Large Order-Invariant Bayesian VARs with Stochastic Volatility*

Joshua C. C. Chan
Purdue University

Gary Koop
University of Strathclyde

Xuewen Yu
Purdue University

This version: June 2023
First version: October 2021

Abstract

Many popular specifications for Vector Autoregressions (VARs) with multivariate stochastic volatility are not invariant to the way the variables are ordered due to the use of a lower triangular parameterization of the error covariance matrix. We show that the order invariance problem in existing approaches is likely to become more serious in large VARs. We propose the use of a specification which avoids the use of this lower triangular parameterization. We show that the presence of multivariate stochastic volatility allows for identification of the proposed model and prove that it is invariant to ordering. We develop a Markov Chain Monte Carlo algorithm which allows for Bayesian estimation and prediction. In exercises involving artificial and real macroeconomic data, we demonstrate that the choice of variable ordering can have non-negligible effects on empirical results when using the non-order invariant approach. In a macroeconomic forecasting exercise involving VARs with 20 variables we find that our order-invariant approach leads to the best forecasts and that some choices of variable ordering can lead to poor forecasts using a conventional, non-order invariant, approach.

*We would like to thank Christiane Baumeister, Robin Braun, Drew Creal, Gergely Gánics, Chenghan Hou, Elmar Mertens, Ulrich Müller, Ivan Petrella, Mikkel Plagborg-Møller, Minchul Shin and Christopher Sims for their constructive comments and suggestions.

Keywords: large vector autoregression, order invariance, stochastic volatility, shrinkage prior

1 Introduction

Vector Autoregressions (VARs) have become one of the most popular models in modern macroeconomics since the seminal work by Sims (1980). In recent years, two prominent developments have occurred in the Baysian VAR literature. First, beginning with Banbura, Giannone, and Reichlin (2010) researchers have been working with large VARs involving dozens or even hundreds of dependent variables.[1] Second, there is an increasing recognition, in most macroeconomic datasets, of the need to allow for time variation in VAR error variances, see, among many others, Clark (2011). Both of these developments have led to VAR specifications which are not invariant to the way the variables are ordered in the VAR. In small VARs, ordering issues can be important, but as we shall show in this paper take on additional importance in large VARs.[2] These considerations motivate the present paper. In it we describe an alternative VAR specification with stochastic volatility (SV) and prove that it is invariant to the ways the variables are ordered. We develop a computationally-efficient Markov Chain Monte Carlo (MCMC) algorithm which is scaleable and allows for Bayesian estimation and forecasting even in high dimensional VARs. We carry out empirical work with artificial and real data which demonstrates the consequences of a lack of order invariance in conventional approaches and that our approach does not suffer from them.

There are two basic insights that underlie our approach. The first is that it is the combination of the use of a lower triangular parameterization of the error covariance matrix $\mathbf{\Sigma}_t$ and the associated priors under this parameterization that lead to order dependence. Beginning with influential early VAR-SV papers such as Cogley and Sargent (2005) and Primiceri (2005), this lower triangular parameterization has been used in numerous papers. However, as noted by Carriero, Clark, and Marcellino (2019), since this multivariate SV specification is constructed based on a lower triangular impact matrix $\mathbf{B}_0$ and priors are independently elicited on $\mathbf{B}_0$ and the SV processes, the implied prior on $\mathbf{\Sigma}_t$ is not order invariant.

[1] In their classic study of the effects of monetary policy, Leeper, Sims, and Zha (1996) estimate a few SVARs of various sizes, including one with 18 variables.

[2] In small VARs, results for a small number of different orderings can be presented as a robustness check. In addition, methods have been proposed for searching over different orderings in papers such as Levy and Lopes (2021) and Wu and Koop (2022). But with a large number of variables a logical ordering of the variables may not present itself and it is not practical to consider every possible ordering.

The second is that SV can be used to identify the impact matrix $\mathbf{B}_0$, see Bertsche and Braun (2020). We adopt a VAR-SV specification that avoids the use of the lower triangular parameterization. If we were working with a homoscedastic VAR, our model would be unidentified. But we let the SV achieve identification. In theory we prove order invariance and, in practice, we show that it works well.

There are a few recent papers that demonstrate the empirical importance of this ordering issue. For example, Bognanni (2018) and Hartwig (2020) show that results from structural analysis based on the reduced-form VARs of Cogley and Sargent (2005) and Primiceri (2005) can be sensitive to how the variables are ordered. Chan, Doucet, León-González, and Strachan (2018) and Chan (2022) demonstrate that estimates of reduced-form error variances based on the lower triangular parameterization can drastically change across different variable orderings. Finally, Arias, Rubio-Ramirez, and Shin (2022) illustrate the empirical relevance of this ordering issue for producing density forecasts using all permutations of 4 macroeconomic variables. The present paper contributes to this growing literature but focuses on the implications for high-dimensional models. In particular, we provide some theoretical analysis to show that this ordering issue becomes much worse in high-dimensional settings. More importantly, we propose a new order-invariant model as well as an efficient algorithm which scales well to high-dimensional VARs.

The remainder of this paper is organized as follows. The next section discusses ordering issues in VAR-SVs. The third section introduces our multivariate SV process and proves that it is identified and order invariant. The fourth section adds the VAR component to the model and discusses Bayesian estimation of the resulting order-invariant VAR-SV. The fifth section uses artificial data to illustrate the performance of our model in terms of accuracy and computation time. The sixth section investigates the practical consequences of using the non-order invariant approach of Cogley and Sargent (2005) using artificial and real macroeconomic data. Finally, section seven carries out a forecasting exercise which demonstrates that our order-invariant approach forecasts well, but that forecasts produced by the model of Cogley and Sargent (2005) are sensitive to the way that the variables are ordered and some ordering choices can lead to poor forecast performance.

2 Ordering Issues in VAR-SVs

Ordering issues in VARs relate to the error covariance matrix, not the VAR coefficients themselves. For instance, Carriero, Clark, and Marcellino (2019) use a VAR-SV which involves a lower triangular parameterization of the error covariance matrix, $\mathbf{\Sigma}_t$. They demonstrate that, conditional on $\mathbf{\Sigma}_t$, the draws of the VAR coefficients are invariant to ordering.[3] Such invariance to ordering of the VAR coefficients holds for all the models discussed in the paper. Accordingly, in this section and the next we focus on the multivariate stochastic volatility process. We say a model is invariant to ordering if the order of the variables does not affect the posterior distribution of the error covariance matrix, subject to permutations. In other words, if the ordering of variables i and j in the VAR is changed, then posterior inference on the error covariances and variances associated with variables i and j is unaffected for all i and j.

Before discussing multivariate stochastic volatility approaches which are not order invariant, it is worthwhile noting that there are a few multivariate SV approaches which are order invariant. For example, some multivariate SV models based on Wishart or inverse-Wishart processes are invariant to ordering; examples include Philipov and Glickman (2006), Asai and McAleer (2009), Chan, Doucet, León-González, and Strachan (2018), Shin and Zhong (2020) and Arias, Rubio-Ramirez, and Shin (2022). Instead of using the Wishart or inverse-Wishart processes, Arias, Rubio-Ramirez, and Shin (2022) introduce the DSC-SV-AH model that uses random walk processes to construct a time-varying correlation matrix based on a new parameterization of correlation matrices proposed in Archakov and Hansen (2021). However, estimation of these SV models is computationally intensive as it typically involves drawing from high-dimensional, non-standard distributions. Consequently, these models are generally not applicable to large datasets. In contrast, the common stochastic volatility models considered in Carriero, Clark, and Marcellino (2016) and Chan (2020) are order invariant and can be estimated quickly even

[3]It is worth noting that one major reason for triangularizing the system is to allow for equation by equation estimation of the model. Any approach which allows for equation by equation estimation can be shown to lead to large computational benefits since there is no need to manipulate the (potentially enormous) posterior covariance matrix of the VAR coefficients in a single block. Triangularization is the most popular equation by equation method and Carriero, Clark, and Marcellino (2019) show how, relative to full system estimation, triangularisation can reduce computational complexity of Bayesian estimation of large VARs from $O(n^6)$ to $O(n^4)$ where n is VAR dimension. This highlights the importance of development of VAR specifications, such as the one in this paper, which allow for equation by equation estimation.

for large systems. However, these models are less flexible as the time-varying error covariance matrix depends on a single SV process (e.g., error variances are always proportional to each other).

Another order invariant approach is based on the discounted Wishart process, which admits efficient filtering and smoothing algorithms for estimation. Multivariate SV models based on this approach are considered in Uhlig (1997), West and Harrison (2006), Prado and West (2010), Bognanni (2018) and Arias, Rubio-Ramirez, and Shin (2022). However, this approach appears to be too tightly parameterized for macroeconomic data and it tends to underperform in terms of both point and density forecasts relative to standard SV models such as Cogley and Sargent (2005) and Primiceri (2005); see, for example, Arias, Rubio-Ramirez, and Shin (2022) for a forecast comparison exercise.

Another model that is potentially order invariant is the factor stochastic volatility model, and recently these models have been applied to large datasets. However, factor models have some well-known drawbacks such as mis-specification concerns associated with assuming a high dimensional volatility process is well approximated with a low dimensional set of factors. Kastner (2019) notes that, if one imposes the usual identification restrictions (e.g., factor loadings are triangular), then the model is not order invariant. Kastner (2019) does not impose those restrictions, arguing that identification of the factor loadings is not necessary for many purposes (e.g. forecasting).

The multivariate stochastic volatility specification that is perhaps most commonly used in macroeconomics is that of Cogley and Sargent (2005). The SV specification of the reduced-form errors, the n-dimensional vector $\mathbf{u}_t$, in Cogley and Sargent (2005) can be written as

$$\mathbf{u}_t = \mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \tag{1}$$

where $\mathbf{B}_0$ is a unit lower triangular matrix with elements b_{ij} and $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}})$ is a diagonal matrix consisting of the log-volatility $\mathbf{h}_t = (h_{1,t}, \ldots, h_{n,t})'$. By assuming that $\mathbf{B}_0$ is lower triangular, the implied covariance matrix on $\mathbf{u}_t$ naturally depends on the order of the elements in $\mathbf{u}_t$. In particular, one can show that under standard assumptions, the variance of the i-th element of $\mathbf{u}_t$, $u_{i,t}$, increases as i increases. Hence, this ordering issue becomes worse in higher dimensional models. In addition, since the stochastic volatility specification in Primiceri (2005) is based on Cogley and Sargent (2005) but with $\mathbf{B}_0$ time-varying, it has a similar ordering issue.

To illustrate this key point, we assume for simplicity that $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. We further assume that the priors on the lower triangular elements of $\mathbf{B}_0$ are independent with prior means 0 and variances 1. Since $\mathbf{B}_0\mathbf{u}_t = \boldsymbol{\varepsilon}_t$, we can obtain the elements of $\mathbf{u}_t$ recursively and compute their variances $\text{Var}(u_{i,t}) = \mathbb{E}u_{i,t}^2$. In particular, we have

$$\begin{aligned}
u_{1,t} &= \varepsilon_{1,t} \\
u_{2,t} &= \varepsilon_{2,t} - b_{2,1}u_{1,t} \\
&\vdots \\
u_{i,t} &= \varepsilon_{i,t} - b_{i,i-1}u_{i-1,t} - \cdots - b_{i,1}u_{1,t}.
\end{aligned}$$

Since $\mathbb{E}(b_{i,j}b_{i,k}) = 0$ for all $j \neq k$, the expectation of any cross-product term in $u_{i,t}^2$ is 0. Hence, we have

$$\begin{aligned}
\mathbb{E}u_{1,t}^2 &= \mathbb{E}\varepsilon_{1,t}^2 = 1 \\
\mathbb{E}u_{2,t}^2 &= \mathbb{E}\varepsilon_{2,t}^2 + \mathbb{E}b_{2,1}^2\mathbb{E}u_{1,t}^2 = 1 + \mathbb{E}u_{1,t}^2 = 2 \\
&\vdots \\
\mathbb{E}u_{i,t}^2 &= \mathbb{E}\varepsilon_{i,t}^2 + \mathbb{E}b_{i,i-1}^2\mathbb{E}u_{i-1,t}^2 + \cdots + \mathbb{E}b_{i,1}^2\mathbb{E}u_{1,t}^2 = 2^{i-1}.
\end{aligned}$$

In other words, the variance of the reduced-form error increases (exponentially) as it is ordered lower in the n-tuple, even though the assumptions about $b_{i,j}$ and $\varepsilon_{i,t}$ are identical across $i = 1, \ldots, n$. In Online Appendix C we provide another illustration using a stochastic volatility specification with a time-varying impact matrix similar to that in Primiceri (2005). There we show that the model has a similar ordering issue.

3 An Order-Invariant Stochastic Volatility Model

3.1 Model Definition

In this section we extend the stochastic volatility specification in Cogley and Sargent (2005) with the goal of constructing an order-invariant model. The key reason why the ordering issue in Cogley and Sargent (2005) occurs is because of the lower triangular assumption for $\mathbf{B}_0$. A simple solution to this problem is to avoid this assumption and

assume that $\mathbf{B}_0$ is an unrestricted square matrix. To that end, let $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ be an $n \times 1$ vector of variables that is observed over the periods $t = 1, \ldots, T$. To fix ideas, consider the following model with zero conditional mean:

$$\mathbf{y}_t = \mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \tag{2}$$

where $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}})$ is diagonal. Here we assume that $\mathbf{B}_0$ is non-singular, but is otherwise unrestricted. Each of the log-volatilities follows a stationary AR(1) process:

$$h_{i,t} = \phi_i h_{i,t-1} + u_{i,t}^h, \quad u_{i,t}^h \sim \mathcal{N}(0, \omega_i^2), \tag{3}$$

for $t = 2, \ldots, T$, where $|\phi_i| < 1$ and the initial condition is specified as $h_{i,1} \sim \mathcal{N}(0, \omega_i^2/(1-\phi_i^2))$. Note that the unconditional mean of the AR(1) process is normalized to be zero for identification. We call this the Order Invariant SV (OI-SV) model.

3.2 Identification and Order-Invariance

The question of identification of our OI-SV model can be dealt with quickly. If there is no SV, it is well-known that $\mathbf{B}_0$ is not identified. In particular, any orthogonal transformation of $\mathbf{B}_0$ gives the same likelihood. But with the presence of SV, Bertsche and Braun (2020) show that $\mathbf{B}_0$ is identified up to permutations and sign changes:

Proposition 1 (Bertsche and Braun (2020))**.** Consider the stochastic volatility model given in (2)-(3) with $\phi_i \neq 0, i = 1, \ldots, n$. Then, $\mathbf{B}_0$ is unique up to permutation of its column and multiplication of its columns by -1.

The assumptions in Proposition 1 are stronger than necessary for identification. In fact, Bertsche and Braun (2020) show that having $n-1$ SV processes is sufficient for identification and derive results relating to partial identification which occurs if there are $n-r$ SV processes with $r > 1$. Intuitively, allowing for time-varying volatility generates non-trivial autocovariances in the squared reduced form errors. These additional moments help identify $\mathbf{B}_0$. We refer readers to Rigobon (2003), Lanne, Lütkepohl, and Maciejowska (2010) and Lewis (2021) for a more detailed discussion of how identification can be achieved in structural VARs via heteroscedasticity.

It is worth emphasizing that Proposition 1 only guarantees $\mathbf{B}_0$ to be identified up to permutations and sign changes. In the estimation we normalize the signs by restricting the diagonal elements of $\mathbf{B}_0$ to be positive. There are different approaches to deal with the permutation or label-switching problem. Here we assume a weakly informative prior on $\mathbf{B}_0$ (e.g., setting the prior mean of $\mathbf{B}_0$ to be an identity matrix) so that column permutations become less likely. Alternatively, one could postprocess the posterior draws to sort them into the correct categories, as suggested in Kaufmann and Schumacher (2019).

Below we outline a few theoretical properties of the stochastic volatility model described in equations (2)–(3). First, we show that the model is order invariant, in the sense that if we permute the order of the dependent variables, the likelihood function implied by this new ordering is the same as that of the original ordering, provided that we permute the parameters accordingly. Stacking $\mathbf{y} = (\mathbf{y}_1', \dots, \mathbf{y}_T')'$ and $\mathbf{h} = (\mathbf{h}_1', \dots, \mathbf{h}_T')'$, note that the likelihood implied by (2) is

$$p(\mathbf{y} \,|\, \mathbf{B}_0, \mathbf{h}) = (2\pi)^{-\frac{nT}{2}} |\det \mathbf{B}_0|^T \prod_{t=1}^{T} |\det \mathbf{D}_t|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T} \mathbf{y}_t' \mathbf{B}_0' \mathbf{D}_t^{-1} \mathbf{B}_0 \mathbf{y}_t},$$

where $|\det \mathbf{C}|$ denotes the absolute value of the determinant of the square matrix $\mathbf{C}$. Now, suppose we permute the order of the dependent variables. More precisely, let $\mathbf{P}$ denote an arbitrary permutation matrix of dimension n, and we define $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$. By the standard change of variable result, the likelihood of $\widetilde{\mathbf{y}} = (\widetilde{\mathbf{y}}_1', \dots, \widetilde{\mathbf{y}}_T')'$ is

$$\begin{aligned}
&|\det \mathbf{P}^{-1}|^T (2\pi)^{-\frac{nT}{2}} |\det \mathbf{B}_0|^T \prod_{t=1}^{T} |\det \mathbf{D}_t|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T} (\mathbf{P}^{-1}\widetilde{\mathbf{y}}_t)' \mathbf{B}_0' \mathbf{D}_t^{-1} \mathbf{B}_0 \mathbf{P}^{-1} \widetilde{\mathbf{y}}_t} \\
&\quad = (2\pi)^{-\frac{nT}{2}} |\det \mathbf{B}_0 \mathbf{P}'|^T \prod_{t=1}^{T} |\det \mathbf{D}_t|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T} \widetilde{\mathbf{y}}_t' \mathbf{P} \mathbf{B}_0' \mathbf{D}_t^{-1} \mathbf{B}_0 \mathbf{P}' \widetilde{\mathbf{y}}_t} \\
&\quad = (2\pi)^{-\frac{nT}{2}} |\det \mathbf{P}\mathbf{B}_0 \mathbf{P}'|^T \prod_{t=1}^{T} |\det \mathbf{P}\mathbf{D}_t \mathbf{P}'|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T} \widetilde{\mathbf{y}}_t' \mathbf{P} \mathbf{B}_0' \mathbf{P}' \mathbf{P} \mathbf{D}_t^{-1} \mathbf{P}' \mathbf{P} \mathbf{B}_0 \mathbf{P}' \widetilde{\mathbf{y}}_t} \\
&\quad = (2\pi)^{-\frac{nT}{2}} |\det \widetilde{\mathbf{B}}_0|^T \prod_{t=1}^{T} |\det \widetilde{\mathbf{D}}_t|^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2}\sum_{t=1}^{T} \widetilde{\mathbf{y}}_t' \widetilde{\mathbf{B}}_0' \widetilde{\mathbf{D}}_t^{-1} \widetilde{\mathbf{B}}_0 \widetilde{\mathbf{y}}_t} = p(\widetilde{\mathbf{y}} \,|\, \widetilde{\mathbf{B}}_0, \widetilde{\mathbf{h}}),
\end{aligned}$$

where $\widetilde{\mathbf{B}}_0 = \mathbf{P}\mathbf{B}_0\mathbf{P}'$, $\widetilde{\mathbf{D}}_t = \mathbf{P}\mathbf{D}_t\mathbf{P}' = \mathrm{diag}(\mathrm{e}^{\widetilde{\mathbf{h}}_t})$ and $\widetilde{\mathbf{h}} = (\widetilde{\mathbf{h}}_1', \dots, \widetilde{\mathbf{h}}_T')$ with $\widetilde{\mathbf{h}}_t = \mathbf{P}\mathbf{h}_t$. Note that we have used the fact that $\mathbf{P}^{-1} = \mathbf{P}'$ and $|\det \mathbf{P}| = 1$, as all permutation matrices are

orthogonal matrices. In addition, since $\mathbf{P}^{-1}\widetilde{\mathbf{y}}_t = \mathbf{y}_t$, the first line in the above derivation is equal to $p(\mathbf{y} \,|\, \mathbf{B}_0, \mathbf{h})$. Hence, we have shown that $p(\mathbf{y} \,|\, \mathbf{B}_0, \mathbf{h}) = p(\widetilde{\mathbf{y}} \,|\, \widetilde{\mathbf{B}}_0, \widetilde{\mathbf{h}})$. We summarize this result in the following proposition.

Proposition 2 (Order Invariance)**.** Let $p(\mathbf{y} \,|\, \mathbf{B}_0, \mathbf{h})$ denote the likelihood of the stochastic volatility model given in (2). Let $\mathbf{P}$ be an arbitrary $n \times n$ permutation matrix and define $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$ and $\widetilde{\mathbf{h}}_t = \mathbf{P}\mathbf{h}_t$. Then, the likelihood with dependent variables $\widetilde{\mathbf{y}}_t$ has exactly the same form but with parameters permuted accordingly. More precisely,

$$p(\mathbf{y} \,|\, \mathbf{B}_0, \mathbf{h}) = p(\widetilde{\mathbf{y}} \,|\, \widetilde{\mathbf{B}}_0, \widetilde{\mathbf{h}}),$$

where $\widetilde{\mathbf{B}}_0 = \mathbf{P}\mathbf{B}_0\mathbf{P}'$ and $\widetilde{\mathbf{h}} = (\widetilde{\mathbf{h}}_1', \ldots, \widetilde{\mathbf{h}}_T')$.

Proposition 2 also makes clear why the usual lower triangular assumption on the impact matrix $\mathbf{B}_0$ implies an order-dependent likelihood: to obtain the same likelihood value, the impact matrix corresponding to the permuted vector $\widetilde{\mathbf{y}}$ should be permuted accordingly via $\widetilde{\mathbf{B}}_0 = \mathbf{P}\mathbf{B}_0\mathbf{P}'$. But if $\mathbf{B}_0$ is lower triangular, in general there does not exist a lower triangular $\widetilde{\mathbf{B}}_0$ such that $\widetilde{\mathbf{B}}_0 = \mathbf{P}\mathbf{B}_0\mathbf{P}'$.

Proposition 2 shows that the likelihood conditional on the log-volatility is order invariant. It turns out that one can also recover the unconditional covariance matrix of the permuted vector $\widetilde{\mathbf{y}}_t$ from the unconditional covariance matrix of $\mathbf{y}_t$ by permuting its rows and columns accordingly. To that end, we next derive the unconditional covariance matrix of the data implied by the stochastic volatility model in (2)-(3). More specifically, given the model parameters $\mathbf{B}_0$, ϕ_i and $\omega_i^2, i = 1, \ldots, n$, suppose we generate the log-volatility processes via (3). It is straightforward to compute the implied unconditional covariance matrix of $\mathbf{y}_t$.

Since the unconditional distribution of $h_{i,t}$ implied by (3) is $\mathcal{N}(0, \omega_i^2/(1-\phi_i^2))$, the unconditional distribution of $\mathrm{e}^{h_{i,t}}$ is log-normal with mean $\mathrm{e}^{\frac{\omega_i^2}{2(1-\phi_i^2)}}$. Hence, the unconditional mean of $\mathbf{D}_t$ is $\mathbf{V}_{\mathbf{D}} = \mathrm{diag}(\mathrm{e}^{\frac{\omega_1^2}{2(1-\phi_1^2)}}, \ldots, \mathrm{e}^{\frac{\omega_n^2}{2(1-\phi_n^2)}})$. It then follows that the unconditional covariance matrix of $\mathbf{y}_t$ can be written as

$$\mathbf{\Sigma} \equiv (\mathbf{B}_0'\mathbf{V}_{\mathbf{D}}^{-1}\mathbf{B}_0)^{-1} = (\overline{\mathbf{B}}_0'\overline{\mathbf{B}}_0)^{-1},$$

where $\overline{\mathbf{B}}_0 = \mathbf{V}_{\mathbf{D}}^{-\frac{1}{2}}\mathbf{B}_0$, i.e., scaling each row of $\mathbf{B}_0$ by $\mathrm{e}^{-\frac{\omega_i^2}{4(1-\phi_i^2)}}$, $i = 1, \ldots, n$. From this expression it is also clear that if we permute the order of the dependent variables via $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$, the unconditional covariance matrix of $\widetilde{\mathbf{y}}_t$ can be obtained by permuting the rows and columns of $\overline{\mathbf{B}}_0$ accordingly. More precisely:

$$\widetilde{\mathbf{\Sigma}} \equiv \mathbf{P}\mathbf{\Sigma}\mathbf{P}' = (\mathbf{P}\overline{\mathbf{B}}_0'\mathbf{P}'\mathbf{P}\overline{\mathbf{B}}_0\mathbf{P}')^{-1} = (\widetilde{\overline{\mathbf{B}}}_0'\widetilde{\overline{\mathbf{B}}}_0)^{-1},$$

where $\widetilde{\overline{\mathbf{B}}}_0 = \mathbf{P}\overline{\mathbf{B}}_0\mathbf{P}'$. Hence, the unconditional variance of any element in $\mathbf{y}_t$ does not depend on its position in the n-tuple.

This establishes order-invariance in the likelihood function defined by the OI-SV. Of course, the posterior will also be order invariant (subject to permutations and sign switches) if the prior is. This will hold in any reasonable case. All it requires is that the prior for the parameters in equation i under one ordering becomes the prior for equation j in a different ordering if variable i in the first ordering becomes variable j in the second ordering.

4 A VAR with an Order-Invariant SV Specification

4.1 Model Definition

We now add the VAR part to our OI-SV model, leading to the OI-VAR-SV which is given by

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_0^{-1}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \tag{4}$$

where $\mathbf{a}$ is an $n \times 1$ vector of intercepts, $\mathbf{A}_1, \ldots, \mathbf{A}_p$ are $n \times n$ matrices of VAR coefficients, $\mathbf{B}_0$ is non-singular but otherwise unrestricted $n \times n$ matrix, and $\mathbf{D}_t = \mathrm{diag}(\mathrm{e}^{h_{1,t}}, \ldots, \mathrm{e}^{h_{n,t}})$ is diagonal.[4] Finally, each of the log-volatility $h_{i,t}$ follows the stationary AR(1) process as specified in (3).

[4]Following the bulk of the literature on large Bayesian VARs, we focus on the case of constant VAR coefficients. The model can be readily extended to accommodate time-varying VAR coefficients, but with additional computational costs.

4.2 Shrinkage Priors

We argued previously that ordering issues are likely to be most important in high dimensional models. With large VARs over-parameterization concerns can be substantial and, for this reason, Bayesian methods involving shrinkage priors are commonly used. In this sub-section, we describe a particular VAR prior with attractive properties. But it is worth noting that any of the standard Bayesian VAR priors (e.g. the Minnesota prior) could have been used and the resulting model would still have been order invariant.

Let $\boldsymbol{\alpha}_i$ denote the VAR coefficients in the i-th equation, $i = 1, \ldots, n$. We consider the Minnesota-type adaptive hierarchical priors proposed in Chan (2021), which have the advantages of both the Minnesota priors (e.g., rich prior beliefs such as cross-variable shrinkage) and modern adaptive hierarchical priors (e.g., heavy-tails and substantial mass around 0). In particular, we construct a Minnesota-type horseshoe prior as follows. For $\alpha_{i,j}$, the j-th coefficient in the i-th equation, let $\kappa_{i,j} = \kappa_1$ if it is a coefficient on an 'own lag' and let $\kappa_{i,j} = \kappa_2$ if it is a coefficient on an 'other lag'. Given the constants $C_{i,j}$ defined below, consider the horseshoe prior on the VAR coefficients (excluding the intercepts) $\alpha_{i,j}, i = 1, \ldots, n, j = 2, \ldots, k$ with $k = np + 1$:

$$(\alpha_{i,j} \,|\, \kappa_1, \kappa_2, \psi_{i,j}) \sim \mathcal{N}(m_{i,j}, \kappa_{i,j}\psi_{i,j}C_{i,j}), \tag{5}$$

$$\sqrt{\psi}_{i,j} \sim \mathcal{C}^+(0, 1), \tag{6}$$

$$\sqrt{\kappa}_1, \sqrt{\kappa}_2 \sim \mathcal{C}^+(0, 1), \tag{7}$$

where $\mathcal{C}^+(0, 1)$ denotes the standard half-Cauchy distribution. κ_1 and κ_2 are the global variance components that are common to, respectively, coefficients of own and other lags, whereas each $\psi_{i,j}$ is a local variance component specific to the coefficient $\alpha_{i,j}$. Lastly, the constants $C_{i,j}$ are obtained as in the Minnesota prior, i.e.,

$$C_{i,j} = \begin{cases} \frac{1}{l^2}, & \text{for the coefficient on the } l\text{-th lag of variable } i, \\ \frac{s_i^2}{l^2 s_j^2}, & \text{for the coefficient on the } l\text{-th lag of variable } j,\ j \neq i, \end{cases} \tag{8}$$

where s_r^2 denotes the sample variance of the residuals from an AR(4) model for the variable $r, r = 1, \ldots, n$. Furthermore, for data in growth rates $m_{i,j}$ is set to be 0; for level data, $m_{i,j}$ is set to be zero as well except for the coefficient associated with the first own lag, which is set to be one.

It is easy to verify that if all the local variances are fixed, i.e., $\psi_{i,j} \equiv 1$, then the Minnesota-type horseshoe prior given in (5)–(7) reduces to the Minnesota prior.[5] Hence, it can be viewed as an extension of the Minnesota prior by introducing a local variance component such that the marginal prior distribution of $\alpha_{i,j}$ has heavy tails. On the other hand, if $m_{i,j} = 0$, $C_{i,j} = 1$ and $\kappa_1 = \kappa_2$, then the Minnesota-type horseshoe prior reduces to the standard horseshoe prior where the coefficients have identical distributions. Therefore, it can also be viewed as an extension of the horseshoe prior (Carvalho, Polson, and Scott, 2010) that incorporates richer prior beliefs on the VAR coefficients, such as cross-variable shrinkage, i.e., shrinking coefficients on own lags differently than other lags.

To facilitate estimation, we follow Makalic and Schmidt (2016) and use the following latent variable representations of the half-Cauchy distributions in (6)-(7):

$$(\psi_{i,j} \,|\, z_{\psi_{i,j}}) \sim \mathcal{IG}(1/2, 1/z_{\psi_{i,j}}), \quad z_{\psi_{i,j}} \sim \mathcal{IG}(1/2, 1), \tag{9}$$

$$(\kappa_l \,|\, z_{\kappa_l}) \sim \mathcal{IG}(1/2, 1/z_{\kappa_l}), \quad z_{\kappa_l} \sim \mathcal{IG}(1/2, 1), \tag{10}$$

for $i = 1, \ldots, n, j = 2, \ldots, k$ and $l = 1, 2$, where $\mathcal{IG}(\alpha, \beta)$ denotes the inverse gamma distribution with density $p(x) = \beta^{\alpha}\Gamma(\alpha)^{-1}x^{-(\alpha+1)}\mathrm{e}^{-\beta/x}$.

Next, we specify the priors on other model parameters. Let $\mathbf{b}_i$ denote the i-th row of $\mathbf{B}_0$ for $i = 1, \ldots, n$, i.e., $\mathbf{B}_0' = (\mathbf{b}_1, \ldots, \mathbf{b}_n)$. We consider independent priors on $\mathbf{b}_i, i = 1, \ldots, n$:[6]

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

For the parameters in the stochastic volatility equations, we assume the priors for $j = 1, \ldots, n$:

$$\phi_j \sim \mathcal{N}(\phi_{0,j}, V_{\phi_j})1(|\phi_j| < 1), \quad \sigma_j^2 \sim \mathcal{IG}(\nu_j, S_j).$$

[5] There are many different versions of "the Minnesota prior". The version that we are using is an independent normal and inverse-Wishart prior that is applicable to VARs with time-varying volatility. See, e.g., Equation (14) in Karlsson (2013) or Equation (13) in Carriero, Clark, and Marcellino (2015).

[6] In the simulations and the empirical application, we set $\mathbf{b}_{0,i} = \mathbf{e}_i$ and $\mathbf{V}_{\mathbf{b}_i} = \mathbf{I}_n$, where $\mathbf{e}_i$ is the i-th column of the identity matrix $\mathbf{I}_n$. This choice of hyperparameters implies an order invariant prior in the sense that if we permute the endogenous variables $\widetilde{\mathbf{y}}_t = \mathbf{P}\mathbf{y}_t$, the prior mean of $\mathbf{P}\mathbf{B}_0\mathbf{P}'$ is also an identity matrix (and all prior variances of the elements of $\mathbf{P}\mathbf{B}_0\mathbf{P}'$ are also 1). Since $\mathbf{B}_0$ is identified up to permutations and sign normalization, this relatively weak prior works well in a range of simulations.

4.3 MCMC Algorithm

In this section we develop a posterior sampler which allows for Bayesian estimation of the order-invariant stochastic volatility model with the Minnesota-type horseshoe prior. For later reference, let $\boldsymbol{\psi}_i = (\psi_{i,1}, \dots, \psi_{i,k})'$ denote the local variance components associated with $\boldsymbol{\alpha}_i$ and define $\boldsymbol{\kappa} = (\kappa_1, \kappa_2)'$. Furthermore, let $\mathbf{z}_{\boldsymbol{\psi}_i} = (z_{\psi_{i,1}}, \dots, z_{\psi_{i,k}})'$ denote the latent variables corresponding to $\boldsymbol{\psi}_i$ and similarly define $\mathbf{z}_{\boldsymbol{\kappa}}$. Next, stack $\mathbf{y} = (\mathbf{y}_1', \dots, \mathbf{y}_T')'$, $\boldsymbol{\alpha} = (\boldsymbol{\alpha}_1', \dots, \boldsymbol{\alpha}_n')'$, $\boldsymbol{\psi} = (\boldsymbol{\psi}_1', \dots, \boldsymbol{\psi}_n')'$, $\mathbf{z}_{\boldsymbol{\psi}} = (\mathbf{z}_{\boldsymbol{\psi}_1}', \dots, \mathbf{z}_{\boldsymbol{\psi}_n}')'$ and $\mathbf{h} = (\mathbf{h}_1', \dots, \mathbf{h}_T')'$ with $\mathbf{h}_t = (h_{1,t}, \dots, h_{n,t})'$. Finally, let $\mathbf{h}_{i,\cdot} = (h_{i,1}, \dots, h_{i,T})'$ represent the vector of log-volatility for the i-th equation, $i = 1, \dots, n$. Then, posterior draws can be obtained by sampling sequentially from each of the full conditional distributions. Below we discuss sampling of $\mathbf{B}_0$ from $p(\mathbf{B}_0 \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\psi}, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = p(\mathbf{B}_0 \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{h})$. The details of the rest of the posterior sampler are given in Online Appendix B. In addition, we report in Online Appendix C the inefficiency factors of the MCMC samples associated with the empirical application. The results indicates that the posterior sampler is efficient in terms of being able to generate MCMC draws with relatively low autocorrelations.

To sample $\mathbf{B}_0$, we adapt the sampling approach in Waggoner and Zha (2003) and Villani (2009) to the setting with stochastic volatility. More specifically, we aim to draw each row of $\mathbf{B}_0$ given all other rows. To that end, we first rewrite (4) as:

$$(\mathbf{Y} - \mathbf{X}\mathbf{A})\mathbf{B}_0' = \mathbf{E}, \tag{11}$$

where $\mathbf{Y}$ is the $T \times n$ matrix of dependent variables, $\mathbf{X}$ is the $T \times k$ matrix of lagged dependent variables with $k = 1 + np$, $\mathbf{A} = (\mathbf{a}, \mathbf{A}_1, \dots, \mathbf{A}_n)'$ is the $k \times n$ matrix of VAR coefficients and $\mathbf{E}$ is the $T \times n$ matrix of errors. Then, for $i = 1, \dots, n$, we have

$$(\mathbf{Y} - \mathbf{X}\mathbf{A})\mathbf{b}_i = \mathbf{E}_i, \quad \mathbf{E}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Omega}_{\mathbf{h}_{i,\cdot}}),$$

where $\mathbf{E}_i$ is the i-th column of $\mathbf{E}$ and $\boldsymbol{\Omega}_{\mathbf{h}_{i,\cdot}} = \text{diag}(\text{e}^{h_{i,1}}, \dots, \text{e}^{h_{i,T}})$. Hence, the full conditional distribution of $\mathbf{b}_i$ is given by

$$\begin{aligned} p(\mathbf{b}_i \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}_{-i}, \mathbf{h}_{i,\cdot}) &\propto |\det \mathbf{B}_0|^T \text{e}^{-\frac{1}{2}\mathbf{b}_i'(\mathbf{Y}-\mathbf{X}\mathbf{A})'\boldsymbol{\Omega}_{\mathbf{h}_{i,\cdot}}^{-1}(\mathbf{Y}-\mathbf{X}\mathbf{A})\mathbf{b}_i} \times \text{e}^{-\frac{1}{2}(\mathbf{b}_i-\mathbf{b}_{0,i})'\mathbf{V}_{\mathbf{b}_i}^{-1}(\mathbf{b}_i-\mathbf{b}_{0,i})} \\ &\propto |\det \mathbf{B}_0|^T \text{e}^{-\frac{1}{2}(\mathbf{b}_i-\widehat{\mathbf{b}}_i)'\mathbf{K}_{\mathbf{b}_i}(\mathbf{b}_i-\widehat{\mathbf{b}}_i)} \end{aligned} \tag{12}$$

where

$$\mathbf{K}_{\mathbf{b}_i} = \mathbf{V}_{\mathbf{b}_i}^{-1} + (\mathbf{Y} - \mathbf{XA})'\mathbf{\Omega}_{\mathbf{h}_{i,\cdot}}^{-1}(\mathbf{Y} - \mathbf{XA}), \quad \widehat{\mathbf{b}}_i = \mathbf{K}_{\mathbf{b}_i}^{-1}\mathbf{V}_{\mathbf{b}_i}^{-1}\mathbf{b}_{0,i}.$$

The above full conditional posterior distribution is non-standard and direct sampling is infeasible. Fortunately, Waggoner and Zha (2003) develop an efficient algorithm to sample $\mathbf{b}_i$ in the case when $\widehat{\mathbf{b}}_i = \mathbf{0}$. Villani (2009) further generalizes this sampling approach for non-zero $\widehat{\mathbf{b}}_i$.

The key idea is to show that $\mathbf{b}_i$ has the same distribution as a linear transformation of an n-vector that consists of one absolute normal and $n-1$ normal random variables. A random variable Z follows the absolute normal distribution $\mathcal{AN}(\mu, \rho)$ if it has the density function

$$f_{\mathcal{AN}}(z; \mu, \rho) = c|z|^{\frac{1}{\rho}} \mathrm{e}^{-\frac{1}{2\rho}(z-\mu)^2}, \quad z \in \mathbb{R}, \rho \in \mathbb{R}^+, \mu \in \mathbb{R},$$

where c is a normalizing constant.

To formally state the results, let $\mathbf{C}_i$ denote the Cholesky factor of $T^{-1}\mathbf{K}_{\mathbf{b}_i}$ such that $\mathbf{K}_{\mathbf{b}_i} = T\mathbf{C}_i\mathbf{C}_i'$. Let $\mathbf{F}_{-i}$ represent the matrix $\mathbf{F}$ with the i-th column deleted and $\mathbf{F}^{\perp}$ be the orthogonal complement of $\mathbf{F}$. Furthermore, define $\mathbf{v}_1 = \mathbf{C}_i^{-1}(\mathbf{B}_{0,-i}')^{\perp}/||\mathbf{C}_i^{-1}(\mathbf{B}_{0,-i}')^{\perp}||$, where $(\mathbf{B}_{0,-i}')^{\perp} \equiv ((\mathbf{B}_0')_{-i})^{\perp}$, and let $(\mathbf{v}_2, \ldots, \mathbf{v}_n) = \mathbf{v}_1^{\perp}$. Then, we claim that

$$(\mathbf{b}_i \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}_{-i}, \mathbf{h}_{i,\cdot}) \overset{\mathrm{d}}{=} (\mathbf{C}_i')^{-1} \sum_{j=1}^{n} \xi_j \mathbf{v}_j,$$

where $\xi_1 \sim \mathcal{AN}(\widehat{\xi}_1, T^{-1})$ and $\xi_j \sim \mathcal{N}(\widehat{\xi}_j, T^{-1}), j = 2, \ldots, n$ with $\widehat{\xi}_j = \widehat{\mathbf{b}}_i'\mathbf{C}_i\mathbf{v}_j$.

To prove the claim, let $\mathbf{b}_i = (\mathbf{C}_i')^{-1}\sum_{j=1}^{n}\xi_j\mathbf{v}_j$. It suffices to show that if we substitute $\mathbf{b}_i$ into (12), $\xi_1, \ldots, \xi_n$ are independent with $\xi_1 \sim \mathcal{AN}(\widehat{\xi}_1, T^{-1})$ and $\xi_j \sim \mathcal{N}(\widehat{\xi}_j, T^{-1}), j = 2, \ldots, n$. To that end, note that by construction, $\mathbf{v}_1, \ldots, \mathbf{v}_n$ forms an orthonormal basis

of $\mathbb{R}^n$, particularly $\sum_{j=1}^n \mathbf{v}_j\mathbf{v}_j' = \mathbf{I}_n$. We first write the quadratic form in (12) as:

$$\begin{aligned}(\mathbf{b}_i - \widehat{\mathbf{b}}_i)'\mathbf{K}_{\mathbf{b}_i}(\mathbf{b}_i - \widehat{\mathbf{b}}_i) &= T\left(\sum_{j=1}^n \xi_j\mathbf{v}_j - \mathbf{C}_i'\widehat{\mathbf{b}}_i\right)'\left(\sum_{j=1}^n \xi_j\mathbf{v}_j - \mathbf{C}_i'\widehat{\mathbf{b}}_i\right)\\ &= T\left(\sum_{j=1}^n \xi_j^2 - 2\sum_{j=1}^n \xi_j\widehat{\mathbf{b}}_i'\mathbf{C}_i\mathbf{v}_j + \widehat{\mathbf{b}}_i'\mathbf{C}_i\left(\sum_{j=1}^n \mathbf{v}_j\mathbf{v}_j'\right)\mathbf{C}_i\widehat{\mathbf{b}}_i\right)\\ &= T\sum_{j=1}^n(\xi_j - \widehat{\xi}_j)^2.\end{aligned}$$

Next, note that by construction, $(\mathbf{v}_2, \ldots, \mathbf{v}_n)$ spans the same space as $\mathbf{B}_{0,-i}'$. Hence, it follows that

$$|\det \mathbf{B}_0| = |\det \mathbf{B}_0'| = |\det(\mathbf{b}_1, \ldots, \mathbf{b}_{i-1}, (\mathbf{C}_i')^{-1}\sum_{j=1}^n \xi_j\mathbf{v}_j, \mathbf{b}_{i+1}, \ldots, \mathbf{b}_n)| \propto |\xi_1|.$$

Finally, substituting the quadratic form and the determinant into (12), we obtain

$$\begin{aligned}p(\mathbf{b}_i \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}_{-i}, \mathbf{h}_{i,\cdot}) \propto& |\xi_1|^T \mathrm{e}^{-\frac{T}{2}\sum_{j=1}^n(\xi_j - \widehat{\xi}_j)^2}\\ =& |\xi_1|^T \mathrm{e}^{-\frac{T}{2}(\xi_1 - \widehat{\xi}_1)^2}\prod_{j=2}^n \mathrm{e}^{-\frac{T}{2}(\xi_j - \widehat{\xi}_j)^2}.\end{aligned}$$

In other words, $\xi_1 \sim \mathcal{AN}(\widehat{\xi}_1, T^{-1})$ and $\xi_j \sim \mathcal{N}(\widehat{\xi}_j, T^{-1}), j = 2, \ldots, n$, and we have proved the claim.

In order to use the above result, we need an efficient way to sample from $\mathcal{AN}(\widehat{\xi}_1, T^{-1})$. This can be done by using the 2-component normal mixture approximation considered in Appendix C of Villani (2009). In addition, the orthogonal complement of $\mathbf{v}_1$, namely, $\mathbf{v}_1^\perp$, can be obtained using the singular value decomposition.[7] In the sampler we fix the sign of the i-th element of $\mathbf{b}_i$ to be positive (so that the diagonal elements of $\mathbf{B}_0$ are positive). This is done by multiplying the draw $\mathbf{b}_i$ by -1 if its i-th element is negative. Details of the other steps of the posterior sampler are given in Online Appendix B.

[7] In MATLAB, the orthogonal complement of $\mathbf{v}_1$ can be obtained using `null`$(\mathbf{v}_1')$.

5 Experiments with Artificial Data

In this section we first present results on two artificial data experiments to illustrate how the new model performs under DGPs with and without stochastic volatility. In the first experiment, we generate a dataset from the VAR in (4) with $n = 3$, $T = 500$, $p = 4$ and the stochastic volatility processes as specified in (3). We then estimate the model using the posterior sampler outlined in the previous section. The first dataset is generated as follows. First, the intercepts are drawn independently from $\mathcal{U}(-10, 10)$, the uniform distribution on the interval $(-10, 10)$. For the VAR coefficients, the diagonal elements of the first VAR coefficient matrix are iid $\mathcal{U}(0, 0.5)$ and the off-diagonal elements are from $\mathcal{U}(-0.2, 0.2)$; all other elements of the j-th ($j > 1$) VAR coefficient matrices are iid $\mathcal{N}(0, 0.1^2/j^2)$. For the impact matrix $\mathbf{B}_0$, the diagonal elements are iid $\mathcal{U}(0.5, 2)$, whereas the off-diagonal elements are iid $\mathcal{N}(0, 1)$. Finally, for the SV processes, we set $\phi_i = 0.95$ and $\omega_i^2 = 0.05, i = 1, \ldots, n$. The results of the artificial data experiments are reported in Figures 1-3.

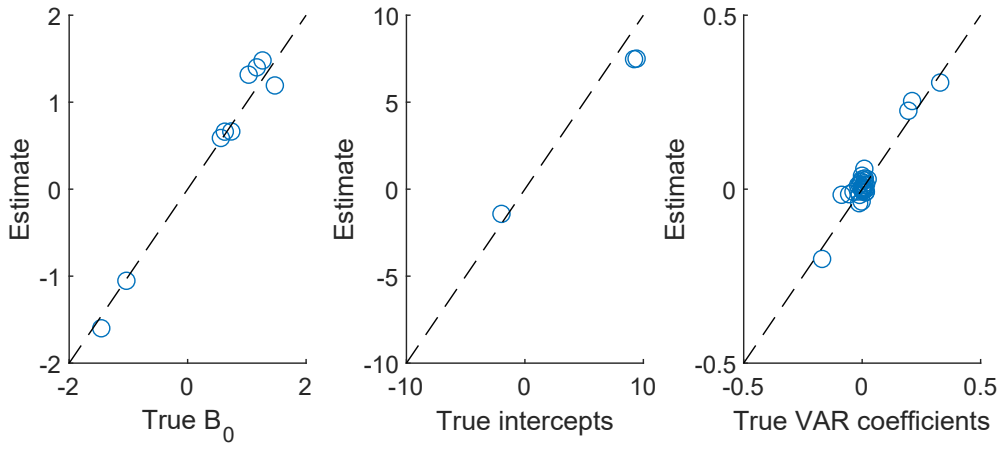

Figure 1: Scatter plot of the posterior means of $\mathbf{B}_0$ (left panel), intercepts (middle panel) and VAR coefficients (right panel) against true values from a DGP with stochastic volatility.

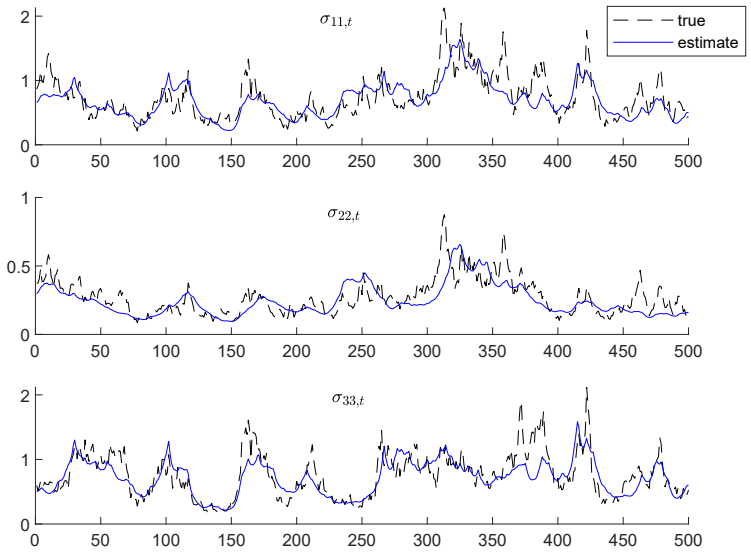

Figure 2: Estimated time-varying reduced-form variances, where $\sigma_{ii,t}$ denote the i-th diagonal element of $\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{D}_t(\mathbf{B}_0^{-1})'$, from a DGP with stochastic volatility.

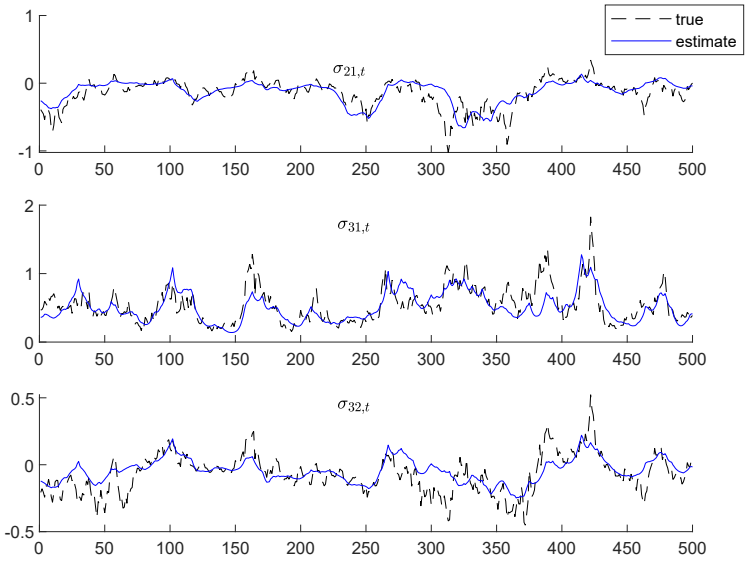

Figure 3: Estimated time-varying reduced-form covariances, where $\sigma_{ij,t}$ denote the (i,j) element of $\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{D}_t(\mathbf{B}_0^{-1})'$, from a DGP with stochastic volatility.

Overall, all the estimates track the true values closely. In particular, the new model is able to recover the time-varying reduced-form covariance matrices. In addition, it is also evident that $\mathbf{B}_0$ can be estimated accurately.

In the second experiment, we generate data from the same VAR but with the stochastic volatility component turned off so that the errors are homoscedastic (in particular, we set $h_{i,t} = 0$). In other words, we have generated data from a homoscedastic, unidentified model. But we are estimating it using the (mis-specified) heteroscedastic model with the posterior sampler outlined in the previous sections. The results are reported in Figures 4-6.

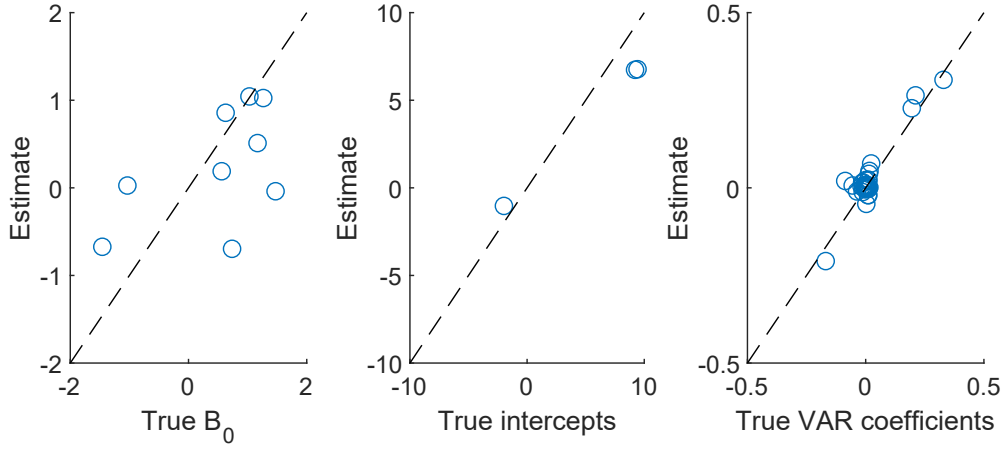

Figure 4: Scatter plot of the posterior means of $\mathbf{B}_0$ (left panel), intercepts (middle panel) and VAR coefficients (right panel) against true values from a DGP without stochastic volatility.

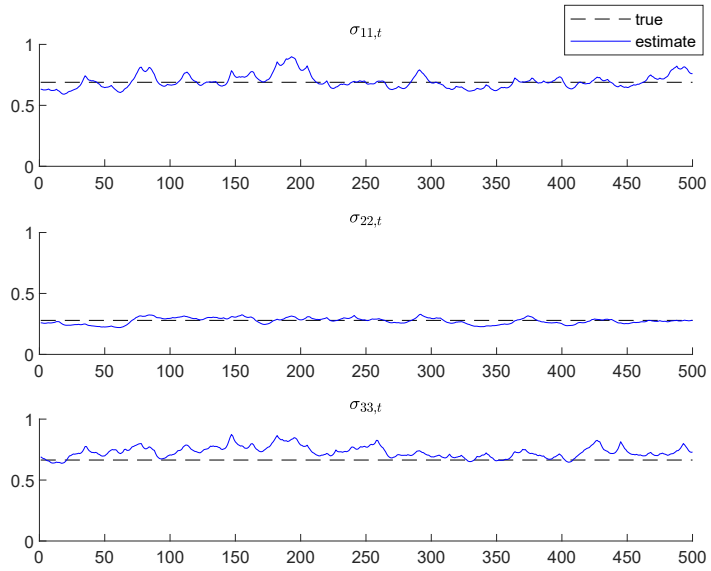

Figure 5: Estimated time-varying reduced-form variances, where $\sigma_{ii,t}$ denote the i-th diagonal element of $\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{D}_t(\mathbf{B}_0^{-1})'$, from a DGP without stochastic volatility

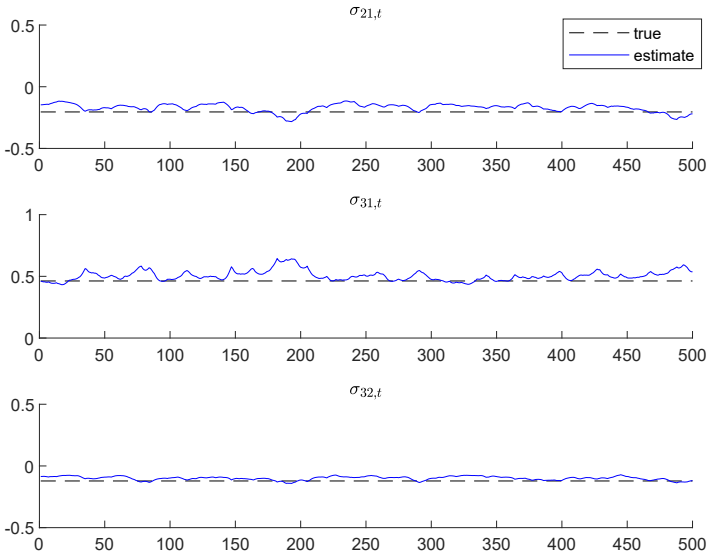

Figure 6: Estimated time-varying reduced-form covariances, where $\sigma_{ij,t}$ denote the (i,j) element of $\mathbf{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{D}_t(\mathbf{B}_0^{-1})'$, from a DGP without stochastic volatility.

When the DGP does not have stochastic volatility, elements of the impact matrix $\mathbf{B}_0$ are much harder to pin down (as expected since these parameters are not identified). But it is interesting to note that the estimates of the time-varying variances are still able to track the true values fairly closely (as these reduced-form error variances are identified). The VAR coefficients, as well, are well-estimated.

Finally, we document the runtimes of estimating OI-VAR-SV models of different dimensions to assess how well the posterior sampler scales to higher dimensions. More specifically, we report in Table 1 the computation times (in minutes) to obtain 10,000 posterior draws from the proposed OI-VAR-SV model of dimensions $n = 10, 20, 50$ and sample sizes $T = 300, 800$. The posterior sampler is implemented in MATLAB on a desktop with an Intel Core i7-7700 @3.60 GHz processor and 64 GB memory. It is evident from the table that for typical applications with 15-30 variables, the OI-VAR-SV model can be estimated quickly. More generally, its estimation time is comparable to that of the triangular model in Carriero, Clark, and Marcellino (2019).

Table 1: The computation times (in minutes) to obtain 10,000 posterior draws from the proposed order invariant VAR-SV model with n variables and T observations. All VARs have $p = 4$ lags.

$T = 300$			$T = 800$		
$n = 10$	$n = 20$	$n = 50$	$n = 10$	$n = 20$	$n = 50$
2.5	17.5	233	6.8	39.7	630

6 Sensitivity to Ordering in the VAR-SV of Cogley and Sargent (2005)

In the preceding sections, we have developed Bayesian methods for order-invariant inference in VARs with SV. In this section, we illustrate the importance of this by showing the degree to which results are sensitive to ordering in the VAR-SV of Cogley and Sargent (2005), hereafter CS-VAR-SV, which is one of the most popular VAR-SV specifications in empirical macroeconomics.[8] We do this using both simulated and macroeconomic data.

[8]In the original model of Cogley and Sargent (2005), the log-volatility follows a unit root process. Here we consider a version in which the log-volatility follows a stationary AR(1) process (with a non-zero mean) so that it is directly comparable to the OI-VAR-SV.

To ensure comparability, we use the same prior for the CS-VAR-SV as for our OI-VAR-SV. In particular, we set the hyperparameters $\phi_{0,j} = 0.95$, $V_{\phi_j} = 0.05^2$, $\nu_j = 3$ and $S_j = 0.1, j = 1, \ldots, n$, for both models. The only exception is the prior on $\mathbf{B}_0$. For the CS-VAR-SV, $\mathbf{B}_0$ is unit lower triangular, and we assume the lower triangular elements have prior distribution $\mathcal{N}(0, 1)$. For OI-VAR-SV, we assume the same $\mathcal{N}(0, 1)$ prior for the off-diagonal elements of $\mathbf{B}_0$, whereas the diagonal elements are iid $\mathcal{N}(1, 1)$. MCMC estimation of the CS-VAR-SV is carried out using the algorithm of Carriero, Chan, Clark, and Marcellino (2022).

6.1 A Simulation Experiment

As discussed in Primiceri (2005), the lower triangular parameterization used in the model of Cogley and Sargent (2005) is unable to accommodate certain co-volatility patterns. To illustrate this point, we simulate $T = 500$ observations from the model in (3) and (4) with $p = 4$ and $n = 3$. To investigate the consequences of the lower triangular restriction we choose a non-triangular data generating process (DGP). We emphasise that the presence of SV in the DGP means that the model is identified. We set

$$\mathbf{B}_0 = \begin{pmatrix} 1 & -0.8 & -0.8 \\ 0.8 & 1 & -0.8 \\ 0.8 & 0.8 & 1 \end{pmatrix}.$$

The VAR coefficients and volatilities are simulated in exactly the same way as described in Section 5. Figure 7 reports the estimated time-varying reduced-form error covariance matrices produced by the CS-VAR-SV and OI-VAR-SV models. It is clear from the figure that the estimates from the CS-VAR-SV model tend to be flat and do not capture the time-variation in the error covariances well. This is perhaps not surprising as the $\mathbf{B}_0$ is restricted to be lower triangular in the CS-VAR-SV. However, the OI-VAR-SV does track the true $\sigma_{ij,t}$ quite well. Since error covariances play an important role in features of interest such as impulse responses, this illustrates the potential negative consequences of working with triangularized systems.

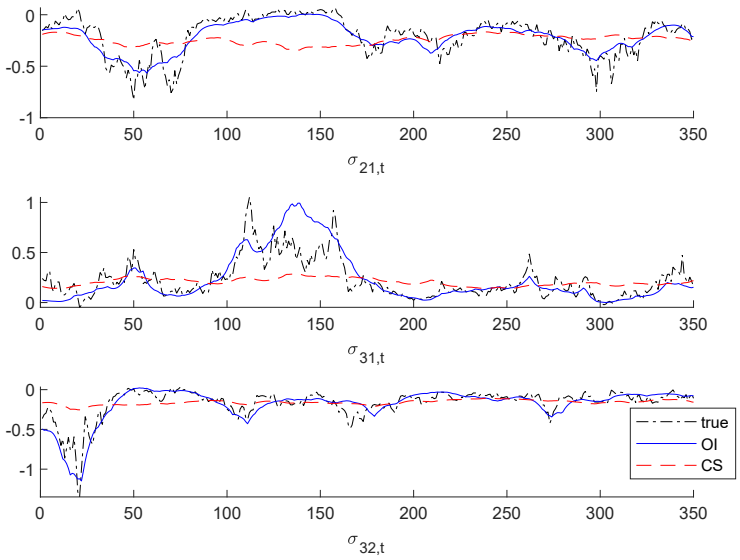

Figure 7: Estimated time-varying reduced-form covariances from the OI-VAR-SV and CS-VAR-SV models, where $\sigma_{ij,t}$ denote the (i,j) element of $\boldsymbol{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{D}_t(\mathbf{B}_0^{-1})'$.

6.2 Ordering Issues in a 20-Variable VAR

Next we provide a simple illustration of how the choice of ordering can influence empirical results in large VARs using a dataset of 20 popular US monthly variables obtained from the FRED-MD database (McCracken and Ng, 2016). The sample period is from 1959:03 to 2019:12. These variables, along with their transformations are listed in Online Appendix A. Four core variables, namely, industrial production, the unemployment rate, PCE inflation and the Federal funds rate, are ranked as the first to the fourth variables, respectively. The remaining 16 variables are ordered as in Carriero, Clark, and Marcellino (2019). We estimate the OI-VAR-SV and CS-VAR-SV models with the variables in this order (we call these models OI-VAR-SV-1 and CS-VAR-SV-1). Then we reverse the order of the variables and re-estimate them (these models are OI-VAR-SV-2 and CS-VAR-SV-2).

Estimates (posterior means) of reduced-form variances and covariances of selected variables are reported in Figure 8 and Figure 9. That is, the first figure reports selected diagonal elements of $\boldsymbol{\Sigma}_t = \mathbf{B}_0^{-1}\mathbf{D}_t(\mathbf{B}_0^{-1})'$ under two different variable orderings, whereas

the second fig

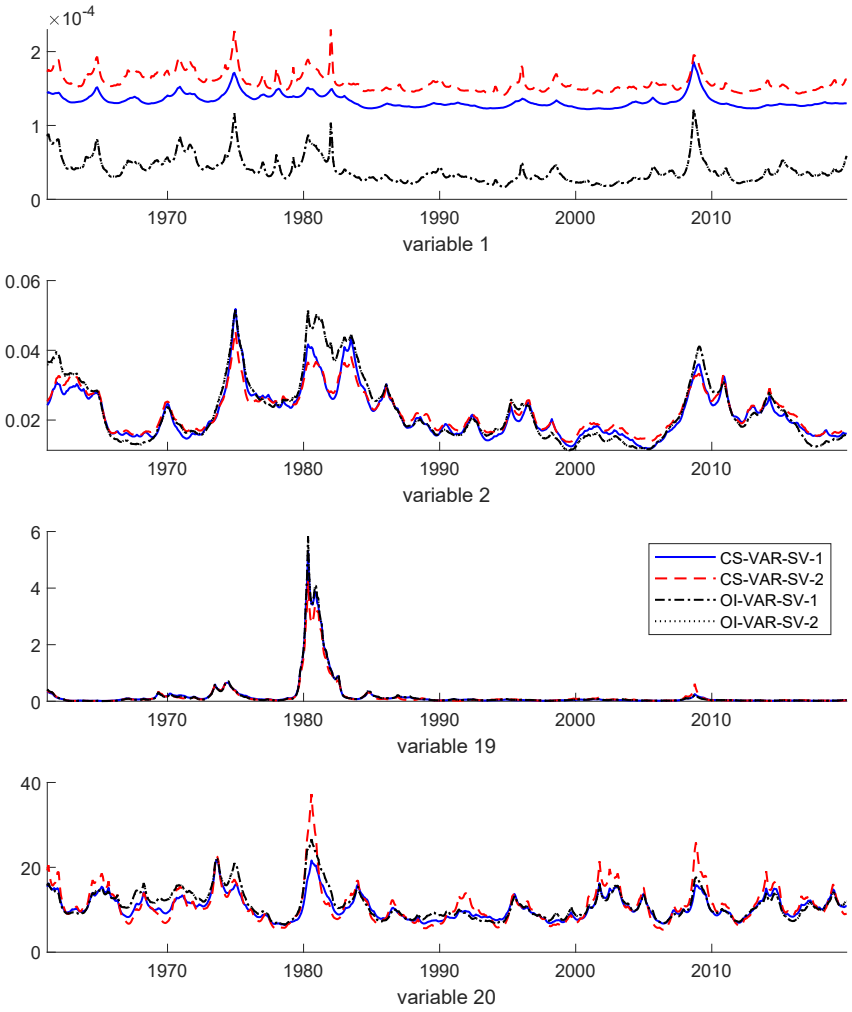

Figure 8: Estimates of time-varying reduced-form error variances from CS-VAR-SV and OI-VAR-SV under two different variable orderings

There are three points to note about these figures. First, as expected, estimates produced by the OI-VAR-SV model under the two different variable orderings are identical (up to MCMC approximation error). Second, estimates produced by the CS-VAR-SV model under the two different variable orderings are often similar to one another, but occasionally differ, particularly in periods of high volatility. This is in line with the results in Hartwig (2020), who shows that the reduced-form covariances in the application of Primiceri (2005) can be substantially different across alternative orderings of variables. Third, estimates from the CS-VAR-SV models are often similar to the OI-VAR-SV models, but

sometimes are substantially different, particularly in the error covariances.

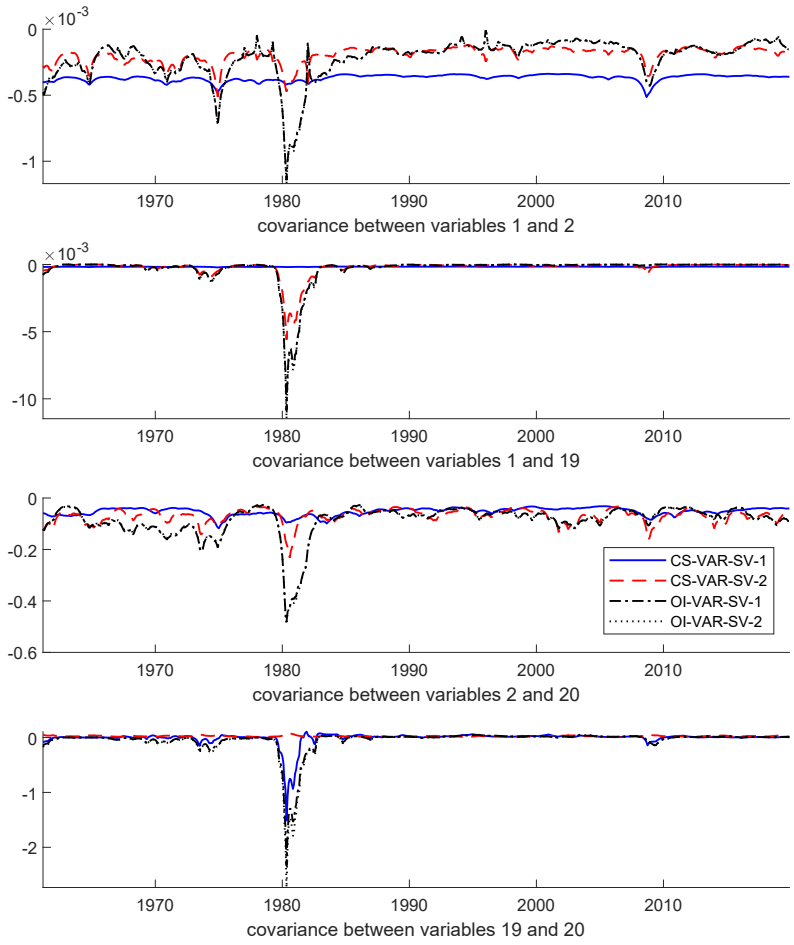

Figure 9: Estimates of time-varying reduced-form error covariances from from CS-VAR-SV and OI-VAR-SV under two different variable orderings

Estimates of other parameters that are not directly related to log-volatility can also be heavily influenced by the choice of ordering. For example, Table 2 reports the posterior means of the shrinkage hyperparameters κ_1 and κ_2 under the two models with two different orderings of the variables. As is clear from the table, the estimates of κ_1 and κ_2 can differ substantially when using the CS-VAR-SV model, depending on how the variables are ordered. For example, the estimate of κ_2 increases by 84% when one reverses the

order of the variables. Of course, under the proposed order-invariant model the estimates are the same (up to MCMC approximation error).

Table 2: Posterior means of κ_1 and κ_2 from the proposed order-invariant stochastic volatility model under two different orderings (OI-VAR-SV-1 and OI-VAR-SV-2), as well as those from the model of Cogley and Sargent (2005) (CS-VAR-SV-1 and CS-VAR-SV-2).

	CS-VAR-SV-1	CS-VAR-SV-2	OI-VAR-SV-1	OI-VAR-SV-2
κ_1	0.0177	0.0170	0.0356	0.0358
κ_2	2.34×10^{-6}	4.30×10^{-6}	2.36×10^{-6}	2.25×10^{-6}

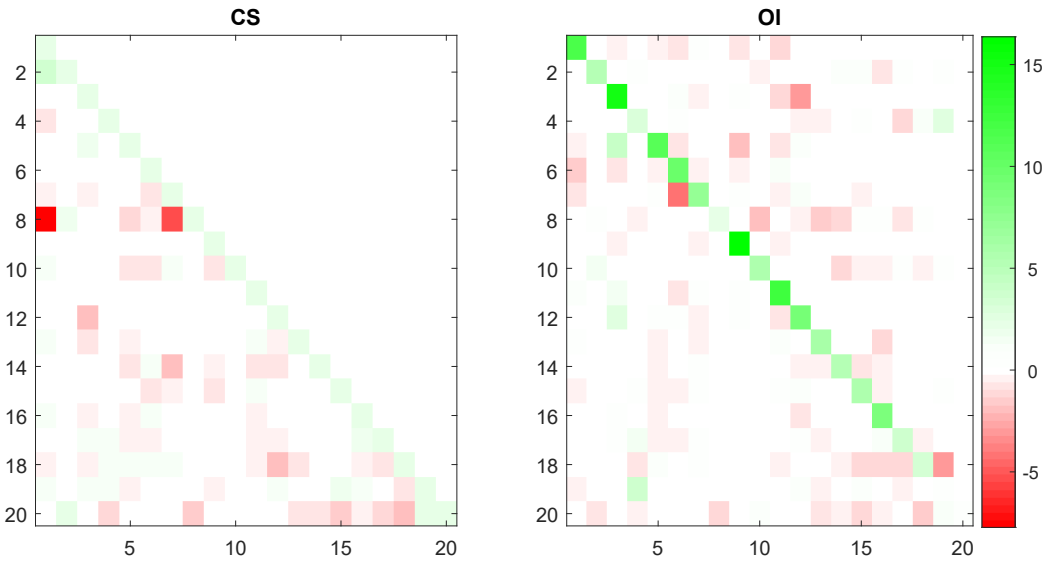

Figure 10: Posterior means of $\mathbf{B}_0$ using CS-VAR-SV-1 (left panel) and OI-VAR-SV-1 (right panel)

We next present results relating to $\mathbf{B}_0$ which highlight the differences between our OI-VAR-SV-1 and the CS-VAR-SV-1.[9] Figure 10 reports the posterior means of $\mathbf{B}_0$ of these two models. Note that the two panels of the figure are very different. Part of this difference is due to the fact that, under OI-VAR-SV, the SV processes are assumed to have zero unconditional means and the diagonal elements of $\mathbf{B}_0$ are unrestricted. Hence,

[9]Note that we are comparing the two models under the first ordering of the variables. The comparable figure using the second ordering reveals similar patterns.

the diagonal elements of $\mathbf{B}_0$ play a key role in adjusting the scale of the error variances. In contrast the diagonal elements of $\mathbf{B}_0$ are restricted to be one for the CS-VAR-SV model. But the key difference lies in the upper-triangular elements of $\mathbf{B}_0$. These are restricted to be zero in the CS-VAR-SV. But several of these upper triangular elements are estimated to be large in magnitude using our OI-VAR-SV model. The data strongly supports an unrestricted impact matrix $\mathbf{B}_0$. In the next subsection, we formally test the hypothesis that $\mathbf{B}_0$ is lower triangular via a Bayesian model comparison exercise.

6.3 Bayesian Model Comparison

In this section we conduct a series of model comparison exercises using the marginal likelihood. First, as highlighted in Herwartz and Lütkepohl (2014), in the presence of heteroskedasticity the lower triangular assumption on $\mathbf{B}_0$ is testable through the lens of the unrestricted OI-VAR-SV model. We test this lower triangular assumption and quantify the impact of different orderings. Another practical issue is the choice of m, the number of stochastic volatility processes, in the model. In the baseline OI-VAR-SV we set $m = n$, but it is plausible that the bias-variance trade-off might favor a smaller value of m. We therefore compare specifications with different values of m to provide guidance for practitioners.

The gold standard for Bayesian model comparison is the marginal likelihood or the marginal data density, defined for model M_k as

$$p(\mathbf{y} \mid M_k) = \int p(\mathbf{y} \mid \boldsymbol{\theta}_k, M_k) p(\boldsymbol{\theta}_k \mid M_k) \mathrm{d}\boldsymbol{\theta}_k,$$

where $\mathbf{y} = (\mathbf{y}_1', \ldots, \mathbf{y}_T')'$ is the full sample, $\boldsymbol{\theta}_k$ is the vector of model-specific parameters and latent variables, $p(\mathbf{y} \mid \boldsymbol{\theta}_k, M_k)$ is the likelihood and $p(\boldsymbol{\theta}_k \mid M_k)$ is the prior distribution. The integration involved in the marginal likelihood computation is typically very high dimensional for large-scale models. Despite recent advances, direct computation of the marginal likelihood for large VARs with stochastic volatility remains challenging. Here we circumvent this computational challenge by approximating the marginal likelihood using a sum of predictive likelihoods, which can be obtained by running the posterior sampler multiple times with an expanding window of data points. More specifically, note

that we can factor the marginal likelihood as

$$p(\mathbf{y} \mid M_k) = p(\mathbf{y}_1 \mid M_k) \prod_{t=1}^{T-1} p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}, M_k),$$

where $\mathbf{y}_{1:t} = (\mathbf{y}_1', \ldots, \mathbf{y}_t')'$ denotes all the data up to time t and $p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}, M_k)$ is the predictive likelihood, i.e., the one-step-ahead density forecast for $\mathbf{y}_{t+1}$ evaluated at the actual observation. Notice that in the above expression, the one-step-ahead density forecast $p(\mathbf{y}_1 \mid M_k)$ depends only on the prior distribution under model M_k. More generally, $p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}, M_k)$ is heavily influenced by the specific choice of the prior when t is small. To reduce the impact of the prior, one often discards part of the initial sample and uses instead the following approximation of the marginal likelihood:

$$p(\mathbf{y} \mid M_k) \approx \prod_{t=t_0}^{T-1} p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}, M_k).$$

We set t_0 so that the first 20 years are used as the training sample. Each predictive likelihood $p(\mathbf{y}_{t+1} \mid \mathbf{y}_{1:t}, M_k)$ is obtained by running a separate MCMC using data $\mathbf{y}_{1:t}$.

To test the lower triangular assumption on $\mathbf{B}_0$, as well as quantifying the impact of different orderings, we compare four models: OI-VAR-SV-1, OI-VAR-SV-2, CS-VAR-SV-1 and CS-VAR-SV-2. The log marginal likelihood estimates are reported in Table 3. Two points are immediately clear from these results. First, the data overwhelmingly favor the proposed OI-VAR-SV relative to the CS-VAR-SV. For instance, under the first ordering, the difference in marginal likelihood values is over 3,500 in log scale. Second, the marginal likelihood values of the proposed OI-VAR-SV under the two different orderings are the same (up to MCMC approximation error), whereas for the CS-VAR-SV, the two orderings give very different marginal likelihood values—a difference of more than 400 in log scale. All in all, these results further highlight the advantages of the proposed OI-VAR-SV: not only is it robust to different variable orderings, it is also more flexible and provides a much better fit of the data.

Table 3: Log marginal likelihood values of the proposed order-invariant stochastic volatility model under two different orderings (OI-VAR-SV-1 and OI-VAR-SV-2), as well as those from the model of Cogley and Sargent (2005) (CS-VAR-SV-1 and CS-VAR-SV-2).

	CS-VAR-SV-1	CS-VAR-SV-2	OI-VAR-SV-1	OI-VAR-SV-2
log ML	−26,301	−26,709	−22,714	−22,712

Next, we investigate the number of SV processes m needed to fit the data well, while accounting for model complexity. To that end, we compare the baseline OI-VAR-SV-1 with different values of m.[10] When $m \leqslant n-2 = 18$, the impact matrix $\mathbf{B}_0$ is only partially identified, but the log marginal likelihood values remain valid. The results in Table 4 show that the data favor the specification with $m = 19$ SV processes. Interestingly, it outperforms both a more restricted version ($m = 18$) as well as a more flexible one ($m = 20$), highlighting the value of conducting a model comparison exercise to select m. In addition, that the best value of m is large suggests that the co-volatility structure in macroeconomic data is complex and requires more a flexible modeling approach.

Table 4: Log marginal likelihood values of the OI-VAR-SV-1 with m stochastic volatility processes.

m	17	18	19	20
log ML	−22,713	−22,701	−22,684	−22,714

7 A Forecasting Exercise

7.1 Comparison of OI-VAR-SV to CS-VAR-SV

Since VARs are commonly used for forecasting, it is interesting to investigate how the sensitivity to ordering of the CS-VAR-SV models affects forecasts and whether working with our order-invariant specification can lead to forecast improvements. Accordingly we carry out a forecasting exercise using the monthly macroeconomic dataset of sub-section 6.2 for the four models (OI-VAR-SV-1, OI-VAR-SV-2, CS-VAR-SV-1 and CS-VAR-SV-2). Forecast performance of the models is evaluated from 1970:03 till the end of the

[10] For $m < n$, the $n \times n$ diagonal matrix $\mathbf{D}_t$ in (2) is set to be $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{m,t}}, 1, \ldots, 1)$.

sample. Root mean squared forecast errors (RMSFEs) are used to evaluate the quality of point forecasts and averages of log predictive likelihoods (ALPLs) are used to evaluate the quality of density forecasts. We focus on four core variables: industrial production, the unemployment rate, PCE inflation and the Federal funds rate. Results for three different forecast horizons ($h = 1, 6, 12$) are presented in Table 5. We use the iterated method of forecasting to compute longer horizon forecasts.

Note first that, as expected, OI-VAR-SV-1 and OI-VAR-SV-2 are producing the same forecasts (up to MCMC approximation error). CS-VAR-SV-1 and CS-VAR-SV-2 are often producing forecasts which are substantially different both from one another and from those produced by the order-invariant approaches. These differences are not that noticeable in the RMSFEs, but are very noticeable in the ALPLs. This suggests ordering issues are more important for density forecasts than for point forecasts. This is not surprising since ordering issues are important for the multivariate stochastic volatility process (which plays an important role in determining predictive variances) but not for the VAR coefficients (which are the key determinants of predictive means). These findings are similar to those found in smaller VAR-SVs by Arias, Rubio-Ramirez, and Shin (2022).

A careful examination of the ALPL results indicates that the best forecasting model is almost always the OI-VAR-SV and in most cases the forecast improvements are statistically significant relative to the CS-VAR-SV-1 benchmark. A comparison of CS-VAR-SV-1 to CS-VAR-SV-2 yields no simple conclusion. Remember that CS-VAR-SV-1 uses a similar ordering of Carriero, Clark, and Marcellino (2019). For most variables and forecast horizons, this ordering is leading to higher ALPLs than the reverse ordering. This finding is also consistent with the results in Arias, Rubio-Ramirez, and Shin (2022). But there are exceptions such as forecasting the Fed funds rate for $h = 6$ and $h = 12$. But the important point is not that, when using CS-VAR-SV models, one ordering is better than the other. The important points are that ordering matters (often in a statistically significant way) and that different variables prefer different orderings. That is, there is no one ordering that forecasts all of the variables best.

Table 5: RMSFE and ALPL of four core macroeconomic time series.

Variables	Models	RMSFE			ALPL		
		$h=1$	$h=6$	$h=12$	$h=1$	$h=6$	$h=12$
Industrial production	CS-VAR-SV-1	0.007	**0.007**	**0.007**	2.071	2.159	2.266
	CS-VAR-SV-2	0.007**	0.007**	0.007**	1.522***	1.651***	1.767***
	OI-VAR-SV-1	**0.007**	0.007	0.007***	**3.660*****	3.458***	**3.360*****
	OI-VAR-SV-2	0.007	0.007	0.007***	3.659***	**3.459*****	3.358***
Unemployment rate	CS-VAR-SV-1	0.159	0.169	0.173	−0.005	−0.689	−0.707
	CS-VAR-SV-2	0.158	0.169	0.173	−0.124	−0.761	−0.742
	OI-VAR-SV-1	**0.158**	**0.167**	**0.173**	**0.463*****	**0.276**	0.152
	OI-VAR-SV-2	0.158	0.167*	0.173	0.463***	0.276	**0.156**
PCE inflation	CS-VAR-SV-1	0.002	**0.002**	**0.002**	2.210	2.311	2.425
	CS-VAR-SV-2	0.002***	0.002	0.002	1.811***	1.919***	2.024***
	OI-VAR-SV-1	**0.002****	0.002	0.002*	4.821***	4.448***	4.249***
	OI-VAR-SV-2	0.002**	0.002	0.002	**4.823*****	**4.453*****	**4.260*****
Federal funds rate	CS-VAR-SV-1	**0.492**	**1.566**	**2.216**	0.222	−8.473	−16.704
	CS-VAR-SV-2	0.497	1.597*	2.264	−0.327***	**−6.708*****	**−12.474*****
	OI-VAR-SV-1	0.493	1.576	2.233	0.294***	−6.956*	−13.466*
	OI-VAR-SV-2	0.494	1.570	2.236	**0.296*****	−7.039	−13.662

Note: The bold figure indicates the best model in each case. *, ** and *** denote, respectively, 0.10, 0.05 and 0.01 significance level for a two-sided Diebold and Mariano (1995) test. The benchmark model in the Diebold Mariano test is CS-VAR-SV-1.

7.2 Adding Time Variation to $\mathbf{B}_0$

So far all the models in this paper have assumed $\mathbf{B}_0$ to be constant over time and we have used the specification of Cogley and Sargent (2005) which maintains this assumption as a benchmark in our investigation of ordering issues. However, following Primiceri (2005) many papers have allowed for time variation in $\mathbf{B}_0$ as well. It is straightforward to extend our model to allow for this, leading to the order-invariant time varying parameter VAR-SV, or OI-TVP-VAR-SV which replaces $\mathbf{B}_0$ by $\mathbf{B}_{0,t}$ and assumes the elements of the latter follow random walks. The definition of the OI-TVP-VAR-SV along with an MCMC algorithm which allows for Bayesian inference is given in the Online Appendix. The Online Appendix also shows that the model of Primiceri (2005), which uses a lower triangular specification for $\mathbf{B}_{0,t}$, is subject to the same sensitivity to ordering issues as in the model of Cogley and Sargent (2005).

One issue that arises with the OI-TVP-VAR-SV is that, unlike the OI-VAR-SV, it is not identified. Another issue with any TVP-VAR is that the computational burden can be very high in large TVP-VARs when using MCMC. But a lack of identification of a model does not preclude good forecast performance. Accordingly, we have repeated our

forecasting exercise using a OI-TVP-VAR-SV involving our four core variables as well as the TVP-VAR-SV which assumes a lower triangular form for $\mathbf{B}_{0,t}$. We consider two versions of the latter with the variables ordered in two different ways (i.e. in the order they appear in Table 5 and in the reverse order. These are labelled Primiceri-1 and Primiceri-2 in the table). Table 6 provides the results and indicates that our OI-TVP-VAR-SV is typically forecasting better than TVP models which use the lower triangular form. The importance of ordering can be seen in the non-order invariant TVP-VAR-SV in that Primiceri-2 is typically forecasting much worse than Primiceri-1.

It is also interesting to compare Table 5 (involving 20 dimensional VARs without TVPs) to Table 6 (involving 4 variable TVP-VARs). The pattern is mixed, but in most cases the high-dimensional OI-VAR-SV is beating its lower-dimensional TVP counterpart. This indicates that the extra variables in the larger model are typically helping improve forecasts more than simply adding TVPs to a smaller model.

Table 6: RMSFE and ALPL of four core macroeconomic time series.

Variables	Models	RMSFE			ALPL		
		$h=1$	$h=6$	$h=12$	$h=1$	$h=6$	$h=12$
Industrial production	OI-TVP-VAR-SV	0.007***	0.008*	0.008	3.465***	2.673***	2.097***
	Primiceri-1	0.009	0.022	0.073	1.757	-0.468	-3.695
	Primiceri-2	0.052***	0.966***	6.217***	-0.344***	-2.685***	-6.085***
Unemployment rate	OI-TVP-VAR-SV	0.203***	0.207*	0.200*	0.249*	-0.839***	-1.697***
	Primiceri-1	0.175	0.194	0.252	0.339	-3.120	-6.878
	Primiceri-2	0.548**	1.304***	5.200***	-1.046***	-5.199***	-9.446**
PCE inflation	OI-TVP-VAR-SV	0.002	0.002	0.002	4.234***	3.842***	3.586
	Primiceri-1	0.009	0.036	0.117	1.573	-0.832	-3.874
	Primiceri-2	0.054**	0.618***	4.549**	-0.222***	-2.576***	-6.042***
Fed funds rate	OI-TVP-VAR-SV	0.596***	1.998	2.692	-0.298	-8.676***	-15.755**
	Primiceri-1	0.529	1.808	3.905	0.048	6.245	-11.650
	Primiceri-2	0.549	4.109***	21.992***	0.058***	-8.501***	-22.344***

Note: *, ** and *** denote, respectively, 0.10, 0.05 and 0.01 significance level for a two-sided Diebold and Mariano (1995) test. The benchmark model for the Diebold-Mariano test is Primiceri-1.

8 Conclusions

In this paper, we have demonstrated, both theoretically and empirically, the consequences of working with VAR-SVs which use a lower triangular parameterization of the error covariance matrices such as that used in Cogley and Sargent (2005) and are thus not order invariant. We have proposed a new specification which is order invariant which involves

working with an unrestricted version of the impact matrix, $\mathbf{B}_0$. Such a model would be unidentified in the homoscedastic VAR but we draw on Bertsche and Braun (2020) to establish that the incorporation of SV identifies the model. We develop an MCMC algorithm which allows for Bayesian inference and prediction in our order-invariant model. In an empirical exercise involving 20 macroeconomic variables we demonstrate the ability of our methods to produce accurate forecasts in a computationally efficient manner.

References

Archakov, I., and P. Hansen (2021): "A new parametrization of correlation matrices," *Econometrica*, 89(4), 1699–1715.

Arias, J. E., J. F. Rubio-Ramirez, and M. Shin (2022): "Macroeconomic forecasting and variable ordering in multivariate stochastic volatility models," *Journal of Econometrics*, forthcoming.

Asai, M., and M. McAleer (2009): "The structure of dynamic correlations in multivariate stochastic volatility models," *Journal of Econometrics*, 150(2), 182–192.

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

Bertsche, D., and R. Braun (2020): "Identification of structural vector autoregressions by stochastic volatility," *Journal of Business & Economic Statistics*, just-accepted, 1–39.

Bognanni, M. (2018): "A class of time-varying parameter structural VARs for inference under exact or set identification," *FRB of Cleveland Working Paper*.

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, 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.

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): "Large hybrid time-varying parameter VARs," *arXiv preprint arXiv:2201.07303*.

Chan, J. C. C., A. Doucet, R. León-González, and R. W. Strachan (2018): "Multivariate stochastic volatility with co-heteroscedasticity," *CAMA Working Paper*.

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.

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.

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.

Diebold, F. X., and R. S. Mariano (1995): "Comparing predictive accuracy," *Journal of Business and Economic Statistics*, 13(3), 253–263.

Hartwig, B. (2020): "Robust inference in time-varying structural VAR models: The DC-Cholesky multivariate stochastic volatility model," *Deutsche Bundesbank Discussion Paper No 34/2020*.

Herwartz, H., and H. Lütkepohl (2014): "Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks," *Journal of Econometrics*, 183(1), 104–116.

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.

Kaufmann, S., and C. Schumacher (2019): "Bayesian estimation of sparse dynamic factor models with order-independent and ex-post mode identification," *Journal of Econometrics*, 210(1), 116–134.

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.

Lanne, M., H. Lütkepohl, and K. Maciejowska (2010): "Structural vector autoregressions with Markov switching," *Journal of Economic Dynamics and Control*, 34(2), 121–131.

Leeper, E. M., C. A. Sims, and T. Zha (1996): "What does monetary policy do?," *Brookings papers on economic activity*, 1996(2), 1–78.

Levy, B., and H. Lopes (2021): "Dynamic ordering learning in multivariate forecasting," *https://arxiv.org/abs/2101.04164*.

Lewis, D. J. (2021): "Identifying shocks via time-varying volatility," *The Review of Economic Studies*.

Makalic, E., and D. F. Schmidt (2016): "A simple sampler for the horseshoe estimator," *IEEE Signal Processing Letters*, 23(1), 179–182.

McCracken, M. W., and S. Ng (2016): "FRED-MD: A monthly database for macroeconomic research," *Journal of Business and Economic Statistics*, 34(4), 574–589.

Philipov, A., and M. Glickman (2006): "Multivariate stochastic volatility via Wishart processes," *Journal of Business and Economic Statistics*, 24, 313–328.

Prado, R., and M. West (2010): *Time Series: Modeling, Computation, and Inference*. Chapman and Hall/CRC.

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

Rigobon, R. (2003): "Identification through heteroskedasticity," *Review of Economics and Statistics*, 85(4), 777–792.

Shin, M., and M. Zhong (2020): "A new approach to identifying the real effects of uncertainty shocks," *Journal of Business and Economic Statistics*, 38(2), 367–379.

Sims, C. A. (1980): "Macroeconomics and reality," *Econometrica*, 48, 1–48.

Uhlig, H. (1997): "Bayesian vector autoregressions with stochastic volatility," *Econometrica*, 65(1), 59–73.

Villani, M. (2009): "Steady-state priors for vector autoregressions," *Journal of Applied Econometrics*, 24(4), 630–650.

Waggoner, D., and T. Zha (2003): "A Gibbs sampler for structural vector autoregressions," *Journal of Economic Dynamics and Control*, 28(2), 349–366.

West, M., and J. Harrison (2006): *Bayesian Forecasting and Dynamic Models*. Springer Science & Business Media.

Wu, P., and G. Koop (2022): "Estimating the ordering of variables in a VAR using a Plackett-Luce Prior," *manuscript*.

Online Appendix A: Data Description

This appendix provides details of the monthly dataset used in the forecasting exercise. The variables and their transformations are the same as in Carriero, Clark, and Marcellino (2019).

Table 7: Monthly dataset of 20 variables from FRED-MD.

Variable	Mnemonic	Transformation
Real personal income	RPI	Δlog
Real PCE	DPCERA3M086SBEA	Δlog
Real manufacturing and trade sales	CMRMTSPLx	Δlog
Industrial production	INDPRO	Δlog
Capacity utilization in manufacturing	CUMFNS	Δ
Civilian unemployment rate	UNRATE	Δ
Total nonfarm employment	PAYEMS	Δlog
Hours worked: goods-producing	CES0600000007	no transformation
Average hourly earnings: goods-producing	CES0600000008	Δlog
PPI for finished goods	WPSFD49207	Δ^2log
PPI for commodities	PPICMM	Δ^2log
PCE price index	PCEPI	Δ^2log
Federal funds rate	FEDFUNDS	no transformation
Total housing starts	HOUST	log
S&P 500 price index	S&P 500	Δlog
U.S.-U.K. exchange rate	EXUSUKx	Δlog
1 yr. Treasury-FEDFUNDS spread	T1YFFM	no transformation
10 yr. Treasury-FEDFUNDS spread	T10YFFM	no transformation
BAA-FEDFUNDS spread	BAAFFM	no transformation
ISM: new orders index	NAPMNOI	no transformation

Online Appendix B: Estimation Details

In this appendix we provide the estimation details of two models: the baseline OISV-VAR and a variant with a time-varying impact matrix $\mathbf{B}_{0,t}$.

Estimation of the Baseline OISV-VAR

For the baseline OISV-VAR, we consider a posterior sampler that sequentially samples from the following full conditional distributions:

1. $p(\mathbf{B}_0 \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\psi}, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = p(\mathbf{B}_0 \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{h})$;
2. $p(\boldsymbol{\alpha} \,|\, \mathbf{y}, \mathbf{B}_0, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\psi}, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = p(\boldsymbol{\alpha} \,|\, \mathbf{y}, \mathbf{B}_0, \mathbf{h}, \boldsymbol{\psi}, \boldsymbol{\kappa})$;
3. $p(\boldsymbol{\psi} \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{B}_0, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = \prod_{i=1}^{n} \prod_{j=2}^{k} p(\psi_{i,j} \,|\, \alpha_{i,j}, \boldsymbol{\kappa}, z_{\psi_{i,j}})$;
4. $p(\boldsymbol{\kappa} \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{B}_0, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\psi}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = \prod_{l=1}^{2} p(\kappa_l \,|\, \boldsymbol{\alpha}, \boldsymbol{\psi}, \mathbf{z}_{\kappa_l})$;
5. $p(\mathbf{z}_{\boldsymbol{\psi}} \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{B}_0, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\kappa}, \boldsymbol{\psi}, \mathbf{z}_{\boldsymbol{\kappa}}) = \prod_{i=1}^{n} \prod_{j=2}^{k} p(z_{\psi_{i,j}} \,|\, \psi_{i,j})$;
6. $p(\mathbf{z}_{\boldsymbol{\kappa}} \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{B}_0, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\kappa}, \boldsymbol{\psi}, \mathbf{z}_{\boldsymbol{\psi}}) = \prod_{l=1}^{2} p(z_{\kappa_l} \,|\, \kappa_l)$.
7. $p(\mathbf{h} \,|\, \mathbf{y}, \mathbf{B}_0, \boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\omega}^2, \boldsymbol{\psi}, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = \prod_{i=1}^{n} p(\mathbf{h}_{i,\cdot} \,|\, \mathbf{y}, \mathbf{B}_0, \boldsymbol{\alpha}, \boldsymbol{\phi}, \boldsymbol{\omega}^2)$;
8. $p(\boldsymbol{\omega}^2 \,|\, \mathbf{y}, \mathbf{B}_0, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\phi}, \boldsymbol{\psi}, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = \prod_{i=1}^{n} p(\omega_i^2 \,|\, \mathbf{h}_{i,\cdot}, \phi_i)$;
9. $p(\boldsymbol{\phi} \,|\, \mathbf{y}, \mathbf{B}_0, \boldsymbol{\alpha}, \mathbf{h}, \boldsymbol{\omega}^2, \boldsymbol{\psi}, \boldsymbol{\kappa}, \mathbf{z}_{\boldsymbol{\psi}}, \mathbf{z}_{\boldsymbol{\kappa}}) = \prod_{i=1}^{n} p(\phi_i \,|\, \mathbf{h}_{i,\cdot}, \sigma_i^2)$;

The main text discusses the computation details of Step 1. Below we outline the remaining steps.

Step 2. We sample the reduced-form VAR coefficients row by row along the lines of Carriero, Clark, and Marcellino (2019). In particular, we extend the triangular algorithm in Carriero, Chan, Clark, and Marcellino (2022) that is designed for a lower triangular impact matrix $\mathbf{B}_0$ to the case where $\mathbf{B}_0$ is a full matrix. To that end, define $\mathbf{A}_{i=0}$ to be a

$k \times n$ matrix that has exactly the same elements as $\mathbf{A}$ except for the i-th column, which is set to be zero, i.e., $\mathbf{A}_{i=0} = (\boldsymbol{\alpha}_1, \ldots, \boldsymbol{\alpha}_{i-1}, \mathbf{0}, \boldsymbol{\alpha}_{i+1}, \ldots, \boldsymbol{\alpha}_n)$. Then, we can rewrite (4) as

$$\mathbf{B}_0(\mathbf{y}_t - \mathbf{A}'_{i=0}\mathbf{x}_t) = (\mathbf{B}_{0,i} \otimes \mathbf{x}'_t)\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t),$$

where $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})'$ and $\mathbf{B}_{0,i}$ is the i-th column of $\mathbf{B}_0$. Let $\mathbf{z}_t^i = \mathbf{B}_0(\mathbf{y}_t - \mathbf{A}'_{i=0}\mathbf{x}_t)$ and stack $\mathbf{z}^i = (\mathbf{z}_1^{i'}, \ldots, \mathbf{z}_T^{i'})'$, we have

$$\mathbf{z}^i = \mathbf{W}^i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}),$$

where $\mathbf{W}^i = \mathbf{X} \otimes \mathbf{B}_{0,i}$ and $\mathbf{D} = \text{diag}(\mathbf{D}_1, \ldots, \mathbf{D}_T)$.

Next, the Minnesota-type horseshoe prior on $\boldsymbol{\alpha}_i$ is conditionally Gaussian given the local variance component $\boldsymbol{\psi}_i$. More specifically, we can rewrite the conditional prior on $\boldsymbol{\alpha}_i$ in (5) as:

$$(\boldsymbol{\alpha}_i \,|\, \boldsymbol{\kappa}, \boldsymbol{\psi}_i) \sim \mathcal{N}(\mathbf{m}_i, \mathbf{V}_{\boldsymbol{\alpha}_i}),$$

where $\mathbf{m}_i = (m_{i,1}, \ldots, m_{i,k})'$ and $\mathbf{V}_{\boldsymbol{\alpha}_i} = \text{diag}(C_{i,1}, \kappa_{i,2}\psi_{i,2}C_{i,2} \ldots, \kappa_{i,k}\psi_{i,k}C_{i,k})$ (the prior on the intercept is Gaussian). Then, by standard linear regression results, we have

$$(\boldsymbol{\alpha}_i \,|\, \mathbf{y}, \mathbf{B}_0, \boldsymbol{\alpha}_{-i}, \mathbf{h}, \boldsymbol{\psi}_i, \boldsymbol{\kappa}) \sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}_i, \mathbf{K}_{\boldsymbol{\alpha}_i}^{-1}),$$

where $\boldsymbol{\alpha}_{-i} = (\boldsymbol{\alpha}'_1, \ldots, \boldsymbol{\alpha}'_{i-1}, \boldsymbol{\alpha}'_{i+1}, \ldots, \boldsymbol{\alpha}'_n)'$,

$$\mathbf{K}_{\boldsymbol{\alpha}_i} = \mathbf{V}_{\boldsymbol{\alpha}_i}^{-1} + \mathbf{W}^{i'}\mathbf{D}^{-1}\mathbf{W}^i, \quad \widehat{\boldsymbol{\alpha}}_i = \mathbf{K}_{\boldsymbol{\alpha}_i}^{-1}\left(\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}\mathbf{m}_i + \mathbf{W}^{i'}\mathbf{D}^{-1}\mathbf{z}^i\right).$$

The computational complexity of this step is the same, i.e., $\mathcal{O}(n^4)$, as in the triangular case considered in Carriero, Chan, Clark, and Marcellino (2022).

Step 3. First, note that the elements of $\boldsymbol{\psi}_i$ are conditionally independent and we can sample them one by one without loss of efficiency. Next, combining (5) and (9), we obtain

$$\begin{aligned} p(\psi_{i,j} \,|\, \alpha_{i,j}, \boldsymbol{\kappa}, z_{\psi_{i,j}}) &\propto \psi_{i,j}^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\kappa_{i,j}C_{i,j}\psi_{i,j}}(\alpha_{i,j}-m_{i,j})^2} \times \psi_{i,j}^{-\frac{3}{2}} \mathrm{e}^{-\frac{1}{\psi_{i,j}z_{\psi_{i,j}}}} \\ &= \psi_{i,j}^{-2} \mathrm{e}^{-\frac{1}{\psi_{i,j}}\left(z_{\psi_{i,j}}^{-1} + \frac{(\alpha_{i,j}-m_{i,j})^2}{2\kappa_{i,j}C_{i,j}}\right)}, \end{aligned}$$

which is the kernel of the following inverse-gamma distribution:

$$(\psi_{i,j} \,|\, \alpha_{i,j}, \boldsymbol{\kappa}, z_{\psi_{i,j}}) \sim \mathcal{IG}\left(1, z_{\psi_{i,j}}^{-1} + \frac{(\alpha_{i,j} - m_{i,j})^2}{2\kappa_{i,j} C_{i,j}}\right).$$

Step 4. Note that κ_1 and κ_2 only appear in their priors in (10) and in (5) — recall $\kappa_{i,j} = \kappa_1$ for coefficients on own lags and $\kappa_{i,j} = \kappa_2$ for coefficients on other lags. To sample κ_1 and κ_2, first define the index set S_{κ_1} that collects all the indexes (i,j) such that $\alpha_{i,j}$ is a coefficient associated with an own lag. That is, $S_{\kappa_1} = \{(i,j) : \alpha_{i,j}$ is a coefficient associated with an own lag$\}$. Similarly, define S_{κ_2} as the set that collects all the indexes (i,j) such that $\alpha_{i,j}$ is a coefficient associated with a lag of other variables. It is easy to check that the numbers of elements in S_{κ_1} and S_{κ_2} are respectively np and $(n-1)np$. Then, we have

$$\begin{aligned} p(\kappa_1 \,|\, \boldsymbol{\alpha}, \boldsymbol{\psi}, z_{\kappa_1}) &\propto \prod_{(i,j)\in S_{\kappa_1}} \kappa_1^{-\frac{1}{2}} \mathrm{e}^{-\frac{1}{2\kappa_1 C_{i,j}\psi_{i,j}}(\alpha_{i,j}-m_{i,j})^2} \times \kappa_1^{-\frac{3}{2}} \mathrm{e}^{-\frac{1}{\kappa_1 z_{\kappa_1}}} \\ &= \kappa_1^{-\left(\frac{np+1}{2}+1\right)} \mathrm{e}^{-\frac{1}{\kappa_1}\left(z_{\kappa_1}^{-1} + \sum_{(i,j)\in S_{\kappa_1}} \frac{(\alpha_{i,j}-m_{i,j})^2}{2\psi_{i,j}C_{i,j}}\right)}, \end{aligned}$$

which is the kernel of the $\mathcal{IG}\left(\frac{np+1}{2}, z_{\kappa_1}^{-1} + \sum_{(i,j)\in S_{\kappa_1}} \frac{(\alpha_{i,j}-m_{i,j})^2}{2\psi_{i,j}C_{i,j}}\right)$ distribution. Similarly, we have

$$(\kappa_2 \,|\, \boldsymbol{\alpha}, \boldsymbol{\psi}, z_{\kappa_2}) \sim \mathcal{IG}\left(\frac{(n-1)np+1}{2}, z_{\kappa_2}^{-1} + \sum_{(i,j)\in S_{\kappa_2}} \frac{(\alpha_{i,j}-m_{i,j})^2}{2\psi_{i,j}C_{i,j}}\right).$$

Steps 5-6. It is straightforward to sample the latent variables $\mathbf{z}_{\boldsymbol{\psi}}$ and $\mathbf{z}_{\boldsymbol{\kappa}}$. In particular, it follows from (9) that $z_{\psi_{i,j}} \sim \mathcal{IG}(1, 1+\psi_{i,j}^{-1}), i = 1, \ldots, n, j = 2, \ldots, k$. Similarly, from (10) we have $z_{\kappa_l} \sim \mathcal{IG}(1, 1+\kappa_l^{-1}), l = 1, 2$.

Step 7. The log-volatility vector for each equation, $\mathbf{h}_{i,\cdot}, i = 1, \ldots, n$, can be sampled using the auxiliary mixture sampler of Kim, Shephard, and Chib (1998) once we obtain the orthogonalized errors. More specifically, we first compute $\mathbf{E}$ using (11). Then, we transform the i-th column of $\mathbf{E}$ via $\mathbf{y}_i^* = (\log E_{1,i}^2, \ldots, \log E_{T,i}^2)'$. Finally we implement the auxiliary mixture sampler in conjunction with the precision sampler of Chan and Jeliazkov (2009) to sample $\mathbf{h}_{i,\cdot}$ using $\mathbf{y}_i^*$ as data.

Steps 8-9. These two steps are standard; see, e.g., Chan and Hsiao (2014).

Estimation of the OISV-VAR with a Time-Varying $\mathbf{B}_{0,t}$

We consider a variant of the OISV-VAR with a time-varying impact matrix $\mathbf{B}_{0,t}$:

$$\mathbf{y}_t = \mathbf{a} + \mathbf{A}_1\mathbf{y}_{t-1} + \cdots + \mathbf{A}_p\mathbf{y}_{t-p} + \mathbf{B}_{0,t}^{-1}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t), \tag{13}$$

where $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}})$ is diagonal and $\mathbf{B}_{0,t}$ is a non-singular but otherwise unrestricted $n \times n$ matrix. Let $\mathbf{b}_{i,t}$ denote the i-th row of $\mathbf{B}_{0,t}$, i.e., $\mathbf{B}_{0,t}' = (\mathbf{b}_{1,t}, \ldots, \mathbf{b}_{n,t})$. Each $\mathbf{b}_{i,t}, i = 1, \ldots, n$, is assumed to evolve according to the following random walk process:

$$\mathbf{b}_{i,t} = \mathbf{b}_{i,t-1} + \mathbf{u}_{i,t}^b, \quad \mathbf{u}_{i,t}^b \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\mathbf{b}_i}) \tag{14}$$

for $t = 2, \ldots, T$, where $\boldsymbol{\Sigma}_{\mathbf{b}_i} = \text{diag}(\sigma_{b,i1}^2, \ldots, \sigma_{b,in}^2)$ is a diagonal matrix. The random walk process is initialized as $\mathbf{b}_{i,1} \sim \mathcal{N}(\mathbf{b}_{i,0}, \mathbf{V}_{\mathbf{b}_i})$, where $\mathbf{b}_{i,0}$ and $\mathbf{V}_{\mathbf{b}_i}$ are fixed hyperparameters. As before, each of the log-volatility $h_{i,t}$ follows a stationary AR(1) process:

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

for $t = 2, \ldots, T$, where $|\phi_i| < 1$ and the initial condition is specified as $h_{i,1} \sim \mathcal{N}(0, \sigma_i^2/(1-\phi_i^2))$. We assume similar priors as in the baseline model. The only new addition is the priors on the error variances in the state equation (14), which we assume to be independent inverse-gamma distributions: $\sigma_{b,ij}^2 \sim \mathcal{IG}(\nu_{b,ij}, S_{b,ij}), i, j = 1, \ldots, n$.

To estimate this model, we modify the posterior sampler for the baseline model to incorporate a time-varying impact matrix. In particular, one only needs to modify Steps 1, 2 and 7 as follows and implement an additional step to sample $\boldsymbol{\Sigma}_{\mathbf{b}_i}$.

Step 1. We sample each $\mathbf{b}_{i,t}$, the i-th row of $\mathbf{B}_{0,t}$, given $\mathbf{B}_{0,s}, s \neq t$ and other rows of $\mathbf{B}_{0,t}$ using the sampling approach in Waggoner and Zha (2003) and Villani (2009) but adopted to our time-varying setting. To that end, let $\mathbf{b}_{-it}$ denote the vector $\mathbf{b}$ with the sub-vector $\mathbf{b}_{i,t}$ removed. First, the i-th equation of (13) can be written as

$$\mathbf{u}_t'\mathbf{b}_{i,t} = \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathcal{N}(0, \text{e}^{h_{i,t}}),$$

where $\mathbf{u}_t = \mathbf{y}_t - \mathbf{a} - \mathbf{A}_1\mathbf{y}_{t-1} - \cdots - \mathbf{A}_p\mathbf{y}_{t-p}$ and $\varepsilon_{i,t}$ is the i-th element of $\boldsymbol{\varepsilon}_t$. Combining the state equation in (14), the full conditional distribution of $\mathbf{b}_{i,t}$, for $2 \leqslant t \leqslant T$, is given by

$$\begin{aligned} p(\mathbf{b}_{i,t} \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}_{-it}, h_{i,t}) \propto & |\det \mathbf{B}_0| \mathrm{e}^{-\frac{1}{2}\mathrm{e}^{-h_{i,t}}\mathbf{b}_{i,t}'\mathbf{u}_t\mathbf{u}_t'\mathbf{b}_{i,t}} \times \mathrm{e}^{-\frac{1}{2}(\mathbf{b}_{i,t}-\mathbf{b}_{i,t-1})'\boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1}(\mathbf{b}_{i,t}-\mathbf{b}_{i,t-1})} \\ & \times \mathrm{e}^{-\frac{1}{2}(\mathbf{b}_{i,t}-\mathbf{b}_{i,t+1})'\boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1}(\mathbf{b}_{i,t}-\mathbf{b}_{i,t+1})} \\ & \propto |\det \mathbf{B}_0| \mathrm{e}^{-\frac{1}{2}(\mathbf{b}_{i,t}-\widehat{\mathbf{b}}_{i,t})'\mathbf{K}_{\mathbf{b}_{i,t}}(\mathbf{b}_{i,t}-\widehat{\mathbf{b}}_{i,t})} \end{aligned} \tag{15}$$

where

$$\mathbf{K}_{\mathbf{b}_{i,t}} = 2\boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1} + \mathrm{e}^{-h_{i,t}}\mathbf{u}_t\mathbf{u}_t', \quad \widehat{\mathbf{b}}_{i,t} = \mathbf{K}_{\mathbf{b}_{i,t}}^{-1}\boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1}(\mathbf{b}_{i,t-1} + \mathbf{b}_{i,t+1}).$$

For $t = 1$ and $t = T$, the full conditional distribution has the same form, but $\mathbf{K}_{\mathbf{b}_{i,t}}$ and $\widehat{\mathbf{b}}_{i,t}$ become

$$\begin{aligned} \mathbf{K}_{\mathbf{b}_{i,1}} &= \mathbf{V}_{\mathbf{b}_i}^{-1} + \boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1} + \mathrm{e}^{-h_{i,1}}\mathbf{u}_1\mathbf{u}_1', \quad \widehat{\mathbf{b}}_{i,1} = \mathbf{K}_{\mathbf{b}_{i,1}}^{-1}\left(\mathbf{V}_{\mathbf{b}_i}^{-1}\mathbf{b}_{i,0} + \boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1}\mathbf{b}_{i,2}\right) \\ \mathbf{K}_{\mathbf{b}_{i,T}} &= \boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1} + \mathrm{e}^{-h_{i,T}}\mathbf{u}_T\mathbf{u}_T', \quad \widehat{\mathbf{b}}_{i,T} = \mathbf{K}_{\mathbf{b}_{i,T}}^{-1}\boldsymbol{\Sigma}_{\mathbf{b}_i}^{-1}\mathbf{b}_{i,T-1}. \end{aligned}$$

The full conditional posterior distribution in (15) is non-standard, but as before one can adopt the approach in Waggoner and Zha (2003) and Villani (2009) to sample $\mathbf{b}_{i,t}$. In particular, $\mathbf{b}_{i,t}$ has the same distribution as a linear transformation of an n-vector that consists of one absolute normal and $n-1$ normal random variables. More specifically, let $\mathbf{C}_{i,t}$ denote the Cholesky factor of $\mathbf{K}_{\mathbf{b}_i}$ such that $\mathbf{K}_{\mathbf{b}_i} = \mathbf{C}_{i,t}\mathbf{C}_{i,t}'$. Let $\mathbf{F}_{-i}$ represent the matrix $\mathbf{F}$ with the i-th column deleted and $\mathbf{F}^{\perp}$ be the orthogonal complement of $\mathbf{F}$. Furthermore, define $\mathbf{v}_1 = \mathbf{C}_{i,t}^{-1}((\mathbf{B}_{0,t}')_{-i})^{\perp}/||\mathbf{C}_{i,t}^{-1}((\mathbf{B}_{0,t}')_{-i})^{\perp}||$ and let $(\mathbf{v}_2, \ldots, \mathbf{v}_n) = \mathbf{v}_1^{\perp}$. Then, we have

$$(\mathbf{b}_{i,t} \,|\, \mathbf{y}, \boldsymbol{\alpha}, \mathbf{b}_{-it}, h_{i,t}) \overset{\mathrm{d}}{=} (\mathbf{C}_{i,t}')^{-1}\sum_{j=1}^{n}\xi_j\mathbf{v}_j,$$

where $\xi_1 \sim \mathcal{AN}(\widehat{\xi}_1, 1)$ and $\xi_j \sim \mathcal{N}(\widehat{\xi}_j, 1), j = 2, \ldots, n$ with $\widehat{\xi}_j = \widehat{\mathbf{b}}_i'\mathbf{C}_{i,t}\mathbf{v}_j$.

To sample from $\mathcal{AN}(\widehat{\xi}_1, 1)$, we use the inverse-transform method instead of the 2-component normal mixture approximation considered in Villani (2009). This is because the normal mixture is designed to approximate $\mathcal{AN}(\mu, \rho)$ well for ρ close to 0. In our time-varying setting, $\rho = 1$ and this mixture approximation appears to be quite different from the target. Finally, in the sampler we fix the sign of the i-th element of $\mathbf{b}_{i,t}$ to be positive (so

that the diagonal elements of $\mathbf{B}_0$ are positive). This is done by multiplying the draw $\mathbf{b}_{i,t}$ by -1 if its i-th is negative.

Step 2. We sample the reduced-form VAR coefficients row by row by extending the triangular algorithm of Carriero, Chan, Clark, and Marcellino (2022) to the case where the impact matrix is unrestricted and time-varying. Let $\mathbf{A}_{i=0} = (\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_{i-1}, \mathbf{0}, \boldsymbol{\alpha}_{i+1}, \dots, \boldsymbol{\alpha}_n)$. Then, we can rewrite (13) as

$$\mathbf{B}_{0,t}(\mathbf{y}_t - \mathbf{A}'_{i=0}\mathbf{x}_t) = \mathbf{B}_{0,t,i}\mathbf{x}'_t\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t),$$

where $\mathbf{x}_t = (1, \mathbf{y}'_{t-1}, \dots, \mathbf{y}'_{t-p})'$ and $\mathbf{B}_{0,t,i}$ is the i-th column of $\mathbf{B}_{0,t}$. Let $\mathbf{z}_t^i = \mathbf{B}_{0,t}(\mathbf{y}_t - \mathbf{A}'_{i=0}\mathbf{x}_t)$ and $\mathbf{W}_t^i = \mathbf{B}_{0,t,i}\mathbf{x}'_t$. Then, stacking $t = 1, \dots, T$, we have

$$\mathbf{z}^i = \mathbf{W}^i\boldsymbol{\alpha}_i + \boldsymbol{\varepsilon}, \quad \boldsymbol{\varepsilon} \sim \mathcal{N}(\mathbf{0}, \mathbf{D}),$$

where $\mathbf{z}^i = (\mathbf{z}_1^{i\prime}, \dots, \mathbf{z}_T^{i\prime})'$, $\mathbf{W}^i = (\mathbf{W}_1^{i\prime}, \dots, \mathbf{W}_T^{i\prime})'$ and $\mathbf{D} = \text{diag}(\mathbf{D}_1, \dots, \mathbf{D}_T)$. The Minnesota-type horseshoe prior on $\boldsymbol{\alpha}_i$ is conditionally Gaussian given the local variance component $\boldsymbol{\psi}_i$. More specifically, we can rewrite the conditional prior on $\boldsymbol{\alpha}_i$ in (5) as:

$$(\boldsymbol{\alpha}_i \,|\, \boldsymbol{\kappa}, \boldsymbol{\psi}_i) \sim \mathcal{N}(\mathbf{m}_i, \mathbf{V}_{\boldsymbol{\alpha}_i}),$$

where $\mathbf{m}_i = (m_{i,1}, \dots, m_{i,k})'$ and $\mathbf{V}_{\boldsymbol{\alpha}_i} = \text{diag}(C_{i,1}, \kappa_{i,2}\psi_{i,2}C_{i,2} \dots, \kappa_{i,k}\psi_{i,k}C_{i,k})$ (the prior on the intercept is Gaussian). Then, by standard linear regression results, we have

$$(\boldsymbol{\alpha}_i \,|\, \mathbf{y}, \mathbf{B}_0, \boldsymbol{\alpha}_{-i}, \mathbf{h}, \boldsymbol{\psi}_i, \boldsymbol{\kappa}) \sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}_i, \mathbf{K}_{\boldsymbol{\alpha}_i}^{-1}),$$

where $\boldsymbol{\alpha}_{-i} = (\boldsymbol{\alpha}'_1, \dots, \boldsymbol{\alpha}'_{i-1}, \boldsymbol{\alpha}'_{i+1}, \dots, \boldsymbol{\alpha}'_n)'$,

$$\mathbf{K}_{\boldsymbol{\alpha}_i} = \mathbf{V}_{\boldsymbol{\alpha}_i}^{-1} + \mathbf{W}^{i\prime}\mathbf{D}^{-1}\mathbf{W}^i, \quad \widehat{\boldsymbol{\alpha}}_i = \mathbf{K}_{\boldsymbol{\alpha}_i}^{-1}\left(\mathbf{V}_{\boldsymbol{\alpha}_i}^{-1}\mathbf{m}_i + \mathbf{W}^{i\prime}\mathbf{D}^{-1}\mathbf{z}^i\right).$$

Step 7. The log-volatility vector for each equation, $\mathbf{h}_{i,\cdot}, i = 1, \dots, n$, can be sampled using the auxiliary mixture sampler of Kim, Shephard, and Chib (1998) once we obtain the $T \times n$ matrix of orthogonalized errors $\mathbf{E}$. That is, the t-th row of $\mathbf{E}$ is $(\mathbf{y}'_t - \mathbf{x}'_t\mathbf{A})\mathbf{B}'_{0,t}$. Then, we transform the i-th column of $\mathbf{E}$ via $\mathbf{y}_i^* = (\log E_{1,i}^2, \dots, \log E_{T,i}^2)'$. Finally we implement the auxiliary mixture sampler in conjunction with the precision sampler of Chan and Jeliazkov (2009) to sample $\mathbf{h}_{i,\cdot}$ using $\mathbf{y}_i^*$ as data.

Step 10. It is straightforward to sample the error variances $\mathbf{\Sigma}_{\mathbf{b}_i} = \text{diag}(\sigma^2_{b,i1}, \ldots, \sigma^2_{b,in})$. In particular, we have

$$\sigma^2_{b,ij} \sim \mathcal{IG}\left(\nu_{b,ij} + (T-1)/2, S_{b,ij} + \sum_{t=2}^{T}(b_{ij,t} - b_{ij,t-1})^2/2\right)$$

for $i, j = 1, \ldots, n$.

Online Appendix C: Additional Results

In this appendix we present additional analytical and simulation results. First, we provide another illustration that shows the reduced-form error variances increase as the variables is ranked lower in $\mathbf{y}_t$, under a time-varying impact matrix $\mathbf{B}_{0,t}$ similar to that in Primiceri (2005). Next, we report some efficiency measures of the posterior sampler for the baseline OISV-VAR-SV.

Consider the following multivariate stochastic volatility specification of the reduced-form errors, the n-dimensional vector $\mathbf{u}_t$, in Primiceri (2005):

$$\mathbf{u}_t = \mathbf{B}_{0,t}^{-1}\boldsymbol{\varepsilon}_t, \quad \boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{D}_t),$$

where $\mathbf{B}_{0,t}$ is a unit lower triangular matrix with elements $b_{ij,t}$ and $\mathbf{D}_t = \text{diag}(\text{e}^{h_{1,t}}, \ldots, \text{e}^{h_{n,t}})$ is a diagonal matrix consisting of the log-volatility $\mathbf{h}_t = (h_{1,t}, \ldots, h_{n,t})'$. Since $\mathbf{B}_{0,t}$ is lower triangular, the implied covariance matrix on $\mathbf{u}_t$ naturally depends on the order of the elements in $\mathbf{u}_t$.

To demonstrate this point, suppose that the vector of the lower triangular elements in $\mathbf{B}_{0,t}$, denoted as $\boldsymbol{\alpha}_t = (b_{21,t}, b_{31,t}, b_{32,t}, \ldots, b_{n(n-1),t})'$, follows the following state transition:

$$\boldsymbol{\alpha}_t = \boldsymbol{\mu} + \boldsymbol{\Phi}(\boldsymbol{\alpha}_{t-1} - \boldsymbol{\mu}) + \mathbf{u}_t^{\boldsymbol{\alpha}}, \quad \mathbf{u}_t^{\boldsymbol{\alpha}} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}_{\boldsymbol{\alpha}}),$$

where $\boldsymbol{\Phi} = \text{diag}(\phi_{21}, \ldots, \phi_{n(n-1)})$ and $\boldsymbol{\Sigma}_{\boldsymbol{\alpha}} = \text{diag}(\sigma^2_{\boldsymbol{\alpha},21}, \ldots, \sigma^2_{\boldsymbol{\alpha},n(n-1)})$ are diagonal with $|\phi_{ij}| < 1$.[11] It follows that unconditionally $b_{ij,t} \sim \mathcal{N}(\mu_{ij}, \sigma^2_{\boldsymbol{\alpha},ij}/(1-\phi^2_{ij}))$. For simplicity we set $\mu_{ij} = 0, \sigma^2_{\boldsymbol{\alpha},ij}/(1-\phi^2_{ij}) = 1$ and $\boldsymbol{\varepsilon}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{I}_n)$. In particular, $\mathbb{E}b_{ij,t} = 0, \mathbb{E}b^2_{ij,t} = 1$ and $\mathbb{E}(b_{ij,t}b_{ik,t}) = 0$ for all $j \neq k$.

Now, we obtain the elements of $\mathbf{u}_t$ recursively and compute their variances $\text{Var}(u_{i,t}) =$

[11] In Primiceri (2005) $\boldsymbol{\alpha}_t$ is assumed to follow a random walk process with $\boldsymbol{\Phi}$ set to be the identity matrix. Consequently, the unconditional mean of $\boldsymbol{\alpha}_t$ does not exist, which in turn implies that the variances of $\mathbf{u}_t$ do not either.

$\mathbb{E}u_{i,t}^2$ as before. Again, we have

$$\begin{aligned}
\mathbb{E}u_{1,t}^2 &= \mathbb{E}\varepsilon_{1,t}^2 = 1 \\
\mathbb{E}u_{2,t}^2 &= \mathbb{E}\varepsilon_{2,t}^2 + \mathbb{E}b_{21,t}^2\mathbb{E}u_{1,t}^2 = 1 + \mathbb{E}u_{1,t}^2 = 2 \\
&\vdots \\
\mathbb{E}u_{i,t}^2 &= \mathbb{E}\varepsilon_{i,t}^2 + \mathbb{E}b_{i(i-1),t}^2\mathbb{E}u_{i-1,t}^2 + \cdots + \mathbb{E}b_{i1,t}^2\mathbb{E}u_{1,t}^2 = 2^{i-1}.
\end{aligned}$$

That is, the variance of the reduced-form error increases as it is ordered lower in the n-tuple, even though the assumptions about $b_{ij,t}$ and $\varepsilon_{i,t}$ are identical across $i = 1, \ldots, n$.

Next, to assess the efficiency of the posterior sampler for the baseline OISV-VAR-SV, we compute the inefficiency factors of the MCMC samples obtained from the posterior sampler. Specifically, Figure 11 plots the inefficiency factors of selected parameters of interest from the empirical application involving a 20-variable VAR. To present the information more succinctly, boxplots of the inefficiency factors are used. Each boxplot summarizes the quantiles of each group of model parameters. The middle line of each box denotes the median, while the lower and upper lines represent, respectively, the 25- and the 75-percentiles. The whiskers extend to the maximum and minimum. For instance, the boxplot associated with $\mathbf{B}_0$ indicates that the median of the inefficiency factors of the impact matrix elements is about 15, and the maximum is about 120. Similarly, the boxplot associated with $\sigma_{ii,t}, i = 1, \ldots, n$, shows that most of the inefficiency factors of the reduced-form variances are below 10, with the maximum being about 40.

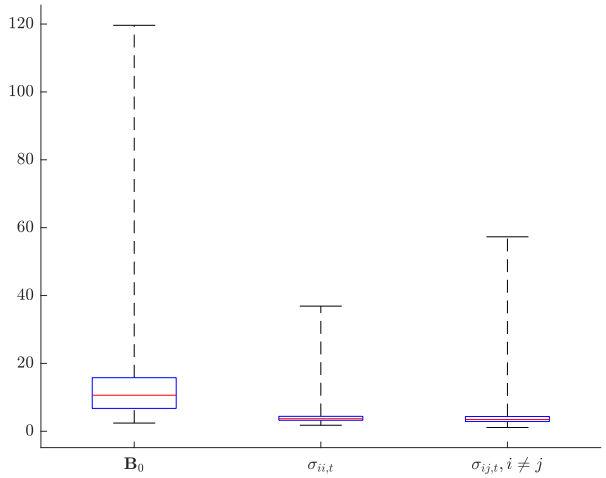

Figure 11: Boxplots of the inefficiency factors corresponding to the posterior draws of the impact matrix $\mathbf{B}_0$ and the reduced-form variances $\sigma_{ii,t}$ and covariances $\sigma_{ij,t}, i \neq j, i, j = 1, \ldots, n$.