the MAR can be interpreted as first constructing linear combinations of GDP, unemployment, etc., across states, and use them as regressors.

2.2.2 Relation to a Factor Model

The MAR also a two-step, hierarchical interpretation analogous to the multilevel interpretation of the matrix factor model discussed in Wang, Liu, and Chen (2019). More specifically, in the first step, suppose that each row of $\mathbf{Y}_t$, denoted as $\mathbf{y}_{t,i,\cdot}$ (as a $k \times 1$

column vector), has the factor structure

$$\mathbf{y}_{t,i,\cdot} = \mathbf{B}_1^{(i)} \mathbf{z}_{t,i,\cdot} + \mathbf{u}_{t,i,\cdot},$$

where $\mathbf{z}_{t,i,\cdot}$ is a $k \times 1$ vector of factors and $\mathbf{B}_1^{(i)}$ is the associated $k \times k$ factor loadings matrix. Next, let $\mathbf{Z}_t = (\mathbf{z}_{t,1,\cdot}, \ldots, \mathbf{z}_{t,n,\cdot})'$ denote the $n \times k$ matrix of factors. Then, in the second step, suppose we construct each column of $\mathbf{Z}_t$, denoted as $\mathbf{z}_{t,\cdot,j}$ for $j = 1, \ldots, k$ as

$$\mathbf{z}_{t,\cdot j} = \mathbf{A}_1^{(j)} \mathbf{y}_{t-1,\cdot j} + \mathbf{v}_{t,\cdot,j},$$

where $\mathbf{y}_{t-1,\cdot,j}$ is the j-th column of $\mathbf{Y}_{t-1}$ and $\mathbf{A}_1^{(j)}$ is an $n \times n$ coefficient matrix. Finally, putting these two stages together and letting $\mathbf{A}_1^{(1)} = \cdots = \mathbf{A}_1^{(k)} \equiv \mathbf{A}_1$ and $\mathbf{B}_1^{(1)} = \cdots = \mathbf{B}_1^{(n)} \equiv \mathbf{B}_1$, we have

$$\mathbf{Y}_t = \mathbf{Z}_t \mathbf{B}_1' + \mathbf{U} = \mathbf{A}_1 \mathbf{Y}_{t-1} \mathbf{B}_1' + \mathbf{E},$$

where $\mathbf{E} = \mathbf{V}\mathbf{B}_1' + \mathbf{U}$.

2.2.3 Relation to a Generalized MAR

Following Hoff (2015) and Chen, Xiao, and Yang (2021), we formulate the MAR in (5) using a bilinear form. However, a more general autoregressive process for matrix-valued data can be represented as

$$\mathbf{Y}_t = \langle \mathcal{C}, \text{vec}(\mathbf{Y}_{t-1}) \rangle + \mathbf{E}_t, \tag{6}$$

where $\mathcal{C} \in \mathbb{R}^{n \times k \times nk}$ is a 3-dimensional array or tensor and $\langle \cdot, \cdot \rangle$ is the generalized inner product (see, e.g., Kolda and Bader, 2009, for a general introduction to tensors and their associated operations). The generalized inner product $\langle \mathcal{C}, \text{vec}(\mathbf{Y}_{t-1}) \rangle$ in this case defines an $n \times k$ matrix, where the (i, j) element is given by

$$\langle \mathcal{C}, \text{vec}(\mathbf{Y}_{t-1}) \rangle_{i,j} = \sum_{l_1=1}^{n} \sum_{l_2=1}^{k} \mathcal{C}_{i,j,(l_2-1)n+l_1} y_{t-1,l_1,l_2}.$$

Note that the third-order tensor $\mathcal{C}$ has n^2k^2 free parameters, and they are difficult to pin down in high-dimensional settings. The bilinear form in (5) may be viewed as a judicious way to constructs $\mathcal{C}$ using matrices $\mathbf{A}_1$ and $\mathbf{B}_1$ with a total of $n^2 + k^2$ free parameters.

Specifically, it can be readily verified that (6) reduces to the MAR in (5) if we set

$$\mathcal{C}_{i,j,(l_2-1)n+l_1} = a_{1,i,l_1} b_{1,j,l_2},$$

where a_{1,i,l_1} is the (i, l_1) element of $\mathbf{A}_1$ and b_{1,j,l_2} is the (j, l_2) element of $\mathbf{B}_1$.

2.2.4 Identification Issues

Finally, we discuss some identification issues that arise in the MAR. First, the parameters $\mathbf{A}_1$ and $\mathbf{B}_1$ are not separately identified, but they are identified up to scale. That is, if $(\mathbf{B}_1 \otimes \mathbf{A}_1)\mathbf{z} = (\widetilde{\mathbf{B}}_1 \otimes \widetilde{\mathbf{A}}_1)\mathbf{z}$, for all $\mathbf{z}$, then $\widetilde{\mathbf{A}}_1 = c\mathbf{A}_1$ and $\widetilde{\mathbf{B}}_1 = c^{-1}\mathbf{B}_1$ for some $c \neq 0$, provided that neither $\mathbf{A}_1$ and $\mathbf{B}_1$ is the zero matrix. Hence, to fix the scale, we normalize the (1,1) element of $\mathbf{B}_1$ to be 1. More generally, for the MAR of order p, we set the (1,1) element of $\mathbf{B}_j$ to be 1, i.e., $b_{j,1,1} = 1, j = 1, \ldots, p$. Similarly, the covariances $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$ are only identified up to scale. We normalize the (1,1) element of $\mathbf{\Sigma}_c$ to be 1: $\sigma_{c,1,1} = 1$.

2.3 Bayesian Shrinkage Priors

We are interested in settings when n or k (or both) is large. In those cases, the matrix autoregression has a large number of parameters, and consequently regularization or shrinkage is vital for obtaining sensible results. In addition, to facilitate fast estimation, we extend the natural conjugate prior (see, e.g., Koop and Korobilis, 2010; Karlsson, 2013) designed for VARs to our setting. To that end, let $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)'$ and $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)'$, so that $\mathbf{A}$ and $\mathbf{B}$ are of dimensions $np \times n$ and $kp \times k$, respectively. We consider the prior of the form $p(\mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c \,|\, \kappa_{\mathbf{A}}, \kappa_{\mathbf{B}}) = p(\mathbf{A}, \mathbf{\Sigma}_r \,|\, \kappa_{\mathbf{A}}) p(\mathbf{B}, \mathbf{\Sigma}_c \,|\, \kappa_{\mathbf{B}})$, where $\kappa_{\mathbf{A}}$ and $\kappa_{\mathbf{B}}$ are some hyperparameters which we treat as unknown.

First, we assume that $(\mathbf{A}, \mathbf{\Sigma}_r)$ has a normal-inverse-Wishart distribution (see, e.g., Kadiyala and Karlsson, 1997; Koop and Korobilis, 2010):

$$\mathbf{\Sigma}_r \sim \mathcal{IW}(\nu_r, \mathbf{S}_r), \quad (\mathrm{vec}(\mathbf{A}) \,|\, \mathbf{\Sigma}_r, \kappa_{\mathbf{A}}) \sim \mathcal{N}(\mathrm{vec}(\mathbf{A}_0), \mathbf{\Sigma}_r \otimes \mathbf{V}_{\mathbf{A}}),$$

where $\mathrm{vec}(\mathbf{A}_0)$ is the prior mean vector and the $np \times np$ prior covariance matrix $\mathbf{V}_{\mathbf{A}}$ is assumed to be diagonal and depend on the unknown hyperparameter $\kappa_{\mathbf{A}}$. The joint

density function of $(\mathbf{A}, \mathbf{\Sigma}_r)$ is thus given by

$$p(\mathbf{A}, \mathbf{\Sigma}_r) \propto |\mathbf{V}_\mathbf{A}|^{-\frac{n}{2}} |\mathbf{\Sigma}_r|^{-\frac{\nu_r+n+np+1}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\mathbf{\Sigma}_r^{-1}\mathbf{S}_r)} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{\Sigma}_r^{-1}(\mathbf{A}-\mathbf{A}_0)'\mathbf{V}_\mathbf{A}^{-1}(\mathbf{A}-\mathbf{A}_0)\right)}, \tag{7}$$

where $\mathrm{tr}(\cdot)$ is the trace operator.

We calibrate $\mathbf{A}_0$ and $\mathbf{V}_\mathbf{A}$ in the spirit of the Minnesota priors pioneered by Doan, Litterman, and Sims (1984) and Litterman (1986). More specifically, $\mathrm{vec}(\mathbf{A}_0)$ is set to be a zero vector for growth rates data. This reflects the prior belief that growth rates data are typically not very persistent, and the coefficient matrix $\mathbf{A}$ is thus shrunk to 0. For levels data, $\mathrm{vec}(\mathbf{A}_0)$ is set to be zero except for the coefficients associated with the first own lag, which are set to be one. This expresses the preference for a random walk specification, reflecting the prior belief that levels data are generally highly persistent.

To calibrate the diagonal elements of $\mathbf{V}_\mathbf{A}$, let $\widehat{s}^2_{i,\bullet} = \sum_{j=1}^k \widehat{s}^2_{i,j}/k$, where $\widehat{s}^2_{i_1,i_2}$ denotes the sample variance of an AR(4) model for the variable y_{t,i_1,i_2}, the (i_1, i_2) element of $\mathbf{Y}_t$. Hence, $\widehat{s}^2_{i,\bullet}$ is the average sample variances of the variables in the i-th row. Then, the j-th diagonal element of $\mathbf{V}_\mathbf{A}$ is assumed to be $v_{\mathbf{A},j,j} = \kappa_\mathbf{A}/(l^2\widehat{s}^2_{i,\bullet})$ for a coefficient associated with lag l of the variable in the i-th row. Intuitively, the prior variance is scaled by $\widehat{s}^2_{i,\bullet}$, and the coefficient associated to a lag l variable is shrunk more heavily to zero as the lag length increases. The overall shrinkage strength is controlled by the hyperparameter $\kappa_\mathbf{A}$, where a smaller value indicates more aggressive shrinkage. We follow the recommendation of Giannone, Lenza, and Primiceri (2015) to estimate $\kappa_\mathbf{A}$ from the data instead of fixing it at some commonly-used subjective value. Finally, we set $\nu_r = n+2$, $\mathbf{S}_r = \mathrm{diag}(\widehat{s}^2_{1,\bullet}, \ldots, \widehat{s}^2_{n,\bullet})$. These hyperparameters are elicited in the spirit of the Minnesota priors. In particular, for $k = 1$, they reduce to those of the standard Minnesota priors (see, e.g., Karlsson, 2013; Carriero, Clark, and Marcellino, 2015).

Similarly, we consider the following normal-inverse-Wishart prior on $(\mathbf{B}, \mathbf{\Sigma}_c)$:

$$\mathbf{\Sigma}_c \sim \mathcal{IW}(\nu_c, \mathbf{S}_c), \quad (\mathrm{vec}(\mathbf{B}) \mid \mathbf{\Sigma}_c, \kappa_\mathbf{B}) \sim \mathcal{N}(\mathrm{vec}(\mathbf{B}_0), \mathbf{\Sigma}_c \otimes \mathbf{V}_\mathbf{B}),$$

where $\mathrm{vec}(\mathbf{B}_0)$ is the prior mean vector and the $kp \times kp$ prior covariance matrix $\mathbf{V}_\mathbf{B}$ is assumed to be diagonal and depend on the unknown hyperparameter $\kappa_\mathbf{B}$. Naturally, one can elicit $\mathbf{B}_0$ and $\mathbf{V}_\mathbf{B}$ to incorporate prior beliefs specific to the application. Below we provide a baseline case that is expected to be applicable to a wide range of cross-region

applications.

As discussed in Section 2.2, when $\mathbf{B}_j = \mathbf{I}_k, j = 1, \ldots, p$, then each column of $\mathbf{Y}_t$, representing observations from a particular region, follows the same VAR with the coefficient matrix $\mathbf{A}$, and there are no interactions among the columns in the conditional mean. We therefore set the prior mean to be $\mathbf{B}_0 = (\mathbf{I}_k, \ldots, \mathbf{I}_k)'$, and shrink $\mathbf{B}$ toward to this simpler setting. This choice of prior mean is also consistent with the identification restrictions that the (1,1) elements of $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are 1. To calibrate the diagonal elements of $\mathbf{V}_\mathbf{B}$, let $\widehat{s}^2_{\bullet,j} = \sum_{i=1}^n \widehat{s}^2_{i,j}/n$ denote the average sample variances of the variables in the j-th column. Then, the i-th diagonal element of $\mathbf{V}_\mathbf{B}$ is assumed to be $v_{\mathbf{B},i,i} = \kappa_\mathbf{B}/(l^2\widehat{s}^2_{\bullet,j})$ for a coefficient associated with lag l of the variable in the j-th column. Hence, a coefficient is shrunk more strongly to its mean if it corresponds to a variable of higher lag, and the prior variance is scaled by $\widehat{s}^2_{\bullet,j}$. The hyperparameter $\kappa_\mathbf{B}$ determines the overall shrinkage strength, which is again estimated from the data. We set $\nu_c = k + 2$, $\mathbf{S}_c = \text{diag}(1, \widehat{s}^2_{\bullet,2}/\widehat{s}^2_{\bullet,1}, \ldots, \widehat{s}^2_{\bullet,k}/\widehat{s}^2_{\bullet,1})$. Here we normalize the scale matrix $\mathbf{S}_c$ so that it is consistent with the identification restriction that the $(1,1)$ element of $\mathbf{\Sigma}_c$ is fixed at one.

Finally, the hyperparameters $\kappa_\mathbf{A}$ and $\kappa_\mathbf{B}$ are assumed to have hierarchical gamma priors: $\kappa_\mathbf{A} \sim \mathcal{G}(c_{\mathbf{A},1}, c_{\mathbf{A},2})$ and $\kappa_\mathbf{B} \sim \mathcal{G}(c_{\mathbf{B},1}, c_{\mathbf{B},2})$.

3 Bayesian Estimation

In this section we provide a general discussion on the estimation of the BMARs specified in (2)-(3). In particular, we develop a fast and simple approach to sample the pairs $(\mathbf{A}, \mathbf{\Sigma}_r)$ and $(\mathbf{B}, \mathbf{\Sigma}_c)$ given the shrinkage hyperparameters $\boldsymbol{\kappa} = (\kappa_\mathbf{A}, \kappa_\mathbf{B})'$ and an arbitrary vector of latent variables $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)'$. In Appendix A we take up various examples of $\boldsymbol{\omega}$ and provide estimation details for tackling each case, as well as the sampling steps for $\boldsymbol{\kappa}$.

We first derive the likelihood function implied by (2)-(3). Letting $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)'$ and $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)'$, note that one can rewrite the mean equation in (2) as:

$$\mathbf{Y}_t = \mathbf{A}'\mathbf{X}_t\mathbf{B} + \mathbf{E}_t,$$

where $\mathbf{X}_t = \text{diag}(\mathbf{Y}_{t-1}, \ldots, \mathbf{Y}_{t-p})$ is an $np \times kp$ block-diagonal matrix of lagged values. Given the covariance structure in (3), the likelihood function can be expressed as:

$$p(\mathbf{Y} \,|\, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\Sigma}_r, \boldsymbol{\omega}) = (2\pi)^{-\frac{Tnk}{2}} |\boldsymbol{\Sigma}_c|^{-\frac{Tn}{2}} |\boldsymbol{\Sigma}_r|^{-\frac{Tk}{2}} \prod_{t=1}^{T} \omega_t^{-\frac{nk}{2}} \mathrm{e}^{-\frac{1}{2\omega_t}\text{tr}\left(\boldsymbol{\Sigma}_c^{-1}(\mathbf{Y}_t - \mathbf{A}'\mathbf{X}_t\mathbf{B})'\boldsymbol{\Sigma}_r^{-1}(\mathbf{Y}_t - \mathbf{A}'\mathbf{X}_t\mathbf{B})\right)}. \tag{8}$$

Assuming the natural conjugate priors for $(\mathbf{A}, \boldsymbol{\Sigma}_r)$ and $(\mathbf{B}, \boldsymbol{\Sigma}_c)$, posterior draws can be obtained by sequentially sampling from: 1) $p(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 2) $p(\mathbf{B}, \boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 3) $p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\omega})$; and 4) $p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa})$. Depending on how one models the latent variables $\boldsymbol{\omega}$, additional blocks might be needed to sample some additional hierarchical parameters. These steps are typically easy to implement as they amount to fitting a univariate time-series model. A variety of examples are given in Appendix A. Below we provide details on implementing Step 1 and Step 2 of sampling from the high-dimensional densities $p(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$ and $p(\mathbf{B}, \boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega})$ efficiently.

More specifically, recall that $\mathbf{A}$ is of dimensions $np \times n$, and sampling $\mathbf{A}$ using conventional methods would involve $\mathcal{O}(n^6)$ elementary operations. Fortunately, it can be shown that $(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$ has a normal-inverse-Wishart distribution, and one can sample $\mathbf{A}$ with computational complexity of the order $\mathcal{O}(n^3)$. To see this, note that it follows from (7) and (8) that

$$\begin{aligned} p(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega}) \propto & |\boldsymbol{\Sigma}_r|^{-\frac{\nu_r + n + np + Tk + 1}{2}} \mathrm{e}^{-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}_r^{-1}\mathbf{S}_r)} \\ & \times \mathrm{e}^{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}_r^{-1}((\mathbf{A}-\mathbf{A}_0)'\mathbf{V}_{\mathbf{A}}^{-1}(\mathbf{A}-\mathbf{A}_0) + \sum_{t=1}^{T}\omega_t^{-1}(\mathbf{Y}_t - \mathbf{A}'\mathbf{X}_t\mathbf{B})\boldsymbol{\Sigma}_c^{-1}(\mathbf{Y}_t - \mathbf{A}'\mathbf{X}_t\mathbf{B})')\right)} \\ = & |\boldsymbol{\Sigma}_r|^{-\frac{\nu_r + n + np + Tk + 1}{2}} \mathrm{e}^{-\frac{1}{2}\text{tr}(\boldsymbol{\Sigma}_r^{-1}\mathbf{S}_r)} \mathrm{e}^{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}_r^{-1}(\mathbf{A}_0'\mathbf{V}_{\mathbf{A}}^{-1}\mathbf{A}_0 + \sum_{t=1}^{T}\omega_t^{-1}\mathbf{Y}_t\boldsymbol{\Sigma}_c^{-1}\mathbf{Y}_t' - \widehat{\mathbf{A}}'\mathbf{K}_{\mathbf{A}}\widehat{\mathbf{A}})\right)} \\ & \times \mathrm{e}^{-\frac{1}{2}\text{tr}\left(\boldsymbol{\Sigma}_r^{-1}(\mathbf{A}-\widehat{\mathbf{A}})'\mathbf{K}_{\mathbf{A}}(\mathbf{A}-\widehat{\mathbf{A}})\right)}, \end{aligned}$$

where

$$\mathbf{K}_{\mathbf{A}} = \mathbf{V}_{\mathbf{A}}^{-1} + \sum_{t=1}^{T}\omega_t^{-1}\mathbf{X}_t\mathbf{B}\boldsymbol{\Sigma}_c^{-1}\mathbf{B}'\mathbf{X}_t', \quad \widehat{\mathbf{A}} = \mathbf{K}_{\mathbf{A}}^{-1}\left(\mathbf{V}_{\mathbf{A}}^{-1}\mathbf{A}_0 + \sum_{t=1}^{T}\omega_t^{-1}\mathbf{X}_t\mathbf{B}\boldsymbol{\Sigma}_c^{-1}\mathbf{Y}_t'\right).$$

In the above derivation we have used the fact that

$$(\mathbf{A}-\mathbf{A}_0)'\mathbf{V}_{\mathbf{A}}^{-1}(\mathbf{A}-\mathbf{A}_0)+\sum_{t=1}^{T}\omega_t^{-1}(\mathbf{Y}_t-\mathbf{A}'\mathbf{X}_t\mathbf{B})\boldsymbol{\Sigma}_c^{-1}(\mathbf{Y}_t-\mathbf{A}'\mathbf{X}_t\mathbf{B})'$$
$$=(\mathbf{A}-\widehat{\mathbf{A}})'\mathbf{K}_{\mathbf{A}}(\mathbf{A}-\widehat{\mathbf{A}})+\mathbf{A}_0'\mathbf{V}_{\mathbf{A}}^{-1}\mathbf{A}_0+\sum_{t=1}^{T}\omega_t^{-1}\mathbf{Y}_t\boldsymbol{\Sigma}_c^{-1}\mathbf{Y}_t'-\widehat{\mathbf{A}}'\mathbf{K}_{\mathbf{A}}\widehat{\mathbf{A}}.$$

In other words, $(\mathbf{A},\boldsymbol{\Sigma}_r\,|\,\mathbf{Y},\mathbf{B},\boldsymbol{\Sigma}_c,\boldsymbol{\kappa},\boldsymbol{\omega})$ has the normal-inverse-Wishart distribution with parameters ν_r+Tk, $\widehat{\mathbf{S}}_r$, $\widehat{\mathbf{A}}$ and $\mathbf{K}_{\mathbf{A}}^{-1}$, where

$$\widehat{\mathbf{S}}_r=\mathbf{S}_r+\mathbf{A}_0'\mathbf{V}_{\mathbf{A}}^{-1}\mathbf{A}_0+\sum_{t=1}^{T}\omega_t^{-1}\mathbf{Y}_t\boldsymbol{\Sigma}_c^{-1}\mathbf{Y}_t'-\widehat{\mathbf{A}}'\mathbf{K}_{\mathbf{A}}\widehat{\mathbf{A}}.$$

Hence, we can sample $(\mathbf{A},\boldsymbol{\Sigma}_r\,|\,\mathbf{Y},\mathbf{B},\boldsymbol{\Sigma}_c,\boldsymbol{\kappa},\boldsymbol{\omega})$ in two steps. First, we sample $\boldsymbol{\Sigma}_r$ marginally from $(\boldsymbol{\Sigma}_r\,|\,\mathbf{Y},\mathbf{B},\boldsymbol{\Sigma}_c,\boldsymbol{\kappa},\boldsymbol{\omega})\sim\mathcal{IW}(\widehat{\mathbf{S}}_r,\nu_r+Tk)$. Then, given the $\boldsymbol{\Sigma}_r$ drawn, we sample

$$(\mathrm{vec}(\mathbf{A})\,|\,\mathbf{Y},\mathbf{B},\boldsymbol{\Sigma}_r,\boldsymbol{\Sigma}_c,\boldsymbol{\kappa},\boldsymbol{\omega})\sim\mathcal{N}\left(\mathrm{vec}(\widehat{\mathbf{A}}),\boldsymbol{\Sigma}_r\otimes\mathbf{K}_{\mathbf{A}}^{-1}\right).$$

Since the covariance matrix is of dimension $n^2p\times n^2p$, and sampling from this high-dimensional density using conventional methods would involve $\mathcal{O}(n^6p^3)$ operations. This can be computationally intensive when n is large. Instead, we adopt an efficient algorithm to sample from the matrix-normal distribution to our setting (e.g., Bauwens, Lubrano, and Richard, 1999, p.320). This algorithm has been used in Carriero, Clark, and Marcellino (2016) and Chan (2020) to estimate various large Bayesian VARs. More specifically, we exploit the Kronecker structure $\boldsymbol{\Sigma}_r\otimes\mathbf{K}_{\mathbf{A}}^{-1}$ to speed up computation. Consequently, the complexity of the problem can be drastically reduced to $\mathcal{O}(n^3p^3)$ operations. We further improve upon this approach by avoiding the computation of the inverse of the $np\times np$ matrix $\mathbf{K}_{\mathbf{A}}$. The computational details are provided in Appendix A.

Similarly, it can be shown that $(\mathbf{B},\boldsymbol{\Sigma}_c\,|\,\mathbf{Y},\mathbf{A},\boldsymbol{\Sigma}_r,\boldsymbol{\kappa},\boldsymbol{\omega})$ has a normal-inverse-Wishart

distribution with parameters $\nu_c + Tn$, $\widehat{\mathbf{S}}_c$, $\widehat{\mathbf{B}}$ and $\mathbf{K}_{\mathbf{B}}^{-1}$, where

$$\mathbf{K}_{\mathbf{B}} = \mathbf{V}_{\mathbf{B}}^{-1} + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{X}_t' \mathbf{A} \boldsymbol{\Sigma}_r^{-1} \mathbf{A}' \mathbf{X}_t, \quad \widehat{\mathbf{B}} = \mathbf{K}_{\mathbf{B}}^{-1} \left(\mathbf{V}_{\mathbf{B}}^{-1} \mathbf{B}_0 + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{X}_t' \mathbf{A} \boldsymbol{\Sigma}_r^{-1} \mathbf{Y}_t \right),$$

$$\widehat{\mathbf{S}}_c = \mathbf{S}_c + \mathbf{B}_0' \mathbf{V}_{\mathbf{B}}^{-1} \mathbf{B}_0 + \sum_{t=1}^{T} \omega_t^{-1} \mathbf{Y}_t' \boldsymbol{\Sigma}_r^{-1} \mathbf{Y}_t - \widehat{\mathbf{B}}' \mathbf{K}_{\mathbf{B}} \widehat{\mathbf{B}}.$$

Again, we can sample $(\mathbf{B}, \boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega})$ in two steps. First, we sample $\boldsymbol{\Sigma}_c$ marginally from $(\boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega}) \sim \mathcal{IW}(\widehat{\mathbf{S}}_c, \nu_c + Tn)$ with the normalization restriction that $\sigma_{c,1,1} = 1$. This can be done using the algorithm in Nobile (2000). Then, given the sampled $\boldsymbol{\Sigma}_c$, we simulate

$$(\mathrm{vec}(\mathbf{B}) \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega}) \sim \mathcal{N}\left(\mathrm{vec}(\widehat{\mathbf{B}}), \boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1}\right)$$

with the normalization restrictions that the (1,1) elements of $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are all 1. Sampling from a Gaussian distribution subjected to linear restrictions can be done efficiently by using Algorithm 2.6 in Rue and Held (2005) or Algorithm 2 in Cong, Chen, and Zhou (2017). We provide the details of this sampling step in Appendix A.

4 Extensions

The MAR modeling framework can be extended along a few directions. For example, one may consider a more flexible covariance matrix for $\mathrm{vec}(\mathbf{E}_t)$. More specifically, instead of the Kronecker product scaled by a single latent variable as specified in (3), we consider a generic $nk \times nk$ error covariance matrix $\boldsymbol{\Omega}_t$:

$$\mathrm{vec}(\mathbf{E}_t) \sim \mathcal{N}(\mathbf{0}_{nk}, \boldsymbol{\Omega}_t). \tag{9}$$

This extended MAR framework can incorporate a wide variety of more flexible multivariate stochastic volatility specifications. For example, the popular Cholesky stochastic volatility model of Cogley and Sargent (2005) can be represented as

$$\boldsymbol{\Omega}_t = \mathbf{B}_0^{-1} \mathbf{D}_t (\mathbf{B}_0^{-1})',$$

where $\mathbf{B}_0$ is an $nk \times nk$ lower triangular matrix with ones on the diagonal, $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{nk,t}})$, and the log-volatilities $h_{1,t}, \ldots, h_{nk,t}$ are specified as independent autoregressive processes. Naturally, various extensions of this model can also be entertained, such as the version in which $\mathbf{B}_0$ is time-varying (Primiceri, 2005); the order-invariant version in which $\mathbf{B}_0$ is a dense matrix (Chan, Koop, and Yu, 2024); and the outlier-augmented version developed by Carriero, Clark, Marcellino, and Mertens (2022).

Another popular class of multivariate stochastic volatility specifications is the family of factor stochastic volatility models (Pitt and Shephard, 1999; Aguilar and West, 2000; Chib, Nardari, and Shephard, 2006; Kastner, 2019), under which the error covariance matrix $\mathbf{\Omega}_t$ is constructed as

$$\mathbf{\Omega}_t = \mathbf{L}\mathbf{G}_t\mathbf{L}' + \mathbf{D}_t,$$

where $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{nk,t}})$ and $\mathbf{G}_t = \text{diag}(\text{e}^{h_{nk+1,t}}, \ldots, \text{e}^{h_{nk+r,t}})$, and $h_{1,t}, \ldots, h_{nk+r,t}$ follow independent AR(1) processes. Typically the number of factors r is set to be small relative to nk.

The error distribution specified in (9) with a generic covariance matrix $\mathbf{\Omega}_t$ is very flexible, but it comes at a cost of substantially more parameters and latent variables when n or k is large. This increases both the computational burden and posterior uncertainty. In particular, the conditional posterior covariance matrices of the MAR coefficient matrices $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)'$ and $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)'$ no longer have a Kronecker structure, when the error covariance matrix deviates from the form specified in (3). Consequently, sampling $\mathbf{A}$ and $\mathbf{B}$ would involve more operations than $\mathcal{O}(n^3)$ and $\mathcal{O}(k^3)$, respectively. Fortunately, using the equation-by-equation estimation approach developed in Carriero, Clark, and Marcellino (2019) and Carriero, Chan, Clark, and Marcellino (2022), sampling $\mathbf{A}$ and $\mathbf{B}$ can still be done in $\mathcal{O}(n^4)$ and $\mathcal{O}(k^4)$ elementary operations, respectively. Appendix A provides some technical details in estimating a BMAR with the factor stochastic volatility.

5 Assessing the MAR Restrictions

As discussed in Section 2, the MAR can be viewed as a special case of the VAR, where each of the VAR coefficient matrix is constructed via the Kronecker product $\mathbf{\Phi}_j = (\mathbf{B}_j \otimes \mathbf{A}_j), j = 1, \ldots, p$. This particular structure is designed to explicitly capture

the intertemporal correlations among the elements in the same row or column. Understandably, one might wish to empirically verify if this modeling choice is appropriate for a particular dataset. Under the Bayesian approach, this can be done using the marginal likelihood. More specifically, assessing these restrictions can be framed as a model comparison exercise comparing an MAR with an unrestricted VAR, and this boils down to computing the marginal likelihoods of the two models.

Below we conduct two Monte Carlo experiments to assess how the marginal likelihood performs in selecting the correct data generating process. In the first Monte Carlo experiment, we generate 500 datasets from a VAR. Each dataset consists of $nk = 50$ time-series, $T = 200$ observations and $p = 2$ lags. Given each dataset, we compute the log marginal likelihoods of the VAR and the MAR. For the VAR, each observation is treated as a 50×1 vector; for the MAR, it is arranged as a 5×10 matrix.[6] In the second experiment, we generate 500 datasets from the MAR with $n = 5, k = 10, T = 200$ and $p = 2$. Again, for each dataset, we compute the log marginal likelihoods of the MAR and the VAR (in the latter case we vectorize the 5×10 matrix observation).

For the VAR, we generate the intercepts from the uniform distribution on $(-10, 10)$, i.e., $\mathcal{U}(-10, 10)$. The diagonal elements of the first VAR coefficient matrix are drawn independently from $\mathcal{U}(0, 0.5)$ and the off-diagonal elements are from $\mathcal{U}(-0.2, 0.2)$; all elements of the second VAR coefficient matrix are independently drawn from $\mathcal{N}(0, 0.05^2)$. The error covariance matrix is generated from the inverse-Wistart distribution $\mathcal{IW}(nk + 5, 0.8 \times \mathbf{I}_{nk} + 0.2 \times \mathbf{1}_{nk}\mathbf{1}'_{nk})$, where $\mathbf{1}_{nk}$ is an $nk \times 1$ column of ones.

For the MAR, the diagonal elements of $\mathbf{A}_1$ are generated independently from $\mathcal{U}(0, 0.5)$ and the off-diagonal elements are from $\mathcal{U}(-0.2, 0.2)$; all elements of $\mathbf{A}_2$ are independently drawn from $\mathcal{N}(0, 0.05^2)$. The diagonal elements of $\mathbf{B}_1$ and $\mathbf{B}_2$ are generated independently from $\mathcal{U}(0.2, 1)$ and the off-diagonal elements are from $\mathcal{U}(0, 0.2)$. Finally, the error covariance matrices $\mathbf{\Sigma}_r$ and $\mathbf{\Sigma}_c$ are generated from $\mathcal{IW}(n + 5, \mathbf{I}_n)$ and $\mathcal{IW}(k + 5, \mathbf{I}_k)$, respectively.

Figure 1 reports the results of these two Monte Carlo experiments. The left panel presents the scatter plot of the log marginal likelihoods of the MAR against those of the VAR when

[6]To make the computation manageable, for the VAR we use the natural conjugate prior on the VAR coefficient matrices and the error covariance matrix. The advantage of the natural conjugate prior is that the associated marginal likelihood is analytically available and no MCMC is required.

the data generating process (DGP) is the VAR. Each point represents the log marginal likelihoods of the two models for each dataset. Since the DGP is the VAR, the log marginal likelihoods of the VAR should be larger than those of the MAR, and the points should lie below the diagonal line. This is indeed the case for all 500 datasets. The right panel reports a similar scatter plot when the DGP is the MAR. For this case, one would expect the points to be above the diagonal line since the log marginal likelihoods of the MAR should be larger. Again, the results show that this is true for all the replications. Overall, these Monte Carlo results show that the marginal likelihood is able to distinguish the two models even for a relatively moderate system ($nk = 50, T = 200$). Therefore, it can be used in practice to assess if the MAR restrictions are appropriate.

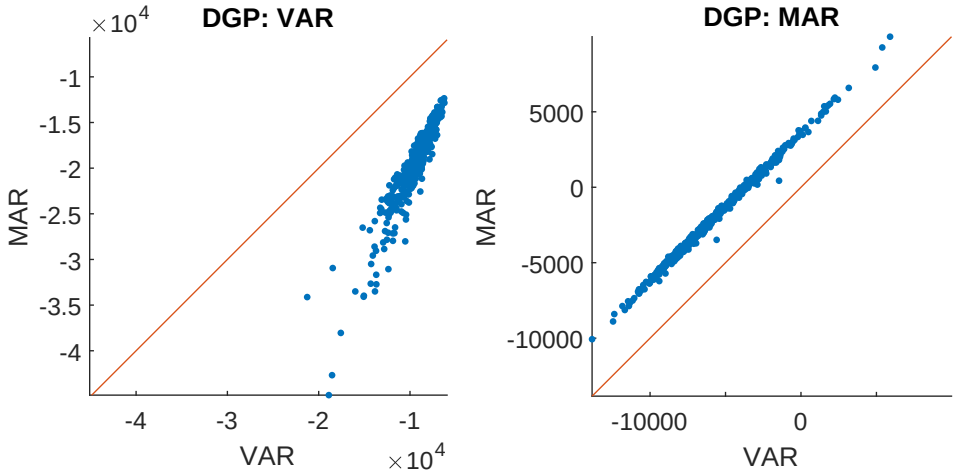

Figure 1: Scatter plots of the log marginal likelihoods of the MAR against those of the VAR when the DGP is the VAR (left panel) and when the DGP is the MAR (right panel).

Relatedly, the marginal likelihood can also be used to select the lag length in MARs, since the MAR has an equivalent VAR representation and it is widely established that the marginal likelihood works well for selecting lag length in VARs. An alternative model selection criterion is the Bayesian information criterion, which can be viewed as the Laplace's approximation to the marginal likelihood.

6 Empirical Application

To illustrate the utility of the proposed models and estimation methods, we consider an application that involves a US state-level dataset. More specifically, for each of the 50 US states, we obtain 6 quarterly time-series sourced from the Bureau of Labor Statistics and the FRED database maintained by the Federal Reserve Bank of St. Louis. These 6 variables are initial unemployment insurance claims, continued unemployment insurance claims, total nonfarm employment, unemployment rate, new housing permits, and real home price index. The sample period is from 1991Q1 to 2023Q1. A detailed description of the variables and their transformations are provided in Appendix B. We represent the data at time t as an $n \times k$ matrix $\mathbf{Y}_t$, where the columns refer to the $k = 50$ states and the rows are the $n = 6$ variables. We first report the full sample estimates in Subsection 6.1. We then present results from a recursive out-of-sample forecasting exercise in Subsection 6.2.

6.1 Full Sample Results

As discussed in Section 2, the MAR has an equivalent VAR representation in which the VAR coefficient matrices have the form $\mathbf{\Phi}_j = (\mathbf{B}_j \otimes \mathbf{A}_j), j = 1, \ldots, p$. We first verify that these restrictions implied by the MAR are appropriate for our state-level dataset. For this purpose, we conduct a model comparison exercise comparing a BMAR against an unrestricted BVAR as outlined in Section 5. The log marginal likelihood of the BMAR is 2461, whereas the value for the unrestricted BVAR is -8631. These results show that there is strong empirical support for the Kronecker product structure.

Next, we report various estimates of interest using the full sample, which includes the COVID-19 pandemic. As widely noted, the COVID-19 pandemic has caused extreme movements in many macroeconomic and financial time-series, and a failure to account for these outliers would result in heavily distorted parameter estimates, as demonstrated in recent papers such as Schorfheide and Song (2021), Lenza and Primiceri (2022) and Bobeica and Hartwig (2023). Therefore, we consider two BMARs that explicitly account for time-varying volatility or potential outliers.

More specifically, both models can be nested within the proposed framework and rep-

resented as the system in (2)–(3). The first Bayesian MAR incorporates the common stochastic volatility specification proposed in Carriero, Clark, and Marcellino (2016) with $\omega_t = \mathrm{e}^{h_t}$, and the log-volatility h_t follows the stationary AR(1) given in (4). This model is referred to as BMAR-CSV. The second model includes the outlier component introduced in Stock and Watson (2016) with $\omega_t = o_t^2$, where o_t follows a 2-part distribution with a point mass at 1 and a uniform distribution on the interval $(2, 20)$. This latter model is referred to as BMAR-O. For both models we set the lag order to be $p = 2$.

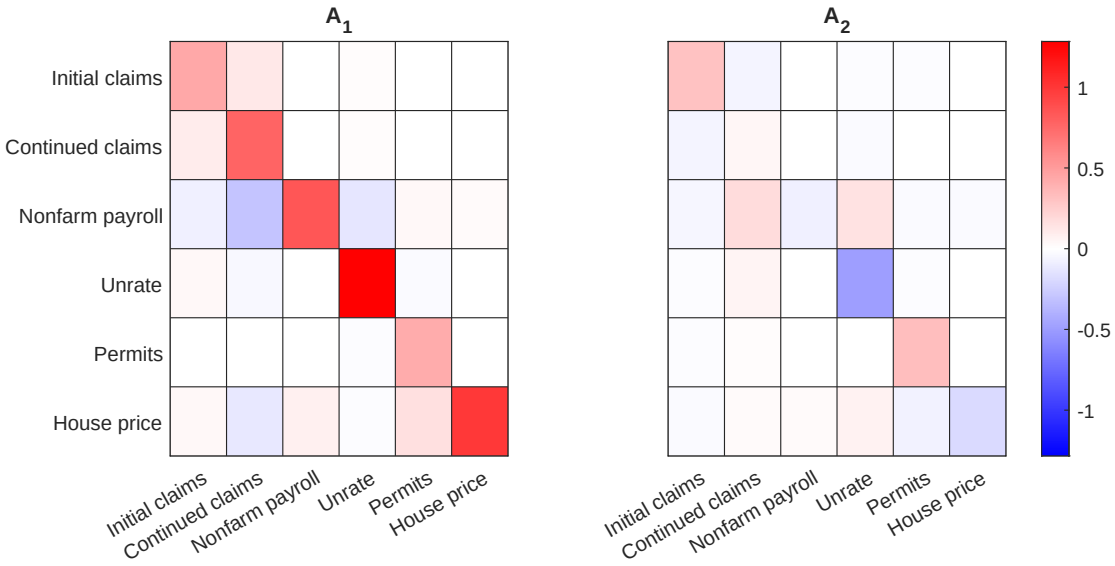

Figure 2: Heatmaps of the posterior means of $\mathbf{A}_1$ and $\mathbf{A}_2$ from the BMAR-CSV.

To visualize the correlation pattern among the variables (rows), we report in Figure 2 heatmaps of the posterior means of $\mathbf{A}_1$ and $\mathbf{A}_2$ from the BMAR-CSV, where red entries denote positive values, blue negative and white zero. First, it is clear that, as expected, all variables are highly persistent. For example, the AR(1) coefficient for nonfarm payroll is estimated to be about 0.85, and the AR(1) and AR(2) coefficient estimates for unemployment rate are, respectively, 1.28 and -0.49. Second, the proposed hierarchical shrinkage prior on $\mathbf{A} = (\mathbf{A}_1, \mathbf{A}_2)'$ strongly shrinks many of the off-diagonal elements to zero. In particular, the shrinkage hyperparameter $\kappa_{\mathbf{A}}$—that controls the overall shrinkage strength on $\mathbf{A}$, where a smaller value indicate more aggressive shrinkage to zero—is estimated to be 0.12 (compared to the prior mean of 1). Despite the shrinkage effects, many off-diagonal elements corresponding to closely related variables remain non-zero.

For example, lagged initial claims have non-negligible impacts on continued claims, and vice versa, reflecting current labor market conditions.

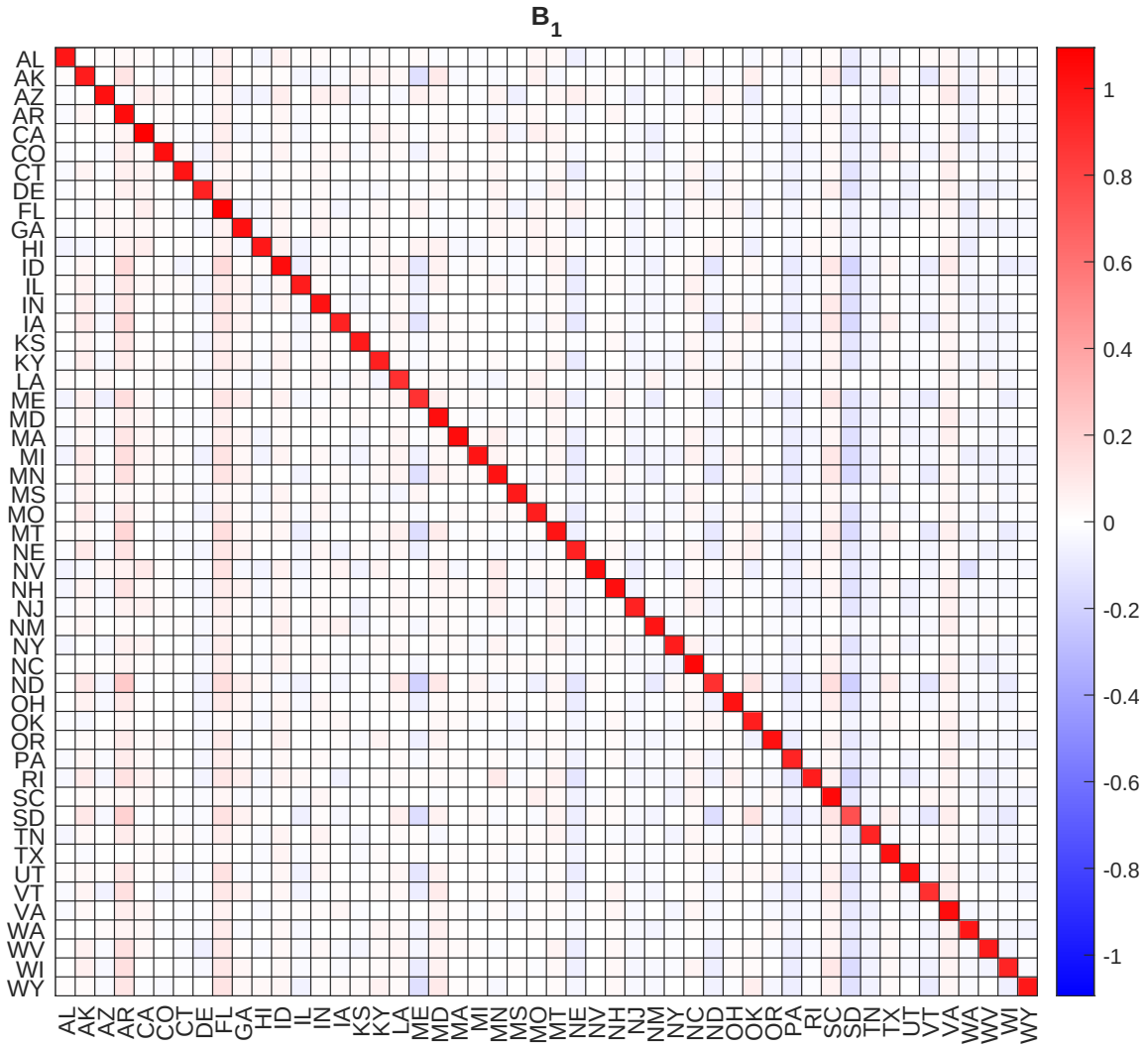

Figure 3: A heatmap of the posterior mean of $\mathbf{B}_1$ from the BMAR-CSV.

Next, Figure 3 reports a heatmap of the posterior mean of $\mathbf{B}_1$ from the BMAR-CSV, which represents the correlation structure among the states (columns). The hierarchical shrinkage prior on $\mathbf{B} = (\mathbf{B}_1, \mathbf{B}_2)'$ strongly shrinks both $\mathbf{B}_1$ and $\mathbf{B}_2$ to the identity matrix; the shrinkage hyperparameter $\kappa_{\mathbf{B}}$ is estimated to be 0.0084 (compared to the prior mean of 1). Even so, some of the diagonal elements of $\mathbf{B}_1$ are substantially different from one and many of the off-diagonal elements are estimated to be non-zero. In particular, the diagonal elements range from 0.75 to 1.1, indicating heterogeneity in variables dynamics across

states. Not surprisingly, many neighboring states or states with similar types of industries show stronger interactions. For instance, lagged variables of Arkansas and Oklahoma most positively impact the variables of South Dakota with corresponding coefficients estimated to be 0.2 and 0.12, respectively.[7] Overall, these results highlight the strong interactions between the states.

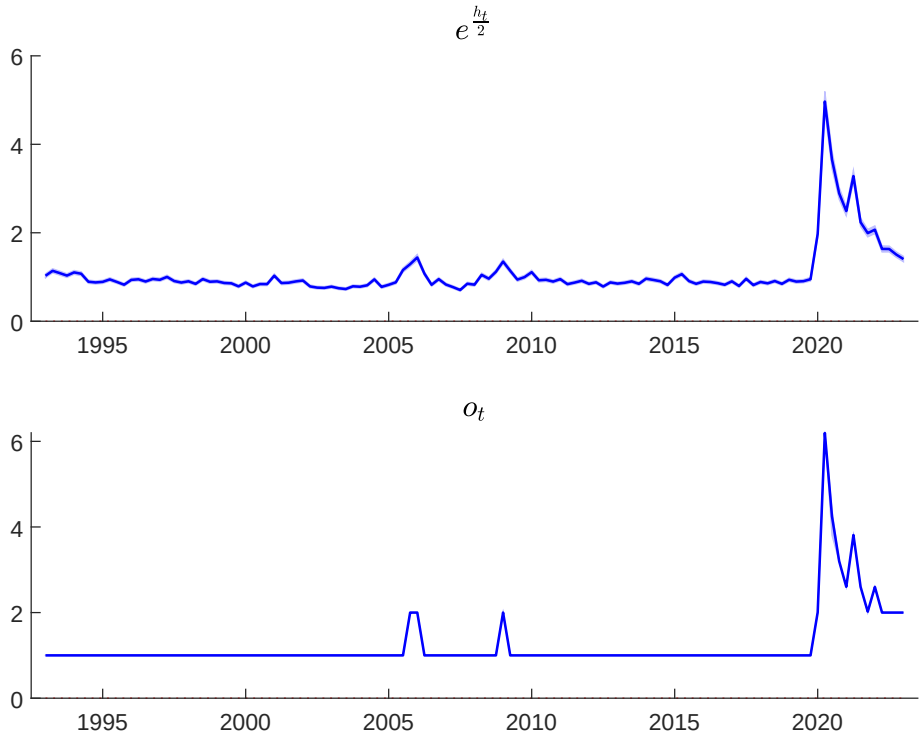

Figure 4: Posterior means of the time-varying standard derivations, $\exp(h_t/2)$ and o_t respectively, from BMAR-CSV (top panel) and BMAR-O (bottom panel).

To assess the extent of time-varying volatility in our sample, we report the posterior means of the error standard derivations from the BMAR-CSV and the BMAR-O in Figure 4. Despite the two very different modeling approaches—the BMAR-CSV prescribes a persistent volatility process whereas the BMAR-O assumes serial independence of the occurrence of outliers—the estimated error standard derivations from the two models are

[7]The diagonal element of $\mathbf{B}_1$ corresponding to South Dakota is estimated to be 0.75. Hence, the impact magnitudes of the lagged variables of Arkansas and Oklahoma are 27% and 16% of those of own state variables.

remarkably similar. In particular, for most of the sample before the onset of the COVID-19 pandemic in 2020Q2, the standard derivations were mostly around 1 (normalized as the size of 'regular' observations). In 2020Q2, the standard derivations jumped to 5 for the BMAR-CSV and 6 for the BMAR-O, and they stayed elevated afterward. These results underscore the empirical relevance of explicitly modeling time-varying volatility or allowing for outliers.

6.2 Forecasting Results

Next, we conduct a recursive out-of-sample forecasting exercise to evaluate the performance of the proposed BMARs. We forecast the 6 quarterly macroeconomic variables for each of the 50 US states. The evaluation period starts from 2000Q1 and ends in 2023Q1. We consider a homoskedastic BMAR with $\omega_t = 1$ (BMAR), BMARs with the common stochastic volatility (BMAR-CSV), an explicit outlier component (BMAR-O) and t errors (BMAR-t). We also include a version with a more general error covariance matrix, namely, the factor stochastic volatility (BMAR-FSV) as specified in Section 4. We set the number of factors to be 6 and they are identified via the 6 macroeconomic variables. To estimate these BMARs, the observations at time t are represented as an $n \times k$ matrix $\mathbf{Y}_t$, where each column denotes a state and each row a macroeconomic variable.

As a comparison, we also consider two dynamic factor models. For identification purposes, the factor loadings matrix is assumed to be lower triangular with ones on the diagonal. For the first dynamic factor model, we set the number of factors to be 6, and stack the observations column by column as $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$. Hence, the factors are identified via the macroeconomic variables. This model is referred to as DFM-6. For the second dynamic factor model, we set the number of factors to be 50 and stack the observations row by row, i.e., $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t')$. In this case the factors are associated with the states. This version is referred to as DFM-50.

To evaluate the performance of jointly forecasting all 300 time-series, we compute the average of log predictive likelihoods for each model over the forecast horizons of one- and four-quarter-ahead. The results are reported in Table 1. A larger value of log predictive likelihood indicates better forecast performance.

Table 1: Joint forecast performance of the proposed BMARs as well as two dynamic factor models with 6 and 50 factors.

	BMAR	BMAR-CSV	BMAR-O	BMAR-t	BMAR-FSV	DFM-6	DFM-50
One-quarter-ahead	−161	22	31	−16	−701	−445	−1,887
Four-quarter-ahead	−2,492	−1,534	−351	−1,110	−4,017	−818	−4,231

The results show that the proposed BMARs perform well compared to the dynamic factor models, especially those with a more structured error covariance matrix. Specifically, BMAR, BMAR-CSV, BMAR-O and BMAR-t all substantially outperform dynamic factor models, and they all assume an error covariance matrix that has a parsimonious Kronecker product structure. In contrast, BMAR-FSV has a much more flexible error covariance matrix constructed from $nk + r = 306$ stochastic volatility processes. But this additional flexibility does not seem to improve density forecast performance.[8] These results suggest that for modeling a large number of state-level variables, the specification in (3)—that assumes a Kronecker product structure for the error covariance matrix—strikes the right balance between flexibility and parsimony.

Among the 4 parsimonious BMARs, the versions with time-varying volatility or heavier tails than Gaussian do substantially better than the homoskedastic Gaussian BMAR. This result is in line with the large empirical literature that demonstrates the superior forecast performance of models with stochastic volatility and heavy tails. Lastly, BMAR-O with an outlier component is the best performing model, highlighting the value of explicitly modeling potential outliers when the sample includes the COVID-19 pandemic.

Next, we look at the point forecast performance for each of the macroeconomic variables. More specifically, we compute the root mean square error for each variable, summing over the 50 states and the evaluation periods. For easy comparison, we report the point forecast performance relative to DFM-6 (which outperforms DFM-50 in the joint forecasting exercise). The results are reported in Tables 2 and 3. Values less than one indicate better forecast performance than the benchmark.

[8]To corroborate this finding, we also investigate the magnitude of the common component in the stochastic volatilities using a similar analysis conducted in Carriero, Clark, and Marcellino (2016). More specifically, we separately fit each of the 300 times-series with an AR(4) model with stochastic volatility. We then compute the first principal component of these 300 estimated stochastic volatility series. We find that the first principal component explains about 74% of the variations, suggesting a dominant common component in the stochastic volatility of the state-level variables.

Table 2: One-quarter-ahead point forecast performance relative to the dynamic factor model with 6 factors.

	BMAR	BMAR-CSV	BMAR-O	BMAR-t	BMAR-FSV	DFM-50
Initial claims	0.99	0.98	0.98	0.98	0.94	0.95
Continued claims	0.94	0.94	0.94	0.94	1.02	1.06
Nonfarm payroll	1.00	0.94	0.93	0.94	0.94	0.94
Unemployment rate	1.19	1.11	1.11	1.11	1.09	0.95
New housing permits	0.86	0.87	0.87	0.87	0.82	0.81
House price	0.74	0.75	0.74	0.75	0.68	0.85

Table 3: Four-quarter-ahead point forecast performance relative to the dynamic factor model with 6 factors.

	BMAR	BMAR-CSV	BMAR-O	BMAR-t	BMAR-FSV	DFM-50
Initial claims	0.98	0.97	0.97	0.97	1.00	1.02
Continued claims	1.01	1.01	1.01	1.01	1.00	1.01
Nonfarm payroll	0.99	0.97	0.97	0.97	0.92	0.98
Unemployment rate	1.03	1.01	1.01	1.01	0.97	0.98
New housing permits	0.93	0.93	0.93	0.93	0.95	0.96
House price	0.98	0.98	0.99	0.98	0.94	0.97

Compared to the joint forecasting exercise, these point forecasting results are more mixed. While the proposed BMARs tend to provide better point forecasts for the majority of the variables compared to DFM-6, DFM-50 performs well in a number of variables. Among the proposed BMARs, BMAR-CSV, BMAR-O, BMAR-t and BMAR-FSV provide very similar point forecasts, and they tend to outperform the homoskedastic Gaussian version. Overall, these results show that the proposed BMARs are competitive relative to dynamic factor models. In addition, point forecast performance can often be improved by allowing for time-varying volatility or an explicit outlier component.

7 Concluding Remarks and Future Research

Two recent developments have motivated our paper: 1) the increasing recognition of the need to allow for flexible time-varying features in modeling most macroeconomic datasets; and 2) the growing availability of a large number of matrix-valued time-series. In response to these developments, we have introduced a new class of matrix autoregressions that

can accommodate time-varying volatility, non-Gaussian errors and COVID-19 outliers. We then developed an efficient, unified approach that scales well to high-dimensional datasets. We illustrated the methodology using a US state-level dataset that involves 300 macroeconomic time-series.

In terms of methodological development, there are multiple lines of future research that are worth pursuing. First, it would be useful to extend the MARs to a mixed-frequency framework—e.g., modeling both quarterly and monthly time-series simultaneously. This can be done, for example, by incorporating the data augmentation approach in Schorfheide and Song (2015) or Chan, Poon, and Zhu (2023) to simulate the missing monthly observations of the quarterly data. An interesting application would be one that aims to construct monthly state-level GDP estimates using both quarterly and monthly variables. Another promising direction is to develop time-varying parameter MARs. In a VAR setting, Chan (2023b) has found evidence that the VAR coefficients in some, but not all, equations are time-varying. The binary indicator approach in Chan (2023b) can be adopted to model the time-varying MAR coefficients. Alternatively, both the dynamic shrinkage approach of Koop and Korobilis (2018) and the dynamic shrinkage with sparsification approach of Huber, Koop, and Onorante (2019) are also promising.

While our application focuses on modeling regional macroeconomic variables, the proposed modeling framework is also well suited for high-dimensional dynamic economic networks, such as financial network or international trade network. For instance, bilateral import/export volumes between countries can be naturally represented as a matrix (e.g., the (i, j) element of $\mathbf{Y}_t$ is the export volume from country i to country j at time t). One complication in such applications is that the diagonal elements of the matrix are missing. In a recent working paper, Chen, Chen, Bolivar, and Chen (2024) use a matrix factor model to study the the pattern and evolution of international trade flow among 24 countries and regions, and they replace the diagonal missing elements with zeros. It would be interesting to investigate the effects of such an imputation scheme or if alternative methods would be more appropriate (e.g., replacing the diagonal elements with GDP values).

Appendix A: Estimation Details

In this appendix we provide estimation details of the proposed Bayesian matrix autoregressions and their extensions. Specifically, Subsections A1-A3 give estimation details of the baseline BMAR as specified by (2) and (3). Subsection A4 provides some technical details for estimating a BMAR with the factor stochastic volatility.

A1: Sampling A and B

For sampling the coefficient matrices $\mathbf{A}$ and $\mathbf{B}$—of dimensions $np \times n$ and $kp \times k$, respectively—from their full conditional distributions, we make use of some standard results on the matrix normal distribution (see, e.g., Bauwens, Lubrano, and Richard, 1999, pp. 301-302). Specifically, an $r \times s$ random matrix $\mathbf{Z}$ is said to have a matrix normal distribution $\mathcal{MN}(\mathbf{M}, \mathbf{S} \otimes \mathbf{R})$ for covariance matrices $\mathbf{R}$ and $\mathbf{S}$ of dimensions $r \times r$ and $s \times s$, respectively, if and only if $\text{vec}(\mathbf{Z}) \sim \mathcal{N}(\text{vec}(\mathbf{M}), \mathbf{S} \otimes \mathbf{R})$. Naturally, a bilinear transformation of a matrix normal random matrix followed by a translation is also a matrix normal random matrix. More precisely, suppose $\mathbf{Z} \sim \mathcal{MN}(\mathbf{M}, \mathbf{S} \otimes \mathbf{R})$ and let $\mathbf{V} = \mathbf{CZD} + \mathbf{E}$. Then, $\mathbf{V} \sim \mathcal{MN}(\mathbf{CMD} + \mathbf{E}, (\mathbf{D}'\mathbf{SD}) \otimes (\mathbf{CRC}'))$.

Now, we can sample $\text{vec}(\mathbf{A}) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}), \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}}^{-1})$ as follows. Let $\mathbf{C}_{\mathbf{K_A}}$ and $\mathbf{C}_{\mathbf{\Sigma}_r}$ be the lower Cholesky factors of $\mathbf{K_A}$ and $\mathbf{\Sigma}_r$, respectively. We claim that if we construct

$$\mathbf{A} = \widehat{\mathbf{A}} + \mathbf{C}_{\mathbf{K_A}}^{-1'} \mathbf{Z} \mathbf{C}'_{\mathbf{\Sigma}_r},$$

where $\mathbf{Z}$ is an $np \times n$ matrix of iid $\mathcal{N}(0,1)$ random variables, then $\text{vec}(\mathbf{A})$ has the desired distribution. To show that, since $\mathbf{Z} \sim \mathcal{MN}(\mathbf{0}, \mathbf{I}_n \otimes \mathbf{I}_{np})$, using the previous result with $\mathbf{C} = \mathbf{C}_{\mathbf{K_A}}^{-1'}$, $\mathbf{D} = \mathbf{C}'_{\mathbf{\Sigma}_r}$ and $\mathbf{E} = \widehat{\mathbf{A}}$, we have $\mathbf{A} \sim \mathcal{MN}(\widehat{\mathbf{A}}, \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}}^{-1})$ and therefore, by definition, $\text{vec}(\mathbf{A}) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{A}}), \mathbf{\Sigma}_r \otimes \mathbf{K}_{\mathbf{A}}^{-1})$. Finally, we note that in the above construction, one can efficiently compute $\mathbf{C}_{\mathbf{K_A}}^{-1'} \mathbf{Z}$ by solving the linear system $\mathbf{C}'_{\mathbf{K_A}} \mathbf{X} = \mathbf{Z}$ for $\mathbf{X}$ without explicitly obtaining the inverse $\mathbf{C}_{\mathbf{K_A}}^{-1'}$.

Next, we outline the sampling of $\text{vec}(\mathbf{B}) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{B}}), \mathbf{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1})$ with the normalization restrictions that the $(1,1)$ elements of $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are all one. To that end, we first represent the normalization restrictions as a system of p linear restrictions: $\mathbf{M}\,\text{vec}(\mathbf{B}) =$

$\mathbf{b}_0$, where $\mathbf{M} = (m_{i,j})$ is a $p \times k^2 p$ selection matrix with $m_{i,(i-1)k+1} = 1$ for $i = 1, \ldots, p$, and $\mathbf{b}_0$ is a $p \times 1$ vector of ones. Then, we can apply Algorithm 2.6 in Rue and Held (2005) or Algorithm 2 in Cong, Chen, and Zhou (2017) to efficiently sample $\text{vec}(\mathbf{B}) \sim \mathcal{N}(\text{vec}(\widehat{\mathbf{B}}), \boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1})$ so that $\mathbf{M}\,\text{vec}(\mathbf{B}) = \mathbf{b}_0$. In particular, one can first sample $\text{vec}(\mathbf{B}_u)$ from the unconstrained conditional posterior distribution using the algorithm discussed earlier, and construct

$$\text{vec}(\mathbf{B}) = \text{vec}(\mathbf{B}_u) + \left(\boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1}\right) \mathbf{M}' \left(\mathbf{M}\left(\boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1}\right) \mathbf{M}'\right)^{-1} (\mathbf{b}_0 - \mathbf{M}\,\text{vec}(\mathbf{B}_u)).$$

Then, $\text{vec}(\mathbf{B})$ has the distribution $\mathcal{N}(\text{vec}(\widehat{\mathbf{B}}), \boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1})$ such that $\mathbf{M}\,\text{vec}(\mathbf{B}) = \mathbf{b}_0$. We summarize the algorithm in Algorithm 1.

Algorithm 1 Sampling $\mathcal{N}(\text{vec}(\widehat{\mathbf{B}}), \boldsymbol{\Sigma}_c \otimes \mathbf{K}_{\mathbf{B}}^{-1})$ such that $\mathbf{M}\,\text{vec}(\mathbf{B}) = \mathbf{b}_0$.

1. Sample $\mathbf{B}_u = \widehat{\mathbf{B}} + \mathbf{C}_{\mathbf{K}_\mathbf{B}}^{-1'} \mathbf{Z} \mathbf{C}_{\boldsymbol{\Sigma}_c}'$, where $\mathbf{Z}$ is a $kp \times k$ matrix of $\mathcal{N}(0,1)$ random variables.
2. Compute $\mathbf{C} = \mathbf{C}_{\boldsymbol{\Sigma}_c^{-1}} \otimes \mathbf{C}_{\mathbf{K}_\mathbf{B}}$, where $\mathbf{C}_{\boldsymbol{\Sigma}_c^{-1}}$ is the lower Cholesky factor of $\boldsymbol{\Sigma}_c^{-1}$.
3. Solve $\mathbf{C}\mathbf{C}'\mathbf{U} = \mathbf{M}'$ for $\mathbf{U}$
4. Solve $\mathbf{M}\mathbf{U}\mathbf{V} = \mathbf{U}'$ for $\mathbf{V}$.
5. Return $\text{vec}(\mathbf{B}) = \text{vec}(\mathbf{B}_u) + \mathbf{V}'(\mathbf{b}_0 - \mathbf{M}\,\text{vec}(\mathbf{B}_u))$.

A2: Sampling $\kappa_\mathbf{A}$ and $\kappa_\mathbf{B}$

Next, we discuss the sampling steps of drawing the hyperparameters $\kappa_\mathbf{A}$ and $\kappa_\mathbf{B}$. First, note that $\kappa_\mathbf{A}$ only appears in two terms: its gamma prior $\kappa_\mathbf{A} \sim \mathcal{G}(c_{\mathbf{A},1}, c_{\mathbf{A},2})$ and $\mathbf{V}_\mathbf{A}$, the prior covariance matrix of $\mathbf{A}$, which is an $np \times np$ diagonal matrix with the i-th diagonal element $v_{\mathbf{A},i,i} = \kappa_\mathbf{A} C_{\mathbf{A},i}$ for some constant $C_{\mathbf{A},i}$. Then, we can express the conditional

distribution of $\kappa_{\mathbf{A}}$ as

$$\begin{aligned} p(\kappa_{\mathbf{A}} \,|\, \mathbf{A}, \boldsymbol{\Sigma}_r) &\propto \kappa_{\mathbf{A}}^{c_{\mathbf{A},1}-1} \mathrm{e}^{-c_{\mathbf{A},2}\kappa_{\mathbf{A}}} \times |\mathbf{V}_{\mathbf{A}}|^{-\frac{n}{2}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\boldsymbol{\Sigma}_r^{-1}(\mathbf{A}-\mathbf{A}_0)'\mathbf{V}_{\mathbf{A}}^{-1}(\mathbf{A}-\mathbf{A}_0)\right)} \\ &\propto \kappa_{\mathbf{A}}^{c_{\mathbf{A},1}-\frac{n^2p}{2}-1} \mathrm{e}^{-c_{\mathbf{A},2}\kappa_{\mathbf{A}}} \mathrm{e}^{-\frac{1}{2}\mathrm{tr}\left(\mathbf{V}_{\mathbf{A}}^{-1}(\mathbf{A}-\mathbf{A}_0)\boldsymbol{\Sigma}_r^{-1}(\mathbf{A}-\mathbf{A}_0)'\right)} \\ &\propto \kappa_{\mathbf{A}}^{c_{\mathbf{A},1}-\frac{n^2p}{2}-1} \mathrm{e}^{-\frac{1}{2}\left(2c_{\mathbf{A},2}\kappa_{\mathbf{A}}+\kappa_{\mathbf{A}}^{-1}\sum_{i=1}^{np} Q_{\mathbf{A},i}/C_{\mathbf{A},i}\right)}, \end{aligned}$$

where $Q_{\mathbf{A},i}$ is the i-th diagonal element of $\mathbf{Q}_{\mathbf{A}} = (\mathbf{A}-\mathbf{A}_0)\boldsymbol{\Sigma}_r^{-1}(\mathbf{A}-\mathbf{A}_0)'$. Note that this is the kernel of the generalized inverse Gaussian distribution

$$\mathcal{GIG}\left(c_{\mathbf{A},1} - \frac{n^2p}{2}, 2c_{\mathbf{A},2}, \sum_{i=1}^{np} Q_{\mathbf{A},i}/C_{\mathbf{A},i}\right).$$

Draws from the generalized inverse Gaussian distribution can be obtained using the algorithm in Devroye (2014).

Similarly, $\kappa_{\mathbf{B}}$ only appears in its gamma prior $\kappa_{\mathbf{B}} \sim \mathcal{G}(c_{\mathbf{B},1}, c_{\mathbf{B},2})$ and $\mathbf{V}_{\mathbf{B}}$, which is a $kp \times kp$ diagonal matrix where the i-th diagonal element is $v_{\mathbf{B},i,i} = \kappa_{\mathbf{B}} C_{\mathbf{B},i}$ for some constant $C_{\mathbf{B},i}$. It can be shown that $(\kappa_{\mathbf{B}} \,|\, \mathbf{B}, \boldsymbol{\Sigma}_c)$ has the generalized inverse Gaussian distribution:

$$\mathcal{GIG}\left(c_{\mathbf{B},1} - \frac{k^2p}{2}, 2c_{\mathbf{B},2}, \sum_{i=1}^{kp} Q_{\mathbf{B},i}/C_{\mathbf{B},i}\right),$$

where $Q_{\mathbf{B},i}$ is the i-th diagonal element of $\mathbf{Q}_{\mathbf{B}} = (\mathbf{B}-\mathbf{B}_0)\boldsymbol{\Sigma}_c^{-1}(\mathbf{B}-\mathbf{B}_0)'$.

A3: Sampling Other Parameters

We now consider a few specific examples of $\boldsymbol{\omega}$ and discuss how one can modify the posterior sampler outlined in the main text to handle each case.

Example 1 Student's t errors

As discussed in Section 2 of the main text, the case of t distributed errors is nested within the proposed framework: the latent variables $\boldsymbol{\omega} = (\omega_1, \ldots, \omega_T)'$ are distributed independently as $(\omega_t \,|\, \nu) \sim \mathcal{IG}(\nu/2, \nu/2)$.

Posterior draws can be obtained by sequentially sampling from: 1) $p(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$;

2) $p(\mathbf{B}, \boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 3) $p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\omega})$; 4) $p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \nu)$; and 5) $p(\nu \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$. Steps 1-2 can be implemented exactly as described in Section 3 of the main text and Step 3 as outlined in Section A2. For Step 4, let $s_t^2 = \text{tr}\left(\boldsymbol{\Sigma}_c^{-1}\mathbf{E}_t'\boldsymbol{\Sigma}_r^{-1}\mathbf{E}_t\right)$, where $\mathbf{E}_t$ can be computed given $\mathbf{Y}_t, \mathbf{A}$ and $\mathbf{B}$ using (2). Then, the conditional distribution of $\boldsymbol{\omega}$ can be expressed as:

$$\begin{aligned} p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \nu) &= \prod_{t=1}^{T} p(\omega_t \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \nu) \\ &\propto \prod_{t=1}^{T} \omega_t^{-(\frac{\nu}{2}+1)} \mathrm{e}^{-\frac{\nu}{2\omega_t}} \times \omega_t^{-\frac{nk}{2}} \mathrm{e}^{-\frac{1}{2\omega_t}s_t^2}. \end{aligned}$$

That is, each ω_t is conditionally independent given the data and other parameters, and has an inverse-gamma distribution: $(\omega_t \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \nu) \sim \mathcal{IG}((nk+\nu)/2, (s_t^2+\nu)/2)$.

Lastly, ν can be sampled by an independence-chain Metropolis-Hastings step with the proposal distribution $\mathcal{N}(\widehat{\nu}, K_\nu^{-1})$, where $\widehat{\nu}$ is the mode of $\log p(\nu \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\omega})$ and K_ν is the negative Hessian evaluated at the mode. For implementation details of this step, see Chan and Hsiao (2014).

Example 2 Outlier detection

The proposed framework can also be used to incorporate the approach in Stock and Watson (2016) and Carriero, Clark, Marcellino, and Mertens (2022) to handle potential outliers. To that end, let $\omega_t = o_t^2$, where o_t follows a mixture distribution that distinguishes between regular observations with $o_t = 1$ and outliers for which $o_t > 2$. More specifically, o_t equals 1 with probability $1 - p_o$; o_t follows a uniform distribution on $(2, 20)$ with probability p_o. The outlier probability p_o is assumed to have a beta prior $\mathcal{B}(a_0, b_0)$, where the hyperparameters a_0 and b_0 are calibrated so that the mean outlier frequency is once every 4 years in quarterly data.

Posterior draws can then be obtained by sequentially sampling from: 1) $p(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 2) $p(\mathbf{B}, \boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 3) $p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\omega}, p_o)$; 4) $p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, p_o)$; and 5) $p(p_o \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$. Steps 1-3 remain the same as before. To implement Step 4, we discretize the distribution using a fine grid as proposed in Stock and Watson (2016). Consequently, each o_t follows a discrete distribution that can be easily sampled

from. In particular, we have

$$p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, p_o) = \prod_{t=1}^{T} p(o_t \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, p_o) \propto \prod_{t=1}^{T} p(o_t \,|\, p_o) o_t^{-nk} \mathrm{e}^{-\frac{s_t^2}{2o_t^2}},$$

where $p(o_t \,|\, p_o)$ is the prior for o_t and $s_t^2 = \operatorname{tr}\left(\boldsymbol{\Sigma}_c^{-1}\mathbf{E}_t'\boldsymbol{\Sigma}_r^{-1}\mathbf{E}_t\right)$. Hence, we can sample each ω_t from its discrete distribution. Lastly, Step 5 can be implemented easily as p_o follows the following beta distribution

$$(p_o \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega}) \sim \mathcal{B}(a_0 + n_o, b_0 + T - n_o),$$

where $n_o = \sum_{t=1}^{T} 1(o_t > 1)$ is the number of outliers.

Example 3 The common stochastic volatility

Next, we incorporate the common stochastic volatility introduced in Carriero, Clark, and Marcellino (2016) to our matrix autoregression with $\omega_t = \mathrm{e}^{h_t}$, where h_t follows an AR(1) process: $h_t = \phi h_{t-1} + \varepsilon_t^h$, where $\varepsilon_t^h \sim \mathcal{N}(0, \sigma_h^2)$. We assume independent truncated normal and inverse-gamma priors for ϕ and σ_h^2: $\phi \sim \mathcal{N}(\phi_0, V_\phi)1(|\phi| < 1)$ and $\sigma_h^2 \sim \mathcal{IG}(\nu_h, S_h)$.

Then, posterior draws can be obtained by sampling from: 1) $p(\mathbf{A}, \boldsymbol{\Sigma}_r \,|\, \mathbf{Y}, \mathbf{B}, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 2) $p(\mathbf{B}, \boldsymbol{\Sigma}_c \,|\, \mathbf{Y}, \mathbf{A}, \boldsymbol{\Sigma}_r, \boldsymbol{\kappa}, \boldsymbol{\omega})$; 3) $p(\boldsymbol{\kappa} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\omega}, \phi, \sigma_h^2)$; 4) $p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \phi, \sigma_h^2)$; 5) $p(\phi \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \Sigma_c, \boldsymbol{\kappa}, \boldsymbol{\omega}, \sigma_h^2)$; and 6) $p(\sigma_h^2 \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \Sigma_c, \boldsymbol{\kappa}, \boldsymbol{\omega}, \phi)$.

Steps 1-2 again can be implemented exactly as described in Section 3 of the main text and Step 3 as outlined in Section A2. For Step 4, we follow the approach outlined in Chan (2017). Specifically, note that

$$\begin{aligned} p(\boldsymbol{\omega} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \phi, \sigma_h^2) &= p(\mathbf{h} \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, \boldsymbol{\kappa}, \phi, \sigma_h^2) \\ &\propto p(\mathbf{h} \,|\, \phi, \sigma_h^2) \prod_{t=1}^{T} p(\mathbf{Y}_t \,|\, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, h_t), \end{aligned}$$

where $p(\mathbf{h} \,|\, \phi, \sigma_h^2)$ is a Gaussian density implied by the state equation,

$$\log p(\mathbf{Y}_t \,|\, \mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_r, \boldsymbol{\Sigma}_c, h_t) = c_t - \frac{nk}{2} h_t - \frac{1}{2} \mathrm{e}^{-h_t} s_t^2$$

and c_t is a normalizing constant that does not dependent on h_t and $s_t^2 = \operatorname{tr}\left(\boldsymbol{\Sigma}_c^{-1}\mathbf{E}_t'\boldsymbol{\Sigma}_r^{-1}\mathbf{E}_t\right)$.

It is easy to check that

$$\frac{\partial}{\partial h_t}\log p(\mathbf{Y}_t \mid \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c, h_t) = -\frac{nk}{2} + \frac{1}{2}\mathrm{e}^{-h_t}s_t^2,$$
$$\frac{\partial^2}{\partial h_t^2}\log p(\mathbf{Y}_t \mid \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c, h_t) = -\frac{1}{2}\mathrm{e}^{-h_t}s_t^2.$$

Then, one can implement a Newton-Raphson algorithm to obtain the mode of the log density $\log p(\mathbf{h} \mid \mathbf{Y}, \mathbf{A}, \mathbf{B}, \mathbf{\Sigma}_r, \mathbf{\Sigma}_c, \phi, \sigma_h^2)$ and the negative Hessian evaluated at the mode, which are denoted as $\widehat{\mathbf{h}}$ and $\mathbf{K}_{\mathbf{h}}$, respectively. Using $\mathcal{N}(\widehat{\mathbf{h}}, \mathbf{K}_{\mathbf{h}}^{-1})$ as a proposal distribution, one can sample $\mathbf{h}$ directly via an acceptance-rejection Metropolis-Hastings step, where the candidate is obtained using the precision-based sampling approach in Chan and Jeliazkov (2009). Building upon earlier work on Gaussian Markov random fields (Rue, 2001) and nonparametric regression (Chib and Jeliazkov, 2006; Chib, Greenberg, and Jeliazkov, 2009), this precision-based sampling approach is generally more efficient than Kalman-filter based methods. Alternatively, one can also sample $\mathbf{h}$ by modifying the auxiliary mixture sampler of Kim, Shephard, and Chib (1998), as implemented in Carriero, Clark, and Marcellino (2016). Finally, Steps 4 and 5 are standard and can be easily implemented (see., e.g., Chan and Hsiao, 2014).

A4: Estimating a BMAR with the Factor Stochastic Volatility

This subsection provides details on estimating a BMAR in which the error covariance matrix of $\mathrm{vec}(\mathbf{E}_t)$ is specified using a factor stochastic volatility model. More specifically, we augment the model with r latent factors $\mathbf{f}_t = (f_{1,t}, \ldots, f_{r,t})'$ and write the errors as

$$\mathrm{vec}(\mathbf{E}_t) = \mathbf{L}\mathbf{f}_t + \mathbf{u}_t, \tag{10}$$

where $\mathbf{L}$ is the $nk \times r$ factor loadings matrix. The latent factors $\mathbf{f}_t$ and $\mathbf{u}_t$ are assumed to be independent at all leads and lags and jointly Gaussian:

$$\begin{pmatrix}\mathbf{u}_t \\ \mathbf{f}_t\end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix}\mathbf{0}_{nk} \\ \mathbf{0}_r\end{pmatrix}, \begin{pmatrix}\mathbf{D}_t & \mathbf{0} \\ \mathbf{0} & \mathbf{G}_t\end{pmatrix}\right),$$

where $\mathbf{D}_t = \mathrm{diag}(\mathrm{e}^{h_{1,t}}, \ldots, \mathrm{e}^{h_{nk,t}})$ and $\mathbf{G}_t = \mathrm{diag}(\mathrm{e}^{h_{nk+1,t}}, \ldots, \mathrm{e}^{h_{nk+r,t}})$.

We first discuss the sampling of $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)'$ and $\mathbf{B} = (\mathbf{B}_1, \ldots, \mathbf{B}_p)'$. Let $\overline{\mathbf{E}}_t = \text{vec}_{n,k}^{-1}(\mathbf{L}\mathbf{f}_t)$ denote the $n \times k$ matrix constructed by reshaping the $nk \times 1$ vector $\mathbf{L}\mathbf{f}_t$ such that $\text{vec}(\overline{\mathbf{E}}_t) = \mathbf{L}\mathbf{f}_t$ and let $\mathbf{U}_t = \text{vec}_{n,k}^{-1}(\mathbf{u}_t)$. Then, (10) can be written as

$$\mathbf{E}_t = \overline{\mathbf{E}}_t + \mathbf{U}_t.$$

Letting $\overline{\mathbf{Y}}_t = \mathbf{Y}_t - \overline{\mathbf{E}}_t$ and substituting $\mathbf{E}_t$ into (2), we have

$$\overline{\mathbf{Y}}_t = \mathbf{A}'\mathbf{X}_t\mathbf{B} + \mathbf{U}_t, \tag{11}$$

where $\text{vec}(\mathbf{U}_t) = \mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_{nk}, \mathbf{D}_t)$. Given a Gaussian prior on $\text{vec}(\mathbf{A})$, one can in principle sample the entire matrix $\mathbf{A}$ in one step; similarly for $\mathbf{B}$. But when n or k is large, it is much faster to sample each column separately. In what follows, we demonstrate both approaches. For our application, $n = 6$ and $k = 50$, and hence we sample $\mathbf{A}$ in one block and draw each column of $\mathbf{B}$ separately.

To sample $\mathbf{A}$, we vectorize the transpose of (11) to obtain

$$\text{vec}(\overline{\mathbf{Y}}_t') = \mathbf{Z}_{\mathbf{B},t}\mathbf{a} + \text{vec}(\mathbf{U}_t'),$$

where $\mathbf{Z}_{\mathbf{B},t} = \mathbf{I}_n \otimes (\mathbf{X}_t\mathbf{B})'$, $\mathbf{a} = \text{vec}(\mathbf{A})$ and $\text{vec}(\mathbf{U}_t') \sim \mathcal{N}(\mathbf{0}_{nk}, \widetilde{\mathbf{D}}_t)$. Here $\widetilde{\mathbf{D}}_t = \mathbf{K}_{n,k}\mathbf{D}_t\mathbf{K}_{n,k}'$ and $\mathbf{K}_{n,k}$ is the commutation matrix such that $\mathbf{K}_{n,k}\text{vec}(\mathbf{U}_t) = \text{vec}(\mathbf{U}_t')$. Consider the Gaussian prior on $\mathbf{a}$:

$$(\mathbf{a} \,|\, \kappa_{\mathbf{a}}) \sim \mathcal{N}(\mathbf{a}_0, \mathbf{V}_{\mathbf{a}}),$$

where $\kappa_{\mathbf{a}}$ is a prior hyperparameter that controls the overall shrinkage strength and $\mathbf{V}_{\mathbf{a}}$ is an $n^2p \times n^2p$ prior covariance matrix that depends on $\kappa_{\mathbf{a}}$. Let $\mathbf{h}$ and $\mathbf{f}$ denote the collections of log-volatilities and latent factors, respectively. Then, the full conditional posterior distribution of $\mathbf{a}$ is

$$(\mathbf{a} \,|\, \mathbf{Y}, \mathbf{B}, \mathbf{L}, \mathbf{h}, \mathbf{f}) \sim \mathcal{N}(\widehat{\mathbf{a}}, \mathbf{K}_{\mathbf{a}}^{-1}),$$

where

$$\mathbf{K}_{\mathbf{a}} = \mathbf{V}_{\mathbf{a}}^{-1} + \sum_{t=1}^{T} \mathbf{Z}_{\mathbf{B},t}'\widetilde{\mathbf{D}}_t^{-1}\mathbf{Z}_{\mathbf{B},t}, \quad \widehat{\mathbf{a}} = \mathbf{K}_{\mathbf{a}}^{-1}\left(\mathbf{V}_{\mathbf{a}}^{-1}\mathbf{a}_0 + \sum_{t=1}^{T} \mathbf{Z}_{\mathbf{B},t}'\widetilde{\mathbf{D}}_t^{-1}\text{vec}(\overline{\mathbf{Y}}_t')\right).$$

The computational complexity of sampling $\mathbf{a}$ in one block is $\mathcal{O}(n^6)$. This is computational feasible when n is relatively small.

Next, we sample each column of $\mathbf{B}$ at a time. To that end, let $\mathbf{b}_i$, $\bar{\mathbf{y}}_{i,t}$ and $\mathbf{u}_{i,t}$ denote the i-th column of $\mathbf{B}$, $\overline{\mathbf{Y}}_t$ and $\mathbf{U}_t$, respectively. Then, the i-th column of (11) can be represented as

$$\bar{\mathbf{y}}_{i,t} = \mathbf{Z}_{\mathbf{A},t}\mathbf{b}_i + \mathbf{u}_{i,t},$$

where $\mathbf{Z}_{\mathbf{A},t} = \mathbf{A}'\mathbf{X}_t$ and $\mathbf{u}_{i,t} \sim \mathcal{N}(\mathbf{0}_n, \mathbf{D}_{i,t})$ with $\mathbf{D}_{i,t} = \text{diag}(\text{e}^{h_{(i-1)n+1,t}}, \ldots, \text{e}^{h_{in,t}})$. If we assume a Gaussian prior on $\mathbf{b}_i$:

$$(\mathbf{b}_i \,|\, \kappa_{\mathbf{b}}) \sim \mathcal{N}(\mathbf{b}_{0,i}, \mathbf{V}_{\mathbf{b}_i}),$$

where $\mathbf{V}_{\mathbf{b}_i}$ is a $kp \times kp$ prior covariance matrix that depends on the hyperparameter $\kappa_{\mathbf{b}}$, then the full conditional posterior distribution of $\mathbf{b}_i$ is

$$(\mathbf{b}_i \,|\, \mathbf{Y}, \mathbf{A}, \mathbf{L}, \mathbf{h}, \mathbf{f}) \sim \mathcal{N}(\widehat{\mathbf{b}}_i, \mathbf{K}_{\mathbf{b}_i}^{-1}),$$

where

$$\mathbf{K}_{\mathbf{b}_i} = \mathbf{V}_{\mathbf{b}_i}^{-1} + \sum_{t=1}^{T} \mathbf{Z}'_{\mathbf{A},t}\mathbf{D}_{i,t}^{-1}\mathbf{Z}_{\mathbf{A},t}, \quad \widehat{\mathbf{b}}_i = \mathbf{V}_{\mathbf{b}_i}^{-1}\mathbf{b}_{0,i} + \sum_{t=1}^{T} \mathbf{Z}'_{\mathbf{A},t}\mathbf{D}_{i,t}^{-1}\bar{\mathbf{y}}_{i,t}.$$

The computational complexity of sampling each $\mathbf{b}_i$ is $\mathcal{O}(k^3)$, iterated over $i = 1, \ldots, k$, with total computational complexity of the order $\mathcal{O}(k^4)$. Hence, sampling each $\mathbf{b}_i$ separately reduces the computational complexity by two orders of magnitude.

Standard algorithms can be applied to sample the factors, log-volatilities and the other parameters. We refer the readers to Chan (2023a) for more details.

Appendix B: Data

The US state-level data are sourced from the Federal Reserve Bank of St. Louis and the Bureau of Labor Statistics. For each of the 50 states, 6 quarterly time-series from 2005Q1 to 2023Q1 are obtained. Table 4 lists the variables and describes how they are transformed.

Table 4: Description of state-level variables in the empirical application.

Variable	Source	Tcode	Seasonal Adjustment
Initial unemployment insurance claims	FRED	3	NSA*
Continued unemployment insurance claims	FRED	3	NSA*
Total nonfarm employment	BLS	2	SA
Unemployment rate	FRED	1	SA
New housing permits	FRED	3	SA
Real home price index	FRED	2	NSA*

Note: Tcode describes the transformation of the variable, where 1 indicates no transformation, 2 denotes year-over-year growth rate, 3 stands for taking natural logarithm. NSA* indicates that the quarterly series has been seasonally adjusted using the X13-ARIMA procedure, and SA indicates that the series is available in seasonally-adjusted form.

References

Aguilar, O., and M. West (2000): "Bayesian dynamic factor models and portfolio allocation," *Journal of Business and Economic Statistics*, 18(3), 338–357.

Assaf, A. G., G. Li, H. Song, and M. G. Tsionas (2019): "Modeling and forecasting regional tourism demand using the Bayesian global vector autoregressive (BGVAR) model," *Journal of Travel Research*, 58(3), 383–397.

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

Barigozzi, M., H. Cho, and H. Maeng (2024): "Tail-robust factor modelling of vector and tensor time series in high dimensions," *arXiv preprint arXiv:2407.09390*.

Barigozzi, M., Y. He, L. Li, and L. Trapani (2023): "Robust tensor factor analysis," *arXiv preprint arXiv:2303.18163*.

Baumeister, C., D. Leiva-León, and E. Sims (2022): "Tracking weekly state-level economic conditions," *Review of Economics and Statistics*.

Bauwens, L., M. Lubrano, and J. Richard (1999): *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press, New York.

Billio, M., R. Casarin, M. Iacopini, and S. Kaufmann (2023): "Bayesian dynamic tensor regression," *Journal of Business & Economic Statistics*, 41(2), 429–439.

Bobeica, E., and B. Hartwig (2023): "The COVID-19 shock and challenges for inflation modelling," *International Journal of Forecasting*, 39(1), 519–539.

Bokun, K. O., L. E. Jackson, K. L. Kliesen, and M. T. Owyang (2023): "FRED-SD: A real-time database for state-level data with forecasting applications," *International Journal of Forecasting*, 39(1), 279–297.

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

——— (2013): "Panel vector autoregressive models: A survey," in *VAR models in macroeconomics–new developments and applications: Essays in honor of Christopher A. Sims*, pp. 205–246. Emerald Publishing Limited.

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

Carriero, A., T. E. Clark, M. Marcellino, and E. Mertens (2022): "Addressing COVID-19 outliers in BVARs with stochastic volatility," *The Review of Economics and Statistics*, Forthcoming.

Carriero, A., T. E. Clark, and M. G. Marcellino (2015): "Bayesian VARs: Specification Choices and Forecast Accuracy," *Journal of Applied Econometrics*, 30(1), 46–73.

——— (2016): "Common drifting volatility in large Bayesian VARs," *Journal of Business and Economic Statistics*, 34(3), 375–390.

——— (2019): "Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *Journal of Econometrics*, 212(1), 137–154.

Carriero, A., G. Kapetanios, and M. Marcellino (2016): "Structural analysis with multivariate autoregressive index models," *Journal of Econometrics*, 192(2), 332–348.

Celani, A., and P. Pagnottoni (2023): "Matrix autoregressive models: generalization and Bayesian estimation," *Studies in Nonlinear Dynamics & Econometrics*.

Chan, J. C. C. (2017): "The Stochastic Volatility in Mean Model with Time-Varying Parameters: An Application to Inflation Modeling," *Journal of Business and Economic Statistics*, 35(1), 17–28.

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

——— (2023a): "Comparing stochastic volatility specifications for large Bayesian VARs," *Journal of Econometrics*, 235(2), 1419–1446.

——— (2023b): "Large hybrid time-varying parameter VARs," *Journal of Business and Economic Statistics*, 41(3), 890–905.

Chan, J. C. C., and E. Eisenstat (2018): "Bayesian Model Comparison for Time-Varying Parameter VARs with Stochastic Volatility," *Journal of Applied Econometrics*, 33(4), 509–532.

Chan, J. C. C., and C. Y. L. Hsiao (2014): "Estimation of Stochastic Volatility Models with Heavy Tails and Serial Dependence," in *Bayesian Inference in the Social Sciences*, ed. by I. Jeliazkov, and X.-S. Yang, pp. 159–180. John Wiley & Sons, Hoboken.

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

Chan, J. C. C., G. Koop, and X. Yu (2024): "Large order-invariant Bayesian VARs with stochastic volatility," *Journal of Business and Economic Statistics*, 42(2), 825–837.

Chan, J. C. C., A. Poon, and D. Zhu (2023): "High-dimensional conditionally Gaussian state space models with missing data," *Journal of Econometrics*, 236(1), 105468.

Chen, B., E. Y. Chen, S. Bolivar, and R. Chen (2024): "Time-varying matrix factor models," *arXiv preprint arXiv:2404.01546*.

Chen, E. Y., R. S. Tsay, and R. Chen (2020): "Constrained factor models for high-dimensional matrix-variate time series," *Journal of the American Statistical Association*, 115(530), 775–793.

Chen, R., H. Xiao, and D. Yang (2021): "Autoregressive models for matrix-valued time series," *Journal of Econometrics*, 222(1), 539–560.

Chib, S., E. Greenberg, and I. Jeliazkov (2009): "Estimation of Semiparametric Models in the Presence of Endogeneity and Sample Selection," *Journal of Computational and Graphical Statistics*, 18, 321–348.

Chib, S., and I. Jeliazkov (2006): "Inference in semiparametric dynamic models for binary longitudinal data," *Journal of the American Statistical Association*, 101(474), 685–700.

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

Chiu, C. J., H. Mumtaz, and G. Pinter (2017): "Forecasting with VAR models: Fat tails and stochastic volatility," *International Journal of Forecasting*, 33(4), 1124–1143.

Clark, T. E. (2011): "Real-time density forecasts from Bayesian vector autoregressions with stochastic volatility," *Journal of Business and Economic Statistics*, 29(3), 327–341.

Clark, T. E., and E. Mertens (2023): "Stochastic volatility in Bayesian vector autoregressions," in *Oxford Research Encyclopedia of Economics and Finance*. Oxford University Press.

Clark, T. E., and F. Ravazzolo (2015): "Macroeconomic forecasting performance under alternative specifications of time-varying volatility," *Journal of Applied Econometrics*, 30(4), 551–575.

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

Cong, Y., B. Chen, and M. Zhou (2017): "Fast simulation of hyperplane-truncated multivariate normal distributions," *Bayesian Analysis*, 12(4), 1017–1037.

Cross, J., and A. Poon (2016): "Forecasting structural change and fat-tailed events in Australian macroeconomic variables," *Economic Modelling*, 58, 34–51.

D'Agostino, A., L. Gambetti, and D. Giannone (2013): "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101.

Devroye, L. (2014): "Random variate generation for the generalized inverse Gaussian distribution," *Statistics and Computing*, 24(2), 239–246.

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

Feldkircher, M., F. Huber, G. Koop, and M. Pfarrhofer (2022): "Approximate Bayesian inference and forecasting in huge-dimensional multicountry VARs," *International Economic Review*, 63(4), 1625–1658.

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

Hamilton, J. D., and M. T. Owyang (2012): "The propagation of regional recessions," *Review of Economics and Statistics*, 94(4), 935–947.

Hartwig, B. (2021): "Bayesian VARs and Prior Calibration in Times of COVID-19," *Available at SSRN 3792070*.

Hoff, P. D. (2015): "Multilinear tensor regression for longitudinal relational data," *The Annals of Applied Statistics*, 9(3), 1169.

Huber, F., G. Koop, and L. Onorante (2019): "Inducing sparsity and shrinkage in time-varying parameter models," *arXiv preprint arXiv:1905.10787*.

Jurado, K., S. C. Ludvigson, and S. Ng (2015): "Measuring uncertainty," *American Economic Review*, 105(3), 1177–1216.

Kadiyala, R. K., and S. Karlsson (1997): "Numerical Methods for Estimation and inference in Bayesian VAR-models," *Journal of Applied Econometrics*, 12(2), 99–132.

Kapetanios, G., L. Serlenga, and Y. Shin (2021): "Estimation and inference for multi-dimensional heterogeneous panel datasets with hierarchical multi-factor error structure," *Journal of Econometrics*, 220(2), 504–531.

Karlsson, S. (2013): "Forecasting with Bayesian vector autoregressions," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmermann, vol. 2 of *Handbook of Economic Forecasting*, pp. 791–897. Elsevier.

Kastner, G. (2019): "Sparse Bayesian time-varying covariance estimation in many dimensions," *Journal of econometrics*, 210(1), 98–115.

Kharrazi, A., E. Rovenskaya, and B. D. Fath (2017): "Network structure impacts global commodity trade growth and resilience," *PloS one*, 12(2), e0171184.

Kim, S., N. Shephard, and S. Chib (1998): "Stochastic Volatility: Likelihood Inference and Comparison with ARCH Models," *Review of Economic Studies*, 65(3), 361–393.

Kolda, T. G., and B. W. Bader (2009): "Tensor decompositions and applications," *SIAM review*, 51(3), 455–500.

Koop, G., and D. Korobilis (2010): "Bayesian Multivariate Time Series Methods for Empirical Macroeconomics," *Foundations and Trends in Econometrics*, 3(4), 267–358.

——— (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177(2), 185–198.

Koop, G., and D. Korobilis (2016): "Model uncertainty in panel vector autoregressive models," *European Economic Review*, 81, 115–131.

Koop, G., and D. Korobilis (2018): "Variational Bayes inference in high-dimensional time-varying parameter models," *Available at SSRN 3246472*.

Koop, G., S. McIntyre, and J. Mitchell (2020): "UK regional nowcasting using a mixed frequency vector auto-regressive model with entropic tilting," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 183(1), 91–119.

Koop, G., S. McIntyre, J. Mitchell, and A. Poon (2020): "Regional output growth in the United Kingdom: More timely and higher frequency estimates from 1970," *Journal of Applied Econometrics*, 35(2), 176–197.

Koop, G., S. G. McIntyre, J. Mitchell, A. Poon, and P. Wu (2023): "Incorporating Short Data into Large Mixed-Frequency VARs for Regional Nowcasting," *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, forthcoming.

Leng, C., and C. Y. Tang (2012): "Sparse matrix graphical models," *Journal of the American Statistical Association*, 107(499), 1187–1200.

Lenza, M., and G. E. Primiceri (2022): "How to Estimate a VAR after March 2020," *Journal of Applied Econometrics*, Forthcoming.

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

Liu, X., and E. Y. Chen (2022): "Identification and estimation of threshold matrix-variate factor models," *Scandinavian Journal of Statistics*, 49(3), 1383–1417.

Lock, E. F. (2018): "Tensor-on-tensor regression," *Journal of Computational and Graphical Statistics*, 27(3), 638–647.

Nobile, A. (2000): "Comment: Bayesian multinomial probit models with a normalization constraint," *Journal of Econometrics*, 99(2), 335–345.

Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004): "Modeling regional interdependencies using a global error-correcting macroeconometric model," *Journal of Business & Economic Statistics*, 22(2), 129–162.

Pitt, M., and N. Shephard (1999): "Time varying covariances: a factor stochastic volatility approach," *Bayesian Statistics*, 6, 547–570.

Primiceri, G. E. (2005): "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72(3), 821–852.

Rue, H. (2001): "Fast sampling of Gaussian Markov random fields with applications," *Journal of the Royal Statistical Society: Series B (Methodological)*, 63(2), 325–338.

Rue, H., and L. Held (2005): *Gaussian Markov Random Fields: Theory and Applications*. CRC press.

Schorfheide, F., and D. Song (2015): "Real-time forecasting with a mixed-frequency VAR," *Journal of Business and Economic Statistics*, 33(3), 366–380.

Schorfheide, F., and D. Song (2021): "Real-time forecasting with a (standard) mixed-frequency VAR during a pandemic," Discussion paper, National Bureau of Economic Research.

Sims, C. A., and T. Zha (2006): "Were There Regime Switches in U.S. Monetary Policy?," *American Economic Review*, 96(1), 54–81.

Stock, J. H., and M. W. Watson (2016): "Core inflation and trend inflation," *Review of Economics and Statistics*, 98(4), 770–784.

Tsionas, M. G., M. Izzeldin, and L. Trapani (2022): "Estimation of large dimensional time varying VARs using copulas," *European Economic Review*, 141, 103952.

Wang, D., X. Liu, and R. Chen (2019): "Factor models for matrix-valued high-dimensional time series," *Journal of Econometrics*, 208(1), 231–248.

Wang, D., Y. Zheng, and G. Li (2024): "High-dimensional low-rank tensor autoregressive time series modeling," *Journal of Econometrics*, 238(1), 105544.

Yu, L., Y. He, X. Kong, and X. Zhang (2022): "Projected estimation for large-dimensional matrix factor models," *Journal of Econometrics*, 229(1), 201–217.