Large Bayesian Tensor VARs with Stochastic Volatility

Joshua C. C. Chan   Yaling Qi

Purdue University   Purdue University

September 2024

Abstract

We consider Bayesian tensor vector autoregressions (TVARs) in which the VAR coefficients are arranged as a three-dimensional array or tensor, and this coefficient tensor is parameterized using a low-rank CP decomposition. We develop a family of TVARs using a general stochastic volatility specification, which includes a wide variety of commonly-used multivariate stochastic volatility and COVID-19 outlier-augmented models. In a forecasting exercise involving 40 US quarterly variables, we show that these TVARs outperform the standard Bayesian VAR with the Minnesota prior. The results also suggest that the parsimonious common stochastic volatility model tends to forecast better than the more flexible Cholesky stochastic volatility model.

Keywords: forecasting, outlier, stochastic volatility, tensor, vector autoregression

JEL classification codes: C11, C53, C55

1 Introduction

Large Bayesian vector autoregressions (BVARs) are now commonly used in macroeconomic forecasting and structural analysis, following early influential papers by Bańbura, Giannone, and Reichlin (2010), Carriero, Kapetanios, and Marcellino (2009) and Koop (2013). The dominant approach to tackle the challenge of parameter proliferation in large systems is to use shrinkage priors to regularize the variations in VAR coefficients; there is now an extensive literature on various shrinkage priors designed for BVARs.[1] In contrast, other dimension reduction techniques are relatively unexplored.

We investigate the usefulness of specifying a low-rank structure on the VAR coefficients for forecasting. More specifically, we follow the approach proposed by Wang, Zheng, Lian, and Li (2022) to treat the VAR coefficients collectively as a three-dimensional array or tensor: for a BVAR with n endogenous variables and p lags, we arrange the $n \times n$ coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ in the third dimension to construct the third-order tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times p}$. We then model $\mathcal{A}$ using a rank-R CP decomposition. We call these BVARs constructed via the CP decomposition tensor VARs or TVARs.

Using this tensor decomposition, the number of free parameters is reduced from $n^2 p$ to $(2n + p)R$. Since the number of free parameters under this tensor decomposition grows linearly in n, it is especially suitable for applications with a large number of variables. This approach is related to the reduced-rank VAR (see, e.g., Carriero, Kapetanios, and Marcellino, 2011), which may be viewed as a special case in which the rank of $\mathcal{A}$ is reduced along one of the three possible dimensions.

Departing from the homoskedastic framework in Wang, Zheng, Lian, and Li (2022), we formulate the TVARs using a general stochastic volatility specification, which can repre-

---

[1] The most widely-used shrinkage priors for BVARs are the family of Minnesota priors developed in a series of papers by Doan, Litterman, and Sims (1984), Litterman (1986) and Kadiyala and Karlsson (1993, 1997). Recent additions to this family include Giannone, Lenza, and Primiceri (2015) and Chan (2022). Another popular family is the adaptive hierarchical shrinkage priors that can be represented as scale mixtures of normals. Examples include the Bayesian Lasso (Park and Casella, 2008; Korobilis and Pettenuzzo, 2019), the normal-gamma prior (Griffin and Brown, 2010; Huber and Feldkircher, 2019), the horseshoe prior (Carvalho, Polson, and Scott, 2010; Follett and Yu, 2019) and the Dirichlet-Laplace prior (Bhattacharya, Pati, Pillai, and Dunson, 2015; Kastner and Huber, 2020). Naturally, one can also combine these adaptive hierarchical priors with the Minnesota priors, as proposed in Chan (2021). Hauzenberger, Huber, and Koop (2024) provide an excellent recent review on the state-of-the-art shrinkage priors developed for macroeconomic forecasting using BVARs.

sent a wide variety of multivariate stochastic volatility and COVID-19 outlier-augmented models commonly used for BVARs. This more general setup is motivated by the increasing recognition of the importance of allowing time-varying volatility for forecasting macroeconomic and financial variables, especially after the extreme economic turbulence triggered by the COVID-19 pandemic (see, e.g., Lenza and Primiceri, 2022; Carriero, Clark, Marcellino, and Mertens, 2022).

We develop efficient estimation procedures for these TVARs with stochastic volatility. In particular, we propose two types of algorithms to sample the components of the CP decomposition of the VAR coefficient tensor $\mathcal{A}$: we can either sample each block of the components jointly as a matrix or sample each column of the component matrix separately. The latter is inspired by the equation-by-equation estimation approach designed for BVARs proposed in Carriero, Clark, and Marcellino (2019) and Carriero, Chan, Clark, and Marcellino (2022), which can drastically reduce the computational burden when the number of endogenous variables n is very large.

We illustrate the methodology using a forecasting exercise that involves 40 US quarterly variables, such as GDP, industrial production, labor market variables and a variety of inflation and interest rates. We consider various TVARs with different stochastic volatility specifications, and compare them to a standard BVAR with the Minnesota prior in an out-of-sample forecasting exercise. The results show that TVARs clearly outperform the standard BVAR, highlighting the usefulness of the low-rank specification for the VAR coefficient tensor. We also find that models with some form of time-varying volatility substantially forecast better than their homoskedastic counterparts. Interesting, the parsimonious common stochastic volatility model of Carriero, Clark, and Marcellino (2016) tends to outperform the more flexible stochastic volatility model of Cogley and Sargent (2005), suggesting that it is foremost important to capture the strong comovements in the macroeconomic volatilities.

Our paper is closely related to the recent work by Luo and Griffin (2023), who also consider a BVAR with a tensor decomposition. While they focus on the Cholesky stochastic volatility model of Cogley and Sargent (2005), we consider a more general setup that can represent a wide range of multivariate stochastic volatility models. Our paper is also related to the emerging literature on modeling multidimensional tensors, where the coefficient tensors are typically parameterized using CP or Tucker decompositions; see,

for example, Billio, Casarin, Iacopini, and Kaufmann (2023) and Wang, Zheng, and Li (2024).

The rest of this paper is organized as follows. Section 2 first introduces a general framework of BVARs with a generic time-varying error covariance matrix and discusses how it can be used to represent a variety of popular stochastic volatility and COVID-19 outlier-augmented models. We then outline the proposed approach of arranging the VAR coefficient matrices as a third-order tensor in Section 3. Next, we introduce the efficient sampling algorithms in Section 4. Section 5 considers a recursive out-of-sample forecasting application that involves 40 US macroeconomic variables. Lastly, Section 6 concludes and briefly discusses some future research directions.

2 Large Bayesian VARs with Stochastic Volatility

Let $\mathbf{y}_t$ be an $n \times 1$ vector of endogenous variables at time t for $t = 1, \ldots, T$. Consider the following VAR(p):

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \cdots + \mathbf{A}_p \mathbf{y}_{t-p} + \mathbf{u}_t,$$

where $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $n \times n$ coefficient matrices. We omit an intercept term for ease of exposition; an intercept or any exogenous variables can be added to the model with minor modifications. Let $\mathbf{A} = (\mathbf{A}_1, \ldots, \mathbf{A}_p)'$ be the $np \times n$ matrix of VAR coefficients and let $\mathbf{x}_t = (\mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ denote a vector of lag variables of dimension $np \times 1$. Then, stacking the observations over $t = 1, \ldots, T$, we can rewrite the VAR more succinctly as

$$\mathbf{Y} = \mathbf{X}\mathbf{A} + \mathbf{U}, \tag{1}$$

where the matrices $\mathbf{Y}$, $\mathbf{X}$ and $\mathbf{U}$ are, respectively, $T \times n$, $T \times np$ and $T \times n$. In a standard homoskedastic VAR, the reduced-form errors $\mathbf{u}_1, \ldots, \mathbf{u}_T$ are assumed to be independent and identically distributed (iid) as $\mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma})$, where $\mathbf{0}_n$ is an $n \times 1$ vector of zeros and $\boldsymbol{\Sigma}$ is an $n \times n$ covariance matrix. However, it is increasingly recognized that some form of time-varying volatility is needed in modeling typical macroeconomic time-series. Early influential papers such as Cogley and Sargent (2005), Primiceri (2005) and Sims and Zha (2006) have highlighted the secular variations in volatility. There is now a large empirical literature that demonstrates the importance of allowing time-varying volatility

in improving model-fit and forecasting performance in the context of Bayesian VARs; examples include Clark (2011), Koop and Korobilis (2013), D'Agostino, Gambetti, and Giannone (2013), Clark and Ravazzolo (2015), Cross and Poon (2016) and Chan and Eisenstat (2018).

Below we outline a few stochastic volatility models suitable for large BVARs. In particular, the innovations of the VAR in (1) are now distributed as

$$\mathbf{u}_t \sim \mathcal{N}(\mathbf{0}_n, \boldsymbol{\Sigma}_t), \tag{2}$$

where $\boldsymbol{\Sigma}_t$ is a generic time-varying covariance matrix.

One of the first stochastic volatility models designed for large BVARs is the common stochastic volatility proposed in Carriero, Clark, and Marcellino (2016). Their model is motivated by the empirical observation that the estimated time-varying error variances of many macroeconomic variables have broadly similar low-frequency movements. A parsimonious way to model these comovements is to introduce a time-varying latent factor to scale the error covariance matrix via

$$\boldsymbol{\Sigma}_t = \mathrm{e}^{h_t} \boldsymbol{\Omega}, \tag{3}$$

where $\boldsymbol{\Omega}$ is a time-invariant covariance matrix. The log-volatility h_t is modeled using a zero-mean stationary AR(1) process:

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

for $t = 2, \ldots, T$, where $|\phi| < 1$ and the process is initialized as $h_1 \sim \mathcal{N}(0, \sigma_h^2/(1-\phi^2))$. The common stochastic volatility model specified in (3)–(4) can be extended to incorporate other useful features. For example, Chan (2020) introduces a general framework that can accommodate heavy-tailed, heteroskedastic and serially dependent innovations.

Another widely-used stochastic volatility specification for BVARs is the Cholesky stochastic volatility model—based on the modified Cholesky decomposition of the covariance matrix—developed in Cogley and Sargent (2005). More specifically, consider the decomposition of $\boldsymbol{\Sigma}_t$ via

$$\boldsymbol{\Sigma}_t = \mathbf{B}_0^{-1} \mathbf{D}_t (\mathbf{B}_0^{-1})', \tag{5}$$

where $\mathbf{B}_0$ is an $n \times n$ lower triangular matrix with ones on the diagonal and $\mathbf{D}_t = \mathrm{diag}(\mathrm{e}^{h_{1,t}}, \ldots, \mathrm{e}^{h_{n,t}})$. Each element of the vector $\mathbf{h}_t = (h_{1,t}, \ldots, h_{n,t})'$ is modeled via an independent autoregressive process:

$$h_{i,t} = \mu_i + \phi_i(h_{i,t-1} - \mu_i) + u_{i,t}^h, \quad u_{i,t}^h \sim \mathcal{N}(0, \sigma_i^2) \tag{6}$$

for $t = 2, \ldots, T$, where $h_{i,1}$ is initialized as $h_{i,1} \sim \mathcal{N}(\mu_i, \sigma_i^2/(1 - \phi_i^2))$ for $i = 1, \ldots, n$.

This Cholesky stochastic volatility specification is more flexible than the common stochastic volatility model in (3), since the former contains n stochastic volatility processes and can accommodate more complex covolatility patterns. This flexibility, however, comes at a cost of higher model complexity. Whether this stochastic volatility specification forecasts better than alternatives in large systems is an empirical question.

Another difference between the Cholesky stochastic volatility and the common stochastic volatility is that the latter is order-invariant—i.e., parameter estimates are invariant to reordering the endogenous variables in $\mathbf{y}_t$—whereas the former is not. One practical implication is that forecasts from the Cholesky stochastic volatility model could differ substantially across different variable orderings, as documented in Arias, Rubio-Ramirez, and Shin (2023) using a similar model of Primiceri (2005).

The reason why the Cholesky stochastic volatility is not order-invariant is partly due to the use of the lower triangular parameterization of $\mathbf{B}_0$ in (5). Motivated by this simple observation, Chan, Koop, and Yu (2024) extend the model by relaxing this lower triangular assumption and instead specify $\mathbf{B}_0$ to be any non-degenerate square matrix. They prove that the model is order invariant. Moreover, based on the results in Bertsche and Braun (2022), $\mathbf{B}_0$ is also identified up to permutations and sign switches.

The extreme movements in many macroeconomic variables at the onset of the COVID-19 pandemic have motivated much recent work on modeling outliers in macroeconomic time-series. An example is the outlier-augmented stochastic volatility model proposed by Carriero, Clark, Marcellino, and Mertens (2022), which builds on the Cholesky stochastic volatility model and the discrete mixture representation for the innovations introduced in Stock and Watson (2016). This outlier-augmented model can also be represented using (2).

There are many other multivariate stochastic volatility models for $\mathbf{\Sigma}_t$. Clark and Mertens

(2023) and Chan (2024) provide two recent reviews on BVARs with a wide range of stochastic volatility and outlier-augmented specifications.

3 Tensor Decomposition of VAR Coefficients

For high-dimensional settings, there are two key challenges related to the proliferation of the VAR coefficients. First, the number of VAR coefficients increases quadratically in n. In many large-scale applications, there are far more VAR coefficients than the number of observations, which makes it necessary to regularize these VAR coefficients. Second, due to the proliferation of VAR coefficients, sampling them tends to be very computational intensive.

These two related challenges are typically tackled separately in the literature. For instance, Bayesian shrinkage priors are widely used to regularize the VAR coefficients. These include the family of Minnesota priors (Doan, Litterman, and Sims, 1984; Litterman, 1986; Kadiyala and Karlsson, 1993, 1997; Giannone, Lenza, and Primiceri, 2015; Chan, 2022) and various adaptive hierarchical shrinkage priors (Huber and Feldkircher, 2019; Korobilis and Pettenuzzo, 2019; Kastner and Huber, 2020; Chan, 2021). The computational challenge is addressed by developing efficient MCMC or variational methods to sample the large number of VAR coefficients (Carriero, Clark, and Marcellino, 2019; Carriero, Chan, Clark, and Marcellino, 2022; Gefang, Koop, and Poon, 2023; Bernardi, Bianchi, and Bianco, 2024). We instead take an alternative approach that tackles these two challenges simultaneously by imposing a low-rank structure on the VAR coefficients.

One possibility is the reduced-rank VAR (see, e.g., Carriero, Kapetanios, and Marcellino, 2011), in which the matrix $\mathbf{A}' = (\mathbf{A}_1, \ldots, \mathbf{A}_p)$ is assumed to have a reduced rank $R < n$. That is, the dimension of the column space of the VAR coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ has rank R. Alternatively, one could impose a low-rank structure on $(\mathbf{A}'_1, \ldots, \mathbf{A}'_p)$ or $(\text{vec}(\mathbf{A}_1), \ldots, \text{vec}(\mathbf{A}_p))$, whose ranks correspond to the dimensions of the row space and vectorized matrix space of the VAR coefficient matrices, respectively. Each of these options essentially reduces the dimension along one of the three different directions.

This motivates us to follow the approach in Wang, Zheng, Lian, and Li (2022) to treat the VAR coefficients collectively as a three-dimensional array or tensor. That is, we

arrange the $n \times n$ coefficient matrices $\mathbf{A}_1, \ldots, \mathbf{A}_p$ in the third dimension to form the third-order tensor $\mathcal{A} \in \mathbb{R}^{n \times n \times p}$. With this tensor representation, we then use a rank-R CP decomposition to construct $\mathcal{A}$:

$$\mathcal{A} = \sum_{r=1}^{R} \boldsymbol{\theta}_1^{(r)} \circ \boldsymbol{\theta}_2^{(r)} \circ \boldsymbol{\theta}_3^{(r)}, \tag{7}$$

where $\boldsymbol{\theta}_1^{(r)}, \boldsymbol{\theta}_2^{(r)} \in \mathbb{R}^n$ and $\boldsymbol{\theta}_3^{(r)} \in \mathbb{R}^p$. Each component $\boldsymbol{\theta}_1^{(r)} \circ \boldsymbol{\theta}_2^{(r)} \circ \boldsymbol{\theta}_3^{(r)}$ is a rank-1 third-order tensor whose (i, j, k) element is the product of the i-th, j-th and k-th elements of $\boldsymbol{\theta}_1^{(r)}, \boldsymbol{\theta}_2^{(r)}$ and $\boldsymbol{\theta}_3^{(r)}$, respectively. Using this tensor decomposition with a small R, the number of parameters is reduced from $n^2 p$ to $(2n + p)R$. We refer the readers to Kolda and Bader (2009) for a general introduction to tensors and their operations.

4   Bayesian Estimation

In this section we describe the priors and outline the posterior simulator. In particular, we focus on the sampling of the components $\boldsymbol{\theta}_1^{(r)}, \boldsymbol{\theta}_2^{(r)}$ and $\boldsymbol{\theta}_3^{(r)}, r = 1, \ldots, R$. To that end, let $\boldsymbol{\Theta}_1 = (\boldsymbol{\theta}_1^{(1)}, \ldots, \boldsymbol{\theta}_1^{(R)})$ and similarly define $\boldsymbol{\Theta}_2$ and $\boldsymbol{\Theta}_3$. The dimensions of $\boldsymbol{\Theta}_1, \boldsymbol{\Theta}_2$ and $\boldsymbol{\Theta}_3$ are, respectively, $n \times R$, $n \times R$ and $p \times R$. For later reference, stack $\boldsymbol{\theta}_1 = \text{vec}(\boldsymbol{\Theta}_1)$, $\boldsymbol{\theta}_2 = \text{vec}(\boldsymbol{\Theta}_2')$ and $\boldsymbol{\theta}_3 = \text{vec}(\boldsymbol{\Theta}_3')$. For reasons that will become transparent later, note that $\boldsymbol{\theta}_1$ is constructed by stacking the columns of $\boldsymbol{\Theta}_1$, whereas $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$ are formed by stacking the rows of $\boldsymbol{\Theta}_2$ and $\boldsymbol{\Theta}_3$, respectively.

Next, consider the following independent Gaussian priors on $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$:

$$\boldsymbol{\theta}_j \sim \mathcal{N}(\boldsymbol{\theta}_{j,0}, \mathbf{V}_{\boldsymbol{\theta}_j}), \quad j = 1, \ldots, 3. \tag{8}$$

Here we consider simple Gaussian priors, but any adaptive hierarchical shrinkage priors that have a conditionally Gaussian representation—such as the normal-gamma prior, the horseshoe prior or the Dirichlet-Laplace prior—can be used.

In what follows, we derive the full conditional posterior distributions of $(\boldsymbol{\theta}_1 \,|\, \mathbf{Y}, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\Sigma})$, $(\boldsymbol{\theta}_2 \,|\, \mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_3, \boldsymbol{\Sigma})$ and $(\boldsymbol{\theta}_3 \,|\, \mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \text{diag}(\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T)$.

8

4.1 Sampling of $\boldsymbol{\theta}_1$

For sampling $\boldsymbol{\theta}_1$, first let

$$\boldsymbol{\Theta}_{-1} = (\boldsymbol{\theta}_3^{(1)} \otimes \boldsymbol{\theta}_2^{(1)}, \dots, \boldsymbol{\theta}_3^{(R)} \otimes \boldsymbol{\theta}_2^{(R)}),$$
$$\boldsymbol{\Theta}_{-2} = (\boldsymbol{\theta}_3^{(1)} \otimes \boldsymbol{\theta}_1^{(1)}, \dots, \boldsymbol{\theta}_3^{(R)} \otimes \boldsymbol{\theta}_1^{(R)}),$$
$$\boldsymbol{\Theta}_{-3} = (\boldsymbol{\theta}_2^{(1)} \otimes \boldsymbol{\theta}_1^{(1)}, \dots, \boldsymbol{\theta}_2^{(R)} \otimes \boldsymbol{\theta}_1^{(R)}).$$

Then, the mode-1 matricization of $\mathcal{A}$ can be written as (see, e.g., Kolda and Bader, 2009):

$$\mathcal{A}_{(1)} = \mathbf{A}' = \boldsymbol{\Theta}_1 \boldsymbol{\Theta}'_{-1}. \tag{9}$$

Furthermore, let $\mathcal{X} \in \mathbb{R}^{T \times n \times p}$ denote the third-order tensor constructed by stacking the $n \times p$ matrices $(\mathbf{y}_{t-1}, \dots, \mathbf{y}_{t-p})$, $t = 1, \dots, T$, along the first dimension so that $\mathrm{vec}(\mathcal{X}_{t,:,:}) = \mathbf{x}_t$. It is easy to verify that its mode-1 matricization is the $T \times np$ matrix $\mathbf{X}$, i.e., $\mathcal{X}_{(1)} = \mathbf{X}$. Hence, combining (1) and (9), we have

$$\mathbf{Y} = \mathcal{X}_{(1)} \boldsymbol{\Theta}_{-1} \boldsymbol{\Theta}'_1 + \mathbf{U}. \tag{10}$$

It follows that

$$\mathrm{vec}(\mathbf{Y}') = (\mathcal{X}_{(1)} \boldsymbol{\Theta}_{-1} \otimes \mathbf{I}_n) \mathrm{vec}(\boldsymbol{\Theta}_1) + \mathrm{vec}(\mathbf{U}'),$$

where $\mathrm{vec}(\mathbf{U}') \sim \mathcal{N}(\mathbf{0}_{Tn}, \boldsymbol{\Sigma})$ with $\boldsymbol{\Sigma} = \mathrm{diag}\,(\boldsymbol{\Sigma}_1, \dots, \boldsymbol{\Sigma}_T)$. Given the Gaussian prior on $\boldsymbol{\theta}_1 = \mathrm{vec}(\boldsymbol{\Theta}_1)$ specified in (8), the full conditional posterior of $\boldsymbol{\theta}_1$ is given by

$$(\boldsymbol{\theta}_1 \mid \mathbf{Y}, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\Sigma}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}_1, \mathbf{K}_{\boldsymbol{\theta}_1}^{-1}\right),$$

where

$$\mathbf{K}_{\boldsymbol{\theta}_1} = \mathbf{V}_{\boldsymbol{\theta}_1}^{-1} + (\boldsymbol{\Theta}'_{-1} \mathcal{X}'_{(1)} \otimes \mathbf{I}_n) \boldsymbol{\Sigma}^{-1} (\mathcal{X}_{(1)} \boldsymbol{\Theta}_{-1} \otimes \mathbf{I}_n)$$
$$\widehat{\boldsymbol{\theta}}_1 = \mathbf{K}_{\boldsymbol{\theta}_1}^{-1}\left(\mathbf{V}_{\boldsymbol{\theta}_1}^{-1} \boldsymbol{\theta}_{1,0} + (\boldsymbol{\Theta}'_{-1} \mathcal{X}'_{(1)} \otimes \mathbf{I}_n) \boldsymbol{\Sigma}^{-1} \mathrm{vec}(\mathbf{Y}')\right).$$

Since $\boldsymbol{\theta}_1$ is of length nR, sampling $\boldsymbol{\theta}_1$ generally involves $\mathcal{O}(n^3 R^3)$ elementary operations. When both n and R are large, this sampling step could be computationally intensive. An alternative is to sample each $\boldsymbol{\theta}_1^{(r)}$ separately, $r = 1, \dots, R$. For a fixed r, this can be done

by substituting

$$\boldsymbol{\Theta}_1\boldsymbol{\Theta}'_{-1} = \sum_{s=1}^{R} \boldsymbol{\theta}_1^{(s)}\boldsymbol{\theta}_{-1}^{(s)\prime},$$

where $\boldsymbol{\theta}_{-1}^{(s)} = \boldsymbol{\theta}_3^{(s)} \otimes \boldsymbol{\theta}_2^{(s)}$, into the transpose of (10) to obtain

$$\mathbf{Y}' = \sum_{s=1}^{R} \boldsymbol{\theta}_1^{(s)}\boldsymbol{\theta}_{-1}^{(s)\prime}\mathcal{X}'_{(1)} + \mathbf{U}'.$$

Next, let $\mathbf{Y}_{1r} = \mathbf{Y}' - \sum_{s\neq r} \boldsymbol{\theta}_1^{(s)}\boldsymbol{\theta}_{-1}^{(s)\prime}\mathcal{X}'_{(1)}$ and vectorize the above equation, we have

$$\mathrm{vec}(\mathbf{Y}_{1r}) = (\mathcal{X}_{(1)}\boldsymbol{\theta}_{-1}^{(r)} \otimes \mathbf{I}_n)\boldsymbol{\theta}_1^{(r)} + \mathrm{vec}(\mathbf{U}').$$

If the marginal prior for $\boldsymbol{\theta}_1^{(r)}$ is

$$\boldsymbol{\theta}_1^{(r)} \sim \mathcal{N}\left(\boldsymbol{\theta}_{1,0}^{(r)}, \mathbf{V}_{\boldsymbol{\theta}_1^{(r)}}\right),$$

the posterior distribution $(\boldsymbol{\theta}_1^{(r)} \,|\, \mathbf{Y}, \{\boldsymbol{\theta}_1^{(s)}\}_{s\neq r}, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\Sigma})$ has the form

$$(\boldsymbol{\theta}_1^{(r)} \,|\, \mathbf{Y}, \{\boldsymbol{\theta}_1^{(s)}\}_{s\neq r}, \boldsymbol{\theta}_2, \boldsymbol{\theta}_3, \boldsymbol{\Sigma}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}_1^{(r)}, \mathbf{K}_{\boldsymbol{\theta}_1^{(r)}}^{-1}\right),$$

where

$$\mathbf{K}_{\boldsymbol{\theta}_1^{(r)}} = (\boldsymbol{\theta}_{-1}^{(r)\prime}\mathcal{X}'_{(1)} \otimes \mathbf{I}_n)\boldsymbol{\Sigma}^{-1}(\mathcal{X}_{(1)}\boldsymbol{\theta}_{-1}^{(r)} \otimes \mathbf{I}_n)$$

$$\widehat{\boldsymbol{\theta}}_1^{(r)} = \mathbf{K}_{\boldsymbol{\theta}_1^{(r)}}^{-1}\left(\mathbf{V}_{\boldsymbol{\theta}_1^{(r)}}^{-1}\boldsymbol{\theta}_{1,0}^{(r)} + (\boldsymbol{\theta}_{-1}^{(r)\prime}\mathcal{X}'_{(1)} \otimes \mathbf{I}_n)\boldsymbol{\Sigma}^{-1}\mathrm{vec}(\mathbf{Y}_{1r})\right).$$

The drawback of sampling each $\boldsymbol{\theta}_1^{(r)}$ separately is that this tends to increase the autocorrelation of the constructed Markov chain. But this approach is substantially faster and remains computationally feasible even when both n and R are large.

4.2 Sampling of $\boldsymbol{\theta}_2$ and $\boldsymbol{\theta}_3$

Next, we derive the conditional distribution of $(\boldsymbol{\theta}_2 \,|\, \mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_3, \boldsymbol{\Sigma})$ and show that it is Gaussian. To start, we aim to write the VAR in (1) as a linear regression in $\boldsymbol{\theta}_2 = \mathrm{vec}(\boldsymbol{\Theta}'_2)$.

We first introduce some notations. Let $\mathbf{e}_i^p$ be the i-th column of $\mathbf{I}_p$. Let $\mathbf{P}$ denote the $np \times np$ commutation matrix so that $\mathbf{P}'\text{vec}(\mathbf{Z}) = \text{vec}(\mathbf{Z}')$ for any $n \times p$ matrix $\mathbf{Z}$. Explicitly, $\mathbf{P}$ can be constructed by setting the (k, l) element to be 1, i.e., $\mathbf{P}_{k,l} = 1$, if there exist i and j such that $k = (i-1)n + j$ and $l = (j-1)p + i$; otherwise, set $\mathbf{P}_{k,l} = 0$.

Proposition 1. The VAR in (1) can be written as

$$\text{vec}(\mathbf{Y}') = \sum_{i=1}^{p} (\mathbf{P}_{i2}\mathcal{X}'_{(2)} \otimes \mathbf{P}_{i1}\boldsymbol{\Theta}_{-2})\boldsymbol{\theta}_2 + \text{vec}(\mathbf{U}'), \tag{11}$$

where $\mathbf{P}_{i1} = (\mathbf{I}_n \otimes (\mathbf{e}_i^p)')\mathbf{P}'$, $\mathbf{P}_{i2} = (\mathbf{e}_i^p)' \otimes \mathbf{I}_T$ and $\mathcal{X}_{(2)}$ is the mode-2 matricization of $\mathcal{X}$.

The proof of this proposition is given in Appendix A. Now, given the representation in (11) and the Gaussian prior on $\boldsymbol{\theta}_2$ specified in (8), by standard linear regression results, one can verify that the full conditional posterior of $\boldsymbol{\theta}_2$ is given by

$$(\boldsymbol{\theta}_2 \mid \mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_3, \boldsymbol{\Sigma}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}_2, \mathbf{K}_{\boldsymbol{\theta}_2}^{-1}\right),$$

where

$$\mathbf{K}_{\boldsymbol{\theta}_2} = \mathbf{V}_{\boldsymbol{\theta}_2}^{-1} + \sum_{i=1}^{p}\sum_{j=1}^{p}(\mathcal{X}_{(2)}\mathbf{P}'_{i2} \otimes \boldsymbol{\Theta}'_{-2}\mathbf{P}'_{i1})\boldsymbol{\Sigma}^{-1}(\mathbf{P}_{j2}\mathcal{X}'_{(2)} \otimes \mathbf{P}_{j1}\boldsymbol{\Theta}_{-2}),$$

$$\widehat{\boldsymbol{\theta}}_2 = \mathbf{K}_{\boldsymbol{\theta}_2}^{-1}\left(\mathbf{V}_{\boldsymbol{\theta}_2}^{-1}\boldsymbol{\theta}_{2,0} + \sum_{i=1}^{p}(\mathcal{X}_{(2)}\mathbf{P}'_{i2} \otimes \boldsymbol{\Theta}'_{-2}\mathbf{P}'_{i1})\boldsymbol{\Sigma}^{-1}\text{vec}(\mathbf{Y}')\right).$$

Therefore, one can sample $\boldsymbol{\theta}_2$ in one block. When n and R are large, it might only be feasible to sample each $\boldsymbol{\theta}_2^{(r)}$ at a time for $r = 1, \ldots, R$. We provide the details of this alternative approach in Appendix C.

Likewise, the conditional distribution of $(\boldsymbol{\theta}_3 \mid \mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\Sigma})$ can be shown to be Gaussian. More specifically, we first write the VAR in (1) as a linear regression in $\boldsymbol{\theta}_3 = \text{vec}(\boldsymbol{\Theta}'_3)$. To that end, let $\mathbf{e}_i^n$ denote the i-th column of $\mathbf{I}_n$, let $\mathcal{X}_{(3)}$ represent the mode-3 matricization of $\mathcal{X}$, and define $\mathbf{Q}$ to be the $n^2 \times n^2$ commutation matrix so that $\mathbf{Q}'\text{vec}(\mathbf{Z}) = \text{vec}(\mathbf{Z}')$ for any $n \times n$ matrix $\mathbf{Z}$. Explicitly, $\mathbf{Q}$ can be constructed by setting the $\mathbf{Q}_{k,l} = 1$ if there exist i and j such that $k = (i-1)n + j$ and $l = (j-1)n + i$; otherwise, set $\mathbf{Q}_{k,l} = 0$.

Proposition 2. The VAR in (1) can be represented as

$$\mathrm{vec}(\mathbf{Y}') = \sum_{i=1}^{n}(\mathbf{Q}_{i2}\mathcal{X}'_{(3)} \otimes \mathbf{Q}_{i1}\boldsymbol{\Theta}_{-3})\boldsymbol{\theta}_3 + \mathrm{vec}(\mathbf{U}'), \tag{12}$$

where $\mathbf{Q}_{i1} = (\mathbf{I}_n \otimes (\mathbf{e}_i^n)')\mathbf{Q}'$ and $\mathbf{Q}_{i2} = (\mathbf{e}_i^n)' \otimes \mathbf{I}_T$.

Given the representation in (12), it is easy to verify that

$$(\boldsymbol{\theta}_3 \mid \mathbf{Y}, \boldsymbol{\theta}_1, \boldsymbol{\theta}_2, \boldsymbol{\Sigma}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}_3, \mathbf{K}_{\boldsymbol{\theta}_3}^{-1}\right),$$

where

$$\mathbf{K}_{\boldsymbol{\theta}_3} = \mathbf{V}_{\boldsymbol{\theta}_3}^{-1} + \sum_{i=1}^{n}\sum_{j=1}^{n}(\mathcal{X}_{(3)}\mathbf{Q}'_{i2} \otimes \boldsymbol{\Theta}'_{-3}\mathbf{Q}'_{i1})\boldsymbol{\Sigma}^{-1}(\mathbf{Q}_{j2}\mathcal{X}'_{(3)} \otimes \mathbf{Q}_{j1}\boldsymbol{\Theta}_{-3})$$

$$\widehat{\boldsymbol{\theta}}_3 = \mathbf{K}_{\boldsymbol{\theta}_3}^{-1}\left(\mathbf{V}_{\boldsymbol{\theta}_3}^{-1}\boldsymbol{\theta}_{3,0} + \sum_{i=1}^{n}(\mathcal{X}_{(3)}\mathbf{Q}'_{i2} \otimes \boldsymbol{\Theta}'_{-3}\mathbf{Q}'_{i1})\boldsymbol{\Sigma}^{-1}\mathrm{vec}(\mathbf{Y}')\right).$$

Alternatively, one can sample each $\boldsymbol{\theta}_3^{(r)}$ at a time for $r = 1, \ldots, R$ when n and R are large as before.

Finally, sampling the time-varying error covariance matrices $\boldsymbol{\Sigma}_1, \ldots, \boldsymbol{\Sigma}_T$ naturally depends on the stochastic volatility specification used. For a wide variety of stochastic volatility specifications commonly employed in applied work, such as the common stochastic volatility and Cholesky stochastic volatility discussed earlier, efficient algorithms are available to sample the latent variables and model parameters. We refer the readers to Chan (2023a) for more details.

5 An Empirical Application

We conduct an out-of-sample forecasting exercise to evaluate the performance of the proposed tensor VARs compared to a standard benchmark. More specifically, we construct a dataset of 40 quarterly macroeconomic and financial variables from the FRED-QD database (McCracken and Ng, 2020). The sample spans from 1969Q1 to 2024Q1. The data include key macroeconomic variables such as GDP, inflation rates, labor market

variables and various interest rates. We refer the readers to Appendix B for the detailed description of the time series and their transformations.

We consider three tensor VARs: a homoskedastic tensor VAR (TVAR), TVARs with the common volatility (TVAR-CSV) and the Cholesky stochastic volatility (TVAR-SV). For now, we set the rank of the VAR coefficient tensor $\mathcal{A}$ to be $R = 1$ for all TVARs. As a benchmark, we also include a standard Bayesian VAR (BVAR) with the Minnesota prior (implemented as the natural conjugate prior). The evaluation period of the forecasting exercise begins in 2010Q1 and ends in 2024Q1.

To assess the performance of jointly forecasting all $n = 40$ time series, we calculate the average log predictive likelihoods for each model over one- and four-quarter-ahead forecast horizons. The results are presented in Table 1. Higher values of log predictive likelihoods signify better forecast performance.

Table 1: Joint density forecast performance of the proposed tensor VARs relative to a standard Bayesian VAR with the Minnesota prior.

	BVAR	TVAR	TVAR-CSV	TVAR-SV
One-quarter-ahead	-8.01	-3.30	-1.31	-2.32
Four-quarter-ahead	-8.53	-3.30	-2.21	-2.66

A few observations can be drawn from these forecasting results. Firstly, comparing the two homoskedastic models, TVAR and BVAR, it is clear that the former substantially outperforms the latter, suggesting that the specification of a low rank structure on the VAR coefficient tensor $\mathcal{A}$ is more appropriate. Secondly, allowing some form of time-varying volatility clearly improves forecast performance. For example, extending a homoskedastic TVAR to a version with the common stochastic volatility increases the average log predictive likelihood from -8 to about -3 for one-step-ahead density forecasts. This finding is line with the large body of empirical evidence that demonstrates the importance of allowing time-varying volatility in macroeconomic forecasting (Clark, 2011; D'Agostino, Gambetti, and Giannone, 2013; Chan, 2023a). Finally, among the two TVARs with stochastic volatility, the version with the parsimonious common stochastic volatility outperforms the one with the more flexible Cholesky stochastic volatility, indicating strong comovements in the macroeconomic volatilities.

Next, we look at the point forecast performance of the TVARs for individual time series.

In particular, Table 2 reports the root mean squared forecast errors (RMSFEs) of a few key macroeconomic variables over the evaluation period relative to the benchmark BVAR. Values less than one indicate better forecast performance than the benchmark.

Table 2: Root mean squared forecast errors of the proposed tensor VARs relative to a standard Bayesian VAR with the Minnesota prior.

	One-quarter-ahead			Four-quarter-ahead		
	TVAR	TVAR-CSV	TVAR-SV	TVAR	TVAR-CSV	TVAR-SV
RPI	0.93	0.93	0.93	0.99	0.99	0.98
GDP	0.65	0.64	0.78	0.82	0.82	0.85
Unemployment	0.68	0.68	0.72	0.86	0.86	0.86
CPI	1.03	1.02	1.14	0.89	0.89	0.93
Fed funds rate	0.83	0.83	1.46	0.60	0.60	0.65
10-year T-bond	0.92	0.92	0.95	0.82	0.82	0.85

The results show that the TVARs tend to outperform the benchmark BVAR for both one- and four-quarter-ahead horizons. For example, the RMSFE of the homoskedastic TVAR for forecasting GDP is only 65% of that of the benchmark. Interestingly, allowing time-varying volatility does not appear to improve point forecasts. In fact, the more flexible TVAR-SV often provides inferior point forecasts compared to the homoskedastic TVAR (though the performance of the more parsimonious TVAR-CSV is virtually identical to TVAR).

So far we have set R, the rank of the VAR coefficient tensor $\mathcal{A}$, to be 1 for all TVARs. Naturally, one might wonder how this choice affects the density and point forecast performance. To investigate the impact of the choice of R, we present in Table 3 the one-quarter-ahead average log predictive likelihoods of the three TVARs for $R = 1, 3, 5, 10$.

Table 3: One-quarter-ahead joint density forecast performance of the proposed tensor VARs for $R = 1, 3, 5, 10$.

	TVAR	TVAR-CSV	TVAR-SV
$R = 1$	-3.30	-1.31	-2.32
$R = 3$	-3.27	-1.39	-2.36
$R = 5$	-3.40	-1.26	-2.48
$R = 10$	-3.53	-1.29	-2.53

It is interesting to note that the choice of R does not seem to have a large impact on

the joint density forecast performance across the three TVARs. In particular, it does not affect the relative ranking of the three models; the two TVARs with time-varying volatility outperform the homoskedastic version regardless of the rank R. Overall, low-rank TVARs tend to work as well as, if not better than, TVARs with $R = 10$.

Table 4 reports the one-quarter-ahead RMSFEs of the three TVARs for forecasting a few key macroeconomic variables. To facilitate comparison, the results for each TVAR are relative to the corresponding TVAR with $R = 1$. With the exception of CPI inflation, increasing the rank R does not appear to substantially improve point forecast performance.

Table 4: One-quarter-ahead RMSFEs of TVARs with $R = 3, 5, 10$ relative to the corresponding TVARs with $R = 1$.

	$R = 3$			$R = 5$			$R = 10$		
	TVAR	TVAR-CSV	TVAR-SV	TVAR	TVAR-CSV	TVAR-SV	TVAR	TVAR-CSV	TVAR-SV
RPI	1.01	1.00	1.01	1.01	1.00	1.02	1.01	1.00	1.03
GDP	1.04	1.02	0.99	1.08	1.05	1.05	1.12	1.07	1.12
Unemployment	1.02	1.01	1.00	1.02	1.01	1.05	1.05	1.01	1.15
CPI inflation	0.97	0.95	1.06	0.96	0.95	0.90	0.89	0.90	0.94
Fed funds rate	1.08	1.11	1.08	1.24	1.21	1.05	1.20	1.18	0.99
10-year T-bond	1.00	1.00	1.22	0.99	1.00	1.17	0.98	0.99	1.16

Overall, these forecasting results demonstrate the benefits of the proposed approach of specifying a low rank structure on the VAR coefficient tensor. In addition, our results highlight the importance of accommodating time-varying volatility in forecasting macroeconomic variables. Notably, in our forecasting exercise, the more parsimonious TVAR-CSV tends to forecast better than the more flexible TVAR-SV.

6   Concluding Remarks and Future Research

We have developed Bayesian tensor VARs in which the VAR coefficients are arranged as a third-order tensor and parameterized using a rank-R CP decomposition. We then introduced efficient sampling algorithms to simulate the components of the tensor decomposition. Through a forecasting exercise, we showed that these TVARs outperform the standard BVAR with the Minnesota prior.

For future research, it would be useful to extend these TVARs to allow for time-varying

coefficients in the mean equations. This tensor framework is especially suitable for developing time variation in the VAR coefficients, as the number of free parameters grows only linear in n. Of course, for large systems one might need additional shrinkage. In those cases, one can consider either the static shrinkage approach in Chan (2023b) or the dynamic shrinkage approaches proposed in Koop and Korobilis (2018) and Huber, Koop, and Onorante (2019).

Appendix A: Proofs of Propositions

In this appendix we provide a proof of Proposition 1. The proof of Proposition 2 is very similar and is thus omitted. We first introduce some useful notations. Let $\mathbf{e}_i^m$ denote the i-th column of $\mathbf{I}_m$; if there is no ambiguity about the dimension, we simply write $\mathbf{e}_i$. For a generic third-order tensor $\mathcal{Z}$, let $\mathcal{Z}_{(k)}$ denote its mode-k matricization for $k = 1, 2, 3$. We use the notation $\mathcal{Z}_{i_1, i_2, i_3}$ to represent the (i_1, i_2, i_3) tensor element of $\mathcal{Z}$ and $\mathcal{Z}_{(k), i, j}$ to denote the (i, j) element of $\mathcal{Z}_{(k)}$.

As an example, consider the third-order tensor $\mathcal{X} \in \mathbb{R}^{T \times n \times p}$. Its mode-2 matricization $\mathcal{X}_{(2)}$ is an $n \times Tp$ matrix, and the (i_1, i_2, i_3) tensor element $\mathcal{X}_{i_1, i_2, i_3}$ maps to the (i_2, j) matrix element $\mathcal{X}_{(2), i_2, j}$, where $j = (i_3 - 1)T + i_1$. Conversely, the (i, j) matrix element maps to the (i_1, i, i_3) tensor element, where $i_1 = \mod(j, T)$ is the remainder after division of j by T, and $i_3 = \lceil j/T \rceil$ is the least integer greater than or equal to j/T. It can be easily verified that $j = (i_3 - 1)T + i_1$.

Naturally, any matrix multiplication involving matrices constructed by mode-k matricization can be written in terms of the elements in the original tensors. For example, consider the (i, j) element of the matrix $\mathbf{E} = \mathcal{X}'_{(2)} \mathcal{A}_{(2)}$. Let $i_3 = \lceil i/T \rceil$ and $i_1 = \mod(i, T)$ so that $i = (i_3 - 1)T + i_1$. Similarly, obtain integers j_1 and j_3 so that $j = (j_3 - 1)n + j_1$. Then, we can write $\mathbf{E}_{i,j}$ in terms of elements in $\mathcal{X}$ and $\mathcal{A}$:

$$
\begin{aligned}
\mathbf{E}_{i,j} &= \sum_{k=1}^{n} \mathcal{X}_{(2), k, i} \mathcal{A}_{(2), k, j} \\
&= \sum_{k=1}^{n} \mathcal{X}_{(2), k, (i_3 - 1)T + i_1} \mathcal{A}_{(2), k, (j_3 - 1)T + j_1} = \sum_{k=1}^{n} \mathcal{X}_{i_1, k, i_3} \mathcal{A}_{j_1, k, j_3}.
\end{aligned}
\tag{13}
$$

Next, we introduce a useful lemma.

Lemma 1. The VAR in (1) can be written as

$$
\mathbf{Y} = \sum_{i=1}^{p} (\mathbf{e}_i' \otimes \mathbf{I}_T) \mathbf{E} \mathbf{P} (\mathbf{I}_n \otimes \mathbf{e}_i) + \mathbf{U},
\tag{14}
$$

where $\mathbf{e}_i \equiv \mathbf{e}_i^p$ is the i-th column of $\mathbf{I}_p$, $\mathbf{E} = \mathcal{X}'_{(2)} \mathcal{A}_{(2)}$, and $\mathbf{P}$ is the $np \times np$ commutation matrix in which $\mathbf{P}_{k,l} = 1$ if there exist integers i and j such that $k = (i - 1)n + j$ and

$l = (j-1)p + i$; otherwise, $\mathbf{P}_{k,l} = 0$.

Proof of Lemma 1. First, note that from

$$\mathbf{Y} = \mathcal{X}_{(1)}\mathcal{A}'_{(1)} + \mathbf{U},$$

the element $\mathbf{Y}_{t,l}$ can be expressed as

$$\mathbf{Y}_{t,l} = \sum_{i=1}^{np} \mathcal{X}_{(1),t,i}\mathcal{A}_{(1),l,i} + \mathbf{U}_{t,l} = \sum_{i_1=1}^{n}\sum_{i_3=1}^{p} \mathcal{X}_{t,i_1,i_3}\mathcal{A}_{l,i_1,i_3} + \mathbf{U}_{t,l},$$

where $i = (i_3 - 1)n + i_1$. Now, it suffices to show that the (t, l) element of the sum on the right-hand side of equation (14) has the same expression and is thus equal to $\mathbf{Y}_{t,l}$.

We start by describing the typical element of the $Tp \times np$ matrix $\mathbf{EP}$. For any (i, j), obtain the pairs of integers (i_1, i_3) and (j_1, j_3) such that $i = (i_3 - 1)T + i_1$ and $j = (j_3 - 1)p + j_1$. Then, by direct computation, we have

$$
\begin{aligned}
(\mathbf{EP})_{i,j} =&(\mathbf{EP})_{(i_3-1)T+i_1,(j_3-1)p+j_1} \\
=&\sum_{s=1}^{np} \mathbf{E}_{(i_3-1)T+i_1,s}\mathbf{P}_{s,(j_3-1)p+j_1} \\
=&\sum_{k=1}^{p}\sum_{l=1}^{n} \mathbf{E}_{(i_3-1)T+i_1,(k-1)n+l}\mathbf{P}_{(k-1)n+l,(j_3-1)p+j_1} \\
=&\sum_{k=1}^{p}\sum_{l=1}^{n} \mathbf{E}_{(i_3-1)T+i_1,(k-1)n+l}1(k = j_1, l = j_3) \\
=&\mathbf{E}_{(i_3-1)T+i_1,(j_1-1)n+j_3}.
\end{aligned}
$$

Then, the (t, l) element of the sum on the right-hand side of equation (14) is given by:

$$\sum_{k=1}^{p}\sum_{i=1}^{Tp}\sum_{j=1}^{np}(\mathbf{e}_k' \otimes \mathbf{I}_T)_{t,i}(\mathbf{EP})_{i,j}(\mathbf{I}_n \otimes \mathbf{e}_k)_{j,l} + \mathbf{U}_{t,l}$$

$$= \sum_{k=1}^{p}\sum_{i_1=1}^{T}\sum_{i_3=1}^{p}\sum_{j_3=1}^{n}\sum_{j_1=1}^{p}(\mathbf{e}_k' \otimes \mathbf{I}_T)_{t,(i_3-1)T+i_1}\mathbf{E}_{(i_3-1)T+i_1,(j_1-1)n+j_3}(\mathbf{I}_n \otimes \mathbf{e}_k)_{(j_3-1)p+j_1,l} + \mathbf{U}_{t,l}$$

$$= \sum_{k=1}^{p}\sum_{i_1=1}^{T}\sum_{i_3=1}^{p}\sum_{j_3=1}^{n}\sum_{j_1=1}^{p}(\mathbf{e}_{k,i_3}'\mathbf{I}_{T,t,i_1})\mathbf{E}_{(i_3-1)T+i_1,(j_1-1)n+j_3}(\mathbf{I}_{n,j_3,l}\mathbf{e}_{k,j_1}) + \mathbf{U}_{t,l}$$

$$= \sum_{k=1}^{p}\sum_{i_1=1}^{T}\sum_{i_3=1}^{p}\sum_{j_3=1}^{n}\sum_{j_1=1}^{p}1(k = i_3, t = i_1)\mathbf{E}_{(i_3-1)T+i_1,(j_1-1)n+j_3}1(j_3 = l, k = j_1) + \mathbf{U}_{t,l}$$

$$= \sum_{k=1}^{p}\mathbf{E}_{(k-1)T+t,(k-1)n+l} + \mathbf{U}_{t,l}$$

$$= \sum_{k=1}^{p}\sum_{s=1}^{n}\mathcal{X}_{t,s,k}\mathcal{A}_{l,s,k} + \mathbf{U}_{t,l},$$

where the last equality holds because of (13). Hence, we have shown that the (t, l) element of the sum on the right-hand side is $\mathbf{Y}_{t,l}$, thus completing the proof. $\qquad\square$

Proof of Proposition 1. We begin by taking the transpose of both sides of equation (14) in Lemma 1:

$$\mathbf{Y}' = \sum_{i=1}^{p}(\mathbf{I}_n \otimes \mathbf{e}_i')\mathbf{P}'\mathbf{E}'(\mathbf{e}_i \otimes \mathbf{I}_T) + \mathbf{U}'.$$

Next, vectorizing both sides of the above equation, we have

$$
\begin{aligned}
\mathrm{vec}(\mathbf{Y}') &= \sum_{i=1}^{p} \left((\mathbf{e}_i' \otimes \mathbf{I}_T) \otimes (\mathbf{I}_n \otimes \mathbf{e}_i') \mathbf{P}' \right) \mathrm{vec}(\mathbf{E}') + \mathrm{vec}(\mathbf{U}') \\
&= \sum_{i=1}^{p} (\mathbf{P}_{i2} \otimes \mathbf{P}_{i1}) \mathrm{vec}(\mathbf{E}') + \mathrm{vec}(\mathbf{U}') \\
&= \sum_{i=1}^{p} (\mathbf{P}_{i2} \otimes \mathbf{P}_{i1}) \mathrm{vec}(\mathcal{A}'_{(2)} \mathcal{X}_{(2)}) + \mathrm{vec}(\mathbf{U}') \\
&= \sum_{i=1}^{p} (\mathbf{P}_{i2} \otimes \mathbf{P}_{i1}) \mathrm{vec}(\mathbf{\Theta}_{-2} \mathbf{\Theta}_2' \mathcal{X}_{(2)}) + \mathrm{vec}(\mathbf{U}') \qquad (15) \\
&= \sum_{i=1}^{p} (\mathbf{P}_{i2} \otimes \mathbf{P}_{i1})(\mathcal{X}'_{(2)} \otimes \mathbf{\Theta}_{-2}) \mathrm{vec}(\mathbf{\Theta}_2') + \mathrm{vec}(\mathbf{U}') \\
&= \sum_{i=1}^{p} (\mathbf{P}_{i2} \mathcal{X}'_{(2)} \otimes \mathbf{P}_{i1} \mathbf{\Theta}_{-2}) \boldsymbol{\theta}_2 + \mathrm{vec}(\mathbf{U}'),
\end{aligned}
$$

where $\mathbf{P}_{i1} = (\mathbf{I}_n \otimes \mathbf{e}_i') \mathbf{P}'$ and $\mathbf{P}_{i2} = \mathbf{e}_i' \otimes \mathbf{I}_T$, $\boldsymbol{\theta}_2 = \mathrm{vec}(\mathbf{\Theta}_2')$. The fourth equality holds because $\mathcal{A}_{(2)} = \mathbf{\Theta}_2 \mathbf{\Theta}_{-2}'$. $\qquad \square$

Appendix B: Data

The quarterly dataset is sourced from the FRED-QD database (McCracken and Ng, 2020) maintained by the Federal Reserve Bank of St. Louis. The sample spans from 1969Q1 to 2024Q1. Table 5 lists the variables and describes how they are transformed. To ensure consistency, each time series is standardized to have zero mean and unit variance.

Table 5: A list of variables and their transformations. Tcode: 1: no transformation; 2: Δx_t; 5: $\Delta \log(x_t)$; 6: $\Delta^2 \log(x_t)$.

Name	Description	T-code
RPI	Real Personal Income	5
INDPRO	IP Index	5
GDP	Real Gross Domestic Product	5
GDPDEFL	GDP deflator	6
DPCERA3M086SBEA	Real PCE	5
CMRMTSPLx	Real M & T Sales	5
HWI	Help-Wanted Index for US	2
HWIURATIO	Help Wanted to Unemployed ratio	2
CLF16OV	Civilian Labor Force	5
UNRATE	Civilian Unemployment Rate	2
PAYEMS	All Employees: Total nonfarm	5
CES0600000007	Hours: Goods-Producing	5
CPIAUCSL	CPI: All Items	6
FEDFUNDS	Effective Federal Funds Rate	2
TB3MS	3-Month T-bill	2
TB6MS	6-Month T-bill	2
GS1	1-Year T-bond	2
GS5	5-Year T-bond	2
GS10	10-Year T-bond	2
AAA	Aaa Corporate Bond Yield	2
BAA	Baa Corporate Bond Yield	2
M1SL	M1 Money Stock	5
M2SL	M2 Money Stock	5
BUSLOANS	Commercial and Industrial Loans	5
NONREVSL	Total Nonrevolving Credit	5
INVEST	Securities in Bank Credit	5
S&P 500	S&P 500	5
S&P div yield	S&P Dividend yield	2
S&P PE ratio	S&P Price/Earnings ratio	5
EXSZUSx	Switzerland / U.S. FX Rate	5
EXJPUSx	Japan / U.S. FX Rate	5
EXUSUKx	U.S. / U.K. FX Rate	5
EXCAUSx	Canada / U.S. FX Rate	5
UEMPMEAN	Average Duration of Unemployment	5
AWHMAN	Hours: Manufacturing	1
ISRATIOx	Inventories to Sales Rati	2
REALLN	Real Estate Loans	5
PPICMM	PPI: Commodities	6
PCEPI	PCE: Chain-type Price Index	6
FPI	Fixed Private Investment	5

Appendix C: Additional Estimation Details

In this appendix we provide technical details on the alternative approach of sampling each $\boldsymbol{\theta}_2^{(r)}$ at a time for $r = 1, \ldots, R$. To that end, we derive the conditional distribution of $\boldsymbol{\theta}_2^{(r)}$ given $\{\boldsymbol{\theta}_2^{(s)}\}_{s \neq r}$ and other model parameters.

We assume the following the marginal prior for $\boldsymbol{\theta}_2^{(r)}$:

$$\boldsymbol{\theta}_2^{(r)} \sim \mathcal{N}\left(\boldsymbol{\theta}_{2,0}^{(r)}, \mathbf{V}_{\boldsymbol{\theta}_2^{(r)}}\right).$$

Next, it follows from equation (15) in the proof of Proposition 1 that

$$\text{vec}(\mathbf{Y}') = \sum_{i=1}^{p} (\mathbf{P}_{i2} \otimes \mathbf{P}_{i1}) \text{vec}(\boldsymbol{\Theta}_{-2}\boldsymbol{\Theta}_2'\mathcal{X}_{(2)}) + \text{vec}(\mathbf{U}')$$

$$= \sum_{i=1}^{p} \text{vec}(\mathbf{P}_{i1}\boldsymbol{\Theta}_{-2}\boldsymbol{\Theta}_2'\mathcal{X}_{(2)}\mathbf{P}_{i2}') + \text{vec}(\mathbf{U}').$$

Noting that $\boldsymbol{\Theta}_2\boldsymbol{\Theta}_{-2}' = \sum_{s=1}^{R} \boldsymbol{\theta}_2^{(s)}\boldsymbol{\theta}_{-2}^{(s)\prime}$, where $\boldsymbol{\theta}_{-2}^{(s)} = \boldsymbol{\theta}_3^{(s)} \otimes \boldsymbol{\theta}_1^{(s)}$, one can express $\mathbf{Y}'$ as:

$$\mathbf{Y}' = \sum_{i=1}^{p} \mathbf{P}_{i1}\boldsymbol{\Theta}_{-2}\boldsymbol{\Theta}_2'\mathcal{X}_{(2)}\mathbf{P}_{i2}' + \mathbf{U}'$$

$$= \sum_{i=1}^{p}\sum_{s=1}^{R} \mathbf{P}_{i1}\boldsymbol{\theta}_{-2}^{(s)}\boldsymbol{\theta}_2^{(s)\prime}\mathcal{X}_{(2)}\mathbf{P}_{i2}' + \mathbf{U}'.$$

Fix r and let $\mathbf{Y}_{2r} = \mathbf{Y}' - \sum_{i=1}^{p}\sum_{s \neq r} \mathbf{P}_{i1}\boldsymbol{\theta}_{-2}^{(s)}\boldsymbol{\theta}_2^{(s)\prime}\mathcal{X}_{(2)}\mathbf{P}_{i2}'$. Vectorize the above equation to get

$$\text{vec}(\mathbf{Y}_{2r}) = \sum_{i=1}^{p} (\mathbf{P}_{i2}\mathcal{X}_{(2)}' \otimes \mathbf{P}_{i1}\boldsymbol{\theta}_{-2}^{(r)})\boldsymbol{\theta}_2^{(r)} + \text{vec}(\mathbf{U}').$$

Therefore, the posterior conditional distribution $(\boldsymbol{\theta}_2^{(r)} \mid \mathbf{Y}, \boldsymbol{\theta}_1, \{\boldsymbol{\theta}_2^{(s)}\}_{s \neq r}, \boldsymbol{\theta}_3, \boldsymbol{\Sigma})$ can be expressed as

$$(\boldsymbol{\theta}_2^{(r)} \mid \mathbf{Y}, \boldsymbol{\theta}_1, \{\boldsymbol{\theta}_2^{(s)}\}_{s \neq r}, \boldsymbol{\theta}_3, \boldsymbol{\Sigma}) \sim \mathcal{N}\left(\widehat{\boldsymbol{\theta}}_2^{(r)}, \mathbf{K}_{\boldsymbol{\theta}_2^{(r)}}^{-1}\right),$$

where

$$\mathbf{K}_{\boldsymbol{\theta}_2^{(r)}} = \mathbf{V}_{\boldsymbol{\theta}_2^{(r)}}^{-1} + \sum_{i=1}^{p}\sum_{j=1}^{p}(\mathcal{X}_{(2)}\mathbf{P}_{i2}' \otimes \boldsymbol{\theta}_{-2}^{(r)\prime}\mathbf{P}_{i1}')\boldsymbol{\Sigma}^{-1}(\mathbf{P}_{j2}\mathcal{X}_{(2)}' \otimes \mathbf{P}_{j1}\boldsymbol{\theta}_{-2}^{(r)})$$

$$\widehat{\boldsymbol{\theta}}_2^{(r)} = \mathbf{K}_{\boldsymbol{\theta}_2^{(r)}}^{-1}\left(\mathbf{V}_{\boldsymbol{\theta}_2^{(r)}}^{-1}\boldsymbol{\theta}_{2,0}^{(r)} + \sum_{i=1}^{p}(\mathcal{X}_{(2)}\mathbf{P}_{i2}' \otimes \boldsymbol{\theta}_{-2}^{(r)\prime}\mathbf{P}_{i1}')\boldsymbol{\Sigma}^{-1}\mathrm{vec}(\mathbf{Y}_{2r})\right).$$

References

ARIAS, J. E., J. F. RUBIO-RAMIREZ, AND M. SHIN (2023): "Macroeconomic Forecasting and Variable Ordering in Multivariate Stochastic Volatility Models," *Journal of Econometrics*, 235(2), 1054–1086.

BAŃBURA, M., D. GIANNONE, AND L. REICHLIN (2010): "Large Bayesian vector auto regressions," *Journal of Applied Econometrics*, 25(1), 71–92.

BERNARDI, M., D. BIANCHI, AND N. BIANCO (2024): "Variational inference for large Bayesian vector autoregressions," *Journal of Business and Economic Statistics*, pp. 1–17.

BERTSCHE, D., AND R. BRAUN (2022): "Identification of structural vector autoregressions by stochastic volatility," *Journal of Business & Economic Statistics*, 40(1), 328–341.

BHATTACHARYA, A., D. PATI, N. S. PILLAI, AND D. B. DUNSON (2015): "Dirichlet–Laplace priors for optimal shrinkage," *Journal of the American Statistical Association*, 110(512), 1479–1490.

BILLIO, M., R. CASARIN, M. IACOPINI, AND S. KAUFMANN (2023): "Bayesian dynamic tensor regression," *Journal of Business & Economic Statistics*, 41(2), 429–439.

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 (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 (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25(2), 400–417.

CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2011): "Forecasting large datasets with Bayesian reduced rank multivariate models," *Journal of Applied Econometrics*, 26(5), 735–761.

CARVALHO, C. M., N. G. POLSON, AND J. G. SCOTT (2010): "The horseshoe estimator for sparse signals," *Biometrika*, 97(2), 465–480.

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

——— (2021): "Minnesota-type adaptive hierarchical priors for large Bayesian VARs," *International Journal of Forecasting*, 37(3), 1212–1226.

——— (2022): "Asymmetric conjugate priors for large Bayesian VARs," *Quantitative Economics*, 13(3), 1145–1169.

——— (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. (2024): "BVARs and stochastic volatility," in *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, ed. by M. P. Clements, and A. B. Galvao. Edward Elgar, Cheltenham, UK.

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., G. KOOP, AND X. YU (2024): "Large order-invariant Bayesian VARs with stochastic volatility," *Journal of Business and Economic Statistics*, 42(2), 825–837.

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.

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.

DOAN, T., R. LITTERMAN, AND C. SIMS (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric reviews*, 3(1), 1–100.

FOLLETT, L., AND C. YU (2019): "Achieving Parsimony in Bayesian VARs with the Horseshoe Prior," *Econometrics and Statistics*, 11, 130–144.

GEFANG, D., G. KOOP, AND A. POON (2023): "Variational Bayesian inference in large Vector Autoregressions with hierarchical shrinkage," *International Journal of Forecasting*, 39(1), 346–363.

GIANNONE, D., M. LENZA, AND G. E. PRIMICERI (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97(2), 436–451.

GRIFFIN, J., AND P. BROWN (2010): "Inference with normal-gamma prior distributions in regression problems," *Bayesian Analysis*, 5(1), 171–188.

HAUZENBERGER, N., F. HUBER, AND G. KOOP (2024): "Macroeconomic forecasting using BVARs," in *Handbook of Research Methods and Applications in Macroeconomic Forecasting*, ed. by M. P. Clements, and A. B. Galvao. Edward Elgar, Cheltenham, UK.

HUBER, F., AND M. FELDKIRCHER (2019): "Adaptive shrinkage in Bayesian vector autoregressive models," *Journal of Business and Economic Statistics*, 37(1), 27–39.

HUBER, F., G. KOOP, AND L. ONORANTE (2019): "Inducing sparsity and shrinkage in time-varying parameter models," *arXiv preprint arXiv:1905.10787*.

KADIYALA, R. K., AND S. KARLSSON (1993): "Forecasting with generalized Bayesian vector auto regressions," *Journal of Forecasting*, 12(3-4), 365–378.

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

KASTNER, G., AND F. HUBER (2020): "Sparse Bayesian vector autoregressions in huge dimensions," *Journal of Forecasting*, 39(7), 1142–1165.

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

KOOP, G. (2013): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28(2), 177–203.

KOOP, G., AND D. KOROBILIS (2013): "Large time-varying parameter VARs," *Journal of Econometrics*, 177(2), 185–198.

——— (2018): "Variational Bayes inference in high-dimensional time-varying parameter models," *Available at SSRN 3246472.*

KOROBILIS, D., AND D. PETTENUZZO (2019): "Adaptive hierarchical priors for high-dimensional vector autoregressions," *Journal of Econometrics*, forthcoming.

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.

LUO, Y., AND J. E. GRIFFIN (2023): "Bayesian inference of vector autoregressions with tensor decompositions," *arXiv preprint arXiv:2211.01727.*

MCCRACKEN, M. W., AND S. NG (2020): "FRED-QD: A quarterly database for macroeconomic research," *Federal Reserve Bank of St. Louis Working Paper*, 2020-005.

PARK, T., AND G. CASELLA (2008): "The Bayesian Lasso," *Journal of the American Statistical Association*, 103(482), 681–686.

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

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.

WANG, D., Y. ZHENG, AND G. LI (2024): "High-dimensional low-rank tensor autoregressive time series modeling," *Journal of Econometrics*, 238(1), 105544.

WANG, D., Y. ZHENG, H. LIAN, AND G. LI (2022): "High-dimensional vector autoregressive time series modeling via tensor decomposition," *Journal of the American Statistical Association*, 117(539), 1338–1356.