KING'S BUSINESS SCHOOL

Data Analytics for Finance & Macro Research Centre

KING'S College LONDON

Bayesian Structural VAR models: a new approach for prior beliefs on impulse responses

M. Bruns, M. Piffer

Working paper No. 2020/2 | August 2020 | ISSN 2516-5933

Bayesian Structural VAR models: a new approach for prior beliefs on impulse responses*

Martin Bruns† Michele Piffer‡

August 25, 2020

Abstract

Structural VAR models are frequently identified using sign restrictions on contemporaneous impulse responses. We develop a methodology that handles a larger set of prior distributions than the one allowed for by traditional methods. We then develop an importance sampler that conveniently explores the posterior distribution. This alleviates the existing trade-off between careful prior selection and tractable posterior sampling. We use this framework to combine sign restrictions with information on the volatility of the variables. Applying our methodology to monetary shocks, we confirm that monetary shocks had a significant effect on the real economy during the Great Moderation.

JEL classification: C32, C11, E50.

Keywords: Sign restrictions, Bayesian inference, Monetary policy shocks.

*We are thankful to Dario Caldara, Marco Del Negro, Lutz Kilian, Toru Kitagawa, Helmut Lütkepohl, Geert Mesters, Haroon Mumtaz, Edoardo Palombo, Gabor Pinter, Malte Rieth and Harald Uhlig for helpful comments and suggestions. We thank the International Association for Applied Econometrics for supporting this article with a travel grant for the IAAE 2018 Annual Conference. Michele Piffer is thankful for the financial support received from the European Union's Horizon 2020 research and innovation program, Marie Sklodowska-Curie grant agreement number 744010. Martin Bruns thanks the German Academic Scholarship Foundation for financial support.

†University of East Anglia, School of Economics, Norwich, NR4 7TJ, United Kingdom.

‡King's Business School, King's College London, Bush House, 30 Aldwych, London, WC2B 4BG, United Kingdom. Corresponding author e-mail: m.b.piffer@gmail.com

1 Introduction

Structural Vector Autoregressive models (SVARs) are extensively used in applied Macroeconomics. To provide results that can be interpreted economically, SVARs require identifying restrictions. It has become popular to introduce identifying restrictions in the form of sign restrictions on selected structural parameters. This is typically implemented using a Bayesian approach with informative prior beliefs that reflect the intended signs (Uhlig, 2005, Baumeister and Hamilton, 2015, Arias et al., 2018).

Sign restrictions present the researcher with a trade-off. There exist infinitely many prior probability distributions that reflect a desired set of sign restrictions. Out of this large class of priors, the literature often limits the analysis to the independent or to the conjugate Normal-inverse-Wishart-(Haar)Uniform priors (hereafter NiWU) in order to ensure a tractable posterior distribution (Uhlig, 2005, Rubio-Ramirez et al., 2010). However, this constrains the type of prior information introduced by the researcher to the one that can be modelled by the NiWU prior. This is a potentially important limitation, given that the results are affected by the specific probability distribution used, even in a large sample. Yet, moving beyond the NiWU prior makes the posterior distribution more challenging to analyse. A trade-off emerges between the flexibility in the selection of the prior distribution used, advocated by Baumeister and Hamilton (2015), and the tractability of the posterior distribution, favoured by Rubio-Ramirez et al. (2010).

The first contribution of the paper consists in developing a methodology that reduces the strength of the above trade-off. Following Arias et al. (2018), we build our approach on a tractable importance sampler that uses the posterior distribution of the NiWU case as an importance distribution. We then show that the sampler can handle a wider class of prior beliefs relative to what studied in Arias et al. (2018) if one builds the sampler in two separate stages: first on the reduced form parameters, and then on the mapping into structural parameters. After acknowledging this point,

one can follow Baumeister and Hamilton (2015) and specify prior beliefs on structural parameters, and use the NiWU approach by Rubio-Ramirez et al. (2010) as the point of departure to explore the posterior associated with the more general prior. We show that the results of the applications in this paper are the same when exploring the posterior distribution using the computationally more demanding sequential approach by Waggoner et al. (2016), further confirming the effectiveness of our sampler.

The second contribution of the paper consists in using the new methodology to propose a new approach for sign restrictions on impulse responses, which are arguably the most important statistic of SVAR models. We parametrize the structural VAR model as in Uhlig (2005), hence in the reduced form autoregressive elements and in the *contemporaneous* impulse responses. We then depart from Uhlig (2005) and specify the prior directly on the contemporaneous impulse responses, rather than using the prior implied by the NiWU approach. We allow for sign restrictions on both contemporaneous and future impulse responses, and for contemporaneous ones, we provide flexibility on the prior distribution used. In offering prior flexibility on the impulse response horizon where flexibility is needed the most (Canova and Pina, 2005 and Canova and Paustian, 2011), our approach offers a balance between prior flexibility on the key structural parameters and conditionally conjugate priors on all the remaining parameters.

Having developed a framework that handles a wide class of priors, we are in a position to confirm that indeed the results in applied work can be quite sensitive to the prior distribution used to model a given set of sign restrictions. When mapping reduced form parameters into structural parameters, the NiWU approach uses orthogonal matrices drawn from the uniform (or Haar) distribution. Building on the analysis by Baumeister and Hamilton (2015), we confirm that this approach can unintentionally imply ill-shaped distributions over the structural parameters. We propose a prior specification that instead features bell-shaped distributions over the structural parameters, and ensures that the prior mass associated with one-standard-deviation shocks

is in line with the scaling of the variables, in a way modelled through a training sample and a set of hyperparameters. With our prior, the implicit mapping from reduced form to structural parameters takes into account the volatility of the variables. Alternative prior specifications are equally compatible with our posterior sampler.

We first show an illustration on the bivariate model by Baumeister and Hamilton (2015) in order to highlight the key intuitions of our approach. We then apply our model to the study of monetary policy shocks. We follow the model by Caldara and Herbst (2019) and use a five-variable VAR model. Caldara and Herbst (2019) identify monetary policy shocks using an external instrument, and show that monetary shocks had an effect on the real economy even during the Great Moderation, despite what is argued, for example, by Boivin et al. (2010). We replicate their analysis using sign restrictions rather than an instrument. We show that the effects on the real economy are quantitatively twice as strong as suggested by Caldara and Herbst (2019), further strengthening their point. Modelling the same sign restrictions through the NiWU confirms that the effects on the real economy are stronger than in Caldara and Herbst (2019), but still delivers quantitatively lower estimates compared to our approach.

From the methodological point of view, we complement the work by Sims and Zha (1998) and Baumeister and Hamilton (2015) and study the case of beliefs on contemporaneous impulse responses, rather than on the contemporaneous relation among variables. Baumeister and Hamilton (2018) combine prior beliefs on contemporaneous relations and contemporaneous impulse responses. Relative to Baumeister and Hamilton (2018), we focus on impulse responses and propose a different prior specification and posterior sampler. Last, we relate to Giacomini and Kitagawa (2015) and Giacomini et al. (2019) in stressing the mapping from reduced form to structural parameters, but we concentrate on a single prior.

The paper is organized as follows. Section 2 outlines the methodology proposed. Section 3 shows an illustrative example on simulated data based on the estimated bivariate VAR model by Baumeister and Hamilton (2015). Section 4 reports the

application to monetary policy shocks in the US. Section 5 concludes.

2 The methodology

In this section we present the structural VAR model and summarize the traditional NiWU approach to sign restrictions. We then outline our methodology and discuss the new importance sampler. Last, we propose one possible prior distribution that can be used with our approach.

2.1 The model

Following Uhlig (2005), we write the structural VAR model as

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\pi}_0 + \sum_{l=1}^{p} \Pi_l \boldsymbol{y}_{t-l} + B\boldsymbol{\epsilon}_t, \\ &= \Pi \boldsymbol{w}_t + B\boldsymbol{\epsilon}_t, \qquad\qquad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, I_k), \end{aligned} \tag{1}$$

where $\boldsymbol{y}_t$ is a $k \times 1$ vector of endogenous variables, $\boldsymbol{\epsilon}_t$ is a $k \times 1$ vector of structural shocks, and $\boldsymbol{w}_t = (1, \boldsymbol{y}'_{t-1}, .., \boldsymbol{y}'_{t-p})'$ is an $m \times 1$ vector of the constant and p lags of the variables, with $m = kp + 1$. The matrix $\Pi = [\boldsymbol{\pi}_0, \Pi_1, .., \Pi_p]$ is of dimension $k \times m$. We normalize the covariance matrix of $\boldsymbol{\epsilon}_t$ to the identity matrix.[1]

Matrix B in equation (1) captures the contemporaneous effects of one-standard-deviation shocks, while future horizons of the impulse responses are calculated using model (1) recursively. Although structural VARs can also be specified in matrix $A = B^{-1}$ rather than in B (see, for example, Sims and Zha, 1998), we use model (1) as in Uhlig (2005) in order to emphasize the key objects of interest for our analysis, which are the contemporaneous impulse responses.[2]

[1] This normalization is frequently used in applications that employ sign restrictions on impulse responses, see for example Canova and De Nicoló (2002), Uhlig (2005), and Benati and Surico (2009).

[2] Whether the model is more conveniently expressed in $A = B^{-1}$ or B (or even in a combined form) depends on whether the identifying restrictions introduced by the researcher are more naturally

The reduced form representation of the structural model is

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\pi}_0 + \sum_{l=1}^{p} \Pi_l \boldsymbol{y}_{t-l} + \boldsymbol{u}_t, \\ &= \Pi \boldsymbol{w}_t + \boldsymbol{u}_t, \qquad\qquad \boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma), \end{aligned} \tag{2}$$

where it holds that $\boldsymbol{u}_t = B\boldsymbol{\epsilon}_t$ and $\Sigma = BB'$. Orthogonal matrices Q, which by construction satisfy $QQ' = I_k$, allow for the mapping from reduced form to structural parameters, with

$$B = h(\Sigma)Q, \tag{3}$$

and $h(\Sigma)$ a factorization of Σ satisfying $h(\Sigma)h(\Sigma)' = \Sigma$, for example the Cholesky factorization. Common estimators for (Π, Σ) are $\hat{\Pi}_T = YW'(WW')^{-1}$ and $\hat{\Sigma}_T = \frac{(Y-\hat{\Pi}_T W)(Y-\hat{\Pi}_T W)'}{T-m}$, with $Y = [\boldsymbol{y}_1, .., \boldsymbol{y}_T]$, $W = [\boldsymbol{w}_1, .., \boldsymbol{w}_T]$, $\boldsymbol{w}_t = (1, \boldsymbol{y}_{t-1}, .., \boldsymbol{y}_{t-p})'$.

2.2 The NiWU approach used in the literature

The most popular approach for sign restricted SVAR models expresses prior beliefs on the parameters $(\boldsymbol{\pi}, \Sigma, Q)$, with $\boldsymbol{\pi} = \text{vec}(\Pi)$ the $km \times 1$ vector that stacks the columns of Π. As already discussed in the literature, when $p(\boldsymbol{\pi}, \Sigma)$ falls within either the independent or the conjugate Normal-inverse-Wishart prior, drawing from the joint posterior distribution $p(\boldsymbol{\pi}, \Sigma|Y)$ is technically convenient (see, for example, Koop and Korobilis, 2010). One can then draw Q matrices from the (Haar)uniform distribution, assess if the implied values for the structural parameters satisfy the sign restrictions, and accept or reject the draw accordingly.

expressed on the contemporaneous relation among variables or on the contemporaneous effects of the shocks, respectively. Restrictions imposed on one form might not be apparent in the other form, due to the nonlinearities in the mapping from one to another. Going through the publications of all top-five journals and the Journal of Monetary Economics between 1998 and 2017, we found that around 13% of the total number of issues checked included at least one application of Structural Vector Autoregressive models. Of the total number of SVAR applications that we found, approximately 15% specifies the model in the A form, 76% specifies the model in the B form, and 9% specifies the model in the hybrid AB form. The detailed list is available at this link.

The convenience of the NiWU approach lies in the existence of efficient algorithms for the sampling of the posterior distribution. In addition, the possibility of discarding undesired draws allows for the straightforward introduction of sign restrictions not only on contemporaneous impulse responses, but also on future horizons, as well as on other structural parameters. The inconvenience lies in the fact that the distribution $p(Q|\Sigma)$ is never updated by the data (see Section C.1.1 of the Online Appendix). This holds true irrespectively on whether the prior is explicitly introduced on $p(Q|\Sigma)$ (as in the NiWU approach), or if $p(Q|\Sigma)$ is determined indirectly by a prior belief on some structural parameter of interest, say B or B^{-1}. In structural VARs, the researcher imposes on the posterior of structural parameters the curvature implied by $p(Q|\Sigma)$, evaluated at Σ approaching $\Sigma_0 = \mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_t')$ at a speed that depends on the sample size. The NiWU approach makes it immediate to appreciate the shape of $p(Q|\Sigma)$, which is flat in Q and can vary only in Σ. However, since $p(Q|\Sigma)$ is flat in Q, the researcher has no control on the implications of $p(Q|\Sigma)$ on the structural parameters of interest given Σ.

To appreciate the importance of the above point, consider for simplicity the case of sign restrictions on the contemporaneous impulse responses. Define SR_B the set of sign restrictions for B, and $p(B)$ the probability distribution used in the analysis. One can introduce SR_B to the analysis by ensuring that $p(B)$ attaches zero mass to the values of B that fail to satisfy SR_B. However, there are infinitely many probability distributions $\{p(B)_1, p(B)_2, p(B)_2, ...\}$ that ensure this property. Since B is not identified, the posterior distributions $\{p(B|Y)_1, p(B|Y)_2, p(B|Y)_3, ...\}$ differ even in a large sample in a way that asymptotically depends only on the distributions $p(Q|\Sigma)$ implied by $p(B)$, i.e. $\{p(Q|\Sigma)_1, p(Q|\Sigma)_2, p(Q|\Sigma)_3, ...\}$. The NiWU approach can flexibly introduce sign restrictions on a wide class of parameters, at the cost of working with an inflexible $p(Q|\Sigma)$.

2.3 The Np(B) approach proposed in this paper

To overcome the limitation summarized in the previous section, we follow the suggestion by Baumeister and Hamilton (2015) and express prior beliefs directly on the structural parameter of interest, which in our case is B. We now discuss this approach and develop an importance sampler for it.

2.3.1 Prior beliefs expressed directly on $(\boldsymbol{\pi}, B)$

We parametrize the model as in equation (1) and express prior beliefs on $(\boldsymbol{\pi}, B)$ as

$$p(\boldsymbol{\pi}, B) \propto \tilde{p}(\boldsymbol{\pi}|B) \cdot \tilde{p}(B) \cdot \mathrm{I}\{\boldsymbol{\pi}, B\}. \tag{4}$$

Since $\boldsymbol{\pi}$ is identified, $\tilde{p}(\boldsymbol{\pi}|B)$ matters less compared to $\tilde{p}(B)$. Hence, we follow the NiWU approach and set

$$\tilde{p}(\boldsymbol{\pi}|B) = \phi(\boldsymbol{\mu}_\pi, V_\pi), \tag{5}$$

where $\phi(\cdot, \cdot)$ represents the Normal probability distribution, and $\boldsymbol{\mu}_\pi$ and V_π can be a function of B. By contrast, we remain general for $\tilde{p}(B)$ and grant the researcher flexibility on the prior beliefs used on B (and, in turn, on the implied distribution $p(Q|\Sigma)$). Additional restrictions can be introduced via the indicator function $\mathrm{I}\{\boldsymbol{\pi}, B\}$, for example sign restrictions on the impulse responses on future horizons, or stationarity of the model.[3]

The above approach strikes a balance between flexibility and tractability. On the one hand, it provides flexibility on the prior beliefs $\tilde{p}(B)$ used for sign restrictions on arguably the most important impulse response horizon, namely the contemporaneous response. On the other hand, as also the NiWU approach, it makes the analysis more tractable by using a normal prior distribution on the reduced form parameters

[3] As in Baumeister and Hamilton (2015) and Baumeister and Hamilton (2019), we require that $\tilde{p}(B)$ is everywhere non-negative, and when integrated over the set of all values of B, it produces a finite positive number.

$\boldsymbol{\pi}$, allowing for a flexible accept/reject algorithm to introduce additional restrictions of interest. Conditioning on B, all structural parameters are identified, making flexibility on the remaining structural parameters less important. The normality on $\boldsymbol{\pi}$ is not restrictive except in small samples, given that $\boldsymbol{\pi}$ is identified. In addition, by parametrizing the model in $\boldsymbol{\pi}$, our approach makes it straightforward to use the prior by Litterman (1986), which is applied directly on $\boldsymbol{\pi}$.

The prior from equations (4)-(5) nests two special cases that are particularly common in the literature, and which we will focus on in the analysis:

$$\textit{Case 1}: \qquad \tilde{p}(\boldsymbol{\pi}, B) = \tilde{p}(\boldsymbol{\pi}) \cdot \tilde{p}(B), \qquad \tilde{p}(\boldsymbol{\pi}) \propto 1, \tag{6}$$

$$\textit{Case 2}: \qquad \tilde{p}(\boldsymbol{\pi}, B) = \tilde{p}(\boldsymbol{\pi}) \cdot \tilde{p}(B), \qquad \tilde{p}(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_\pi, V_\pi). \tag{7}$$

Case 1 uses a flat prior on $\boldsymbol{\pi}$. It is also suitable to implement the Minnesota prior through dummy observations, as in the application in Section 4. *Case 2* keeps the proper Normal prior on $\boldsymbol{\pi}$, but introduces prior independence. It allows for the introduction of a more flexible specification of the Minnesota prior. As we show in Section C of the Online Appendix, the joint posterior distribution associated with the general prior (4)-(5) satisfies

$$\begin{aligned} p(\boldsymbol{\pi}, B|Y) &\propto \tilde{p}(\boldsymbol{\pi}, B|Y) \cdot \mathrm{I}\{\boldsymbol{\pi}, B\}, \\ &\propto \tilde{p}(\boldsymbol{\pi}|B, Y) \cdot \tilde{p}(B|Y) \cdot \mathrm{I}\{\boldsymbol{\pi}, B\}, \end{aligned} \tag{8}$$

$$\tilde{p}(\boldsymbol{\pi}|B, Y) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \tag{9}$$

$$\begin{aligned} \tilde{p}(B|Y) \propto \tilde{p}(B) \cdot |det(B)|^{-T} \cdot |det(V_\pi)|^{-\frac{1}{2}} \cdot |det(V_\pi^*)|^{\frac{1}{2}} \cdot \\ \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T \otimes (BB')^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1} \boldsymbol{\mu}_\pi\right\}}, \end{aligned} \tag{10}$$

with $\tilde{\boldsymbol{y}}$, W, $\boldsymbol{\mu}_\pi^*$ and V_π^* defined in Section C.1.1 of the Online Appendix. Equation

(10) implies

$$\tilde{p}(\Sigma|Y) \propto v_{\{B\to\Sigma,Q\}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot \tag{11}$$
$$\cdot\, e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}-\boldsymbol{\mu}_\pi^{*\prime}V_\pi^{*^{-1}}\boldsymbol{\mu}_\pi^*+\boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\right\}} \cdot \int_{O(k)} \tilde{p}\big(h(\Sigma)Q\big)dQ,$$

$$\propto |\det(\Sigma)|^{-\frac{T+1}{2}} \cdot |\det(V_\pi)|^{-\frac{1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot \tag{12}$$
$$\cdot\, e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}-\boldsymbol{\mu}_\pi^{*\prime}V_\pi^{*^{-1}}\boldsymbol{\mu}_\pi^*+\boldsymbol{\mu}_\pi'V_\pi^{-1}\boldsymbol{\mu}_\pi\right\}} \cdot \int_{O(k)} \tilde{p}\big(h(\Sigma)Q\big)dQ,$$

with $v_{\{B\to\Sigma,Q\}} = |\det(\Sigma)|^{-\frac{1}{2}}$ the volume element when mapping B to (Σ, Q), and $O(k)$ the space of orthogonal matrices of dimensions $k\times k$. The results further simplify when considering *Cases 1* and *2* from equations (6)-(7), see Section C.1.2-Section C.1.3 of the Online Appendix.

Equations (8)-(10) suggest that as long as we can obtain draws from $\tilde{p}(B|Y)$, generating draws from $p(\boldsymbol{\pi}, B|Y)$ is straightforward. However, to become practical in applied work, our approach requires a convenient algorithm to draw from $\tilde{p}(B|Y)$. We turn to this issue in the next section.

2.3.2 A new posterior sampler for $p(B|Y)$

We use two approaches to explore $\tilde{p}(B|Y)$. We first use our importance sampler, which we discuss in this section. We then use the Dynamic Striated Metropolis-Hastings (DSMH) algorithm by Waggoner et al. (2016), which is computationally more demanding, but it can handle irregularly shaped posterior distributions and a large number of parameters. We use the draws from the DSMH algorithm to approximate the true posterior distribution of interest used as a benchmark to assess the performance of our sampler. Section D of the Appendix discusses how we implement the algorithm by Waggoner et al. (2016).

We build our sampling procedure on importance sampling techniques. Consider a vector of parameters of interest, $\boldsymbol{\theta}$. Suppose we are interested in sampling from the

target distribution $p(\boldsymbol{\theta})$, and we cannot draw from $p(\boldsymbol{\theta})$ directly, but can evaluate it. Suppose that we can extract proposal draws from the importance distribution $q(\boldsymbol{\theta})$. To the extent that the importance distribution fully covers the support of $p(\boldsymbol{\theta})$, we can obtain draws from $q(\boldsymbol{\theta})$ by resampling with replacement the draws $\{\boldsymbol{\theta}_i\}$ obtained from the importance distribution using weights $w(\boldsymbol{\theta}_i) = \frac{p(\boldsymbol{\theta}=\boldsymbol{\theta}_i)}{q(\boldsymbol{\theta}=\boldsymbol{\theta}_i)}$ (see for example Koop, 2003, chapter 4).

Define $\tilde{p}(B|Y)_{Np(B)}$ as the posterior distribution associated with the general prior $\tilde{p}(B)$ from our Np(B) approach (equation 10), and $p(B|Y)_{NiWU}$ as the posterior distribution implied on B by the NiWU approach. Since sampling from $p(B|Y)_{NiWU}$ is not challenging, one could set $\boldsymbol{\theta} = B$, $p(\boldsymbol{\theta}) = \tilde{p}(B|Y)_{Np(B)}$ and $q(\boldsymbol{\theta}) = p(B|Y)_{NiWU}$. Arias et al. (2018) show that this approach works successfully if the target distribution $\tilde{p}(B|Y)_{Np(B)}$ does not differ too much from $p(B|Y)_{NiWU}$. However, this procedure does not work in a general framework, because one cannot ensure that $p(B|Y)_{NiWU}$ sufficiently covers the support of $\tilde{p}(B|Y)_{Np(B)}$.

We circumvent the above challenge by exploring $\tilde{p}(B|Y)_{Np(B)}$ indirectly. First, define the following functions:

- $\tilde{p}(\Sigma|Y)_{Np(B)}$: posterior distribution on Σ implied by $\tilde{p}(B|Y)_{Np(B)}$, equation (12);
- $\tilde{p}(Q|Y,\Sigma)_{Np(B)}$: conditional distribution on Q implied by $\tilde{p}(B|Y)_{Np(B)}$;
- $p(\Sigma|Y)_{NiWU}$: posterior distribution on Σ corresponding to the NiWU approach;
- $p(Q)_U$: (Haar)uniform distribution on Q employed in the NiWU approach;

Then, notice that drawing from $\tilde{p}(B|Y)_{Np(B)}$ is equivalent to drawing from $\tilde{p}(\Sigma|Y)_{Np(B)}$ and mapping Σ into B using draws of Q from $\tilde{p}(Q|Y,\Sigma)_{Np(B)}$. Accordingly, consider the following importance sampling procedure. First, explore $\tilde{p}(\Sigma|Y)_{Np(B)}$ using $p(\Sigma|Y)_{NiWU}$ as an importance distribution, with weights

$$w^{\text{stage A}} = \frac{v_{B\to\Sigma,Q}\cdot\tilde{p}(B|Y)_{Np(B)}}{q(\Sigma|Y)} = \frac{\tilde{p}(\Sigma|Y)_{Np(B)}}{p(\Sigma|Y)_{NiWU}}. \tag{13}$$

Since Σ is identified, $p(\Sigma|Y)_{NiWU}$ and $\tilde{p}(\Sigma|Y)_{Np(B)}$ are close to each other except in small samples, making $p(\Sigma|Y)_{NiWU}$ a candidate importance function for $\tilde{p}(\Sigma|Y)_{Np(B)}$. Then, use $p(Q)_U$ as a proposal function for $\tilde{p}(Q|Y,\Sigma)_{Np(B)}$ to turn draws from $\tilde{p}(\Sigma|Y)_{Np(B)}$ into draws from $\tilde{p}(B|Y)_{Np(B)}$. Since $p(Q)_U$ is uniform in $O(k)$, it ensures zero probability that the parameter space covered by $p(Q|Y,\Sigma)_{Np(B)}$ is not explored.

Case 1 from equation (6) is particularly convenient to further build intuition behind our sampler. As shown in Section C.1.2 of the Online Appendix, under *Case 1* the mode of the target distribution $\tilde{p}(\Sigma|Y)_{Np(B)}$ is implicitly defined by the equality

$$\Sigma = \frac{T-m}{T-m+1}\hat{\Sigma} - \frac{2}{T-m+1}\frac{d}{d\Sigma^{-1}} log\Big(\int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ\Big). \tag{14}$$

By contrast, if proposal draws are generated from a Conjugate Normal inverse Wishart prior that is flat in $\boldsymbol{\pi}$, the mode of the proposal distribution equals $\frac{1}{d+T+k+1-m}S + \frac{T-m}{d+T+k+1-m}\hat{\Sigma}_T$, see Section B.1 of the Online Appendix. As the sample size increases, both modes approach $\hat{\Sigma}_T$, which approaches the population moment $\Sigma_0 = \mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_t')$. This result is not new and it plays a central role in our algorithm work: since Σ is identified, as the sample size increases differences in prior beliefs on Σ become irrelevant in the posterior, provided the priors are strictly positive in the neighbourhood of Σ_0. Draws from $p(\Sigma|Y)_{NiWU}$ are easy to generate and offer the starting point for the importance sampler.

The crucial step of the algorithm is to select the hyperparameters in the prior proposal distribution $q(\Sigma)$ to make the weights from equation (13) as balanced as possible, while still ensuring that the proposal draws are easy to generate from the posterior $q(\Sigma|Y)$. Building on this intuition, the appendix of the paper provides the derivations for the following algorithm:

Algorithm: Sign restrictions

Stage A: generate draws from $\tilde{p}(\Sigma|Y)_{Np(B)}$:

1. Select the proposal prior beliefs $q(\boldsymbol{\pi}, \Sigma) = q(\boldsymbol{\pi}) \cdot q(\Sigma)$ as

$$\begin{aligned} &\textit{Case 1}: && q(\boldsymbol{\pi}) \propto 1, && q(\Sigma) = |\det(\Sigma)|^{-\frac{1}{2}}, && \text{or} \\ &\textit{Case 2}: && q(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_\pi, V_\pi), && q(\Sigma) = |\det(\Sigma)|^{-\frac{1}{2}}; \end{aligned}$$

2. Generate m_1 proposal draws $\{\Sigma^{(d)}\}_{d=1}^{m_1}$ through either direct sampling or Gibbs sampling, using

$$\begin{aligned} &\textit{Case 1}: && \Sigma|Y \sim \text{iW}(S^*, d^*), && \text{or} \\ &\textit{Case 2}: && \Sigma|Y, \Pi \sim \text{iW}(S^*, d^*), && \boldsymbol{\pi}|Y, \Sigma = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*), \end{aligned}$$

through *Algorithms A* or *C* discussed in Section B of the Online Appendix, with $(S^*, d^*, V_\pi^*, \boldsymbol{\mu}_\pi^*)$ specified in Table 2 in the appendix of the paper.

3. for each $\Sigma^{(d)}$,

   3a. extract one matrix Q_c using the algorithm by Rubio-Ramirez et al. (2010);

   3b. compute the candidate matrix $B_c = h(\Sigma^{(d)})Q_c$. If B_c satisfies the sign restrictions on B, store $\left(Q_i(\Sigma^{(d)}), B_i(\Sigma^{(d)}), w_i(\Sigma^{(d)})^{\text{stage B}}\right) = \left(Q_c, B_c, w_{dc}^{\text{stage B}}\right)$, with $w_{dc}^{\text{stage B}}$ equal to $\tilde{p}(B)$ evaluated at B_c, otherwise move back to Step 3a;

   3c. repeat Steps 3a-3b until m_2 draws $\left\{Q_i(\Sigma^{(d)}), B_i(\Sigma^{(d)}), w_i(\Sigma^{(d)})^{\text{stage B}}\right\}_{i=1}^{m_2}$ are stored. Store the number of attempts $m_3(\Sigma^{(d)})$ required to generate m_2 successful draws. Check that the effective sample size $\text{ESS}_d^{\text{B}} = \left(\sum_i \left(w_i(\Sigma^{(d)})^{\text{stage B}}\right)^2\right)^{-1}$ is sufficiently high, otherwise increase m_2;

3d. compute

$$w(\Sigma^{(d)})^{\text{stage A}} = \frac{\tilde{p}(\Sigma|Y)_{Np(B)}}{p(\Sigma|Y)_{NiWU}} \approx \frac{\sum_{i=1}^{m_2} \tilde{p}\big(B_i\big)}{m_3(\Sigma^{(d)})}, \qquad (15)$$

and assess if the relative effective sample size

$$\text{rel ESS}^{\text{A}} = \Big(\sum_d \big(w(\Sigma^{(d)})^{\text{stage A}} / \sum_d \big(w(\Sigma^{(d)})^{\text{stage A}}\big)\big)^2\Big)^{-1}, \quad (16)$$

is sufficiently high;

4. generate a new set $\big\{\Sigma^{(d)}\big\}_{d=1}^{m_4}$ by drawing from $\big\{\Sigma^{(d)}\big\}_{d=1}^{m_1}$ with replacements using weights $\big\{w(\Sigma^{(d)})^{\text{stage A}}\big\}_{d=1}^{m_1}$;

Stage B: move from $\tilde{p}(\Sigma|Y)_{Np(B)}$ *to* $p(\boldsymbol{\pi}, B|Y)_{Np(B)}$:

5. randomly select Σ_c from $\big\{\Sigma^{(d)}\big\}_{d=1}^{m_4}$ generated in Step 4;
6. randomly select one matrix B_c out of the set $\big\{B_i(\Sigma_c)\big\}_{i=1}^{m_2}$ stored from Step 3 using weights $\{w_i(\Sigma_c)^{\text{stage B}}\}_{i=1}^{m_2}$;
7. compute $(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ associated with B_c and generate one draw $\boldsymbol{\pi}_c$ from $\tilde{p}(\boldsymbol{\pi}|Y, B) = \phi(\boldsymbol{\mu}_\pi^*, V_\pi^*)$;
8. if $\text{I}\{\boldsymbol{\pi}_c, B_c\} = 1$, store $(\boldsymbol{\pi}_c, B_c)$, otherwise move back to Step 5;
9. repeat until m_5 draws are successfully generated.

Our algorithm effectively resamples the posterior draws from the NiWU approach and makes them representative of the posterior distribution associated with the prior beliefs $\tilde{p}(B)$ from our approach. We design the algorithm in two stages to stress that Stage B should be carried out only if the relative effective sample size in Stage A is satisfactory. In the rest of the paper we document that the sampling time of our algorithm is roughly 4 minutes in the bivariate simulation exercise and 40 minutes in the monetary application. The relative effective sample size is very high and equals

more than 0.95 in the simulation exercise and around 0.80 in the monetary application. Figure F1 in the Online Appendix report the statistics that document the performance of our sampler in the applications of the paper, and Section C.2 discusses a number of additional diagnostic tests.

2.4 Proposing one possible prior $\tilde{p}(B)$

The paper has so far developed an approach that is general in the prior distribution $\tilde{p}(B)$. We conclude the section on the methodology by discussing one possible prior specification for $\tilde{p}(B)$. Other prior beliefs are also possible, and must ultimately be chosen by the applied researcher.

Specifying prior beliefs $\tilde{p}(B)$ is challenging. The literature still provides limited guidance on explicit prior beliefs on structural parameters. Baumeister and Hamilton (2015) introduce restrictions on B^{-1} rather than on B, and use the existing literature to form prior beliefs on the contemporaneous elasticities among variables. However, as discussed by Kilian and Lütkepohl (2017) and Uhlig (2017), researchers' beliefs do not frequently go beyond the sign of contemporaneous impulse responses. Such beliefs, in turn, require taking a stand on expected magnitudes of the responses, on which it is indeed challenging to form beliefs. As an example, one may entertain the belief that an exogenous, one-standard-deviation monetary increase in the interest rate decreases inflation, but lacks prior beliefs on the scale of such a decrease.

To overcome this challenge, we propose a prior specification that builds on the Minnesota prior. With the Minnesota prior, one first approximates a reasonable scale s_i of each variable variable y_i (see, for example, the discussion in Canova, 2007 and Kilian and Lütkepohl, 2017). This is frequently done using a training sample, setting s_i either equal to the standard deviation of the residual of univariate AR(1) processes estimated on each variable, or equal to the standard deviation of the variable. Once $\{s_i\}_{i=1}^{k}$ is set, Bayesian shrinkage is introduced through a set of hyperparameters that shrink the elements in $\boldsymbol{\pi}$ towards the random walk or the white noise process, taking

the relative scale of the variables into account. We extend this approach as follows. Call b_{ij} the entry of B capturing the effect of a one-standard-deviation shock j to variable i, and call γ_i the reasonable scale of such effect. γ_i can be set equal to the same statistic s_i from the the Minnesota prior. Alternatively, one can use the training sample to estimate Σ and then set $\gamma_i = \hat{\Sigma}^{0.5}_{ii,\text{training}}$ since it can be shown that the covariance restrictions $\Sigma = BB'$ imply

$$-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}, \tag{17}$$

with Σ_{ii} the $i-th$ element of the diagonal of Σ.[4] Given $\{\gamma_i\}_{i=1}^{k}$, we set $\tilde{p}(B) = \prod_i \prod_j \tilde{p}(b_{ij}|\gamma_i, \psi_1, \psi_2)$, with $\tilde{p}(b_{ij}|\gamma_i, \psi_1, \psi_2) = I(b_{ij}) \cdot \phi(\mu_{ij}, \sigma_{ij})$ the probability distribution of (potentially truncated) normal distributions $\phi(\mu_{ij}, \sigma_{ij})$. $I(b_{ij})$ takes value 1 if the sign restrictions on b_{ij} are satisfied. The underlying hyperparameters are set as follows:

1. if no sign restriction is imposed on b_{ij}, set $\mu_{ij} = 0$ and $\sigma_{ij} = \psi_2\gamma_i/1.96$, so that the distribution is symmetric around 0, with 95% prior mass in the space $(-\psi_2\gamma_i,\ \psi_2\gamma_i)$;

2. if b_{ij} is restricted to be positive, start from a normal distribution with $\mu_{ij} = \psi_1\gamma_i$ and calibrate the variance such that the truncated distribution has 95% prior mass in the space $(0,\ \psi_2\gamma_i)$;

3. if b_{ij} is restricted to be negative, start from a normal distribution with $\mu_{ij} = -\psi_1\gamma_i$ and calibrate the variance such that the truncated distribution has 95% prior mass in the space $(-\psi_2\gamma_i,\ 0)$.

With this prior, the researcher sets a plausible upper bound γ_i for the effect of the shock, and then introduces Bayesian shrinkage through the hyperparameters ψ_1 and

[4] Given $\Sigma = BB'$, the equations corresponding to the diagonal elements of Σ are $\Sigma_{ii} = b_{i1}^2 + b_{i2}^2 + ... + b_{ik}^2$. Since Σ_{ii} is non-negative and since $b_{ij}^2 \geq 0$, each element b_{ij} must satisfy $-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}$. See also equation (33) in Baumeister and Hamilton (2015).

ψ_2. ψ_1 controls for the first moment of the prior, ψ_2 controls for the second moment. Our prior is a generalization of the Generalized Normal prior in Arias et al. (2018), with the advantage that the implied distribution $p(Q|\Sigma)$ is not necessarily flat in Q. We refer to Section E of the Online Appendix for more details.

3 An illustrative example

In this section we outline the intuition for our approach using simulations on a bivariate VAR model. We then discuss what drives the difference between our Np(B) approach and the traditional NiWU approach.

3.1 Simulation exercise

We build our simulation exercise on the model by Baumeister and Hamilton (2015). We first employ ordinary least squares to estimate their bivariate reduced form VAR model, which uses data on the growth rates of the US real labour compensation and of total employment. The model includes a constant and 8 lags, and covers the period 1970Q1 through 2014Q4. We then use the estimated reduced form VAR as the data generating process. We generate a dataset of 680 observations, initializing the data from the estimated unconditional mean. We discard the first 100 observations to make the data less dependent on the initial point, and store the next 100 observations to use as a training sample. We then divide the remaining 480 observations into four pseudo datasets, including up to the first 60, 120, 240 and 480 observations. We use the same training sample for all datasets to improve the comparison, and to avoid an unreasonably short training sample for the dataset of smaller size.

We estimate the structural VAR model from equation (1) by introducing sign restrictions on the contemporaneous impulse responses. We identify the demand shock and the supply shock as the structural shocks that move wages and employment in the same and in the opposite direction, respectively. For each dataset, we estimate

the model multiple times, always introducing the same sign restrictions, but modelling them with different approaches. For the Np(B) approach proposed in our paper, we use *Case 1* for the main illustration, featuring a flat improper prior on $\boldsymbol{\pi}$ as discussed in Section 2.3. We specify $\tilde{p}(B)$ as from Section 2.4, setting $\psi_1 = 0.8$ and $\psi_2 = 1.5$ for the illustration, using $\gamma_i = \hat{\Sigma}^{0.5}_{ii,\text{training}}$ with $\hat{\Sigma}_{\text{training}}$ the estimate of Σ on the training sample. For the NiWU approach, we use the most popular specifications and parametrizations used in the literature, see Table 2 of the Appendix of the paper. All models include a constant term and 8 lags, in accordance with the DGP. Section D of the Online appendix discusses how we implement the Dynamic Striated Metropolis-Hastings algorithm, which we use to further assess the effectiveness of our sampler. Since we introduce no additional restrictions through the indicator function $\text{I}\{\boldsymbol{\pi}, B\}$, there is no distinction between $\tilde{p}(B|Y)$ and $p(B|Y)$. Table F1 in the Online Appendix reports the values of the tuning parameters for the algorithms used.

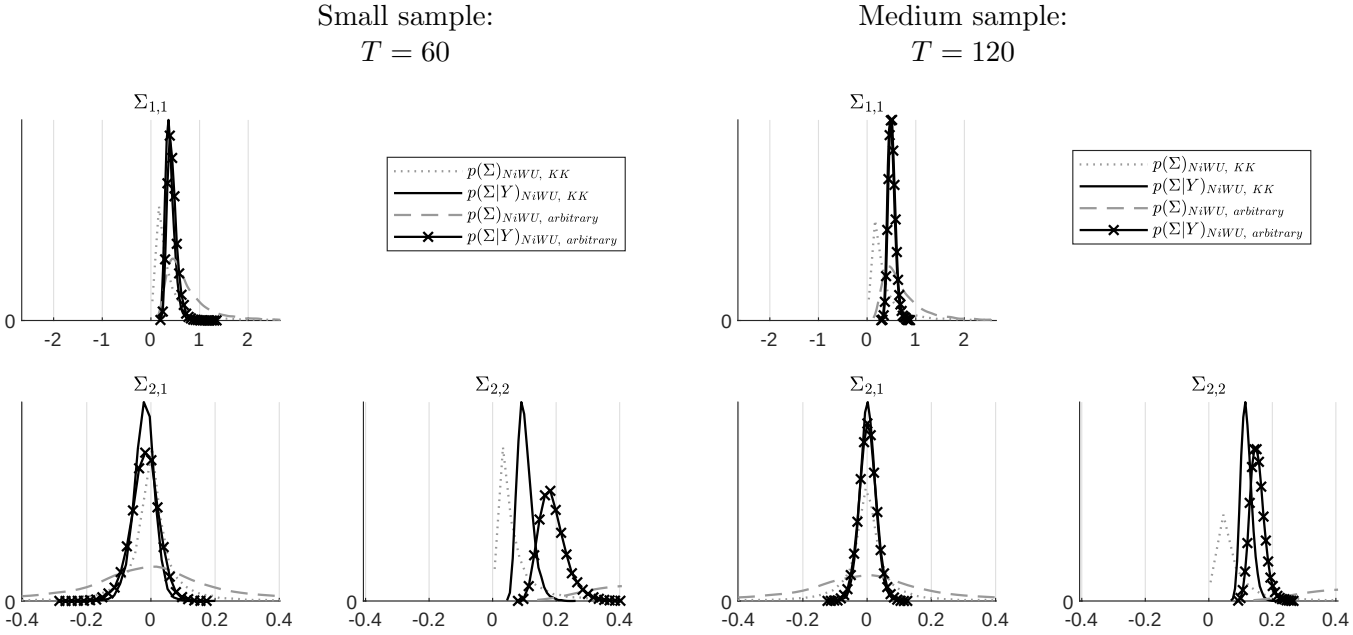
3.2 The intuition behind our importance sampler

We illustrate the intuition behind our posterior sampler in two steps. First, we show graphically that proposal distributions are, in principle, easier to find for $\tilde{p}(\Sigma|Y)_{Np(B)}$ then for $\tilde{p}(B|Y)_{Np(B)}$. Second, we argue that the hyperparameters in the prior distribution $q(\Sigma)$ behind the actual proposal distribution $q(\Sigma|Y)$ can be selected to heuristically minimize the distance between $q(\Sigma|Y)$ and the target function $\tilde{p}(\Sigma|Y)_{Np(B)}$.

Figure 1 starts from two alternative parametrizations of an inverse Wishart distribution on Σ, one in line with Kadiyala and Karlsson (1997), the other purely arbitrary for the sake of illustration.[5] These priors on Σ are then separately used in the NiWU approach with the prior $p(\boldsymbol{\pi}, \Sigma) = p(\boldsymbol{\pi}) \cdot p(\Sigma)$, $p(\boldsymbol{\pi}) \propto 1$. The figure shows the posterior distributions corresponding to the sample sizes $T = 60$ and $T = 120$, sampled

[5] The inverse Wishart distribution $p(\Sigma) \propto |\det(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}S\right]}$ requires specifying the hyperparameters (S, d). We parametrize (S, d) either as in Kadiyala and Karlsson (1997), or setting $S = 4 \cdot I_k$ and $d = 4 \cdot k$ as arbitrary values that help the illustration.

Figure 1: On prior beliefs and sample size for Σ

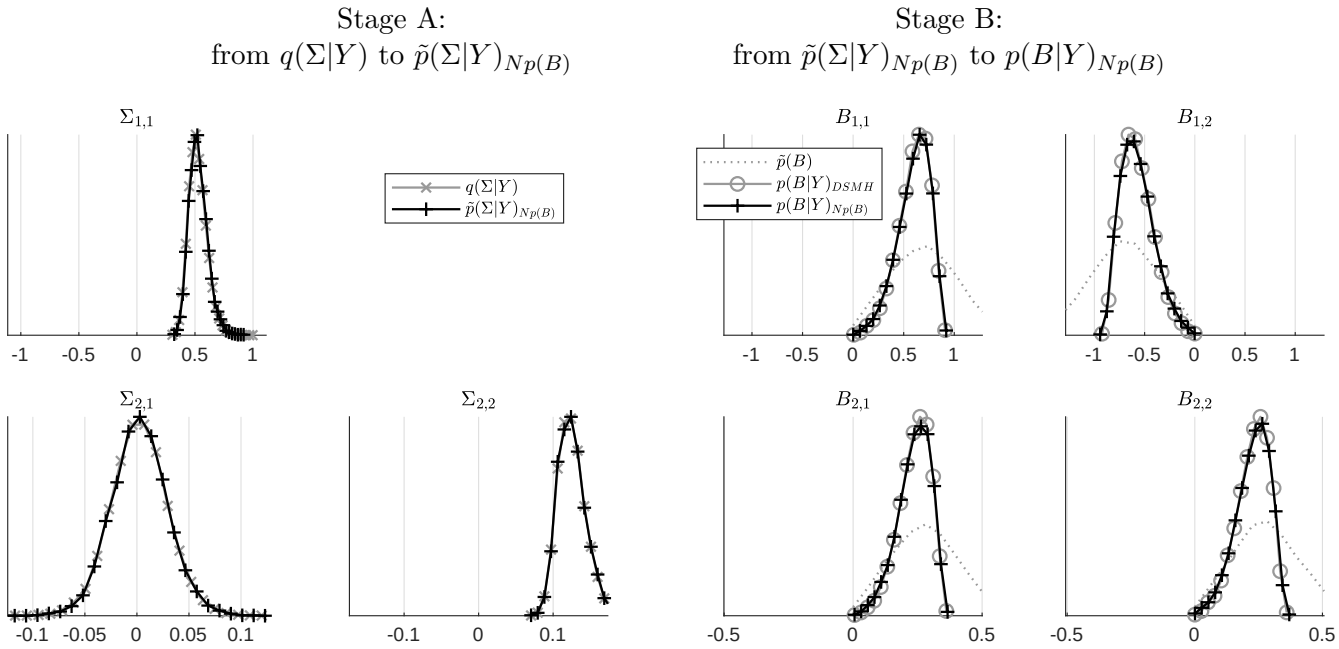


Note: The draws for this figure are generated using *Algorithm B* discussed in Section B of the Online Appendix. See Table F1 for how we set the tuning parameters.

using *Algorithm B* discussed in Section B.1 of the Online Appendix. As shown in the figure, the two parametrizations of the inverse Wishart distribution imply very different priors for Σ, one being much tighter than the other. The posteriors, instead, are very close to each other, a difference that becomes even more negligible for larger datasets. This confirms the well-known result that for identified parameter, differences in prior beliefs do not matter asymptotically, provided the prior gives non-zero mass in the area that is relevant asymptotically. This suggests that exploring the distribution $\tilde{p}(\Sigma|Y)_{Np(B)}$ associated with our general prior $\tilde{p}(B)$ can, at least in principle, take advantage of some proposal prior $q(\Sigma)$ such that the corresponding posterior $q(\Sigma|Y)$ is sufficiently close to the distribution of interest $\tilde{p}(\Sigma|Y)_{Np(B)}$, given the size of the sample at hand.

Figure 2 builds on the above intuition and reports the update from our Np(B) approach for $T = 120$. Proposal draws are obtained from the posterior distribution $q(\Sigma|Y)$ associated with the proposal prior $q(\boldsymbol{\pi}) \propto 1$, $q(\Sigma) = |\det(\Sigma)|^{-\frac{1}{2}}$ discussed in the

Figure 2: Performance of our sampler ($T = 120$)



Note: The draws for this figure are generated using our algorithm, discussed in Section 2.3.2. See Table F1 in the Online Appendix for how we set the tuning parameters.

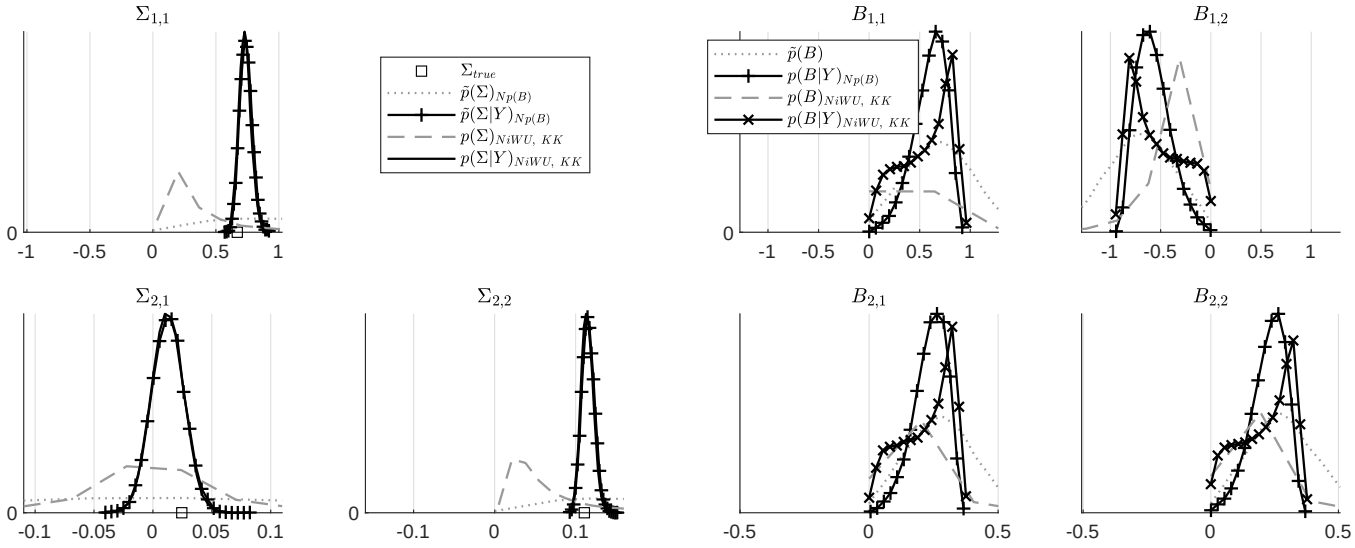
appendix of this paper. The right panel of the figure shows the posterior distribution of interest $p(B|Y)_{Np(B)} = \tilde{p}(B|Y)_{Np(B)}$, sampled either through Stage B of our algorithm or using the Dynamic Striated Metropolis-Hastings algorithm. It is encouraging to see that these distributions overlap, suggesting that our sampler is correctly sampling the posterior of interest.

Table 1: Relative effective sample size from Stage A for different proposal priors $q(\Sigma)$

Proposal prior $q(\Sigma)$	$T = 60$	$T = 120$	$T = 240$	$T = 480$
Case 1)				
1a) $\Sigma \sim \text{iW}(S_{KK}, d_{KK})$	0.4134	0.7166	0.8588	0.9269
1b) $q(\Sigma) \propto \lvert\det(\Sigma)\rvert^{-\frac{1}{2}}$ (*our approach*)	0.9595	0.9822	0.9873	0.9881
Case 2)				
2a) $\Sigma \sim \text{iW}(S_{KK}, d_{KK})$	0.5031	0.7134	0.8614	0.9302
2b) $q(\Sigma) \propto \lvert\det(\Sigma)\rvert^{-\frac{1}{2}}$ (*our approach*)	0.9682	0.9823	0.9875	0.9880

Note: The relative effective sample size is defined in equation (16). See Table 2 in the appendix of the paper for the proposal distributions $q(\Sigma|Y)$ used in the sampler, and Table F1 in the Online Appendix for how we set the tuning parameters.

Figure 3: Comparison to the NiWU approach ($T = 480$)



Note: The draws for this figure are generated using *Algorithm B* discussed in Section B of the Online Appendix, our algorithm discussed in Section 2.3.2, and their Dynamic Striated Metropolis-Hastings algorithm discussed in Section D of the Online Appendix. See Table F1 for how we set the tuning parameters.

Table 1 further elaborates on the role played by the selection of the prior distribution associated with the proposal posterior distribution $p(\Sigma|Y)$. The table reports the relative effective sample size from Stage A of our sampler for different sizes of the dataset. *Case 1a* shows the illustrative case in which the proposal prior distribution specifies the hyperparameters as in Kadiyala and Karlsson (1997). While it is confirmed that the effective sample size increases as T increases, for $T = 60$ the effective sample size is still too low, suggesting that the proposal posterior distribution is still far from the distribution of interest. By contrast, our selected proposal prior implies a relative effective sample size above 0.95 already for $T = 60$. This intuition is confirmed when our prior on $\boldsymbol{\pi}$ is replaced with *Case 2*, as shown in the second half of the table.[6] The results suggest that a careful selection of the proposal prior distribution make the algorithm work.

[6] The illustration for *Case 2* specifies the hyperparameters in the Minnesota prior for $\boldsymbol{\pi}$ as in Canova (2007), leaving the prior for B unchanged.

3.3 Comparison to the NiWU approach

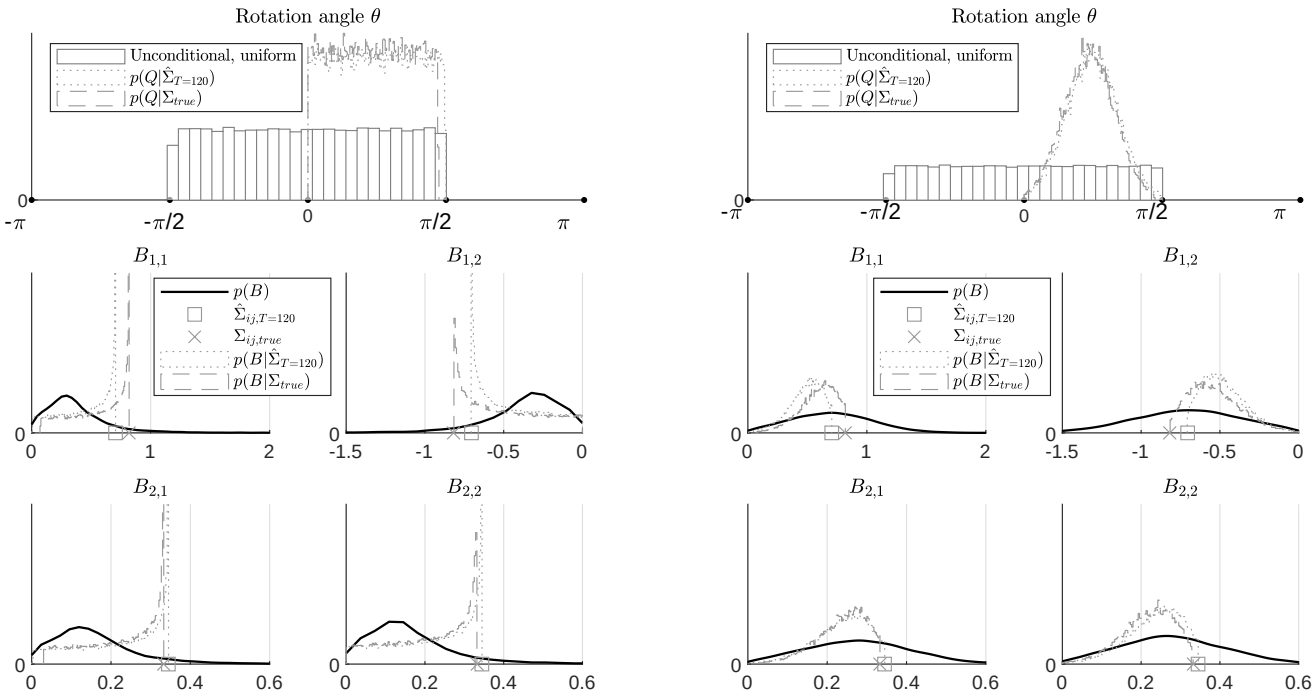
Having discussed the key intuition of the sampler, we now illustrate what drives the difference between the Np(B) and the NiWU approach. Figure 3 compares our Np(B) approach with the NiWU approach, where the latter is calibrated as in Kadiyala and Karlsson (1997).[7] The left panel shows that the prior distributions on Σ are quite different when comparing the explicit prior introduced by the NiWU approach (labelled as $p(\Sigma)_{NiWU,\ KK}$), and the distribution on Σ implied by our prior $\tilde{p}(B)$ (labelled as $\tilde{p}(\Sigma)_{Np(B)}$). Yet, the posteriors are effectively identical. Again, this finding is due to Σ being identified, and is in line with the discussion of Figure 1. By contrast, the posterior distributions on B remain different even as the size of the dataset increases, due to B being not identified. In this illustration, the marginal distributions $p(B_{i,j}|Y)$ associated with $p(B|Y)$ have fatter tails under the NiWU approach than under our approach, an issue which we now inspect further.

While tempting, it is not advisable to interpret the differences the posteriors $p(B|Y)_{Np(B)}$ and $p(B|Y)_{NiWU}$ by eyeballing differences in the corresponding priors on B. To appreciate why, note that our prior on B is wider than the prior implied by the NiWU approach shown. However, our posterior is actually tighter than the posterior from the NiWU. This finding is due to the fact that, as the sample size increases, differences in $p(B|Y)$ are driven by differences in the distribution $p(Q|Y,\Sigma)$, where it holds that $p(Q|Y,\Sigma) = p(Q|\Sigma)$. Yet, inspecting the implications of $p(Q|\Sigma)$ on B is not straightforward when looking at $p(B)$, neither in our approach nor in the NiWU approach.

The bivariate case used in this illustration makes the comparison of $p(Q|\Sigma)_{NiWU}$ and $p(Q|\Sigma)_{Np(B)}$ feasible graphically, further helping assess the difference between our approach and the NiWU approach. When Q is of dimension 2×2, distributions on Q can be shown graphically as the distribution on the corresponding rotation angle θ of Givens transformations matrices (see, for example, Fry and Pagan, 2011, as well

[7] The results are the same with alternative parametrizations of the inverse Wishart distribution.

Figure 4: Implications of $p(Q|\Sigma)$ on B



Note: The draws for this figure are generated using *Algorithm F* discussed in Section C.1.1 of the Online Appendix. See Table F1 for how we set the tuning parameters.

as the analysis in Baumeister and Hamilton, 2015). Uniformity on Q is equivalent to uniformity on θ. The top-left plot of Figure 4 shows that indeed the angle of the rotation matrices that replicate draws of Q from the algorithm by Rubio-Ramirez et al. (2010) is uniformly distributed in the support $[-\pi/2, \pi/2]$. Conditioning on Σ, the rotation angles consistent with the sign restrictions are uniform in a subset of the space $[-\pi/2, \pi/2]$. While the NiWU approach treats such angles as equally plausible, the Np(B) approach does not, taking an explicit stand on the part of the structural parameter space that is considered more in line with the scaling of the variables. The remaining panels of Figure 4 show the implications of $p(Q|\Sigma)_{NiWU}$ and $p(Q|\Sigma)_{Np(B)}$ on B by showing $p(B|\Sigma)$. Given the constraint from equation (17), no draw of b_{ij} is obtained outside of the interval $[-\Sigma_{i,i}^{0.5}, +\Sigma_{i,i}^{0.5}]$. The NiWU approach implies a distribution that is skewed towards such bounds (see also equation (33)

in Baumeister and Hamilton, 2015 and their Figure 1), while the Np(B) approach features a bell-shaped distribution within these bounds. This clarifies the origin of the update of the prior distributions. Our prior on B is actually wider than the prior on B implied by the NiWU approach. Yet, what either prior actually imposes upon the data is $p(B|\Sigma)$, which is very wide and ill-shaped for the NiWU approach, and smoother and tighter for our approach.

All applications of the Np(B) approach are set to generate 25,000 posterior draws and take approximately four minutes to run. The algorithm for the NiWU approach was run to generate the same number of posterior draws and takes less than a minute. The Dynamic Striated Metropolis Hastings algorithm run for less then 90 minutes. See Table F2 in the Online Appendix for the details.

4 Application to monetary policy shocks

In this last section we apply our methodology to study the effects of monetary policy shocks on the real economy. We use the model by Caldara and Herbst (2019), which includes five variables. In their analysis, Caldara and Herbst (2019) show that monetary policy shocks in the US had an economically significant effect on the real economy during the Great Moderation, contrary to what was found, for example, by Boivin et al. (2010). Using sign restrictions, we find that the effects on the real economy are even larger than what is documented by Caldara and Herbst (2019). This result is confirmed when applying sign restrictions through the traditional NiWU approach, an approach that yet delivers a smaller quantitative difference from the results in Caldara and Herbst (2019).

4.1 Model, prior specifications, and posterior sampler

We specify the model as in Caldara and Herbst (2019), except that we identify monetary policy shocks using sign restrictions instead of an external instrument. We include

the average federal funds rate over the last week of each month, the log of manufacturing industrial production, the unemployment rate, the log of the produced price index for finished goods. We include a constant and 12 lags of the endogenous variables. As in Caldara and Herbst (2019), the model uses the period 1990M1 through 1993M12 as training sample and the period 1994M1 through 2007M7 as estimation sample.

We use the same prior beliefs for $\boldsymbol{\pi}$ as Caldara and Herbst (2019). We set $p(\boldsymbol{\pi}) \propto 1$, which corresponds to *Case 1* discussed in Section 2.3. We then add dummy variables to model the Minnesota Prior as discussed in Del Negro and Schorfheide (2011). We use the same hyperparameter values for the Bayesian shrinkage on $\boldsymbol{\pi}$ as Caldara and Herbst (2019).[8] While Caldara and Herbst (2019) introduce dummy variables also to model prior beliefs on Σ, we follow our approach and specify the prior on B. We introduce the sign restrictions that a contractionary monetary policy shock does not decrease unemployment and corporate credit spreads, and does not increase the federal funds rate, industrial production and prices. We introduce these restrictions in the month when the shock hits, and up to three months after the shock. We model the prior $\tilde{p}(B)$ as discussed in Section 2.4. We set $\gamma_i = \hat{\Sigma}^{0.5}_{ii,\text{training}}$, $\psi_1 = 0.8$ and $\psi_2 = 1.5$. This gives prior probability mass also above the estimated upper bound for b_{ij}, making the prior less dogmatic. Last, given that the model is partially identified, we impose non-repetition of the sign of the identified column of B compared to its remaining columns. Due to the shortness of the training sample relative to the number of parameters of the model, $\hat{\Sigma}_{\text{training}}$ is estimated with a VAR including one lag.

As in the illustration from Section 3, we sample the posterior distribution $\tilde{p}(B|Y)$ using both the Dynamic Striated Metropolis-Hastings algorithm and our algorithm, always checking that the results for $p(B|Y)$ are virtually identical. We set our sampler using proposal draws as discussed in the appendix of the paper. The sampler delivers

[8] The approach by Caldara and Herbst (2019) strictly requires no prior beliefs on $\boldsymbol{\pi}$, but on the corresponding structural autoregressive parameters. Yet, they specify such prior to indirectly imply the Minnesota Prior on $\boldsymbol{\pi}$. Having specified the model directly in $\boldsymbol{\pi}$, our approach makes it more natural to introduce prior beliefs on $\boldsymbol{\pi}$.

a relative effective sample size for Stage A of 0.80. The computational time of our sampler is 37 minutes, compared to the DSMH algorithm, which took approximately 5 hours. The NiWU approach took 4 minutes to generate the same number of posterior draws compared to our sampler. See Table F1 for how we set the tuning parameters and Table F2 in the Online Appendix for the details of the computational time.

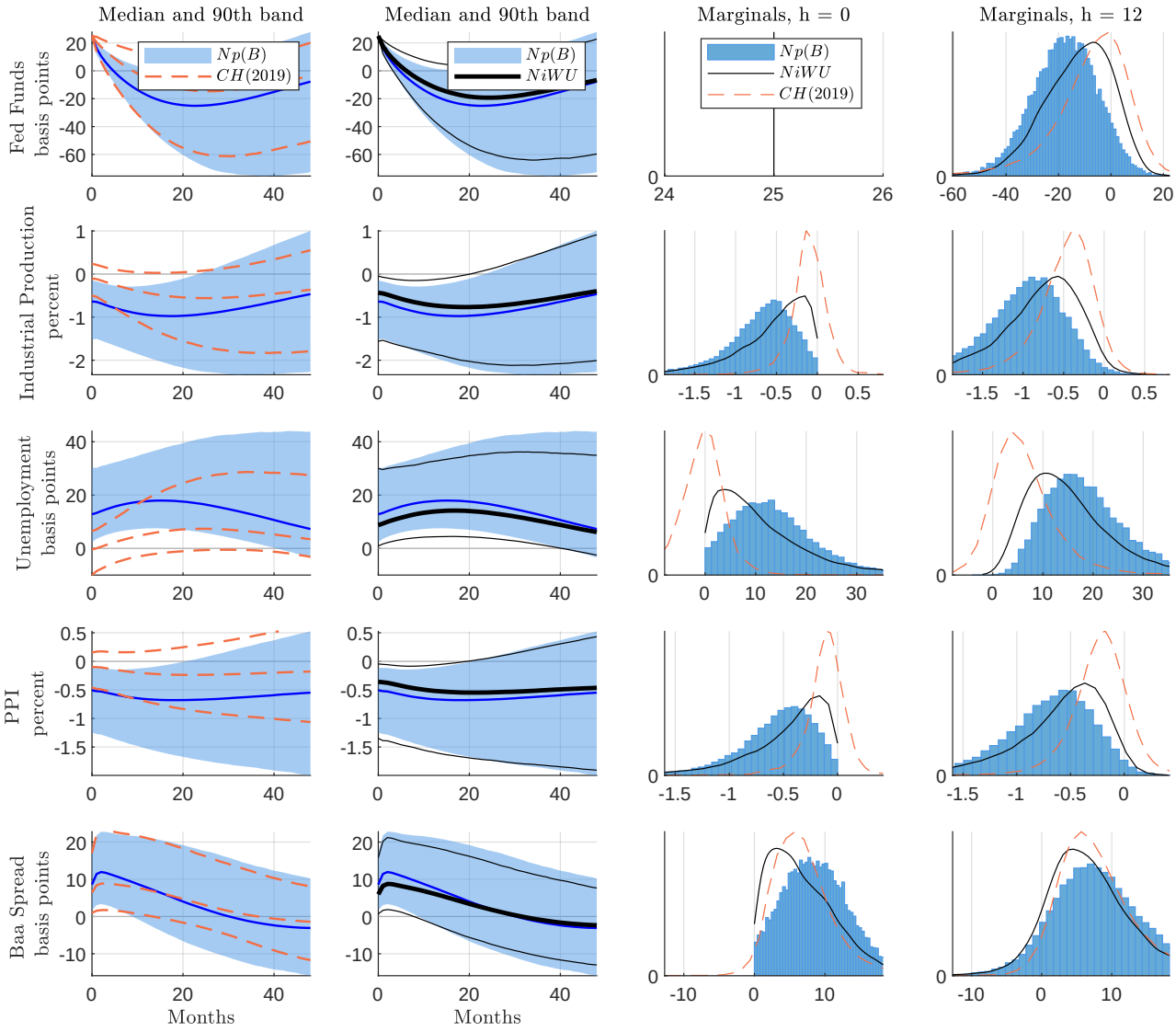
4.2 Results

Figure 5 shows the results for the impulse responses to a contractionary monetary policy shock normalized to increase the federal funds rate by 25 basis points. The left two columns report the pointwise median and the pointwise 90% credible band. The remaining columns complement the illustration by plotting the marginal posterior distribution of the same normalized impulse responses contemporaneously or after 12 months. The figure shows the results from Caldara and Herbst (2019) adjusted to the same normalization of the impulse responses, as well as when applying sign restrictions either through our Np(B) approach or through the traditional NiWU approach.

While sign restrictions make it fruitless to assess if a restricted effect is different from zero or not, they can assess if the effects are quantitatively stronger or weaker than what found by Caldara and Herbst (2019), and can highlight qualitative differences at longer horizons. The first column of Figure 5 documents that sign restrictions broadly confirm the qualitative results by Caldara and Herbst (2019). In response to a monetary tightening, the real economy is affected through a decrease in industrial production and an increase in unemployment for up to three years. In addition, prices decrease for well after the periods for which sign restrictions are introduced. However, the effects are quantitatively stronger when identifying monetary policy shocks through sign restrictions. While delivering very similar results for the policy rate and the credit spreads, the results from sign restrictions for the remaining variables suggest a response that is double in size as the one from Caldara and Herbst (2019).

We now turn to study the differences when applying sign restrictions either through

Figure 5: Monetary policy shocks: Impulse Responses



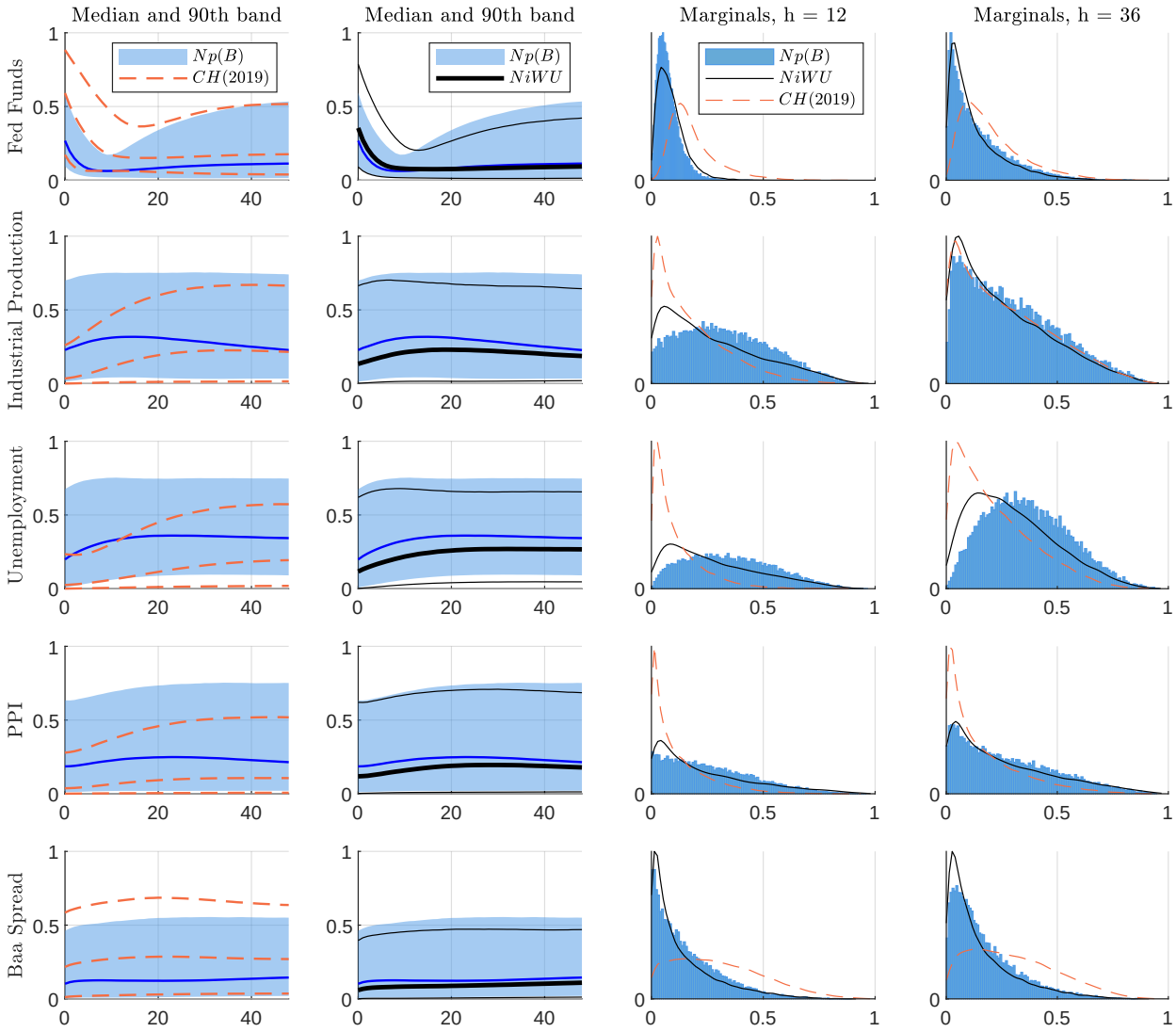
Note: The monetary shock is normalized to imply an impact 25 basis points increase in the policy rate. For the comparison to the responses by Caldara and Herbst (2019), we run their codes, apply the same normalization of the impulse responses as in our analysis, and compute the pointwise percentiles as well as the marginal distributions.

the traditional NiWU approach or through our approach. When inspecting only the pointwise median and 90th band, the results seem to be limited to a milder effect on the pointwise median responses from the NiWU approach. The remaining columns of

Figure 5 suggest, instead, that there is a relevant difference in the results depending on how sign restrictions are modelled. The NiWU approach delivers results that imply a weaker response of all variables broadly on all horizons within the first year from the shock, yet still confirming that sign restrictions imply stronger results than in Caldara and Herbst (2019). According to the NiWU approach, the mode of the impact response of unemployment is +5 basis points. Modelling the same sign restrictions taking into account the volatility of the data changes the results. According to our Np(B) approach, the mode of the same contemporaneous responses of unemployment is more than doubled and equals around +11 basis points. The same holds for industrial production.

A similar result holds for the forecast error variance decomposition, which is shown in Figure 6. The illustration is organized in the same way as with the impulse responses, except that the marginal distributions are now shown for 12 and 36 months after the shock, as stressed in Caldara and Herbst (2019). Caldara and Herbst (2019) document that monetary shock play a non-negligible role in explaining the forecast error variance of industrial production and unemployment. Our approach suggests an even stronger effect, as can be best seen from the last two columns of Figure 6. The mode of the marginal distributions in Caldara and Herbst (2019) suggest that less than 10% of the forecast error variance of industrial production and unemployment is explained by monetary shocks. Applying sign restrictions through our approach increases this statistic to around 25%. As for the analysis of impulse responses, sign restrictions applied without taking into account the volatility of the data deliver results that are quantitatively in-between the results by Caldara and Herbst (2019) and our results. As also in Caldara and Herbst (2019), we stress that posterior uncertainty on forecast error variance decompositions remains high.

Figure 6: Monetary policy shocks: Forecast Error Variance Decomposition



5 Conclusions

Structural Vector Autoregressive models are frequently identified using sign restrictions on the impulse response of selected structural shocks of interest. However, it is not clear how this identification approach should be implemented in practice. On the one hand, it is convenient to use the independent or the conjugate Normal-inverse-

Wishart-Uniform prior employed in the literature, as this makes posterior sampling highly tractable given the existing methodologies. On the other hand, it is important to retain flexibility on the prior beliefs implied for the key structural parameters of interest, since such prior affects the statistics of interest even in a large sample.

We propose an approach that offers flexibility for the prior specification on the contemporaneous impulse responses, while ensuring that the joint posterior distribution is tractable. We illustrate the intuition of our approach using simulations on the bivariate demand and supply model by Baumeister and Hamilton (2015). We then develop an application to study the effects of monetary policy shocks in the US during the Great Moderation. We confirm the results by Caldara and Herbst (2019) that monetary shocks were effective at moving the real economy, and document that the effects on unemployment and industrial production are actually twice as big as in Caldara and Herbst (2019) when identifying the monetary shocks using relatively conventional sign restrictions on the impulse responses.

References

Arias, J. E., Rubio-Ramírez, J. F. and Waggoner, D. F. (2018), 'Inference based on Structural Vector Autoregressions identified with sign and zero restrictions: Theory and applications', *Econometrica* **86**(2), 685–720.

Baumeister, C. and Hamilton, J. D. (2015), 'Sign restrictions, structural vector autoregressions, and useful prior information', *Econometrica* **83**(5), 1963–1999.

Baumeister, C. J. and Hamilton, J. D. (2018), 'Inference in structural vector autoregressions when the identifying assumptions are not fully believed: Re-evaluating the role of monetary policy in economic fluctuations', *Journal of Monetary Economics* **100**, 48–65.

Baumeister, C. J. and Hamilton, J. D. (2019), 'Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks', *American Economic Review* **109**(5), 1873–1910.

Benati, L. and Surico, P. (2009), 'VAR analysis and the great moderation', *American Economic Review* **99**(4), 1636–52.

Boivin, J., Kiley, M. T. and Mishkin, F. S. (2010), How has the monetary transmission mechanism evolved over time?, *in* 'Handbook of Monetary Economics', Vol. 3, Elsevier, pp. 369–422.

Caldara, D. and Herbst, E. (2019), 'Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy svars', *American Economic Journal: Macroeconomics* **11**(1), 157–92.

Canova, F. (2007), *Methods for applied macroeconomic research*, Vol. 13, Princeton University Press.

Canova, F. and De Nicoló, G. (2002), 'Monetary disturbances matter for business fluctuations in the G-7', *Journal of Monetary Economics* **49**(6), 1131–1159.

Canova, F. and Paustian, M. (2011), 'Business cycle measurement with some theory', *Journal of Monetary Economics* **58**(4), 345–361.

Canova, F. and Pina, J. P. (2005), What VAR tell us about DSGE models?, *in* 'New Trends in Macroeconomics', Springer, pp. 89–123.

Del Negro, M. and Schorfheide, F. (2011), *Bayesian Macroeconometrics*, ed. Herman van Dijk, Gary Koop and John Geweke, pages 293-389. Oxford University Press.

Fry, R. and Pagan, A. (2011), 'Sign restrictions in structural vector autoregressions: A critical review', *Journal of Economic Literature* **49**(4), 938–960.

Giacomini, R. and Kitagawa, T. (2015), 'Robust inference about partially identified SVARs', *Manuscript, UCL* .

Giacomini, R., Kitagawa, T. and Uhlig, H. (2019), Estimation under ambiguity, Technical report, cemmap working paper.

Kadiyala, K. R. and Karlsson, S. (1997), 'Numerical methods for estimation and inference in Bayesian VAR-models', *Journal of Applied Econometrics* pp. 99–132.

Kilian, L. and Lütkepohl, H. (2017), *Structural vector autoregressive analysis*, Cambridge University Press.

Koop, G. (2003), *Bayesian Econometrics*, John Wiley & Sons Ltd.

Koop, G. and Korobilis, D. (2010), 'Bayesian multivariate time series methods for empirical macroeconomics', *Foundations and Trends in Econometrics* **3**(4), 267–358.

Litterman, R. B. (1986), 'Forecasting with Bayesian vector autoregressions five years of experience', *Journal of Business & Economic Statistics* **4**(1), 25–38.

Rubio-Ramirez, J. F., Waggoner, D. F. and Zha, T. (2010), 'Structural vector autoregressions: Theory of identification and algorithms for inference', *The Review of Economic Studies* **77**(2), 665–696.

Sims, C. A. and Zha, T. (1998), 'Bayesian methods for dynamic multivariate models', *International Economic Review* pp. 949–968.

Uhlig, H. (2005), 'What are the effects of monetary policy on output? Results from an agnostic identification procedure', *Journal of Monetary Economics* **52**(2), 381–419.

Uhlig, H. (2017), Shocks, sign restrictions, and identification, *in* 'Advances in Economics and Econometrics', Cambridge University Press.

Waggoner, D. F., Wu, H. and Zha, T. (2016), 'Striated Metropolis–Hastings sampler for high-dimensional models', *Journal of Econometrics* **192**(2), 406–420.

Appendix

As shown in Section B.1 of the Online Appendix, the proposal prior distributions

$$p(\boldsymbol{\pi}, \Sigma)_{prop} = q(\boldsymbol{\pi}) \cdot q(\Sigma), \tag{18}$$

$$q(\boldsymbol{\pi}) \propto 1, \tag{19}$$

$$q(\Sigma) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}S\right]}, \tag{20}$$

allows for proposal draws directly generated from the proposal posterior distribution

$$\Sigma|Y \sim \text{iW}(S^*, d^*), \tag{21}$$

$$d^* = T - m - c - k - 1, \tag{22}$$

$$S^* = S + \hat{\Sigma}_T(T-m). \tag{23}$$

Under *Case 1* considered in the paper (equation 6, derived analytically in Section C.1.2 of the Online Appendix), the weights in Stage A of our algorithm equal

$$w_d^{\text{stage A}} = \frac{\tilde{p}(\Sigma|Y)_{Np(B)}}{p(\Sigma|Y)_{NiWU}}, \tag{24}$$

$$= \frac{v_{\{B\to\Sigma,Q\}} \cdot |\det(\Sigma)|^{-\frac{T-m}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\hat{\Sigma}_T(T-m)\right]} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ}{|\det(\Sigma)|^{-\frac{T-m-c}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(S+\hat{\Sigma}_T(T-m)\right)\right]}}, \tag{25}$$

$$= \frac{|\det(\Sigma)|^{-\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T-m}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\hat{\Sigma}_T(T-m)\right]} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ}{|\det(\Sigma)|^{-\frac{T-m-c}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(S+\hat{\Sigma}_T(T-m)\right)\right]}}, \tag{26}$$

$$= |\det(\Sigma)|^{-\frac{c+1}{2}} \cdot e^{\frac{1}{2}\text{trace}\left[\Sigma^{-1}S\right]} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ. \tag{27}$$

The integral in equation (27) can be evaluated using *Algorithm E*, see Section C.1.1 of the Online Appendix or Step 3d of our algorithm. In our applications, setting

$(c, S) = (-1, 0 \cdot I_k)$ is sufficient to reduce the volatility of the weights considerably, delivering a sufficiently high effective sample size.

Consider now *Case 2* (equation 7, derived analytically in Section C.1.3 of the Online Appendix). The weights in Stage A corresponding to the same proposal draws from equation (21) now equal

$$w_d^{\text{stage A}} = \frac{\tilde{p}(\Sigma|Y)_{Np(B)}}{p(\Sigma|Y)_{NiWU}}, \tag{28}$$

$$= \frac{v_{\{B\to\Sigma,Q\}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}-\boldsymbol{\mu}_\pi^{*\prime}V_\pi^{*-1}\boldsymbol{\mu}_\pi^*\right\}} \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ}{|\det(\Sigma)|^{-\frac{T-m-c}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(S+\hat{\Sigma}_T(T-m)\right)\right]}}, \tag{29}$$

$$= \frac{|\det(\Sigma)|^{-\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}-\boldsymbol{\mu}_\pi^{*\prime}V_\pi^{*-1}\boldsymbol{\mu}_\pi^*\right\}} \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ}{|\det(\Sigma)|^{-\frac{T-m-c}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}\left(S+\hat{\Sigma}_T(T-m)\right)\right]}}, \tag{30}$$

$$= |\det(\Sigma)|^{-\frac{m+c+1}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{\frac{1}{2}\text{trace}\left[\Sigma^{-1}S\right]} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T\otimes\Sigma^{-1}\right)\tilde{\boldsymbol{y}}-\boldsymbol{\mu}_\pi^{*\prime}V_\pi^{*-1}\boldsymbol{\mu}_\pi^*\right\}} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ. \tag{31}$$

Since this function is less convenient to evaluate compared to equation (27), we replace the proposal prior beliefs (18)-(20) with

$$p(\boldsymbol{\pi}, \Sigma)_{prop} = q(\boldsymbol{\pi}) \cdot q(\Sigma), \tag{32}$$

$$q(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_\pi, V_\pi), \tag{33}$$

$$q(\Sigma) \propto |\det(\Sigma)|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\text{trace}\left[\Sigma^{-1}S\right]}. \tag{34}$$

where the prior mean and variance equal the ones used in our prior $\tilde{p}(\boldsymbol{\pi})$ in equation (7). As shown in Section B.2 of the Online Appendix, the proposal draws obtained

from a Gibbs sampler built on the conditional distributions

$$\boldsymbol{\pi}|Y,\Sigma \sim \mathrm{N}\big(\boldsymbol{\mu}_\pi^*, V_\pi^*\big), \tag{35}$$

$$\Sigma|Y,\Pi \sim \mathrm{iW}(S^*, d^*), \tag{36}$$

$$V_\pi^* = \big(V_\pi^{-1} + WW' \otimes \Sigma^{-1}\big)^{-1}, \tag{37}$$

$$\boldsymbol{\mu}_\pi^* = V_\pi^*\big(V_\pi^{-1}\boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T\big), \tag{38}$$

$$d^* = T - c - k - 1, \tag{39}$$

$$S^* = S + (Y - \Pi W)(Y - \Pi W)', \tag{40}$$

are representative of the marginal posterior distribution

$$p(\Sigma|Y) \propto |\det(V_\pi^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2}\left\{\mathrm{trace}\left[\Sigma^{-1}S\right] + \tilde{\boldsymbol{y}}'\left(I_T \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} {V_\pi^*}^{-1} \boldsymbol{\mu}_\pi^*\right\}}. \tag{41}$$

Hence, the weights for Stage A,

$$w_d^{\text{stage A}} = \frac{\tilde{p}(\Sigma|Y)_{Np(B)}}{p(\Sigma|Y)_{NiWU}}, \tag{42}$$

$$= \frac{v_{\{B\to\Sigma,Q\}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} {V_\pi^*}^{-1} \boldsymbol{\mu}_\pi^*\right\}} \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ}{|\det(V_\pi^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2}\left\{\mathrm{trace}\left[\Sigma^{-1}S\right] + \tilde{\boldsymbol{y}}'\left(I_T \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} {V_\pi^*}^{-1} \boldsymbol{\mu}_\pi^*\right\}}}, \tag{43}$$

$$= \frac{|\det(\Sigma)|^{-\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T}{2}} \cdot |\det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} {V_\pi^*}^{-1} \boldsymbol{\mu}_\pi^*\right\}} \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ}{|\det(V_\pi^*)|^{\frac{1}{2}} \cdot |\det(\Sigma)|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2}\left\{\mathrm{trace}\left[\Sigma^{-1}S\right] + \tilde{\boldsymbol{y}}'\left(I_T \otimes \Sigma^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} {V_\pi^*}^{-1} \boldsymbol{\mu}_\pi^*\right\}}}, \tag{44}$$

$$= |\det(\Sigma)|^{-\frac{c+1}{2}} \cdot e^{\frac{1}{2}\mathrm{trace}\left[\Sigma^{-1}S\right]} \cdot \int_{O(k)} p_B\big(h(\Sigma)Q\big)dQ, \tag{45}$$

are now the same as from equation (27), and become particularly tractable to evaluate for $(c, S) = (-1, 0 \cdot I_k)$. In our applications, we find that the relative effective sample

size for *Case 2* is Section 3 is much higher when using the proposal prior distribution (32)-(34) rather than (18)-(20).

Table 2 summarizes the key results needed to our the sampler. For each case considered for the prior beliefs, the table indicates a candidate proposal prior distribution, the corresponding proposal posterior distribution, and the corresponding weights for Stage A of our sampler.

Table 2: Proposal prior and posterior distributions for our sampler and comparison to the NiWU approach

	Proposal prior $q(\boldsymbol{\pi}, \Sigma)$	Proposal posterior $q(\boldsymbol{\pi}, \Sigma \vert Y)$ and $q(\Sigma \vert Y)$	Weights in Stage A of our algorithm	NiWU approach for the comparison
Case 1)	$\tilde{p}(\boldsymbol{\pi}) \propto 1$ (Flat prior or Minnesota prior via dummies observations)			
1a)	$q(\boldsymbol{\pi}, \Sigma) = q(\boldsymbol{\pi}) \cdot q(\Sigma)$ $q(\boldsymbol{\pi}) \propto 1$ $q(\Sigma) \propto \lvert\det(\Sigma)\rvert^{\frac{c}{2}} \cdot$ $\cdot\, e^{-\frac{1}{2}\mathrm{trace}(\Sigma^{-1}S)}$	$\Sigma \vert Y \sim iW(S^*, d^*)$ $d^* = T - m - c - k - 1$ $S^* = S + \hat{\Sigma}_T(T-m)$ $T \geq c + m + 2k + 1$	$\propto \lvert\det(\Sigma)\rvert^{-\frac{c+1}{2}} \cdot e^{\frac{1}{2}\mathrm{trace}\left[\Sigma^{-1}S\right]} \cdot$ $\cdot \int_{O(k)} \tilde{p}\big(h(\Sigma)Q\big)dQ$	$p(\boldsymbol{\pi}, \Sigma) = p(\boldsymbol{\pi}) \cdot p(\Sigma)$ $p(\boldsymbol{\pi}) \propto 1$ either $p(\Sigma) \propto 1$ or $\Sigma \sim \mathrm{iW}(S, d)$
	Special case for $(c, S) = (-1, 0 \cdot I_k)$, used in our algorithm			
1b)	$q(\boldsymbol{\pi}, \Sigma) = q(\boldsymbol{\pi}) \cdot q(\Sigma)$ $q(\boldsymbol{\pi}) \propto 1$ $q(\Sigma) \propto \lvert\det(\Sigma)\rvert^{-\frac{1}{2}}$	$\Sigma \vert Y \sim iW(S^*, d^*)$ $d^* = T - m - k$ $S^* = \hat{\Sigma}_T(T-m)$ $T \geq 2k + m$	$\propto \int_{O(k)} \tilde{p}\big(h(\Sigma)Q\big)dQ$	(same as above)
Case 2)	$\tilde{p}(\boldsymbol{\pi} \vert B) = \tilde{p}(\boldsymbol{\pi})\ N(\boldsymbol{\mu}_\pi, V_\pi)$ (Flexible Minnesota prior)			
2a)	$q(\boldsymbol{\pi}, \Sigma) = q(\boldsymbol{\pi}) \cdot q(\Sigma)$ $\boldsymbol{\pi}_{prop} \sim N(\boldsymbol{\mu}_\pi, V_\pi)$ $q(\Sigma) \propto \lvert\det(\Sigma)\rvert^{\frac{c}{2}} \cdot$ $\cdot\, e^{-\frac{1}{2}\cdot\mathrm{trace}(\Sigma^{-1}S)}$	$\mathrm{vec}(\Pi) \vert Y, \Sigma \sim N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ $\Sigma \vert Y, \Pi \sim iW(d^*, S^*)$ $d^* = T - c - k - 1$ $S^* = S + (Y - \Pi W)(Y - \Pi W)'$ $T \geq c + 2k + 1$ $V_\pi^* = (V_\pi^{-1} + WW' \otimes \Sigma^{-1})^{-1}$ $\boldsymbol{\mu}_\pi^* = V_\pi^*(V_\pi^{-1}\boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T)$	$\propto \lvert\det(\Sigma)\rvert^{-\frac{c+1}{2}} \cdot e^{\frac{1}{2}\mathrm{trace}\left[\Sigma^{-1}S\right]} \cdot$ $\cdot \int_{O(k)} \tilde{p}\big(h(\Sigma)Q\big)dQ$	$p(\boldsymbol{\pi}, \Sigma) = p(\boldsymbol{\pi}) \cdot p(\Sigma)$ $\boldsymbol{\pi} \sim N(\boldsymbol{\mu}_\pi, V_\pi)$ either $p(\Sigma) \propto 1$ or $\Sigma \sim \mathrm{iW}(S, d)$
	Special case for $(c, S) = (-1, 0 \cdot I_k)$, used in our algorithm			
2b)	$q(\boldsymbol{\pi}) \sim N(\boldsymbol{\mu}_\pi, V_\pi)$ $q(\Sigma) \propto \lvert\det(\Sigma)\rvert^{-\frac{1}{2}}$	$\mathrm{vec}(\Pi) \vert Y, \Sigma \sim N(\boldsymbol{\mu}_\pi^*, V_\pi^*)$ $\Sigma \vert Y, \Pi \sim iW(d^*, S^*)$ $d^* = T - k$ $S^* = (Y - \Pi W)(Y - \Pi W)'$ $T \geq 2k$ $V_\pi^* = (V_\pi^{-1} + WW' \otimes \Sigma^{-1})^{-1}$ $\boldsymbol{\mu}_\pi^* = V_\pi^*(V_\pi^{-1}\boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T)$	$\propto \int_{O(k)} \tilde{p}\big(h(\Sigma)Q\big)dQ$	(same as above)

Note: Our algorithm requires a proposal distribution $q(\Sigma \vert Y)$. The table lists different possible proposal distributions depending on the corresponding proposal prior $q(\Sigma)$, and reports the corresponding weights for Step 3d of our algorithm. The last column shows which parametrization of the NiWU approach is used as a comparison in the analysis of the paper.

The Data Analytics for Finance and Macro (DAFM) Research Centre
Forecasting trends is more important now than it has ever been. Our quantitative financial and macroeconomic research helps central banks and statistical agencies better understand their markets. Having our base in central London means we can focus on issues that really matter to the City. Our emphasis on digitalisation and the use of big data analytics can help people understand economic behaviour and adapt to changes in the economy.

Find out more at kcl.ac.uk/dafm

KING'S BUSINESS SCHOOL

Data Analytics for Finance & Macro Research Centre