# Asymmetric Conjugate Priors for Large Bayesian VARs

Joshua C.C. Chan[*]

Purdue University

This version: October 2021
First version: July 2019

## Abstract

Large Bayesian VARs are now widely used in empirical macroeconomics. One popular shrinkage prior in this setting is the natural conjugate prior as it facilitates posterior simulation and leads to a range of useful analytical results. This is, however, at the expense of modeling flexibility, as it rules out cross-variable shrinkage — i.e., shrinking coefficients on lags of other variables more aggressively than those on own lags. We develop a prior that has the best of both worlds: it can accommodate cross-variable shrinkage, while maintaining many useful analytical results, such as a closed-form expression of the marginal likelihood. This new prior also leads to fast posterior simulation — for a BVAR with 100 variables and 4 lags, obtaining 10,000 posterior draws takes less than half a minute on a standard desktop. We demonstrate the usefulness of the new prior via a structural analysis using a 15-variable VAR with sign restrictions to identify 5 structural shocks.

Keywords: shrinkage prior, marginal likelihood, optimal hyperparameters, structural VAR, sign restrictions

JEL classifications: C11, C52, C55, E44

# 1 Introduction

Large Bayesian vector autoregressions (BVARs) have become increasingly popular in empirical macroeconomics for forecasting and structural analysis since the influential work by Banbura, Giannone, and Reichlin (2010). Prominent examples include Carriero, Kapetanios, and Marcellino (2009), Koop (2013), Koop and Korobilis (2013) and Korobilis and Pettenuzzo (2019). VARs tend to have a lot of parameters, and the key that makes these highly parameterized VARs useful is the introduction of shrinkage priors. For large BVARs, one commonly adopted prior is the natural conjugate prior, which has a few advantages over alternatives. First, this prior is conjugate, and consequently it gives rise to a range of useful analytical results, including a closed-form expression of the marginal likelihood.[1] Second, the posterior covariance matrix of the VAR coefficients under this prior has a Kronecker product structure, which can be used to speed up computations.

On the other hand, a key limitation of the natural conjugate prior is that the prior covariance matrix of the VAR coefficients needs to have a Kronecker product structure, which implies cross-equation restrictions that might not be reasonable. In particular, this Kronecker structure requires symmetric treatment of own lags and lags of other variables. In many applications one might wish to shrink the coefficients on other variables' lags more strongly to zero than those of own lags. This cross-variable shrinkage, however, cannot be implemented using the natural conjugate prior due to this Kronecker structure. Carriero, Clark, and Marcellino (2015) summarize this dilemma between computational convenience and prior flexibility as: "While the pioneering work of Litterman (1986) suggested it was useful to have cross-variable shrinkage, it has become more common to estimate larger models without cross-variable shrinkage, in order to have a Kronecker structure that speeds up computations and facilitates simulation."

We develop a prior that solves this dilemma — this new prior allows asymmetric treatment between own lags and lags of other variables, while it maintains many useful analytical results, such as a closed-form expression of the marginal likelihood. In addition, we

---

[1]An analytical expression for the marginal likelihood is valuable for many purposes. First, it is useful for model selection — e.g., choosing the lag length in BVARs. Second, it can be used to select prior hyperparameters that control the degree of shrinkage. Examples include Del Negro and Schorfheide (2004), Schorfheide and Song (2015) and Carriero, Clark, and Marcellino (2015). This approach of selecting hyperparameters is incorporated in the BEAR MATLAB toolbox developed by the European Central Bank (Dieppe, Legrand, and Van Roye, 2016).

exploit these analytical results to develop an efficient method to simulate directly from the posterior distribution — we obtain *independent* posterior draws and avoid Markov chain Monte Carlo (MCMC) methods altogether. For a BVAR with 100 variables and 4 lags, simulating 10,000 posterior draws under this new asymmetric conjugate prior takes less than 30 seconds.

To develop this asymmetric conjugate prior, we first write the BVAR in a recursive structural form, under which the error covariance matrix is diagonal. We then adopt an equation-by-equation estimation approach in the spirit of Carriero, Clark, and Marcellino (2019). In particular, we assume that the parameters are *a priori* independent across equations — i.e., the joint prior density is a product of densities, each for the set of parameters in each equation. Under this setup, we show that if the VAR coefficients and the error variance in each equation follows a normal-inverse-gamma prior, the posterior distribution has the same form — i.e., it is a product of normal-inverse-gamma densities. It is useful to emphasize that this particular structural-form parameterization is intended to be a computational device and it does not commit the user to a recursive identification scheme. In particular, one can recover the reduced-form parameters from the structural-form parameter, and the posterior sampler can be used in conjunction with other identification schemes, as demonstrated in the application.

To help elicit the hyperparameters in this asymmetric conjugate prior, we prove that if we assume a standard inverse-Wishart prior on the reduced-form error covariance matrix, the implied prior on the structural-form impact matrix and error variances is a product of normal-inverse-gamma densities; and vice versa. Hence, using this proposition, we can first elicit the hyperparameters in the reduced-form prior, which is often more natural, and then obtain the implied hyperparameters in the structural-form prior. Since we can directly specify prior beliefs on the reduced-form error covariance matrix, this proposition implies that the proposed prior — with carefully chosen hyperparameters — is independent of the order of the variables.

We demonstrate the usefulness of the proposed asymmetric conjugate prior and the associated posterior sampler by revisiting the empirical study in Furlanetto, Ravazzolo, and Sarferaz (2019), who use 6-variable VAR to identify 5 structural shocks using a set of sign restrictions on the contemporaneous impact matrix. Here we augment their system with a set of additional variables and consider a 15-variable VAR, which is one of the largest

VARs identified with sign restrictions considered in the literature so far. Even though there are good reasons to consider larger systems — such as concerns of informational deficiency and non-unique mapping from economic variables to data — empirical works that impose sign restrictions typically use small or medium VARs (e.g., up to 6 or 7 variables) due to the computational burden. With more variables and sign restrictions, it is clear that conventional approaches that use standard Gibbs samplers are too computationally expensive. In contrast, using the proposed asymmetric conjugate prior and the associated efficient sampler, it is feasible to conduct structural analysis with a large number of sign restrictions, which helps sharpen inference.

The rest of the paper is organized as follows. We first introduce in Section 2 a reparameterization of the reduced-form BVAR and the new asymmetric conjugate prior. We then derive the associated posterior distribution and the marginal likelihood. Section 3 discusses a few extensions of the standard BVAR, and outlines the corresponding sampling schemes. It is followed by a structural analysis using sign restrictions to illustrate the usefulness of the proposed prior in Section 4. Lastly, Section 5 concludes and briefly discusses some future research directions.

## 2  Bayesian VARs and Conjugate Priors

Let $\mathbf{y}_t = (y_{1,t}, \ldots, y_{n,t})'$ be an $n \times 1$ vector of endogenous variables at time $t$. A standard VAR can be written as:

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

where $\widetilde{\mathbf{b}}$ is an $n \times 1$ vector of intercepts, $\widetilde{\mathbf{B}}_1, \ldots, \widetilde{\mathbf{B}}_p$ are $n \times n$ VAR coefficient matrices and $\widetilde{\boldsymbol{\Sigma}}$ is a full covariance matrix.

The parameters in this model can be naturally divided into two blocks: the error covariance matrix $\widetilde{\boldsymbol{\Sigma}}$ and the matrix of intercepts and VAR coefficients, i.e., $\widetilde{\mathbf{B}} = (\widetilde{\mathbf{b}}, \widetilde{\mathbf{B}}_1, \cdots, \widetilde{\mathbf{B}}_p)'$. Under this parameterization, there is a conjugate prior on $(\widetilde{\mathbf{B}}, \widetilde{\boldsymbol{\Sigma}})$, namely, the normal-inverse-Wishart distribution:

$$\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\widetilde{\nu}_0, \widetilde{\mathbf{S}}_0), \quad (\text{vec}(\widetilde{\mathbf{B}}) \,|\, \widetilde{\boldsymbol{\Sigma}}) \sim \mathcal{N}(\text{vec}(\widetilde{\mathbf{B}}_0), \widetilde{\boldsymbol{\Sigma}} \otimes \widetilde{\mathbf{V}}),$$

where $\otimes$ denotes the Kronecker product, $\text{vec}(\cdot)$ vectorizes a matrix by stacking the columns from left to right and $\mathcal{IW}$ denotes the inverse-Wishart distribution. This prior is commonly called the natural conjugate prior and can be traced back to Zellner (1971). For textbook treatment of this prior and the associated posterior distribution, see, e.g., Koop and Korobilis (2010), Karlsson (2013) or Chan (2020b).

The main advantage of the natural conjugate prior is that it gives rise to a range of analytical results. For example, the associated posterior and one-step-ahead predictive distributions are both known; the marginal likelihood is also available in closed-form. These analytical results are useful for a variety of purposes. For instance, the closed-form expression of the marginal likelihood under the natural conjugate prior can be used to calculate optimal hyperparameters, as is done in Del Negro and Schorfheide (2004), Schorfheide and Song (2015) and Carriero, Clark, and Marcellino (2015). The direct sampling algorithm to draw from the posterior distribution of $(\text{vec}(\widetilde{\mathbf{B}}), \widetilde{\boldsymbol{\Sigma}})$ can be used to develop efficient posterior samplers to estimate large, heteroscedastic Bayesian VARs. Examples include Carriero, Clark, and Marcellino (2016) and Chan (2020a).

On the other hand, one key drawback of the natural conjugate prior is that the prior covariance matrix of $\text{vec}(\widetilde{\mathbf{B}})$ is restrictive — to be conjugate it needs to have the Kronecker product structure $\widetilde{\boldsymbol{\Sigma}} \otimes \widetilde{\mathbf{V}}$, which implies cross-equation restrictions on the covariance matrix. In particular, this structure requires symmetric treatment of own lags and lags of other variables. In many situations one might want to shrink the coefficients on lags of other variables more strongly to zero than those of own lags. This prior belief, however, cannot be implemented using the natural conjugate prior due to the Kronecker structure.

Here we develop a prior that solves this dilemma: this new prior allows asymmetric treatment between own lags and lags of other variables, while it maintains many useful analytical results. In what follows, we first consider a reparameterization of the reduced-form VAR in (1). We introduce in Section 2.2 the new asymmetric conjugate prior and discuss its properties. We then derive the associated posterior distribution and discuss an efficient sampling scheme in Section 2.3. Finally, we give an analytical expression of the marginal likelihood in Section 2.4.

## 2.1 The Bayesian VAR in Structural Form

In this section we introduce a reparameterization of the reduced-form VAR in (1) and derive the associated likelihood function. To that end, we first write the VAR in the following structural form:

$$\mathbf{A}\mathbf{y}_t = \mathbf{b} + \mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \boldsymbol{\varepsilon}_t^y, \quad \boldsymbol{\varepsilon}_t^y \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}), \tag{2}$$

where $\mathbf{b}$ is an $n \times 1$ vector of intercepts, $\mathbf{B}_1, \ldots, \mathbf{B}_p$ are $n \times n$ VAR coefficient matrices, $\mathbf{A}$ is an $n \times n$ lower triangular matrix with ones on the diagonal and $\boldsymbol{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_n^2)$ is diagonal. Since the covariance matrix $\boldsymbol{\Sigma}$ is diagonal, we can estimate this recursive system equation by equation without loss of efficiency.[2] It is easy to see that we can recover the reduced-form parameters by setting $\widetilde{\mathbf{b}} = \mathbf{A}^{-1}\mathbf{b}$, $\widetilde{\mathbf{B}}_j = \mathbf{A}^{-1}\mathbf{B}_j, j = 1, \ldots, p$ and $\widetilde{\boldsymbol{\Sigma}} = \mathbf{A}^{-1}\boldsymbol{\Sigma}(\mathbf{A}^{-1})'$.

For later reference, we introduce some notations. Let $b_i$ denote the $i$-th element of $\mathbf{b}$ and let $\mathbf{b}_{j,i}$ represent the $i$-th row of $\mathbf{B}_j$. Then, $\boldsymbol{\beta}_i = (b_i, \mathbf{b}_{1,i}, \ldots, \mathbf{b}_{p,i})'$ is the intercept and VAR coefficients for the $i$-th equation. Furthermore, let $\boldsymbol{\alpha}_i$ denote the free elements in the $i$-th row of the impact matrix $\mathbf{A}$, i.e., $\boldsymbol{\alpha}_i = (A_{i,1}, \ldots, A_{i,i-1})'$. We then follow Chan and Eisenstat (2018) to rewrite the $i$-th equation of the system in (2) as:

$$y_{i,t} = \widetilde{\mathbf{w}}_{i,t}\boldsymbol{\alpha}_i + \widetilde{\mathbf{x}}_t\boldsymbol{\beta}_i + \varepsilon_{i,t}^y, \quad \varepsilon_{i,t}^y \sim \mathcal{N}(0, \sigma_i^2),$$

where $\widetilde{\mathbf{w}}_{i,t} = (-y_{1,t}, \ldots, -y_{i-1,t})$ and $\widetilde{\mathbf{x}}_t = (1, \mathbf{y}'_{t-1}, \ldots, \mathbf{y}'_{t-p})$. Note that $y_{i,t}$ depends on the contemporaneous variables $y_{1,t}, \ldots, y_{i-1,t}$. But since the system is triangular, when we perform the change of variables from $\boldsymbol{\varepsilon}_t^y$ to $\mathbf{y}_t$ to obtain the likelihood function, the corresponding Jacobian has unit determinant and the likelihood function has the usual Gaussian form.

---

[2]Carriero, Clark, and Marcellino (2019) pioneer a similar equation-by-equation estimation approach to estimate a large VAR with a standard stochastic volatility specification. However, they use the reduced-form parameterization in (1), whereas here we use the structural form in (2). As we will see below, the latter parameterization has the advantage of having a convenient representation as $n$ independent regressions and it consequently leads to a more efficient sampling scheme. Ando and Zellner (2010) also consider a similar reparameterization of the reduced-form VAR that allows equation-by-equation estimation. But in their implementation they need to switch between two parameterizations, which makes estimation more cumbersome.

If we let $\mathbf{x}_{i,t} = (\widetilde{\mathbf{w}}_{i,t}, \widetilde{\mathbf{x}}_t)$, we can further simplify the $i$-th equation as:

$$y_{i,t} = \mathbf{x}_{i,t}\boldsymbol{\theta}_i + \varepsilon_{i,t}^y, \quad \varepsilon_{i,t}^y \sim \mathcal{N}(0, \sigma_i^2),$$

where $\boldsymbol{\theta}_i = (\boldsymbol{\beta}_i', \boldsymbol{\alpha}_i')'$ is of dimension $k_i = np + i$. Hence, we have rewritten the structural VAR in (2) as a system of $n$ independent regressions. Moreover, by stacking the elements of the impact matrix $\boldsymbol{\alpha}_i$ and the VAR coefficients $\boldsymbol{\beta}_i$, we can sample them together to improve efficiency.[3]

To derive the likelihood function, we further stack $\mathbf{y}_i = (y_{i,1}, \ldots, y_{i,T})'$ and define $\mathbf{X}_i$ and $\boldsymbol{\varepsilon}_i^y$ similarly. Hence, we can rewrite the above equation as follows:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\theta}_i + \boldsymbol{\varepsilon}_i^y, \quad \boldsymbol{\varepsilon}_i^y \sim \mathcal{N}(\mathbf{0}, \sigma_i^2\mathbf{I}_T).$$

Finally, let $\boldsymbol{\theta} = (\boldsymbol{\theta}_1', \ldots, \boldsymbol{\theta}_n')'$ and $\boldsymbol{\sigma}^2 = (\sigma_1^2, \ldots, \sigma_n^2)'$. Then, the likelihood function of the VAR in (2) is given by

$$p(\mathbf{y} \,|\, \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^n p(\mathbf{y}_i \,|\, \boldsymbol{\theta}_i, \sigma_i^2) = \prod_{i=1}^n (2\pi\sigma_i^2)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_i^2}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)'(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\theta}_i)}. \tag{3}$$

In other words, the likelihood function is the product of $n$ Gaussian densities.

## 2.2 Asymmetric Conjugate Priors

Next we introduce a conjugate prior on $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ that allows differential treatment between prior variances on own lags versus others. We assume that the parameters are *a priori* independent across equations, i.e., $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^n p(\boldsymbol{\theta}_i, \sigma_i^2)$. Furthermore, we consider a normal-inverse-gamma prior for each pair $(\boldsymbol{\theta}_i, \sigma_i^2), i = 1, \ldots, n$:

$$(\boldsymbol{\theta}_i \,|\, \sigma_i^2) \sim \mathcal{N}(\mathbf{m}_i, \sigma_i^2\mathbf{V}_i), \qquad \sigma_i^2 \sim \mathcal{IG}(\nu_i, S_i), \tag{4}$$

---

[3]This more efficient blocking scheme has been used previously in the literature. For example, Eisenstat, Chan, and Strachan (2016) use it to speed up computations in the context of time-varying parameter VARs with stochastic volatility.

and we write $(\boldsymbol{\theta}_i, \sigma_i^2) \sim \mathcal{NIG}(\mathbf{m}_i, \mathbf{V}_i, \nu_i, S_i)$. In other words, the prior density of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{n} c_i (\sigma_i^2)^{-\left(\nu_i + 1 + \frac{k_i}{2}\right)} \mathrm{e}^{-\frac{1}{\sigma_i^2}\left(S_i + \frac{1}{2}(\boldsymbol{\theta}_i - \mathbf{m}_i)' \mathbf{V}_i^{-1}(\boldsymbol{\theta}_i - \mathbf{m}_i)\right)}, \tag{5}$$

where $c_i = (2\pi)^{-\frac{k_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} S_i^{\nu_i} / \Gamma(\nu_i)$.[4]

Since the prior variance of each element of $\boldsymbol{\theta}_i$ is controlled by the corresponding diagonal element of $\mathbf{V}_i$, it is obvious that this prior can accommodate different prior variances between own lags versus others. As we will show in the next section, this prior is also conjugate. To distinguish this from the natural conjugate prior, we call the prior in (4) the asymmetric conjugate prior. The hyperparameters of the asymmetric conjugate prior are $\mathbf{m}_i, \mathbf{V}_i, \nu_i$ and $S_i$, $i = 1, \ldots, n$. Next we describe how one can elicit these hyperparameters.

First partition $\mathbf{m}_i = (\mathbf{m}'_{\boldsymbol{\beta},i}, \mathbf{m}'_{\boldsymbol{\alpha},i})$ and $\mathbf{V}_i = \mathrm{diag}(\mathbf{V}_{\boldsymbol{\beta},i}, \mathbf{V}_{\boldsymbol{\alpha},i})$, where $\mathbf{m}_{\boldsymbol{\alpha},i}$ and $\mathbf{V}_{\boldsymbol{\alpha},i}$ are the hyperparameters corresponding to $\boldsymbol{\alpha}_i$, whereas $\mathbf{m}_{\boldsymbol{\beta},i}$ and $\mathbf{V}_{\boldsymbol{\beta},i}$ are those associated with $\boldsymbol{\beta}_i$. In what follows, we first discuss eliciting the hyperparameters associated with $\boldsymbol{\alpha}_i$ and $\sigma_i^2$ — i.e., $\mathbf{m}_{\boldsymbol{\alpha},i}, \mathbf{V}_{\boldsymbol{\alpha},i}, \nu_i$ and $S_i$. We then introduce two Minnesota-type shrinkage priors for the VAR coefficients $\boldsymbol{\beta}_i$.

Since $\boldsymbol{\alpha}_i$ and $\sigma_i^2$ control the reduced-form error covariance matrix $\widetilde{\boldsymbol{\Sigma}} = \mathbf{A}^{-1} \boldsymbol{\Sigma} (\mathbf{A}^{-1})'$, one concern is that an arbitrary choice of the hyperparameters for $\boldsymbol{\alpha}_i$ and $\sigma_i^2$ would induce some unreasonable prior on $\widetilde{\boldsymbol{\Sigma}}$. For example, the implied prior on the $i$-th diagonal element of $\widetilde{\boldsymbol{\Sigma}}$ might mechanically depend on its position in the $n$-tuple — i.e., the induced prior on $\widetilde{\boldsymbol{\Sigma}}$ depends on the order of the variables and is not invariant to reordering. This problem is especially acute for large systems due to the fact that $\mathbf{A}^{-1}$ is lower triangular.

To avoid this potential non-invariance problem, we instead specify a prior on the reduced-form error covariance matrix $\widetilde{\boldsymbol{\Sigma}}$. And given this prior on $\widetilde{\boldsymbol{\Sigma}}$, we then derive the implied prior on $\boldsymbol{\alpha}_i$ and $\sigma_i^2$, $i = 1, \ldots, n$. Since we can directly elicit prior beliefs on the elements

---

[4]The proposed prior is related to the work of Baumeister and Hamilton (2015), who also consider equation-specific conjugate priors on the structural VAR coefficients and error variances. But since they consider a more general setting with a possibly non-triangular impact matrix $\mathbf{A}$, they need a Metropolis-Hastings step to explore the marginal posterior distribution $\mathbf{A}$. In contrast, in our special case with a lower triangular $\mathbf{A}$, the proposed prior can be shown to be conjugate and direct sampling of all the parameters is available.

of $\widetilde{\boldsymbol{\Sigma}}$, as a result these prior beliefs do not depend on the parameters' position in $\widetilde{\boldsymbol{\Sigma}}$. To that end, we consider a standard inverse-Wishart prior on $\widetilde{\boldsymbol{\Sigma}}$ centered around $\mathbf{S} = \text{diag}(s_1^2, \ldots, s_n^2)$, where $s_i^2$ denotes the sample variance of the residuals from an AR(4) model for the variable $i, i = 1, \ldots, n$. More precisely, $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{S})$ with $\nu_0 = n + 2$. This prior on $\widetilde{\boldsymbol{\Sigma}}$ is commonly used in the literature (e.g., in Kadiyala and Karlsson, 1997; Carriero, Clark, and Marcellino, 2015). It turns out that, quite remarkably, the implied prior on $\boldsymbol{\alpha}_i$ and $\sigma_i^2$ is normal-inverse-gamma. The following proposition and corollary summarize this result.[5]

**Proposition 1.** *Consider the following normal-inverse-gamma priors on the diagonal elements of $\boldsymbol{\Sigma}$ and the lower triangular elements of $\mathbf{A}$:*

$$\sigma_i^2 \sim \mathcal{IG}\left(\frac{\nu_0 + i - n}{2}, \frac{s_i^2}{2}\right), \quad i = 1, \ldots, n, \tag{6}$$

$$(A_{i,j} \,|\, \sigma_i^2) \sim \mathcal{N}\left(0, \frac{\sigma_i^2}{s_j^2}\right), \quad 1 \leqslant j < i \leqslant n,\ i = 2, \ldots, n. \tag{7}$$

*Then, $\widetilde{\boldsymbol{\Sigma}}^{-1} = \mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}$ has the Wishart distribution $\widetilde{\boldsymbol{\Sigma}}^{-1} \sim \mathcal{W}(\nu_0, \mathbf{S}^{-1})$, where $\mathbf{S} = \text{diag}(s_1^2, \ldots, s_n^2)$. It follows that $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{S})$.*

The proof is given in Appendix C. Since the mapping $\widetilde{\boldsymbol{\Sigma}}^{-1} = \mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}$ is one-to-one, the converse of Proposition 1 is also true.

**Corollary 1.** *Using the same notations as in Proposition 1, if $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{S})$, then the implied priors on $A_{i,j}$ and $\sigma_i^2$, $i = 1, \ldots, n, j = 1, \ldots, i-1$, are the normal-inverse-gamma distributions given in (6) and (7).*

The proof is given in Appendix C. With these results, we can now make precise the claim that the proposed normal-inverse-gamma prior is independent of the order of the variables. Suppose the covariance matrix of the reduced-form error vector $\widetilde{\boldsymbol{\varepsilon}}_t^y$ is $\widetilde{\boldsymbol{\Sigma}}$, which has the prior $\mathcal{IW}(\nu_0, \mathbf{S})$. We claim that if we change the order of the reduced-form errors, we can use the normal-inverse-gamma prior to specify an appropriate inverse-Wishart prior — we

---

[5]In the context of a general structural VAR, Sims and Zha (1998) suggest precisely this approach of deriving the implied prior on the impact matrix from a natural prior (inverse-Wishart) on the error covariance matrix. However, under their parameterization (Cholesky factorization), the implied prior on the impact matrix is non-standard. In contrast, we use a modified Cholesky factorization and the implied prior can be shown to be of the form of a normal-inverse-gamma distribution.

simply need to reorder the associated hyperparameters accordingly. More specifically, let $\pi$ denote an arbitrary permutation of $n$ elements with the associated permutation matrix $\mathbf{P}_\pi$. If we permute the order of the dependent variables via $\mathbf{P}_\pi \mathbf{y}_t = (y_{\pi(1),t}, \ldots, y_{\pi(n),t})'$, its error covariance matrix is then $\mathbf{P}_\pi \widetilde{\boldsymbol{\Sigma}} \mathbf{P}_\pi'$ and has the $\mathcal{IW}(\nu_0, \mathbf{S}_\pi)$ distribution, where $\mathbf{S}_\pi = \mathbf{P}_\pi \mathbf{S} \mathbf{P}_\pi' = \mathrm{diag}(s^2_{\pi(1)}, \ldots, s^2_{\pi(n)})$. Using Proposition 1, we can find a set of normal-inverse-gamma priors that induces this inverse-Wishart prior $\mathcal{IW}(\nu_0, \mathbf{S}_\pi)$. We summarize this result in the following corollary.

**Corollary 2.** *Suppose* $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{S})$ *and let* $\pi$ *denote a permutation of* $n$ *elements with the associated permutation matrix* $\mathbf{P}_\pi$. *It follows that* $\mathbf{P}_\pi \widetilde{\boldsymbol{\Sigma}} \mathbf{P}_\pi' \sim \mathcal{IW}(\nu_0, \mathbf{S}_\pi)$, *where* $\mathbf{S}_\pi = diag(s^2_{\pi(1)}, \ldots, s^2_{\pi(n)})$. *Consider the following normal-inverse-gamma priors on the diagonal elements of* $\boldsymbol{\Sigma}$ *and the lower triangular elements of* $\mathbf{A}$:

$$\sigma_i^2 \sim \mathcal{IG}\left(\frac{\nu_0 + i - n}{2}, \frac{s^2_{\pi(i)}}{2}\right), \quad i = 1, \ldots, n,$$

$$(A_{i,j} \mid \sigma_i^2) \sim \mathcal{N}\left(0, \frac{\sigma_i^2}{s^2_{\pi(j)}}\right), \quad 1 \leqslant j < i \leqslant n, \ i = 2, \ldots, n.$$

*Then,* $\breve{\boldsymbol{\Sigma}}^{-1} = \mathbf{A}' \boldsymbol{\Sigma}^{-1} \mathbf{A}$ *has the Wishart distribution* $\mathcal{W}(\nu_0, \mathbf{S}_\pi^{-1})$. *That is,* $\breve{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{S}_\pi)$.

The proof follows directly from Proposition 1. In addition, Proposition 1 holds for the more general case where $\mathbf{S}$ is any symmetric positive definite matrix. That is, the induced priors on the structural-form variances are independent gamma distributions and the conditional priors of the free elements of $\mathbf{A}$ are normal distributions. But unlike the previous case with diagonal $\mathbf{S}$, here the free elements in the same row of $\mathbf{A}$ are correlated. We summarize the results in the following corollary.[6] Its proof is given in Appendix C.

**Corollary 3.** *Suppose* $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{R})$, *where* $\mathbf{R}$ *is a symmetric positive definite matrix. Factor* $\widetilde{\boldsymbol{\Sigma}}^{-1} = \mathbf{C}' \boldsymbol{\Sigma}^{-1} \mathbf{C}$ *and* $\mathbf{R}^{-1} = \mathbf{L}' \mathbf{S}^{-1} \mathbf{L}$, *where* $\mathbf{C}$ *and* $\mathbf{L}$ *are lower triangular matrices with ones on the main diagonal,* $\boldsymbol{\Sigma} = diag(\sigma_1^2, \ldots, \sigma_n^2)$ *and* $\mathbf{S} = diag(s_1^2, \ldots, s_n^2)$ *are diagonal matrices. Let* $\mathbf{c}_i$ *denote the free elements of the* $i$-*th row of* $\mathbf{C}$, *i.e.,* $\mathbf{c}_i = (C_{i,1}, \ldots, C_{i,i-1})'$, *and let* $\mathbf{L}_{1:i-1}$ *denote the* $(i-1) \times (i-1)$ *matrix that consists of the first* $(i-1)$ *rows and columns of* $\mathbf{L}$. *Similarly define* $\mathbf{l}_i$ *and* $\mathbf{S}_{1:i-1}$. *Then, the implied priors*

---

[6]Chan and Jeliazkov (2009) have shown a similar result. However, their proof is not sufficiently constructive and they did not give an explicit mapping between the inverse-Wishart parameters and the normal-inverse-gamma parameters.

*on $\mathbf{c}_i$ and $\sigma_i^2$ are*

$$\sigma_i^2 \sim \mathcal{IG}\left(\frac{\nu_0 + i - n}{2}, \frac{s_i^2}{2}\right), \quad i = 1, \ldots, n, \tag{8}$$

$$(\mathbf{c}_i \,|\, \sigma_i^2) \sim \mathcal{N}\left(\mathbf{l}_i, \sigma_i^2 \mathbf{L}'_{1:i-1} \mathbf{S}^{-1}_{1:i-1} \mathbf{L}_{1:i-1}\right), \quad i = 2, \ldots, n. \tag{9}$$

*Since the mapping $\widetilde{\boldsymbol{\Sigma}}^{-1} = \mathbf{C}' \boldsymbol{\Sigma}^{-1} \mathbf{C}$ is one-to-one, the converse is also true. That is, if we assume that $(\mathbf{c}_i, \sigma_i^2)$ follows the normal-inverse-gamma distributions given in (8) and (9), then the implied prior on $\widetilde{\boldsymbol{\Sigma}}$ is inverse-Wishart: $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{R})$.*

Hence, the above proposition and corollaries give us a guide to elicit the hyperparameters associated with $\boldsymbol{\alpha}_i$ and $\sigma_i^2$. In our baseline case, we set $\nu_i = 1 + i/2$, $S_i = s_i^2/2$, $\mathbf{m}_{\boldsymbol{\alpha},i} = \mathbf{0}$ and $\mathbf{V}_{\boldsymbol{\alpha},i} = \mathrm{diag}(1/s_1^2, \ldots, 1/s_{i-1}^2)$. These values imply that the prior on the reduced-form error covariance matrix is $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(n+2, \mathbf{S})$ with prior mean $\mathbf{S}$. Furthermore, if we wish to have an inverse-Wishart prior on $\widetilde{\boldsymbol{\Sigma}}$ with a non-diagonal prior mean, we can simply use Corollary 3 to elicit the associated values for $\mathbf{m}_{\boldsymbol{\alpha},i}$ and $\mathbf{V}_{\boldsymbol{\alpha},i}$.

Next, we describe two ways to set the hyperparameters $\mathbf{m}_{\boldsymbol{\beta},i}$ and $\mathbf{V}_{\boldsymbol{\beta},i}$. The first approach directly elicits prior beliefs on the structural VAR coefficients $\boldsymbol{\beta}_i$. The idea of cross-variable shrinkage in this setting has been previously explored in Leeper, Sims, and Zha (1996) and Sims and Zha (1998). We therefore follow Sims and Zha (1998), who consider Minnesota-type shrinkage priors for VAR coefficients in the structural form.[7] More specifically, we set $\mathbf{m}_{\boldsymbol{\beta},i} = \mathbf{0}$ to shrink the VAR coefficients to zero for growth rates data; for level data, $\mathbf{m}_{\boldsymbol{\beta},i}$ is set to be zero as well except the coefficient associated with the first own lag, which is set to be one.

Recall that $\mathbf{V}_{\boldsymbol{\beta},i}$ is the ratio of the prior covariance matrix of $\boldsymbol{\beta}_i$ relative to the error variance $\sigma_i^2$. Similar to the Minnesota prior, here we assume $\mathbf{V}_{\boldsymbol{\beta},i}$ to be diagonal with the

---

[7]As noted by Sims and Zha (1998), a general structural VAR is a simultaneous equations model and there is no dependent variable in an equation (other than an arbitrary normalization). Hence, the distinction between 'own' lags vs 'others' is not meaningful. However, our setup is the special case of a recursive system, where there is a natural dependent variable in each equation. Hence, we can distinguish between 'own' vs 'other' lags.

$k$-th diagonal element $(\mathbf{V}_{\boldsymbol{\beta},i})_k$ set to be:

$$(\mathbf{V}_{\boldsymbol{\beta},i})_k = \begin{cases} \frac{\kappa_1}{l^2 s_i^2}, & \text{for the coefficient on the } l\text{-th lag of variable } i, \\ \frac{\kappa_2}{l^2 s_j^2}, & \text{for the coefficient on the } l\text{-th lag of variable } j, j \neq i, \\ \kappa_3, & \text{for the intercept.} \end{cases}$$

The hyperparameter $\kappa_1$ controls the overall shrinkage strength for coefficients on own lags, whereas $\kappa_2$ controls those on lags of other variables. These two hyperparameters will play a key role in the empirical analysis, and we will select them optimally by maximizing the associated marginal likelihood. We set $\kappa_3 = 100$, which implies essentially no shrinkage for the intercepts.[8]

By contrast, in the second approach we first elicit the prior means and variances on the reduced-form VAR coefficients. We then derive the implied prior means and variances on the structural-form VAR coefficients. Since the prior beliefs are elicited on the reduced-form parameters, this approach is comparable to commonly-used Minnesota priors on the reduced-form parameters. More specifically, let $\widetilde{\mathbf{m}}_{\boldsymbol{\beta},i}$ and $\sigma_i^2 \widetilde{\mathbf{V}}_{\boldsymbol{\beta},i}$ denote the prior mean vector and covariance matrix of the reduced-form parameters $\widetilde{\boldsymbol{\beta}}_i$ — these hyperparameters can be elicited similarly as above. For later reference, we let $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ denote the corresponding shrinkage hyperparameters. Finally, let $\mathbf{m}_{\boldsymbol{\beta},i}$ and $\sigma_i^2 \mathbf{V}_{\boldsymbol{\beta},i}$ denote the corresponding hyperparameters of the structural-form parameters $\boldsymbol{\beta}_i$. Appendix D provides the details of the derivation and the explicit formulas of $\mathbf{m}_{\boldsymbol{\beta},i}$ and $\mathbf{V}_{\boldsymbol{\beta},i}$ using $\widetilde{\mathbf{m}}_{\boldsymbol{\beta},i}$ and $\widetilde{\mathbf{V}}_{\boldsymbol{\beta},i}$ as inputs. Both approaches give a normal-inverse-gamma prior of the form given in (4).

## 2.3 Posterior Distribution and Efficient Sampling

In this section we first derive the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ under the asymmetric conjugate prior and show that it has indeed the same form as the prior. Then, we describe an efficient method for posterior simulation.

---

[8]In principle one can select $\kappa_3$ optimally as well, but the corresponding optimization is more costly to solve. More generally, high-dimensional numerical optimization using derivative-free methods is time consuming. One feasible alternative is to use Automatic Differentiation to obtain the relevant partial derivatives, which are then fed to numerical optimization routines that use these partial derivatives to more efficiently find the maximizer. See Chan, Jacobi, and Zhu (2020) for an example.

Since both the likelihood in (3) and the prior in (5) have the product form, we can estimate each pair $(\boldsymbol{\theta}_i, \sigma_i^2)$ separately. More specifically, the posterior distribution of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ is given by:

$$p(\boldsymbol{\theta},\boldsymbol{\sigma}^2 \,|\, \mathbf{y}) \propto p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2) p(\mathbf{y} \,|\, \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{n} p(\boldsymbol{\theta}_i, \sigma_i^2) p(\mathbf{y}_i \,|\, \boldsymbol{\theta}_i, \sigma_i^2)$$

$$= \prod_{i=1}^{n} c_i (\sigma_i^2)^{-\left(\nu_i + 1 + \frac{k_i}{2}\right)} \mathrm{e}^{-\frac{1}{\sigma_i^2}\left(S_i + \frac{1}{2}(\boldsymbol{\theta}_i - \mathbf{m}_i)' \mathbf{V}_i^{-1}(\boldsymbol{\theta}_i - \mathbf{m}_i)\right)} \times \left(2\pi\sigma_i^2\right)^{-\frac{T}{2}} \mathrm{e}^{-\frac{1}{2\sigma_i^2}(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_i)'(\mathbf{y}_i - \mathbf{X}_i \boldsymbol{\theta}_i)}$$

$$= \prod_{i=1}^{n} c_i (2\pi)^{-\frac{T}{2}} (\sigma_i^2)^{-\left(\nu_i + \frac{T+k_i}{2} + 1\right)} \mathrm{e}^{-\frac{1}{\sigma_i^2}\left(S_i + \frac{1}{2}\left(\boldsymbol{\theta}_i'(\mathbf{V}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i)\boldsymbol{\theta}_i - 2\boldsymbol{\theta}_i'(\mathbf{V}_i^{-1}\mathbf{m}_i + \mathbf{X}_i'\mathbf{y}_i) + \mathbf{m}_i'\mathbf{V}_i^{-1}\mathbf{m}_i + \mathbf{y}_i'\mathbf{y}_i\right)\right)}$$

$$= \prod_{i=1}^{n} c_i (2\pi)^{-\frac{T}{2}} (\sigma_i^2)^{-\left(\nu_i + \frac{T+k_i}{2} + 1\right)} \mathrm{e}^{-\frac{1}{\sigma_i^2}\left(\widehat{S}_i + \frac{1}{2}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i)' \mathbf{K}_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i)\right)},$$

where $\mathbf{K}_{\boldsymbol{\theta}_i} = \mathbf{V}_i^{-1} + \mathbf{X}_i'\mathbf{X}_i$, $\widehat{\boldsymbol{\theta}}_i = \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}(\mathbf{V}_i^{-1}\mathbf{m}_i + \mathbf{X}_i'\mathbf{y}_i)$ and $\widehat{S}_i = S_i + (\mathbf{y}_i'\mathbf{y}_i + \mathbf{m}_i'\mathbf{V}_i^{-1}\mathbf{m}_i - \widehat{\boldsymbol{\theta}}_i'\mathbf{K}_{\boldsymbol{\theta}_i}\widehat{\boldsymbol{\theta}}_i)/2$. Hence, the posterior distribution is a product of $n$ normal-inverse-gamma distributions and we have:

$$(\boldsymbol{\theta}_i, \sigma_i^2 \,|\, \mathbf{y}) \sim \mathcal{NIG}\left(\widehat{\boldsymbol{\theta}}_i, \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}, \nu_i + \frac{T}{2}, \widehat{S}_i\right), \quad i = 1, \ldots, n. \tag{10}$$

Using properties of the normal-inverse-gamma distribution, it is easy to see that the posterior means of $\boldsymbol{\theta}_i$ and $\sigma_i^2$ are respectively $\widehat{\boldsymbol{\theta}}_i$ and $\widehat{S}_i/(\nu_i + T/2 - 1)$. Other posterior moments can also be obtained by using similar properties of the normal-inverse-gamma distribution. For other quantities of interest where analytical results are not available, we can estimate them by posterior simulation. For example, the $h$-step-ahead predictive distribution of $\mathbf{y}_{T+h}$ is non-standard. But we can obtain posterior draws from $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 \,|\, \mathbf{y})$ to construct the $h$-step-ahead predictive distribution.

In what follows, we outline an efficient method to simulate a sample of size $M$ from the posterior distribution. Here we can directly generate *independent* draws from the posterior distribution as opposed to MCMC draws that are correlated by construction. First, note that $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 \,|\, \mathbf{y})$ is a product of $n$ normal-inverse-gamma distributions as given in (10). Thus we can sample each pair $(\boldsymbol{\theta}_i, \sigma_i^2 \,|\, \mathbf{y})$ individually. Next, we can sample $(\boldsymbol{\theta}_i, \sigma_i^2 \,|\, \mathbf{y})$ in two steps. First, we draw $\sigma_i^2$ marginally from $(\sigma_i^2 \,|\, \mathbf{y}) \sim \mathcal{IG}(\nu_i + T/2, \widehat{S}_i)$.

Then, given the $\sigma_i^2$ sampled, we obtain $\boldsymbol{\theta}_i$ from the conditional distribution

$$(\boldsymbol{\theta}_i \,|\, \mathbf{y}, \sigma_i^2) \sim \mathcal{N}(\widehat{\boldsymbol{\theta}}_i, \sigma_i^2 \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}).$$

Here the covariance matrix $\sigma_i^2 \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}$ is of dimension $k_i = np + i$. When $n$ is large, sampling from this normal distribution using conventional methods — based on the Cholesky factor of $\sigma_i^2 \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}$ — is computationally intensive for two reasons. First, inverting the $k_i \times k_i$ matrix $\mathbf{K}_{\boldsymbol{\theta}_i}$ to obtain the covariance matrix $\sigma_i^2 \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}$ is computationally costly. Second, the Cholesky factor of the covariance matrix needs to be computed $M$ times — once for each draw of $\sigma_i^2$ from the marginal distribution. It turns out that both of these computationally intensive steps can be avoided.

To that end, we introduce the following notations: given a non-singular square matrix $\mathbf{F}$ and a conformable vector $\mathbf{d}$, let $\mathbf{F} \backslash \mathbf{d}$ denote the unique solution to the linear system $\mathbf{F}\mathbf{z} = \mathbf{d}$, i.e., $\mathbf{F} \backslash \mathbf{d} = \mathbf{F}^{-1}\mathbf{d}$. When $\mathbf{F}$ is lower triangular, this linear system can be solved quickly by forward substitution; when $\mathbf{F}$ is upper triangular, it can be solved by backward substitution.[9] Now, compute the Cholesky factor $\mathbf{C}_{\mathbf{K}_{\boldsymbol{\theta}_i}}$ of $\mathbf{K}_{\boldsymbol{\theta}_i}$ such that $\mathbf{K}_{\boldsymbol{\theta}_i} = \mathbf{C}_{\mathbf{K}_{\boldsymbol{\theta}_i}} \mathbf{C}'_{\mathbf{K}_{\boldsymbol{\theta}_i}}$. Note that this needs to be done only once. Let $\mathbf{u}$ be a $k_i \times 1$ vector of independent sample from $\mathcal{N}(0, \sigma_i^2)$. Then, return

$$\widehat{\boldsymbol{\theta}}_i + \mathbf{C}'_{\mathbf{K}_{\boldsymbol{\theta}}} \backslash \mathbf{u},$$

which has the $\mathcal{N}(\widehat{\boldsymbol{\theta}}_i, \sigma_i^2 \mathbf{K}_{\boldsymbol{\theta}_i}^{-1})$ distribution.[10] Finally, we can further speed up the computations by vectorizing all operations to obtain $M$ posterior draws instead of using for-loops.

This sampling scheme is more efficient than the method in Carriero, Clark, and Marcellino (2019), who propose estimating the reduced-form parameters equation-by-equation. The main reason is that their method requires computing the Cholesky factor of every sampled reduced-form error covariance matrix (e.g., a total of $M$ times for $M$ draws as they use MCMC). In contrast, the proposed method needs to compute the Cholesky factor of the error covariance matrix only once, because here it does not depend on the sampled coefficients (i.e., the proposed sampler is not a Gibbs sampler). This difference becomes

---

[9]Forward and backward substitutions are implemented in standard packages such as MATLAB, GAUSS and R. In MATLAB, for example, it is done by `mldivide(F, d)` or simply `F\d`.

[10]Note that $\widehat{\boldsymbol{\theta}}_i$ can be obtained similarly without explicitly computing the inverse of $\mathbf{K}_{\boldsymbol{\theta}_i}$. Specifically, it is easy to see that $\widehat{\boldsymbol{\theta}}_i$ can be calculated as: $\mathbf{C}'_{\mathbf{K}_{\boldsymbol{\theta}_i}} \backslash (\mathbf{C}_{\mathbf{K}_{\boldsymbol{\theta}_i}} \backslash (\mathbf{V}_i^{-1}\mathbf{m}_i + \mathbf{X}'_i\mathbf{y}_i))$ by forward then backward substitution. Also note that since $\mathbf{V}_i^{-1}$ is diagonal, its inverse is straightforward to compute.

more important when $n$ becomes larger, as number of operations for computing the Cholesky factor of an $n \times n$ matrix is $\mathcal{O}(n^3)$.

To get a sense of how long it takes to obtain posterior draws using the proposed algorithm, we fit Bayesian VARs of different sizes, each with $p = 4$ lags. The algorithm is implemented using MATLAB on a desktop with an Intel Core i7-7700 @3.60 GHz processor and 64GB memory. The computation times (in seconds) to obtain 10,000 posterior draws of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ are reported in Table 1. As it is evident from the table, the proposed method is fast and scales well. It also compares favorably to the algorithm in Carriero, Clark, and Marcellino (2019), especially when $n$ is large. For example, for a large BVAR with $n = 100$ variables, the proposed method takes about half a minute to obtain 10,000 posterior draws. In comparison, using the algorithm in Carriero, Clark, and Marcellino (2019) takes about 43 minutes.[11]

Table 1: The computation times (in seconds) to obtain 10,000 posterior draws of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ using the proposed method compared to the method in Carriero, Clark, and Marcellino (2019). All BVARs have $p = 4$ lags.

|  | $n = 25$ | $n = 50$ | $n = 100$ |
|---|---|---|---|
| proposed method | 1.3 | 6.8 | 28 |
| CCM | 58 | 238 | 2,574 |

## 2.4 The Marginal Likelihood

In this section we provide an analytical expression of the marginal likelihood. This closed-form expression is useful for a range of purposes, such as obtaining optimal hyperparamaters or designing efficient estimation algorithms for more flexible Bayesian VARs.

To prevent arithmetic underflow and overflow, we evaluate the marginal likelihood in log scale. Given the likelihood function in (3) and the asymmetric conjugate prior in (5), the

---

[11] As recently pointed out in Bognanni (2021), the original algorithm in Carriero, Clark, and Marcellino (2019) is not exact and can only be viewed as an approximation. The corrected algorithm described in Carriero, Chan, Clark, and Marcellino (2021) is about 3-10 times slower than the original algorithm, depending on the size of the system. Hence, the advantage of our posterior sampler is even more substantial when the corrected algorithm is used.

associated log marginal likelihood of the VAR has the following analytical expression:

$$
\begin{aligned}
\log p(\mathbf{y}) = &-\frac{Tn}{2}\log(2\pi) + \sum_{i=1}^{n}\left[-\frac{1}{2}\left(\log|\mathbf{V}_i| + \log|\mathbf{K}_{\boldsymbol{\theta}_i}|\right)\right.\\
&\left.+\log\Gamma\left(\nu_i + \frac{T}{2}\right) + \nu_i\log S_i - \log\Gamma(\nu_i) - \left(\nu_i + \frac{T}{2}\right)\log\widehat{S}_i\right].
\end{aligned}
\tag{11}
$$

The details of the derivation are given in Appendix B. The above expression is straightforward to evaluate. We only note that to compute the log determinant $\log|\mathbf{K}_{\boldsymbol{\theta}_i}|$, it is numerically more stable to first compute its Cholesky factor $\mathbf{C}_{\mathbf{K}_{\boldsymbol{\theta}_i}}$ and return $2\sum\log c_{ii}$, where $c_{ii}$ is the $i$-th diagonal element of the $\mathbf{C}_{\mathbf{K}_{\boldsymbol{\theta}_i}}$.

# 3   Extensions

In this section we briefly discuss how we can use the above analytical results and the efficient sampling scheme in more general settings. Suppose we augment our BVAR in (3) to the model $p(\mathbf{y}\,|\,\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma})$ with the additional parameter vector $\boldsymbol{\gamma}$. Further, consider the prior $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma}) = p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2\,|\,\boldsymbol{\gamma})p(\boldsymbol{\gamma})$, where $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2\,|\,\boldsymbol{\gamma})$ is the asymmetric conjugate prior that could potentially depend on $\boldsymbol{\gamma}$ and the marginal prior $p(\boldsymbol{\gamma})$ is left unspecified for now. Before we discuss some efficient posterior samplers, we first give two examples that fit into this framework.

In our first example, we augment the BVAR by treating the hyperparameters $\kappa_1$ and $\kappa_2$ as parameters to be estimated. That is, $\boldsymbol{\gamma} = (\kappa_1, \kappa_2)'$. This extension is useful as it takes into account the parameter uncertainty of $\kappa_1$ and $\kappa_2$ (see also Giannone, Lenza, and Primiceri, 2015). This extension is considered in the empirical application. In our second example, we extend the BVAR by adding an MA(1) component to each equation:

$$
\begin{aligned}
y_{i,t} &= \mathbf{x}_{i,t}\boldsymbol{\theta}_i + \varepsilon_{i,t}^y,\\
\varepsilon_{i,t}^y &= u_{i,t} + \psi_i u_{i,t-1},
\end{aligned}
$$

where $u_{i,t} \sim \mathcal{N}(0, \sigma_i^2), t = 1, \ldots, T, i = 1, \ldots, n$. In this case, $\boldsymbol{\gamma} = (\psi_1, \ldots, \psi_n)'$. This extension is motivated by the empirical finding that allowing for moving average errors often improves forecast performance (see, e.g., Chan, 2013, 2020a).

Both examples fit into the framework with likelihood $p(\mathbf{y} \mid \boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma})$ and prior $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2, \boldsymbol{\gamma})$. One natural posterior sampler is to construct a Markov chain by sequentially sampling from $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}, \boldsymbol{\gamma})$ and $p(\boldsymbol{\gamma} \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)$. The first density is a product of normal-inverse-gamma densities, and we can efficiently obtain a draw from it as described before. The second density depends on the model, but it is often easy to sample from.[12]

Alternatively, a more efficient approach is the collapsed sampler that samples from $p(\boldsymbol{\gamma} \mid \mathbf{y})$. This sampling scheme is typically more efficient as it integrates out the high-dimensional parameters $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ analytically. The density $p(\boldsymbol{\gamma} \mid \mathbf{y})$ can be evaluated quickly since

$$p(\boldsymbol{\gamma} \mid \mathbf{y}) \propto p(\mathbf{y} \mid \boldsymbol{\gamma}) p(\boldsymbol{\gamma}),$$

where $p(\mathbf{y} \mid \boldsymbol{\gamma})$ is the 'marginal likelihood' of the standard BVAR. For instance, for the first example with $\boldsymbol{\gamma} = (\kappa_1, \kappa_2)'$, the quantity $p(\mathbf{y} \mid \boldsymbol{\gamma})$ is exactly as the analytical expression given in (11). Finally, given the posterior draws of $\boldsymbol{\gamma}$, we can obtain the posterior draws of $(\boldsymbol{\theta}, \boldsymbol{\sigma}^2)$ from $p(\boldsymbol{\theta}, \boldsymbol{\sigma}^2 \mid \mathbf{y}, \boldsymbol{\gamma})$.

# 4    Application: Identifying Financial Shocks

In this section we illustrate the usefulness of the proposed asymmetric conjugate prior by revisiting the empirical study in Furlanetto, Ravazzolo, and Sarferaz (2019), who identify financial shocks in a VAR using sign restrictions. More specifically, for their baseline they consider a 6-variable VAR — using US data on GDP, prices, interest rate, investment, stock prices and credit spread — to identify demand, supply, monetary, investment and financial shocks using a set of sign restrictions on the contemporaneous impact matrix. Here we revisit their empirical study by using a larger set of variables and consider a 15-variable VAR.

There are a few reasons in favor of using a larger set of macroeconomic and financial variables. First, there is the concern of informational deficiency of using a limited information set. As pointed out in the seminal papers by Hansen and Sargent (1991) and

---

[12]For our first example with a low-dimensional $\boldsymbol{\gamma} = (\kappa_1, \kappa_2)'$, an independent-chain Metropolis-Hastings algorithm can be easily constructed. For our second example with $\boldsymbol{\gamma} = (\psi_1, \dots, \psi_n)'$, it turns out that we can factor $p(\boldsymbol{\gamma} \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = p(\boldsymbol{\psi} \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2) = \prod_{i=1}^{n} p(\psi_i \mid \mathbf{y}, \boldsymbol{\theta}, \boldsymbol{\sigma}^2)$. Then, each $\psi_i$ can be simulated using the method in Chan (2013).

Lippi and Reichlin (1993, 1994), when the econometrician considers a narrower set of variables than the economic agent, the underlying model used by the econometrician is non-fundamental, in the sense that current and past observations of the variables do not span the same space spanned by the structural shocks. Consequently, structural shocks cannot be recovered from the model. Using a larger set of relevant variables can alleviate this concern (see, e.g., Gambetti, 2021, for a recent review on non-fundamentalness).

Second, the mapping from variables in an economic model to the data is typically not unique. For example, take the economic variable inflation. Should one match it to the data based on the CPI, PCE, or the GDP deflator (Loria, Matthes, and Wang, 2020)? One natural way to circumvent an arbitrary choice is to include multiple data series corresponding to the same economic variable in the analysis. An added benefit of this approach is that by using more data series (and the corresponding sign restrictions), one often obtains sharper inference.

While Furlanetto, Ravazzolo, and Sarferaz (2019) consider a non-informative prior for their 6-variable VAR, shrinkage is essential here for our much larger system. In particular, we use the proposed asymmetric conjugate prior to shrink the VAR coefficients in a data-based manner. After describing the macroeconomic dataset in Section 4.1, we first present the estimates of the shrinkage hyperparameters in Section 4.2, highlighting the empirical relevance of allowing for different levels of shrinkage on own lags and other lags. We then conduct a structural analysis with sign restrictions in Section 4.3.

## 4.1   Data

The dataset for our empirical application consists of 15 US quarterly variables and the sample period is from 1985:Q1 to 2019:Q4. These variables are constructed from raw time-series obtained from various sources, including the FRED database at the Federal Reserve Bank of St. Louis and the Federal Reserve Bank of Philadelphia. The complete list of these time-series and their sources are given in Appendix A.

In addition to the 6 variables used in the baseline model in Furlanetto, Ravazzolo, and Sarferaz (2019) — namely, real GDP, GDP deflator, 3-month treasury rate, ratio of private investment over output, S&P 500 index and a credit spread defined as the difference between Moody's baa corporate bond yield and the federal funds rate— we include 9

additional macroeconomic and financial variables, such as ratio of total credit over real estate value, industrial production, mortgage rates, as well as other measures of inflation, interest rates and stock prices. These 15 variables are listed in Table 3.

## 4.2 Optimal Shrinkage Hyperparameters

In this section we fit the 15-variable VAR with the proposed asymmetric conjugate prior. More specifically, we implement the version that elicits prior beliefs on the reduced-form parameters $\widetilde{\boldsymbol{\beta}}_i$ with hyperparameters $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$. We obtain the optimal hyperparameter values by maximizing the log marginal likelihood. The results are reported in Table 2.

For comparison, we also consider two useful benchmarks. In the first case we set $\widetilde{\kappa}_1 = \widetilde{\kappa}_2 = \widetilde{\kappa}$ and maximize the log marginal likelihood with respect to $\widetilde{\kappa}$ only. This benchmark mimics the standard practice of using the natural conjugate prior that does not distinguish between own lags and lags of other variables. We refer to this version as the symmetric prior. The second benchmark is a set of subjectively chosen values that apply cross-variable shrinkage. In particular, we follow Carriero, Clark, and Marcellino (2015) and consider $\widetilde{\kappa}_1 = 0.04$ and $\widetilde{\kappa}_2 = 0.0016$. This second benchmark is referred to as the subjective prior.

Table 2: Optimal values of the hyperparameters under the symmetric prior ($\widetilde{\kappa}_1 = \widetilde{\kappa}_2$), the subjective prior (Carriero, Clark, and Marcellino, 2015) and the proposed asymmetric prior (elicited on the reduced-form parameters).

|  | Symmetric prior | Subjective prior | Asymmetric prior |
|---|---|---|---|
| $\widetilde{\kappa}_1$ | 0.008 | 0.040 | 0.058 |
| $\widetilde{\kappa}_2$ | 0.008 | 0.0016 | 0.0043 |
| log-ML | 4,333.3 | 4,329.8 | 4,341.6 |

Under the symmetric prior with the restriction that $\widetilde{\kappa}_1 = \widetilde{\kappa}_2$, the optimal hyperparameter value is 0.008. If we allow $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ to be different, we obtain very different results: the optimal value for $\widetilde{\kappa}_1$ increases more than 7 times to 0.058, whereas the optimal value of $\widetilde{\kappa}_2$ reduces by about half to 0.0043. These results suggest that the data prefers shrinking the coefficients on lags of other variables much more aggressively to zero than those on own lags. This makes intuitive sense as one would expect, on average, a variable's own lags

would contain more information about its future evolution than lags of other variables. By relaxing the restriction that $\widetilde{\kappa}_1 = \widetilde{\kappa}_2$, the marginal likelihood value increases by about 4,000. If we were to test the hypothesis that $\widetilde{\kappa}_1 = \widetilde{\kappa}_2$, this large difference in marginal likelihood values would have decidedly reject it.

In addition, the optimal values of $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ under the asymmetric prior are also quite different from those of the subjective prior with $\widetilde{\kappa}_1 = 0.04$ and $\widetilde{\kappa}_2 = 0.0016$. By selecting the values of $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ optimally, one can increase the marginal likelihood value by more than 13,000. These results suggest that the subjective prior shrinks both the coefficients on own and other lags too aggressively for our dataset.

We have so far taken the empirical Bayes approach of choosing hyperparameter values by maximizing the log marginal likelihood. A fully Bayesian approach would specify proper priors on $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ and obtain the corresponding posterior distribution. The latter approach has the additional advantage of being able to quantity parameter uncertainty of $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$. In view of this, we take a fully Bayesian approach and treat $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ as parameters to be estimated. Specifically, we assume a uniform prior on the unit square $(0, 1) \times (0, 1)$ for $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$, and compute the marginal posterior distribution of $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$. The contour plot of the joint posterior density is reported in Figure 1.

As the contour plot in the right panel shows, most of the mass of $\widetilde{\kappa}_1$ lies between 0.03 and 0.09, whereas the mass of $\widetilde{\kappa}_2$ is mostly between 0.002 and 0.007. That is, $\widetilde{\kappa}_1$ tends to be an order of magnitude larger than $\widetilde{\kappa}_2$. These results confirm the conclusion that one should shrink the coefficients on other lags much more aggressively to zero than those on own lags. Moreover, since there is virtually no mass along the diagonal line $\widetilde{\kappa}_1 = \widetilde{\kappa}_2$, requiring them to be the same as in the natural conjugate prior appears to be too restrictive. As a comparison, we also plot the values of $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$ under the symmetric and subjective priors. As is evident from the figure, the values under both priors are far from the high-density region of the posterior distribution.

Overall, the estimation results indicate that the optimal hyperparameter values could be very different from some subjectively chosen values commonly used in empirical work. In addition, they also highlight the importance of allowing for different levels of shrinkage on own versus other lags — and therefore the empirical relevance of the proposed asymmetric conjugate prior.
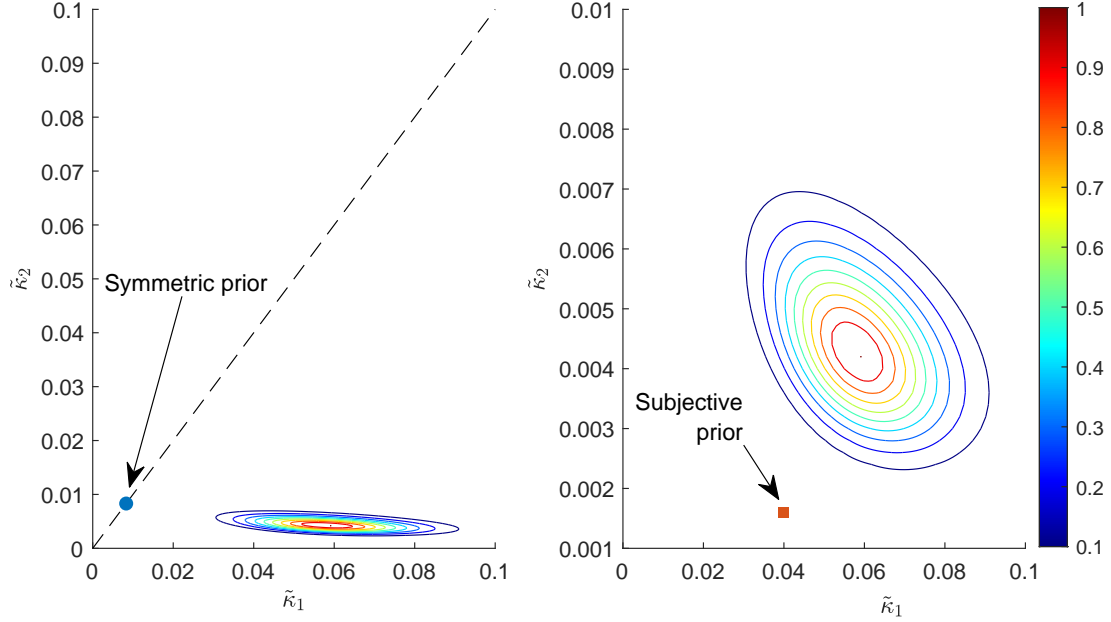
Figure 1: Contour plot of the joint posterior density of $\widetilde{\kappa}_1$ and $\widetilde{\kappa}_2$. The mode of the density is normalized to one for easy comparison.

## 4.3 Sign Restrictions and Impulse Responses

Next, we revisit the empirical application in Furlanetto, Ravazzolo, and Sarferaz (2019) that identifies financial shocks using sign restrictions on the contemporaneous impact matrix. We first replicate their baseline results from a 6-variable VAR to identify 5 structural shocks, namely, demand, supply, monetary, investment and financial shocks. We then consider a larger, 15-variable VAR with the asymmetric conjugate prior to identify the same structural shocks.

For the replication exercise, we use the same 6 variables and the associated sign restrictions considered in Furlanetto, Ravazzolo, and Sarferaz (2019), which are listed in the first 6 rows of Table 3. The sign restrictions to identify supply, demand and monetary shocks are standard and are consistent with a wide range of dynamic stochastic general equilibrium models (see, e.g., Canova and Paustian, 2011). To disentangle investment and financial shocks from demand shocks, they are assumed to have different effects on the ratio of investment over output. More specifically, positive investment and financial

shocks have a positive effect on the ratio, which is in line with the idea that these shocks create investment booms. In contrast, positive demand shocks reduce the ratio of investment over output — investment level can still increase in response to demand shocks, but not as much as other components of the output.

Finally, stock prices are used to disentangle investment shocks from financial shocks, following the influential paper by Christiano, Motto, and Rostagno (2014). In particular, investment shocks are assumed to have a negative impact on the stock prices, whereas financial shocks have a positive impact. The first assumption captures the idea that investment shocks are shocks to the supply of capital, which generate negative comovements between the stock of capital and its price, where the price of capital is viewed as a proxy of the value of equity. In contrast, financial shocks are stocks to the demand for capital, which generate positive comovements between the stock of capital and its price.

We then augment the 6-variable VAR with 9 additional variables. Some of the new variables are alternative time series corresponding to a particular economic variable (e.g., CPI and PCE as prices; the Dow Jones Industrial Average as stock prices). In those cases the same sign restrictions for the economic variable as discussed above are imposed. In addition, we also include a few other seemingly relevant variables to alleviate the concern of informational deficiency. The additional variables and the corresponding sign restrictions are listed in rows 7-15 of Table 3.

Given the posterior draws of the parameters, the algorithm proposed in Rubio-Ramirez, Waggoner, and Zha (2010) is used to incorporate the sign restrictions to construct impulse responses. This procedure can be computational intensive when the number of variables and the number of identified shocks are large. In particular, when there are a large number of sign restrictions, it is not uncommon to require millions of posterior draws to obtain enough samples that satisfy all the sign restrictions.[13] For instance, Furlanetto, Ravazzolo, and Sarferaz (2019) consider a non-informative prior and construct impulse

---

[13]There are different implementations of the algorithm in Rubio-Ramirez, Waggoner, and Zha (2010). One common approach, as is done in Furlanetto, Ravazzolo, and Sarferaz (2019), is to combine every uniform draw from the orthogonal group O($n$) with a new posterior draw to generate candidate impulse responses. If the impulse responses do not satisfy all the sign restrictions, a new orthogonal matrix and a new posterior draw are used to generate candidate impulse responses. Alternatively, one can reuse the posterior draw and only generate a new orthogonal matrix. In this case the number of posterior draws required can be reduced. But to ensure the algorithm terminates, one typically sets an upper bound on the number of times a posterior draw can be reused.

responses using a standard Gibbs sampler in conjunction with the algorithm in Rubio-Ramirez, Waggoner, and Zha (2010). They report estimation time of about a week for the 6-variable VAR using a 12-core workstation. For our 15-variable VAR with more sign restrictions, it would be extremely computationally expensive to use a standard Gibbs sampler to obtain enough posterior draws. To make estimation feasible, we adopt the proposed asymmetric conjugate prior and the associated efficient sampler to generate posterior draws of the structural parameters. These structural parameters are then transformed to the corresponding reduced-form parameters, which are then used to construct candidate impulse responses.

Table 3: Sign restrictions and identified shocks.

| | Supply | Demand | Monetary | Investment | Financial |
|---|---|---|---|---|---|
| GDP | + | + | + | + | + |
| GDP deflator | − | + | + | + | + |
| 3-month tbill rate | NA | + | − | + | + |
| Investment/output | NA | − | NA | + | + |
| S&P 500 | + | NA | NA | − | + |
| Spread | NA | NA | NA | NA | NA |
| Spread 2 | NA | NA | NA | NA | NA |
| Credit/Real estate value | NA | NA | NA | NA | NA |
| Mortgage rates | NA | NA | NA | NA | NA |
| CPI | − | + | + | + | + |
| PCE | − | + | + | + | + |
| employment | NA | NA | NA | NA | NA |
| Industrial production | + | + | + | + | + |
| 1-year tbill rate | NA | + | − | + | + |
| DJIA | + | NA | NA | − | + |

Note: Spread is defined as the difference between Moody's baa corporate bond yield and the federal funds rate; Spread 2 is the difference between Moody's baa corporate bond yield and 10-year treasury yield.

We first report results from the baseline model in Furlanetto, Ravazzolo, and Sarferaz (2019) using the asymmetric conjugate prior elicited on the reduced-form parameters and an updated dataset. Since an improper (non-informative) prior is used in the original study, to make our results comparable, we consider a proper but relatively vague prior by setting $\widetilde{\kappa}_1 = \widetilde{\kappa}_2 = 1$. Figure 2 plots the impulse responses of the 6 variables to an one-standard-deviation financial shock.

Despite using a different prior and dataset, our results are remarkably similar to those reported in Figure 1 of Furlanetto, Ravazzolo, and Sarferaz (2019). In particular, we also find a large effect on GDP and a relatively smaller impact on prices. The responses of investment and stock prices are persistent. In particular, the credible intervals exclude 0 for the first 7 quarters after impact for both variables. It is also interesting to note that even though the responses of the credit spread are unrestricted in the estimation, they are strongly counter-cyclical — this highlights the fact that the estimated structural shock behaves like a financial stock. The only noticeable difference is that our median responses of prices are all positive, whereas those in Furlanetto, Ravazzolo, and Sarferaz (2019) have mixed signs (although the credible intervals in both cases are relatively wide and include 0 for some periods).
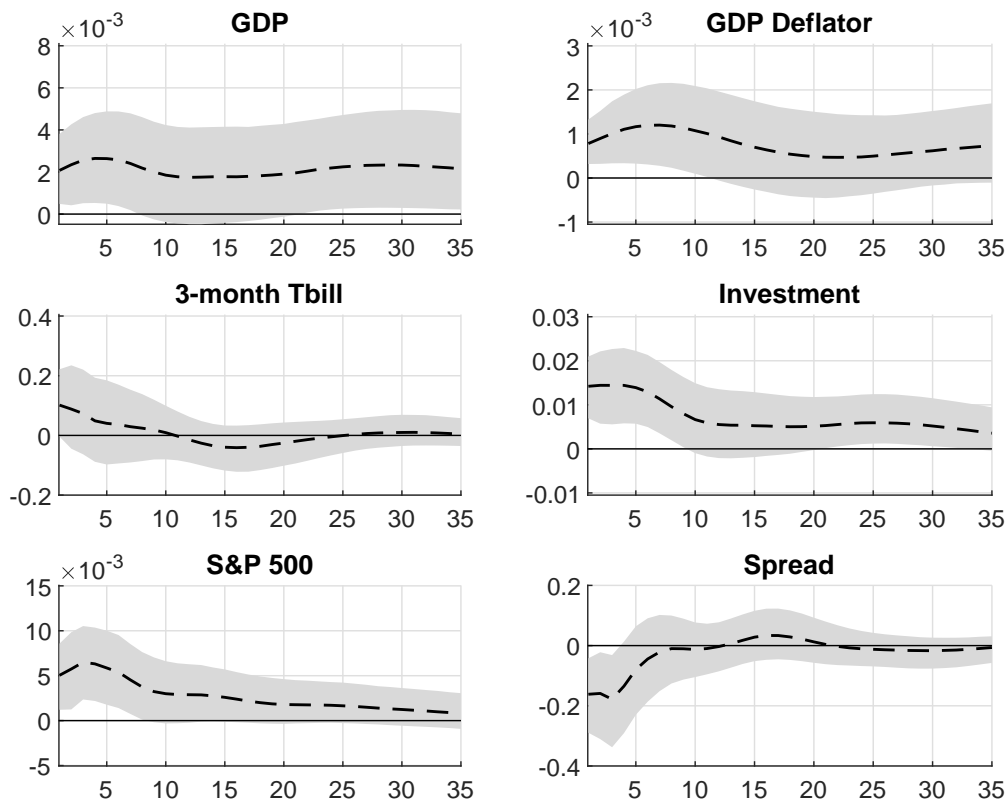


Figure 2: Impulse responses from a 6-variable VAR with the asymmetric conjugate prior to an one-standard-deviation financial shock. The shaded region represents the 16-th and 84-th percentiles.

As a comparison, we also compute the corresponding impulse responses under a standard independent normal and inverse-Wishart priors: a Minnesota prior on the VAR coefficients with hyperparameters $\widetilde{\kappa}_1 = \widetilde{\kappa}_2 = 1$ and an inverse-Wishart prior on $\widetilde{\Sigma}$. The results are almost identical to those obtained under the asymmetric conjugate prior (results are reported in Appendix E), suggesting any minor differences from Furlanetto, Ravazzolo, and Sarferaz (2019) are likely due to a slightly different dataset used.
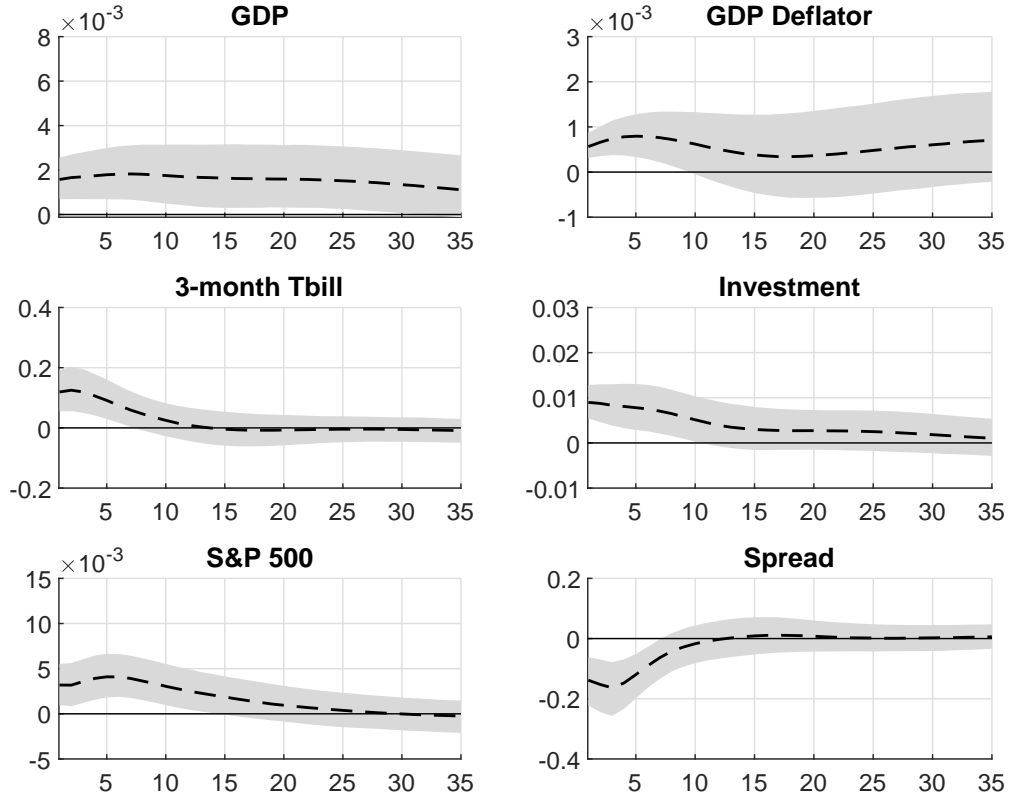


Figure 3: Impulse responses from a 15-variable VAR with the asymmetric conjugate prior to an one-standard-deviation financial shock. The shaded region represents the 16-th and 84-th percentiles.

Next, we construct the impulse responses from the 15-variable VAR. Given the large number of VAR coefficients, appropriate shrinkage is vital in this case. We therefore obtain the optimal shrinkage hyperparameters by maximizing the marginal likelihood of the VAR with the asymmetric conjugate prior. As reported in Section 4.2, the optimal

shrinkage hyperparameters are found to be $\widetilde{\kappa}_1 = 0.058$ and $\widetilde{\kappa}_2 = 0.0043$. The following results are based on these hyperparameter values.

Figure 3 reports the impulse responses of the same 6 variables to an one-standard-deviation financial shock. Compared to the results from the 6-variable VAR depicted in Figure 2, the median responses are mostly the same, but the credible intervals are substantially narrower. For example, the responses of GDP are much more precisely estimated — the credible intervals exclude 0 for the first 30 quarters after impact. These results highlight the usefulness of the proposed asymmetric conjugate prior and the value of including more variables and sign restrictions in identifying structural shocks.

As mentioned earlier, there are typically multiple data series corresponding to the same economic variable. And it is often unclear which one should be used, if only one variable is to be selected. In our application, for example, the time series GDP deflator, CPI and PCE are all good candidates for the economic variable prices. Instead of picking one out of the three candidates, we include them all in the analysis. Figure 4 reports impulse responses of the three measure of prices to an one-standard-deviation financial shock. The median responses of the three variables are all positive and have very similar shapes. However, their credible intervals are somewhat different — only in the case of CPI do the credible intervals mostly exclude 0. If one had only used CPI, one might have drawn a stronger conclusion than it is warranted. This again highlights the benefit of including more variables and performing a more comprehensive structural analysis.
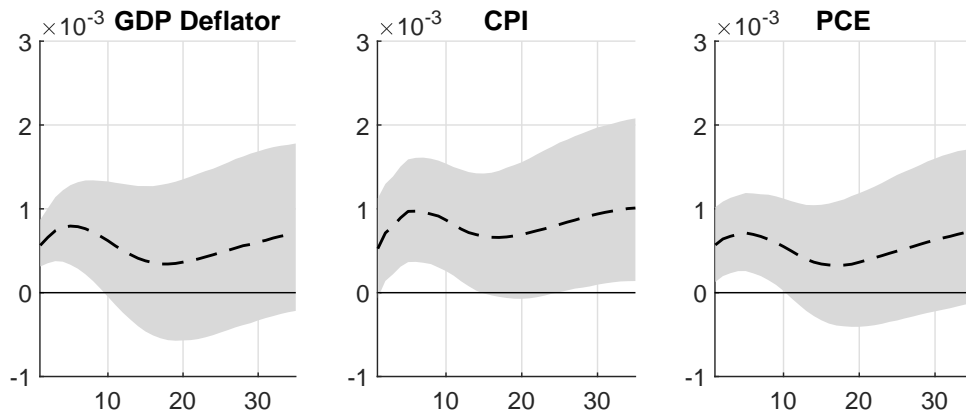


Figure 4: Impulse responses of prices to an one-standard-deviation financial shock. The shaded region represents the 16-th and 84-th percentiles.

# 5 Concluding Remarks and Future Research

We developed a new asymmetric conjugate prior for large BVARs that can accommodate cross-variable shrinkage, while maintaining many useful analytical results as the natural conjugate prior. Using the new prior and the associated efficient sampler, we were able to identify 5 structural shocks using a 15-variable VAR with a set of sign restrictions on the contemporaneous impact matrix. We showed that the larger number of variables and sign restrictions — when used in conjunction with the asymmetric conjugate prior — can sharpen inference on impulse responses.

There is now a large empirical literature that shows that models with stochastic volatility tend to forecast substantially better (Clark, 2011; D'Agostino, Gambetti, and Giannone, 2013; Cross and Poon, 2016). In future work, it would be useful to develop similar efficient posterior samplers for large BVARs with stochastic volatility, such as the models in Carriero, Clark, and Marcellino (2019) and Chan, Eisenstat, and Strachan (2020). In addition, it would be interesting to use Proposition 1 to construct a multivariate stochastic volatility model that is invariant to reordering of the variables.

# Appendix A: Data

This appendix provides the details of the raw data used to construct the variables in the empirical application. In particular, Table 4 lists the variables and their sources. The sample period is from 1985:Q1 to 2019:Q4.

Table 4: Description of variables used in the empirical application.

| Variable | Description | Source |
| --- | --- | --- |
| GDP | Log of real GNP/GDP | Federal Reserve Bank of Philadelphia |
| GDP Deflator | Log of price index of GNP/GDP | Federal Reserve Bank of Philadelphia |
| 3-month treasury bill | 3-month treasury bill rate | Federal Reserve Bank of St. Louis |
| Investment | Log of real gross private domestic investment | Federal Reserve Bank of St. Louis |
| S&P 500 | Log of S&P 500 | Yahoo Finance |
| Total credit | Log of loans to non-financial private sector | Board of Governors of the Federal Reserve System |
| Mortgages | Log of home mortgages of households and non-profit organizations | Federal Reserve Bank of St. Louis |
| Real estate value | Log of real estate at market value of households and non-profit organizations | Federal Reserve Bank of St. Louis |
| Corporate bond yield | Moody's baa corporate bond yield | Federal Reserve Bank of St. Louis |
| 10-year treasury note | 10-year treasury constant maturity rate | Federal Reserve Bank of St. Louis |
| Federal funds rate | Federal funds rate | Federal Reserve Bank of St. Louis |
| Mortgage rate | 30-year fixed rate mortgage average | Federal Reserve Bank of St. Louis |
| CPI | Log of consumer price index | Federal Reserve Bank of St. Louis |
| PCE | Log of price index of personal consumption expenditure | Federal Reserve Bank of St. Louis |
| Employment | Log of employment level | Federal Reserve Bank of St. Louis |
| Industrial production | Log of industrial production index | Federal Reserve Bank of St. Louis |
| 1-year treasury bill | 1-year treasury constant maturity rate | Federal Reserve Bank of St. Louis |
| Dow Jones Industrial Average | Log of Dow Jones Industrial Average index | Google Finance |

# Appendix B: Derivation of the Marginal Likelihood

In this appendix we prove that the marginal likelihood of the VAR($p$) under the asymmetric conjugate prior in (4) has the following expression:

$$p(\mathbf{y}) = \prod_{i=1}^{n} (2\pi)^{-\frac{T}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} |\mathbf{K}_{\boldsymbol{\theta}_i}|^{-\frac{1}{2}} \frac{\Gamma(\nu_i + T/2) S_i^{\nu_i}}{\Gamma(\nu_i) \widehat{S}_i^{\nu_i + \frac{T}{2}}}.$$

This result follows from direct computation:

$$
\begin{aligned}
p(\mathbf{y}) = \prod_{i=1}^{n} p(\mathbf{y}_i) &= \prod_{i=1}^{n} \int p(\boldsymbol{\theta}_i, \sigma_i^2) p(\mathbf{y}_i \mid \boldsymbol{\theta}_i, \sigma^2) \mathrm{d}(\boldsymbol{\theta}_i, \sigma_i^2) \\
&= \prod_{i=1}^{n} c_i (2\pi)^{-\frac{T}{2}} \int (\sigma_i^2)^{-\left(\nu_i + \frac{T+k_i}{2} + 1\right)} \mathrm{e}^{-\frac{1}{\sigma_i^2}\left(\widehat{S}_i + \frac{1}{2}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i)' \mathbf{K}_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i)\right)} \mathrm{d}(\boldsymbol{\theta}_i, \sigma_i^2) \\
&= \prod_{i=1}^{n} c_i (2\pi)^{-\frac{T}{2}} (2\pi)^{\frac{k_i}{2}} |\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}|^{\frac{1}{2}} \frac{\Gamma(\nu_i + T/2)}{\widehat{S}_i^{\nu_i + \frac{T}{2}}} \\
&= \prod_{i=1}^{n} (2\pi)^{-\frac{T}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} |\mathbf{K}_{\boldsymbol{\theta}_i}|^{-\frac{1}{2}} \frac{\Gamma(\nu_i + T/2) S_i^{\nu_i}}{\Gamma(\nu_i) \widehat{S}_i^{\nu_i + \frac{T}{2}}}.
\end{aligned}
$$

where $c_i = (2\pi)^{-\frac{k_i}{2}} |\mathbf{V}_i|^{-\frac{1}{2}} S_i^{\nu_i} / \Gamma(\nu_i)$, $\mathbf{K}_{\boldsymbol{\theta}_i} = \mathbf{V}_i^{-1} + \mathbf{X}_i' \mathbf{X}_i$, $\widehat{\boldsymbol{\theta}}_i = \mathbf{K}_{\boldsymbol{\theta}_i}^{-1}(\mathbf{V}_i^{-1} \mathbf{m}_i + \mathbf{X}_i' \mathbf{y}_i)$ and $\widehat{S}_i = S_i + (\mathbf{y}_i' \mathbf{y}_i + \mathbf{m}_i' \mathbf{V}_i^{-1} \mathbf{m}_i - \widehat{\boldsymbol{\theta}}_i' \mathbf{K}_{\boldsymbol{\theta}_i} \widehat{\boldsymbol{\theta}}_i)/2$. In the above derivation we have used the fact that

$$\int (\sigma_i^2)^{-\left(\nu_i + \frac{T+k_i}{2} + 1\right)} \mathrm{e}^{-\frac{1}{\sigma_i^2}\left(\widehat{S}_i + \frac{1}{2}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i)' \mathbf{K}_{\boldsymbol{\theta}_i}(\boldsymbol{\theta}_i - \widehat{\boldsymbol{\theta}}_i)\right)} = (2\pi)^{\frac{k_i}{2}} |\mathbf{K}_{\boldsymbol{\theta}_i}^{-1}|^{\frac{1}{2}} \frac{\Gamma(\nu_i + T/2)}{\widehat{S}_i^{\nu_i + \frac{T}{2}}}.$$

The above equality holds because the quantity on the right-hand side is the normalizing constant of the $(\boldsymbol{\theta}_i, \sigma_i^2) \sim \mathcal{NIG}(\mathbf{m}_i, \mathbf{V}_i, \nu_i, S_i)$ distribution.

# Appendix C: Proofs of Proposition and Corollaries

This appendix provides proofs of Proposition 1 and two corollaries in the main text. We first record in the following lemma the determinant of the Jacobian of transformation from the structural-form parameterization to the reduced-form parameterization.[14] This lemma was proved in Chan and Jeliazkov (2009), and we include it here for convenience. The proof uses the differential forms approach that is equivalent to calculating the Jacobian (see, e.g., Theorem 2.1.1 in Muirhead, 1982).

**Lemma 1.** Suppose $\mathbf{W}$ is a $n \times n$ positive definite matrix and $\mathbf{W} = \mathbf{T}'\widetilde{\mathbf{T}}\mathbf{T}$, where $\mathbf{T}$ is a lower triangular matrix with ones on the main diagonal and $\widetilde{\mathbf{T}}$ is a diagonal matrix with positive diagonal elements. Denote the lower diagonal elements of $\mathbf{T}$ by $t_{ij}$, $1 \leqslant j < i \leqslant n$, and the diagonal elements of $\widetilde{\mathbf{T}}$ by $t_{ii}$, $i = 1, \ldots, n,$. Let $(\mathrm{d}\mathbf{W})$ denote the differential form $(\mathrm{d}\mathbf{W}) \equiv \underset{i \geqslant j}{\wedge} \mathrm{d}w_{ij}$ and similarly define $(\mathrm{d}\mathbf{T}) \equiv \underset{i \geqslant j}{\wedge} \mathrm{d}t_{ij}$. Then we have

$$(\mathrm{d}\mathbf{W}) = \prod_{i=1}^{n} t_{ii}^{i-1}(\mathrm{d}\mathbf{T}).$$

In other words, the determinant of the Jacobian of the transformation from $\mathbf{T}'\widetilde{\mathbf{T}}\mathbf{T}$ to $\mathbf{W}$ is $\prod_{i=1}^{n} t_{ii}^{-i+1}$.

**Proof of the lemma**: By the definition $\mathbf{W} = \mathbf{T}'\widetilde{\mathbf{T}}\mathbf{T}$, we have

$$
\begin{pmatrix}
w_{11} & w_{21} & \ldots & w_{n1} \\
w_{21} & w_{22} & \ldots & \vdots \\
\vdots & \vdots & \ddots & \vdots \\
w_{n1} & w_{n2} & \ldots & w_{nn}
\end{pmatrix}
=
\begin{pmatrix}
1 & t_{21} & \ldots & t_{n1} \\
0 & 1 & \ldots & t_{n2} \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & 1
\end{pmatrix}
\begin{pmatrix}
t_{11} & 0 & \ldots & 0 \\
0 & t_{22} & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
0 & 0 & \ldots & t_{nn}
\end{pmatrix}
\begin{pmatrix}
1 & 0 & \ldots & 0 \\
t_{21} & 1 & \ldots & 0 \\
\vdots & \vdots & \ddots & \vdots \\
t_{n1} & t_{n2} & \ldots & 1
\end{pmatrix}.
$$

---

[14]For a different structural-form parameterization where the error covariance matrix is the identity matrix and the impact matrix is triangular with free diagonal elements, Zha (1999) derives the determinant of the Jacobian of transformation from the reduced-form parameterization.

Hence, we can express each $w_{ij}$ in terms of $\{t_{ij}\}$:

$$w_{ii} = t_{ii} + \sum_{j=i+1}^{n} t_{ji}^2 t_{jj}, \quad i = 1, \ldots, n, \tag{12}$$

$$w_{ij} = t_{ij}t_{ii} + \sum_{k=i+1}^{n} t_{ki}t_{kj}t_{kk}, \quad 1 \leqslant j < i \leqslant n. \tag{13}$$

Next, we take differentials of these two equations so that we can write the differential form $(\mathrm{d}\mathbf{W})$ in terms of $(\mathrm{d}\mathbf{T})$. Since we are going to take the exterior product of these differentials and the exterior products of repeated differentials are zero, there is no need to keep track of differentials in $t_{ij}$ that have previously occurred. Therefore, we take differentials of (12) and (13) and ignore those that have previously occurred:

$$\mathrm{d}w_{nn} = \mathrm{d}t_{nn}$$

$$\mathrm{d}w_{n,n-1} = \mathrm{d}t_{nn}\mathrm{d}t_{n,n-1} + \ldots$$

$$\vdots$$

$$\mathrm{d}w_{n1} = t_{nn}\mathrm{d}t_{n1} + \ldots$$

$$\mathrm{d}w_{n-1,n-1} = \mathrm{d}t_{n-1,n-1} + \ldots$$

$$\vdots$$

$$\mathrm{d}w_{11} = \mathrm{d}t_{11} + \ldots$$

Finally, taking exterior products gives

$$\bigwedge_{i \geqslant j} \mathrm{d}w_{ij} = t_{nn}^{n-1}t_{n-1,n-1}^{n-2} \cdots t_{22} \bigwedge_{i \geqslant j} \mathrm{d}t_{ij}$$

as claimed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

**Proof of Proposition 1**: Assume the same notation as in Lemma 1. To prove Proposition 1, we consider the case where

$$t_{ii} \sim \mathcal{G}\left(\frac{\nu_0 + i - n}{2}, \frac{s_i^2}{2}\right), \quad i = 1, \ldots, n, \tag{14}$$

$$(t_{ij} \mid t_{ii}) \sim \mathcal{N}\left(0, \frac{t_{ii}^{-1}}{s_j^2}\right), \quad 1 \leqslant j < i \leqslant n, \ i = 2, \ldots, n. \tag{15}$$

More specifically, we will show that the density of $\mathbf{W} = \mathbf{T}'\widetilde{\mathbf{T}}\mathbf{T}$ is the same as that of the Wishart distribution $\mathcal{W}(\nu, \mathbf{S}^{-1})$, where $\mathbf{S} = \text{diag}(s_1^2, \ldots, s_n^2)$. Then, if we let $t_{ii} = 1/\sigma_i^2$ and $A_{i,j} = t_{ij}$, we have $\widetilde{\boldsymbol{\Sigma}}^{-1} = \mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A} \sim \mathcal{W}(\nu_0, \mathbf{S}^{-1})$.

To prove the proposition, we first compute the determinant of $\mathbf{W}$ and the trace $\text{tr}(\mathbf{SW})$. Since the determinant of $\mathbf{T}$ is 1, we have

$$|\mathbf{W}| = |\widetilde{\mathbf{T}}| = \prod_{i=1}^{n} t_{ii}.$$

Next, using (12), we have

$$\text{tr}(\mathbf{SW}) = \sum_{i=1}^{n} w_{ii}s_i^2$$

$$= \sum_{i=1}^{n} t_{ii}s_i^2 + \sum_{i=1}^{n}\sum_{j=i+1}^{n} t_{ji}^2 t_{jj}s_i^2$$

$$= \sum_{i=1}^{n} t_{ii}s_i^2 + \sum_{j=2}^{n}\sum_{i=1}^{j-1} t_{ji}^2 t_{jj}s_i^2$$

$$= \sum_{i=1}^{n} t_{ii}s_i^2 + \sum_{i=2}^{n}\sum_{j=1}^{i-1} t_{ij}^2 t_{ii}s_j^2,$$

where we change the order of the double summations in the third equality and interchange the dummy indices $i$ and $j$ in the last equality.

Now, it follows from the distributional assumptions in (14) and (15) that the kernel of the joint density of $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ is

$$\prod_{i=1}^{n} t_{ii}^{\frac{\nu_0+i-n}{2}-1} e^{-\frac{s_i^2}{2}t_{ii}} \times \prod_{i=2}^{n} t_{ii}^{\frac{i-1}{2}} e^{-\frac{1}{2}\sum_{j=1}^{i-1} t_{ij}^2 t_{ii}s_j^2}$$

$$= \left(\prod_{i=1}^{n} t_{ii}^{\frac{\nu_0-n-1}{2}+(i-1)}\right) e^{-\frac{1}{2}\left(\sum_{i=1}^{n} t_{ii}s_i^2 + \sum_{i=2}^{n}\sum_{j=1}^{i-1} t_{ij}^2 t_{ii}s_j^2\right)}.$$

Next, we derive the kernel of the density of $\mathbf{W}$. By the lemma, the determinant of the Jacobian is $\prod_{i=1}^{n} t_{ii}^{-i+1}$. Substituting $\text{tr}(\mathbf{W})$ and $|\mathbf{W}|$ into the above expression and

multiplying the determinant of the Jacobian, we obtain the kernel of the density of $\mathbf{W}$:

$$|\mathbf{W}|^{\frac{\nu_0-n-1}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\mathbf{SW})},$$

which is the kernel of the Wishart density $\mathcal{W}(\nu_0,\mathbf{S}^{-1})$. $\qquad\qquad\square$

**Proof of Corollary 1**: Here we use the same notation as in Proposition 1. Assume $\mathbf{W}\sim\mathcal{W}(\nu,\mathbf{S}^{-1})$, and let $\mathbf{W}=\mathbf{T}'\widetilde{\mathbf{T}}\mathbf{T}$, where $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ are given in Lemma 1. If we can show that $t_{ii}$ and $(t_{ij}\,|\,t_{ii})$ follow the same normal-gamma distributions given in (14) and (15), respectively, then we are done. Since the transformation between $\mathbf{W}$ and $\mathbf{T}'\widetilde{\mathbf{T}}\mathbf{T}$ is one-to-one, the proof essentially just "reverses" the equalities given in Proposition 1. More specifically, in the proof of Proposition 1 we showed that $|\mathbf{W}|=\prod_{i=1}^n t_{ii}$ and

$$\mathrm{tr}(\mathbf{SW})=\sum_{i=1}^n t_{ii}s_i^2+\sum_{i=2}^n\sum_{j=1}^{i-1}t_{ij}^2 t_{ii}s_j^2.$$

Also, by Lemma 1, the determinant of the Jacobian of transformation is $\prod_{i=1}^n t_{ii}^{i-1}$. Hence, the kernel of the joint distribution of $t_{ii},i=1,\ldots,n$, and $t_{ij},1\leqslant j<i\leqslant n,\ i=2,\ldots,n$ is given by:

$$\begin{aligned}
|\mathbf{W}|^{\frac{\nu_0-n-1}{2}}\mathrm{e}^{-\frac{1}{2}\mathrm{tr}(\mathbf{SW})}&\times\prod_{i=1}^n t_{ii}^{i-1}\\
&=\left(\prod_{i=1}^n t_{ii}^{\frac{\nu_0-n-1}{2}+(i-1)}\right)\mathrm{e}^{-\frac{1}{2}\left(\sum_{i=1}^n t_{ii}s_i^2+\sum_{i=2}^n\sum_{j=1}^{i-1}t_{ij}^2 t_{ii}s_j^2\right)}\\
&=\prod_{i=1}^n t_{ii}^{\frac{\nu_0+i-n}{2}-1}\mathrm{e}^{-\frac{s_i^2}{2}t_{ii}}\times\prod_{i=2}^n t_{ii}^{\frac{i-1}{2}}\mathrm{e}^{-\frac{1}{2}\sum_{j=1}^{i-1}t_{ij}^2 t_{ii}s_j^2}.
\end{aligned}$$

It follows that $t_{ii},i=1,\ldots,n$, are independent gamma random variables given in (14). Moreover, conditional on $t_{ii}$, $t_{ij},1\leqslant j<i$, are independent normal variables given in (15).

**Proof of Corollary 3**: Suppose $\widetilde{\boldsymbol{\Sigma}}\sim\mathcal{IW}(\nu_0,\mathbf{R})$, where $\mathbf{R}$ is a symmetric positive definite matrix. Factor $\mathbf{R}^{-1}=\mathbf{L}'\mathbf{S}^{-1}\mathbf{L}$, where $\mathbf{L}$ is lower triangular with ones on the main diagonal and $\mathbf{S}$ is diagonal. Since $\widetilde{\boldsymbol{\Sigma}}^{-1}\sim\mathcal{W}(\nu_0,\mathbf{R}^{-1})$, by the properties of the Wishart distribution, we have $(\mathbf{L}')^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{L}^{-1}\sim\mathcal{W}(\nu_0,\mathbf{S}^{-1})$. Now, applying Corollary 1, we obtain $(\mathbf{L}')^{-1}\widetilde{\boldsymbol{\Sigma}}^{-1}\mathbf{L}^{-1}=\mathbf{A}'\boldsymbol{\Sigma}^{-1}\mathbf{A}$, where $\mathbf{A}$ is lower triangular with ones on the main diagonal and $\boldsymbol{\Sigma}$ is diagonal. The diagonal elements of $\boldsymbol{\Sigma}$ and the lower triangular elements

33

of $\mathbf{A}$ follow the normal-inverse-gamma distributions:

$$\sigma_i^2 \sim \mathcal{IG}\left(\frac{\nu_0 + i - n}{2}, \frac{s_i^2}{2}\right), \quad i = 1, \ldots, n,$$

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

Letting $\mathbf{C} = \mathbf{AL}$, we can write $\widetilde{\boldsymbol{\Sigma}}^{-1} = \mathbf{C}'\boldsymbol{\Sigma}^{-1}\mathbf{C}$. Since both $\mathbf{A}$ and $\mathbf{L}$ are lower triangular with ones on the main diagonal, so is $\mathbf{C}$. It remains to show that $\mathbf{c}_i$, the free elements of the $i$-th row of $\mathbf{C}$, follows the normal distribution in (9). Since $\mathbf{C}' = \mathbf{L}'\mathbf{A}'$, we can write $\mathbf{c}_i$ in terms of $\mathbf{A}$ and $\mathbf{L}$ as:

$$\mathbf{c}_i = \mathbf{l}_i + \mathbf{L}'_{1:i-1}\mathbf{a}_i,$$

where $\mathbf{l}_i$ and $\mathbf{a}_i$ are respectively the free elements of the $i$-th row of $\mathbf{L}$ and $\mathbf{A}$, and $\mathbf{L}_{1:i-1}$ is the $(i-1) \times (i-1)$ matrix that consists of the first $(i-1)$ rows and columns of $\mathbf{L}$. Since $\mathbf{c}_i$ is an affine transformation of the normal vector $\mathbf{a}_i$, conditional on $\sigma_i^2$, $\mathbf{c}_i$ is normally distributed with mean vector $\mathbf{l}_i$ and covariance matrix $\sigma_i^2 \mathbf{L}'_{1:i-1}\mathbf{S}_{1:i-1}^{-1}\mathbf{L}_{1:i-1}$, where $\mathbf{S}_{1:i-1}$ is the submatrix consisting of the first $(i-1)$ rows and columns of $\mathbf{S} = \mathrm{diag}(s_1^2, \ldots, s_n^2)$.

$\square$

# Appendix D: Derivation of the Implied Structural-Form Hyperparameters

This appendix derives the implied prior hyperparameters of the structural-form VAR coefficients from the hyperparameters of the reduced-form VAR coefficients. Let $(B_k)_{ij}$ and $(\widetilde{B}_k)_{ij}$ denote the $(i,j)$-th element of the structural-form coefficient matrix $\mathbf{B}_k$ and the reduced-form coefficient matrix $\widetilde{\mathbf{B}}_k$, respectively, for $i, j = 1, \ldots, n, k = 1, \ldots, p$. Further, let $\mathbf{m}_{\widetilde{\boldsymbol{\beta}},i}$ and $\sigma_i^2 \mathbf{V}_{\widetilde{\boldsymbol{\beta}},i}$ denote the prior mean vector and covariance matrix of the reduced-form parameters $\widetilde{\boldsymbol{\beta}}_i = (\widetilde{b}_i, (\widetilde{B}_1)_{i1}, \ldots, (\widetilde{B}_1)_{in}, \ldots, (\widetilde{B}_p)_{i1}, \ldots, (\widetilde{B}_p)_{in})'$. The hyperparameters $\mathbf{m}_{\widetilde{\boldsymbol{\beta}},i}$ and $\mathbf{V}_{\widetilde{\boldsymbol{\beta}},i}$ can be elicited in a standard way — e.g., following the Minnesota prior in Doan, Litterman, and Sims (1984) and Litterman (1986). We further assume that $\mathbf{V}_{\widetilde{\boldsymbol{\beta}},i}$ is diagonal as is usually done in the literature. The goal is to derive the structural-form hyperparameters $\mathbf{m}_{\boldsymbol{\beta},i}$ and $\mathbf{V}_{\boldsymbol{\beta},i}$ given $\mathbf{m}_{\widetilde{\boldsymbol{\beta}},i}$ and $\mathbf{V}_{\widetilde{\boldsymbol{\beta}},i}$.

To that end, recall that the structural-form and reduced-form coefficients are related via

$$(B_k)_{ij} = (\widetilde{B}_k)_{ij} + \sum_{l=1}^{i-1} A_{i,l}(\widetilde{B}_k)_{lj}.$$

Using the law of iterated expectations and the assumption that the prior means of $\boldsymbol{\alpha}_i = (A_{i,1}, \ldots, A_{i,i-1})'$ are zero (see Proposition 1), we obtain

$$\mathbb{E}(B_k)_{ij} = \mathbb{E}\left[\mathbb{E}((B_k)_{ij} \mid \widetilde{\mathbf{B}}_k)\right] = \mathbb{E}(\widetilde{B}_k)_{ij}.$$

Hence, we have $\mathbf{m}_{\boldsymbol{\beta},i} = \mathbf{m}_{\widetilde{\boldsymbol{\beta}},i}$.

Next, we elicit the prior covariance matrix $\mathbf{V}_{\boldsymbol{\beta},i}$, which is in general a full matrix. Here for simplicity we ignore the correlations between the structural-form coefficients and set $\mathbf{V}_{\boldsymbol{\beta},i}$ to be diagonal. One could work out all the covariances following a similar derivation presented below. Now, we derive the variance of a generic element $(B_k)_{ij}$. Let $f(j,k) = (k-1)n + j + 1$ be an integer-value function keeping track of the indices. Further, let $(V_{\widetilde{\boldsymbol{\beta}},i})_{f(j,k)}$ and $(m_{\widetilde{\boldsymbol{\beta}},i})_{f(j,k)}$ denote the $f(j,k)$-th diagonal element of $\mathbf{V}_{\widetilde{\boldsymbol{\beta}},i}$ and

the $f(j,k)$-th element of $\mathbf{m}_{\widetilde{\beta},i}$, respectively. Then, we have

$$
\begin{aligned}
\mathrm{Var}((B_k)_{ij}) &= \mathbb{E}\left[\mathrm{Var}((B_k)_{ij} \mid \mathbf{A})\right] + \mathrm{Var}\left[\mathbb{E}((B_k)_{ij} \mid \mathbf{A})\right] \\
&= \mathbb{E}\left[\mathrm{Var}((\widetilde{B}_k)_{ij}) + \sum_{l=1}^{i-1} A_{i,l}^2 \mathrm{Var}((\widetilde{B}_k)_{lj})\right] + \mathrm{Var}\left[\mathbb{E}(\widetilde{B}_k)_{ij} + \sum_{l=1}^{i-1} A_{i,l}\mathbb{E}(\widetilde{B}_k)_{lj}\right] \\
&= \mathrm{Var}((\widetilde{B}_k)_{ij}) + \sum_{l=1}^{i-1} \mathrm{Var}(A_{i,l})\mathrm{Var}((\widetilde{B}_k)_{lj}) + \sum_{l=1}^{i-1} \mathrm{Var}(A_{i,l})(\mathbb{E}(\widetilde{B}_k)_{lj})^2 \\
&= \sigma_i^2 (V_{\widetilde{\beta},i})_{f(j,k)} + \sum_{l=1}^{i-1} \frac{\sigma_i^2}{s_l^2}\left[\sigma_l^2 (V_{\widetilde{\beta},l})_{f(j,k)} + (m_{\widetilde{\beta},l})_{f(j,k)}^2\right].
\end{aligned}
$$

Note that this variance involves the error variances of other equations, namely, $\sigma_1^2, \ldots, \sigma_{i-1}^2$. To avoid this issue, we replace $\sigma_l^2$ by its prior mean $s_l^2$. Using this simplification, we obtain:

$$
\mathrm{Var}((B_k)_{ij}) \approx \sigma_i^2 \left[(V_{\widetilde{\beta},i})_{f(j,k)} + \sum_{l=1}^{i-1}\left((V_{\widetilde{\beta},l})_{f(j,k)} + s_l^{-2}(m_{\widetilde{\beta},l})_{f(j,k)}^2\right)\right].
$$

Given this approximation, we set the diagonal element in $\mathbf{V}_{\beta,i}$ associated with $(B_k)_{ij}$ to be $(V_{\widetilde{\beta},i})_{f(j,k)} + \sum_{l=1}^{i-1}((V_{\widetilde{\beta},l})_{f(j,k)} + s_l^{-2}(m_{\widetilde{\beta},l})_{f(j,k)}^2)$.

# Appendix E: Comparison with Independent Normal and Inverse-Wishart Prior

This appendix reports impulse responses of the 6-variable VAR described in Section 4 under the independent normal and inverse-Wishart priors. More specifically, we consider a standard Minnesota prior on the VAR coefficients with hyperparameters $\widetilde{\kappa}_1 = \widetilde{\kappa}_2 = 1$ and an inverse-Wishart prior on the reduced-form error covariance matrix $\widetilde{\boldsymbol{\Sigma}} \sim \mathcal{IW}(\nu_0, \mathbf{S})$. Given the posterior draws of the structural-form parameters, we transform them to the corresponding reduced-form parameters. We then use the algorithm described in Rubio-Ramirez, Waggoner, and Zha (2010) to incorporate the sign restrictions to construct impulse responses.
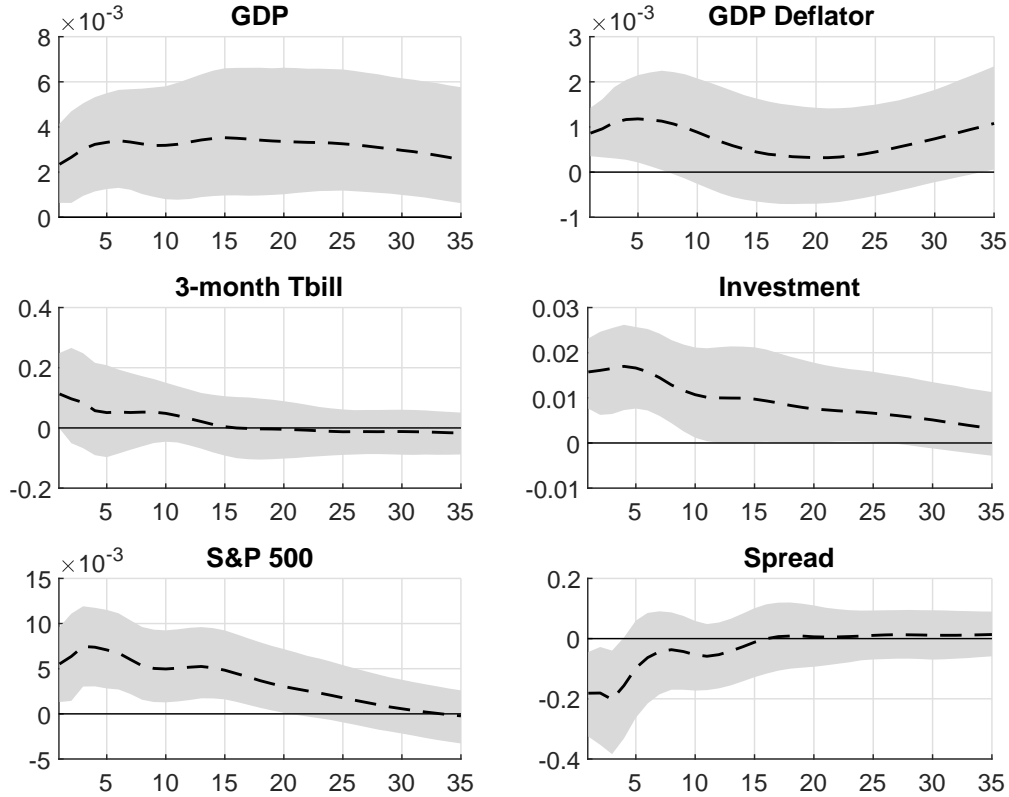


Figure 5: Impulse responses from a 6-variable VAR with the indpendent normal and inverse-Wishart priors to an one-standard-deviation financial shock. The shaded region represents the 16-th and 84-th percentiles.

Figure 5 reports the impulse responses to an one-standard-deviation financial shock. These results are almost identical to those obtained under the asymmetric conjugate prior with the same hyperparameters $\widetilde{\kappa}_1 = \widetilde{\kappa}_2 = 1$ reported in Figure 2 in the main text.

# References

ANDO, T., AND A. ZELLNER (2010): "Hierarchical Bayesian analysis of the seemingly unrelated regression and simultaneous equations models using a combination of direct Monte Carlo and importance sampling techniques," *Bayesian Analysis*, 5(1), 65–95.

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

BAUMEISTER, C., AND J. D. HAMILTON (2015): "Sign restrictions, structural vector autoregressions, and useful prior information," *Econometrica*, 83(5), 1963–1999.

BOGNANNI, M. (2021): "Comment on 'Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors'," *Journal of Econometrics*.

CANOVA, F., AND M. PAUSTIAN (2011): "Business cycle measurement with some theory," *Journal of Monetary Economics*, 58(4), 345–361.

CARRIERO, A., J. C. C. CHAN, T. E. CLARK, AND M. G. MARCELLINO (2021): "Corrigendum to: Large Bayesian vector autoregressions with stochastic volatility and non-conjugate priors," *working paper*.

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*, forthcoming.

CARRIERO, A., G. KAPETANIOS, AND M. MARCELLINO (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25(2), 400–417.

CHAN, J. C. C. (2013): "Moving Average Stochastic Volatility Models with Application to Inflation Forecast," *Journal of Econometrics*, 176(2), 162–172.

——— (2020a): "Large Bayesian VARs: A Flexible Kronecker Error Covariance Structure," *Journal of Business and Economic Statistics*, 38(1), 68–79.

——— (2020b): "Large Bayesian Vector Autoregressions," in *Macroeconomic Forecasting in the Era of Big Data*, ed. by P. Fuleky, pp. 95–125. Springer.

CHAN, J. C. C., AND E. EISENSTAT (2018): "Comparing Hybrid Time-Varying Parameter VARs," *Economics Letters*, 171, 1–5.

CHAN, J. C. C., E. EISENSTAT, AND R. W. STRACHAN (2020): "Reducing the State Space Dimension in a Large TVP-VAR," *Journal of Econometrics*, 218(1), 105–118.

CHAN, J. C. C., L. JACOBI, AND D. ZHU (2020): "Efficient Selection of Hyperparameters in Large Bayesian VARs Using Automatic Differentiation," *Journal of Forecasting*, 39(6), 934–943.

CHAN, J. C. C., AND I. JELIAZKOV (2009): "MCMC Estimation of Restricted Covariance Matrix," *Journal of Computational and Graphical Statistics*, 18, 457–480.

CHRISTIANO, L. J., R. MOTTO, AND M. ROSTAGNO (2014): "Risk Shocks," *American Economic Review*, 104(1), 27–65.

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.

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

D'AGOSTINO, A., L. GAMBETTI, AND D. GIANNONE (2013): "Macroeconomic forecasting and structural change," *Journal of Applied Econometrics*, 28, 82–101.

DEL NEGRO, M., AND F. SCHORFHEIDE (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–673.

DIEPPE, A., R. LEGRAND, AND B. VAN ROYE (2016): "The BEAR toolbox," *ECB Working Paper 1934*.

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

EISENSTAT, E., J. C. C. CHAN, AND R. W. STRACHAN (2016): "Stochastic Model Specification Search for Time-Varying Parameter VARs," *Econometric Reviews*, 35(8-10), 1638–1665.

FURLANETTO, F., F. RAVAZZOLO, AND S. SARFERAZ (2019): "Identification of financial factors in economic fluctuations," *The Economic Journal*, 129(617), 311–337.

GAMBETTI, L. (2021): "Shocks, Information, and Structural VARs," in *Oxford Research Encyclopedia of Economics and Finance*.

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

HANSEN, L. P., AND T. J. SARGENT (1991): "Two difficulties in interpreting vector autoregressions," in *Rational Expectations Econometrics*, ed. by L. P. Hansen, and T. J. Sargent, pp. 77–119. Westview Press.

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

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.

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

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

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

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

LEEPER, E. M., C. A. SIMS, AND T. ZHA (1996): "What does monetary policy do?," *Brookings papers on economic activity*, 1996(2), 1–78.

LIPPI, M., AND L. REICHLIN (1993): "The dynamic effects of aggregate demand and supply disturbances: Comment," *The American Economic Review*, 83(3), 644–652.

——— (1994): "VAR analysis, nonfundamental representations, Blaschke matrices," *Journal of Econometrics*, 63(1), 307–325.

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

LORIA, F., C. MATTHES, AND M.-C. WANG (2020): "Economic Theories and Macroeconomic Reality," *Working Paper*.

MUIRHEAD, R. J. (1982): *Aspects of Multivariate Statistical Theory*. John Wiley & Sons, Inc.

RUBIO-RAMIREZ, J. F., D. F. WAGGONER, AND T. ZHA (2010): "Structural vector autoregressions: Theory of identification and algorithms for inference," *The Review of Economic Studies*, 77(2), 665–696.

SCHORFHEIDE, F., AND D. SONG (2015): "Real-Time Forecasting With a Mixed-Frequency VAR," *Journal of Business and Economic Statistics*, 33(3), 366–380.

SIMS, C. A., AND T. ZHA (1998): "Bayesian methods for dynamic multivariate models," *International Economic Review*, 39(4), 949–968.

ZELLNER, A. (1971): *An Introduction to Bayesian Inference in Econometrics.* Wiley, New York.

ZHA, T. (1999): "Block recursion and structural vector autoregressions," *Journal of Econometrics*, 90(2), 291–316.