Check for updates

Quantitative Economics 14 (2023), 1221–1250 1759-7331/20231221

A new posterior sampler for Bayesian structural vector autoregressive models

Martin Bruns
School of Economics, University of East Anglia

Michele Piffer
King's Business School, King's College London

We develop an importance sampler for sign restricted Bayesian structural vector autoregressive models. The algorithm nests as a special case the sampler associated with the popular Normal inverse Wishart Uniform prior, while allowing to move beyond such prior in medium sized models. We then propose a prior on contemporaneous impulse responses that provides flexibility on the magnitude and shape of the impact responses. We illustrate the quantitative relevance of the choice of the prior in an application to US monetary policy shocks. We find that the real effects of monetary policy shocks are stronger under our proposed prior than in the Normal inverse Wishart Uniform setup.

Keywords. Sign restrictions, Bayesian inference, monetary policy shocks.

JEL classification. C11, C32, E50.

1. Introduction

Bayesian Structural Vector Autoregressive (B-SVAR) models are popular in empirical macroeconomics. However, there is an ongoing search in the literature for efficient posterior samplers for such models. Such efficient algorithms exist to explore the posterior distribution when the prior is restricted to the so-called Normal inverse Wishart Uniform (NiWU) prior (Uhlig (2005) and Rubio-Ramirez, Waggoner, and Zha (2010)). However, it is not clear how one should sample from the posterior distribution when the prior does not fall within the NiWU special case. Demand for such samplers has increased since it was documented that the NiWU prior can impose unintended features on key statistics of the model (Baumeister and Hamilton (2015) and Wolf (2020)).

This paper develops an importance sampler for underidentified SVARs, for example, SVARs that identify structural shocks using sign restrictions. To address the main

Martin Bruns: martin.j.bruns@gmail.com
Michele Piffer: m.b.piffer@gmail.com

We thank 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 744010. Martin Bruns thanks the German Academic Scholarship Foundation for financial support.

© 2023 The Authors. Licensed under the Creative Commons Attribution-NonCommercial License 4.0. Available at http://qeconomics.org. https://doi.org/10.3982/QE2207

challenge of importance samplers, namely the choice of a suitable proposal function, our sampler builds on two key features. First, the sampler is set up in two stages, initially on the reduced form parameters and then on the mapping from the reduced form to the structural parameters. Second, for both stages it generates proposal draws from the posterior distribution associated with the NiWU prior. While importance samplers for SVARs have been explored in the past (Sims and Zha (1998), Leeper, Sims, Zha, Hall, and Bernanke (1996) and Zha (1999)), the established practice of generating proposal draws directly for the structural parameters using Normal or t-distributions has been documented to perform poorly (Waggoner and Zha (2003)). We show that building the sampler in two steps using proposal draws from the posterior associated with the NiWU prior leads to a good performance of importance samplers in B-SVARs.

Our sampler nests the sampler from the NiWU approach as a special case. We show that our sampler recovers the same posterior distribution as the Dynamic Striated Metropolis–Hastings (DSMH) algorithm by Waggoner, Wu, and Zha (2016), yet with less computational effort. The DSMH algorithm is designed for a wider class of models than the SVAR models, is more computionally demanding and requires comparatively more involved coding, but works well to establish a benchmark. We use Monte Carlo simulations to document that our sampler works efficiently, and to illustrate its key properties.

We then apply our sampler to study the implications of using the NiWU prior instead of a different prior. Prior beliefs on structural parameters in underidentified models matter also asymptotically. Since more than one prior distribution can be used to model any given set of sign restrictions, the question arises to what extent the results are driven by the data and the sign restrictions, or by the specific prior distribution used to model the sign restrictions (Baumeister and Hamilton (2015)). Having developed an algorithm that can handle a different prior from the NiWU prior, we specify a set of sign restrictions in an application to US monetary policy shocks, and compare the posterior distribution when modeling the same sign restrictions using different priors. To do so, we formulate a new flexible prior which the researcher can use to express prior beliefs on contemporaneous impulse responses. One advantage of our prior is that it provides flexibility not only on the sign, but also on the magnitude and shape of the contemporaneous impulse responses.

We build the above application on the model by Caldara and Herbst (2019), who identify monetary policy shocks using an external instrument. We use their five-variable model and apply standard sign restrictions up to 3 months from the shock. We simulate a monetary contraction that increases the policy rate by 25 basis points. We find that the NiWU prior implies a posterior suggesting a 9 basis points impact increase in the unemployment rate. We show that changing from the NiWU prior can make a significant difference for the results. For example, applying the same sign restrictions with our alternative prior leads to estimate an impact effect that is 50% higher, that is, up to a 13 basis points of an increase in unemployment. Similarly, the NiWU prior can lead to a posterior distribution suggesting a relatively strong role of monetary policy shocks in driving the forecast error variance of the policy rate (up to 35%). Modeling the same sign restrictions with our prior can lead to an effect up to 30% lower for at least half a year, suggesting a stronger role for the systematic component of monetary policy.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

A SVAR model potentially allows for sign restrictions over related but conceptually very different dimensions. Kilian and Lütkepohl (2017) take the view that applied researchers can frequently express prior beliefs on impulse responses. Baumeister and Hamilton (2015, forthcoming) show cases in which sign restrictions are naturally expressed on the contemporaneous response of the variables of the model, for instance, structural elasticities. Baumeister and Hamilton (2018) propose a hybrid approach that jointly accommodates sign restrictions on both the impulse responses and the structural relations among variables. Our algorithm works in any of the above scenarios. We derive and then apply our algorithm to a case featuring only sign restrictions on the impulse responses for illustrative simplicity. Our additional contribution of a new prior distribution on the sign, shape, and magnitude of the contemporaneous impulse responses can be used jointly with restrictions on the contemporaneous relations, substituting the prior in Baumeister and Hamilton (2018) with our prior. It should be stressed that, being an importance sampler, our sampler can work well in some applications and poorly in others, and it is unlikely to work in large models or in relatively small samples. Our prior only provides flexibility on the contemporaneous impulse responses. See Canova, Kociecki, and Piffer (2023) for an approach offering flexibility on the shape over multiple horizons of the impulse response.

The paper also relates to the active literature on posterior sampling in B-SVARs. Kociecki, Rubaszek, and Ca'Zorzi (2012) and Chan (2022) show that direct sampling can be used in recursive B-SVARs. Kociecki (2010) works with prior beliefs on impulse responses, but works under the recursive identification approach. Arias, Rubio-Ramírez, and Waggoner (2018) develop an importance sampler for sign and zero restrictions. Their approach converges back to the NiWU prior when only sign restrictions are considered, which is the case of interest for our paper. Korobilis (2022) proposes a new algorithm for sign restrictions using a factor structure for the shocks. We follow the more traditional specification of SVARs, and use an accept/reject feature to introduce sign restrictions on contemporaneous and future horizons of the impulse responses. We relate to Giacomini, Kitagawa, and Uhlig (2019) and Giacomini and Kitagawa (2021) in stressing the mapping from reduced form to structural parameters, but we concentrate on a single prior. Others have focused on how to shrink posterior bands associated with the NiWU prior; see, for instance, Antolín-Díaz and Rubio-Ramírez (2018), Amir-Ahmadi and Drautzburg (2021), and Volpicella (2021).

The paper is organized as follows. Section 2 outlines the methodology proposed. Section 3 shows a simulation exercise based on the estimated bivariate VAR model by Baumeister and Hamilton (2015). Section 4 presents the empirical application to US monetary policy shocks. Section 5 concludes.

2. The methodology

In this section, we present the structural VAR model, discuss how different prior beliefs relate to each other in SVARs, and outline our posterior sampler. Last, we propose one possible prior distribution that can be used with our approach. As notation, we use capital letters for matrices, lower case letters for scalars, and lower case in bold for vectors.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.1 *The model*

The structural VAR(p) model for the $k \times 1$ vector $\boldsymbol{y}_t$ in the so-called B-form (see, e.g., Uhlig (2005)) is given by

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

with p the number of lags of the model, $\boldsymbol{\pi}_0$ an intercept, and $\boldsymbol{\epsilon}_t$ a $k \times 1$ vector of structural shocks, whose covariance matrix is normalized to the identity matrix. The matrix $\Pi = [\boldsymbol{\pi}_0, \Pi_1, \ldots, \Pi_p]$ is of dimension $k \times m$ with $m = kp + 1$, and collects the intercept and the autoregressive parameters of the model. Matrix B in equation (1) captures the contemporaneous effects of one-standard-deviation shocks, while future horizons of the impulse responses can be calculated using equation (1) recursively. Alternatively, the SVAR can be specified by highlighting the relation between reduced form and structural parameters, as in Rubio-Ramirez, Waggoner, and Zha (2010) and Arias, Rubio-Ramírez, and Waggoner (2018),

$$\boldsymbol{y}_t = \boldsymbol{\pi}_0 + \sum_{l=1}^{p} \Pi_l \boldsymbol{y}_{t-l} + \boldsymbol{u}_t, \tag{2}$$

$$\boldsymbol{u}_t \sim N(\mathbf{0}, \Sigma), \tag{3}$$

$$\boldsymbol{u}_t = h(\Sigma) Q \boldsymbol{\epsilon}_t, \tag{4}$$

with $h(\Sigma)$ a factorization of Σ satisfying $h(\Sigma)h(\Sigma)' = \Sigma$, for example, the Cholesky factorization. Equations (2)–(3) capture the reduced form VAR, while equation (4) shows the mapping from structural to reduced form shocks. Q is an orthogonal matrix.

Model (1) parametrizes the SVAR in (Π, B), while model (2)–(4) parametrizes the SVAR in (Π, Σ, Q), where it holds that

$$B = h(\Sigma)Q, \tag{5}$$

$$\Sigma = BB', \tag{6}$$

$$Q = Bh\left(BB'\right)^{-1}. \tag{7}$$

Other parametrizations exist, for instance, the A-form and the AB-form (Amisano and Giannini (2012)and Arias, Rubio-Ramírez, and Waggoner (2018)). No parametrization is better than others, although each parametrization better highlights some features of the model and not others.[1] 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 $Y = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_T]$, $W = [\boldsymbol{w}_1, \ldots, \boldsymbol{w}_T]$, $\boldsymbol{w}_t = (1, \boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p})'$.

[1] Structural VARs can also be specified in matrix $A = B^{-1}$ rather than in B (see, e.g., Sims and Zha (1998)). Whether the model is more conveniently expressed in A or B (or even in a combined form) depends on whether the identifying restrictions introduced by the researcher are more naturally 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

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

2.2 *Priors and posteriors*

Because the model admits more than one parametrization, prior beliefs formulated on one parametrization imply prior beliefs on the other parametrization (Baumeister and Hamilton (2015)). The same holds for the posterior, a point that plays a crucial role in our sampler. We first discuss the case in which priors do not impose any sign restriction, as this simplifies the discussion and highlights the intuition behind our sampler. We then generalize the analysis to sign restrictions.

The illustrative case of no sign restrictions Define $\boldsymbol{\pi} = \text{vec}(\Pi)$ of dimension $km \times 1$, where $\text{vec}(\cdot)$ is the vectorization operator stacking columns vertically. Consider the prior

$$p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}|B) \cdot p(B), \tag{8}$$

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

$$p(B), \tag{10}$$

with $\phi(\boldsymbol{\mu}_\pi, V_\pi)$ the probability density function of the Normal distribution with expected value $\boldsymbol{\mu}_\pi$ and variance V_π, which can depend on B. $p(\boldsymbol{\pi}, B)$ is restricted to imply a Normal $p(\boldsymbol{\pi}|B)$. By contrast, $p(B)$ is freely selected by the researcher. Because of the mapping (6)–(7), the prior $p(\boldsymbol{\pi}, B)$ from (8)–(10) implies prior beliefs $p(\boldsymbol{\pi}, \Sigma, Q)$. Such joint prior can be decomposed in different ways, including

$$p(\boldsymbol{\pi}, \Sigma, Q) = p(\boldsymbol{\pi}|\Sigma, Q) \cdot p(Q|\Sigma) \cdot p(\Sigma),$$

$$p(\boldsymbol{\pi}|\Sigma, Q) = \phi(\boldsymbol{\mu}_\pi, V_\pi),$$

$$p(Q|\Sigma), \tag{11}$$

$$p(\Sigma). \tag{12}$$

$p(\boldsymbol{\pi}|\Sigma, Q)$ is by construction Normal and coincides with $p(\boldsymbol{\pi}|B)$ from (9). $p(Q|\Sigma)$, $p(\Sigma)$ from (11)–(12) are not free; they are implied by (10).

The priors $p(\boldsymbol{\pi}, B)$ and $p(\boldsymbol{\pi}, \Sigma, Q)$ from equations (8)–(12) are effectively two different sides of the same prior. One can, for instance, numerically explore $p(\boldsymbol{\pi}, B)$ directly by drawing from $p(B)$ and $p(\boldsymbol{\pi}|B)$ or indirectly by drawing from $p(\Sigma)$, $p(Q|\Sigma)$, $p(\boldsymbol{\pi}|\Sigma, Q)$, and then mapping the draws from $(\boldsymbol{\pi}, \Sigma, Q)$ to $(\boldsymbol{\pi}, B)$ using (5). Which one is numerically less costly needs to be assessed on a case-by-case basis. Define $p(\boldsymbol{\pi}, B|Y)$ and $p(\boldsymbol{\pi}, \Sigma, Q|Y)$ the joint posterior distributions associated with the joint priors $p(\boldsymbol{\pi}, B)$ and $p(\boldsymbol{\pi}, \Sigma, Q)$. $p(\boldsymbol{\pi}, B|Y)$ and $p(\boldsymbol{\pi}, \Sigma, Q|Y)$ are effectively two sides of the same posterior, so either can be used to explore the posterior.

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% specify the model in the *A* form, 76% specify the model in the *B* form, and 9% specify the model in the hybrid *AB* form. The detailed list is available at https://drive.google.com/open?id=1aTAi4bCveThY2RxEYUhDY0E29kO5xLC0.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

The prior (8)–(12) nests the popular Normal-inverse-Wishart-(Haar)Uniform (NiWU) prior as a special case. We use notation $p_{NiWU,c}(\cdot)$ to refer to the special case of the conjugate NiWU prior, which starts from the parametrization $(\boldsymbol{\pi}, \Sigma, Q)$ and introduces the priors

$$p_{NiWU,c}(\boldsymbol{\pi}, \Sigma, Q) = p_{NiWU,c}(\boldsymbol{\pi}|\Sigma, Q) \cdot p_{NiWU,c}(Q|\Sigma) \cdot p_{NiWU,c}(\Sigma), \tag{13}$$

$$p_{NiWU,c}(\boldsymbol{\pi}|\Sigma, Q) = p_{NiWU,c}(\boldsymbol{\pi}|\Sigma) = \phi(\boldsymbol{\mu}_\pi, \bar{V}_\pi \otimes \Sigma), \tag{14}$$

$$p_{NiWU,c}(Q|\Sigma) = p_{NiWU,c}(Q) = U_{O(k)} \propto 1, \tag{15}$$

$$p_{NiWU,c}(\Sigma) = iW(S, d). \tag{16}$$

$p_{NiWU,c}(\boldsymbol{\pi}|\Sigma)$ is restricted to the Normal density. $p_{NiWU,c}(Q|\Sigma)$ is restricted to the (Haar)Uniform distribution, which we indicate with $U_{O(k)}$, where $p_{NiWU,c}(Q|\Sigma) = p_{NiWU,c}(Q)$ in the absence of sign restrictions. $p_{NiWU,c}(\Sigma)$ is restricted to the inverse Wishart density, with $iW(S, d)$ indicating the corresponding probability density function. Because of the mapping from (5), the joint prior (13)–(16) implies a prior on $(\boldsymbol{\pi}, B)$,

$$p_{NiWU,c}(\boldsymbol{\pi}, B) = p_{NiWU,c}(\boldsymbol{\pi}|B) \cdot p_{NiWU,c}(B),$$

$$p_{NiWU,c}(\boldsymbol{\pi}|B) = \phi(\boldsymbol{\mu}_\pi, V_\pi), \tag{17}$$

$$p_{NiWU,c}(B). \tag{18}$$

Since $p_{NiWU,c}(\boldsymbol{\pi}|\Sigma)$ from (14) is Normal, $p_{NiWU,c}(\boldsymbol{\pi}|B)$ from (17) is Normal and coincides with $p_{NiWU,c}(\boldsymbol{\pi}|\Sigma)$. $p_{NiWU,c}(B)$ from (18) is not free. It is implied by (15)–(16), and it can be characterized analytically; see, for instance, Kociecki (2017) and Arias, Rubio-Ramírez, and Waggoner (2018). $p_{NiWU,c}(\boldsymbol{\pi}, \Sigma, Q)$ and $p_{NiWU,c}(\boldsymbol{\pi}, B)$ are two equivalent representations of the same NiWU prior, and are associated with the posteriors $p_{NiWU,c}(\boldsymbol{\pi}, \Sigma, Q|Y)$, $p_{NiWU,c}(\boldsymbol{\pi}, B|Y)$. Because of the conjugate nature of the prior, $p_{NiWU,c}(\boldsymbol{\pi}|Y, B)$ is Normal and $p_{NiWU,c}(\Sigma|Y)$ is inverse Wishart. The conjugate version of the NiWU prior simplifies the exposition, but we will consider the independent version of the prior further below, replacing notation $_{NiWU,c}$ with $_{NiWU,i}$.

To introduce the key steps of the sampler, suppose we need to generate draws from $p(B)$, but cannot directly draw from it. An alternative consists of indirectly drawing from $p(B)$ by sequentially drawing from $p(\Sigma)$, $p(Q|\Sigma)$, and then mapping draws from (Σ, Q) into B. The same can be said for the posterior, where draws for $p(B|Y)$ can, in principle, be generated indirectly from $p(\Sigma|Y)$, $p(Q|Y, \Sigma)$. Our paper analyzes to what extent the special case of a NiWU prior can be used in an importance sampler that uses draws from $p_{NiWU,c}(\Sigma|Y)$, $p_{NiWU,c}(Q|Y, \Sigma)$, or their equivalent under an independent prior (which can both be drawn from at a low computational cost), as proposal draws for $p(\Sigma|Y)$, $p(Q|Y, \Sigma)$.

We begin with the relationship between the priors $p(\Sigma)$ and $p_{NiWU,c}(\Sigma)$. If $p(B)$ coincides with $p_{NiWU,c}(B)$, then $p(\Sigma)$ and $p_{NiWU,c}(\Sigma)$ coincide, and $p(\Sigma)$ is an inverse Wishart. If $p(B)$ is sufficiently close to $p_{NiWU,c}(B)$, then $p(\Sigma)$ and $p_{NiWU,c}(\Sigma)$ are similar, and $p(\Sigma)$ is approximately an inverse Wishart. One could then in principle explore $p(\Sigma)$ using an importance sampler with proposal draws from $p_{NiWU,c}(\Sigma)$. If, instead,

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$p(B)$ is sufficiently different from $p_{NiWU,c}(B)$, then $p(\Sigma)$ and $p_{NiWU,c}(\Sigma)$ might differ enough that proposal draws from $p_{NiWU,c}(\Sigma)$ do not work in an importance sampler aimed at exploring $p(\Sigma)$. However, since Σ is identified, the *posterior* distributions $p(\Sigma|Y)$ and $p_{NiWU,c}(\Sigma|Y)$ will be similar in a large sample even if the priors are not, because priors for identified parameters are known to vanish as the sample size increases. Whether proposal draws from $p_{NiWU,c}(\Sigma|Y)$ are suitable in an importance sampler for $p(\Sigma|Y)$ depends on the specific prior $p(B)$ and on the application. But the fact that $p(\Sigma|Y)$ and $p_{NiWU,c}(\Sigma|Y)$ converge to the same mass point as the sample size increases suggests an importance sampler for $p(\Sigma|Y)$ that builds on the convenient special case of $p_{NiWU,c}(\Sigma|Y)$.[2]

Consider now the relationship between the posterior $p(Q|Y,\Sigma)$ and the prior $p_{NiWU,c}(Q)$. $p(Q|Y,\Sigma)$ is not necessarily flat in the parameter space of orthogonal matrices. However, the (Haar) Uniform prior $p_{NiWU,c}(Q)$ by construction explores the full parameter space of orthogonal matrices. Hence, exploring $p(Q|Y,\Sigma)$ using $p_{NiWU,c}(Q)$ is generally feasible. The computational cost depends on the application. Before further inspecting the above intuition we generalize the exposition and discuss the posterior sampler in detail.

The case of sign restrictions Define $\mathrm{I}\{\boldsymbol{\pi}, B\}$ and $\mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\}$ two indicator functions taking value 1 if the structural parameters associated with (Π, B) or (Π, Σ, Q) satisfy the intended restrictions. This allows introducing restrictions on contemporaneous as well as future horizons of the impulse responses, on the stationarity of the model, and more. We work with the class of priors

$$\tilde{p}(\boldsymbol{\pi}, B) = \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B), \tag{19a}$$

$$p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}|B) \cdot p(B), \tag{19b}$$

$$p(\boldsymbol{\pi}|B) = \phi(\boldsymbol{\mu}_{\pi}, V_{\pi}), \tag{19c}$$

$$p(B). \tag{19d}$$

$p(\boldsymbol{\pi}, B)$ is restricted to imply a Normal density $p(\boldsymbol{\pi}|B)$. By contrast, $p(B)$ is freely selected by the researcher. The prior $p(\boldsymbol{\pi}, B)$ implies prior beliefs $p(\boldsymbol{\pi}, \Sigma, Q)$, which we omit here for convenience. As in the previous section, $p(\boldsymbol{\pi}, B)$ and $p(\boldsymbol{\pi}, \Sigma, Q)$ are effectively two sides of the same prior. In its more general independent form, the special case

[2] Section S.3.1.2 in the Supplementary Material shows that when the prior $p(\boldsymbol{\pi}, \Sigma)$ simplifies to $p(\boldsymbol{\pi}) \propto 1$, the mode of the posterior (12) is implicitly defined by

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

As the sample size increases, the mode approaches $\hat{\Sigma}_T$, which approaches the population moment $\Sigma_0 = \mathrm{E}(\boldsymbol{u}_t\boldsymbol{u}_t')$. Similarly, if proposal draws are generated from the posterior associated with a conjugate Normal inverse Wishart prior that is flat in $\boldsymbol{\pi}$, the mode of the proposal distribution equals $\frac{1}{d+T+k+1-m}S + \frac{T-m}{d+T+k+1-m}\hat{\Sigma}_T$, which also approaches $\hat{\Sigma}_T$ (Section S.2.1). This result is not new: 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 neighborhood of Σ_0 (Poirier (1998)).

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

of the NiWU prior uses

$$\tilde{p}_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q) = \mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q), \tag{20a}$$

$$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q) = p_{NiWU,i}(\boldsymbol{\pi}|\Sigma, Q) \cdot p_{NiWU,i}(Q|\Sigma) \cdot p_{NiWU,i}(\Sigma), \tag{20b}$$

$$p_{NiWU,i}(\boldsymbol{\pi}|\Sigma, Q) = p_{NiWU,i}(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_{\pi}, V_{\pi}), \tag{20c}$$

$$p_{NiWU,i}(Q|\Sigma) = p_{NiWU,i}(Q) = U_{O(k)} \propto 1, \tag{20d}$$

$$p_{NiWU,i}(\Sigma) = iW(S, d). \tag{20e}$$

$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q)$ implies a specific prior $p_{NiWU,i}(\boldsymbol{\pi}, B)$, which we omit here for convenience. $p_{NiWU,i}(\boldsymbol{\pi}, \Sigma, Q)$ and $p_{NiWU,i}(\boldsymbol{\pi}, B)$ are effectively two sides of the same prior, which is a special case of the prior associated with $p(\boldsymbol{\pi}, \Sigma, Q)$ and $p(\boldsymbol{\pi}, B)$. Contrary to $p(\boldsymbol{\pi}, B)$ and $p(\boldsymbol{\pi}, \Sigma, Q)$, the distributions $\tilde{p}(\boldsymbol{\pi}, B)$ and $\tilde{p}(\boldsymbol{\pi}, \Sigma, Q)$ only attach positive probability mass to the part of the parameter space in which the additional restrictions are satisfied, for instance, on the stationarity of the model or on the sign of the response at longer horizons.[3]

We now discuss the posterior distributions associated with the generic prior $p(\boldsymbol{\pi}, B)$. As we show in the Supplementary Material, the prior distribution $p(\boldsymbol{\pi}, B)$ from equation (19) leads to the posterior

$$\tilde{p}(\boldsymbol{\pi}, B|Y) \propto \mathrm{I}\{\boldsymbol{\pi}, B\} \cdot p(\boldsymbol{\pi}, B|Y), \tag{21}$$

$$p(\boldsymbol{\pi}, B|Y) = p(\boldsymbol{\pi}|Y, B) \cdot p(B|Y),$$

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

$$p(B|Y) \propto p(B) \cdot \left|\det(B)\right|^{-T} \cdot \left|\det(V_{\pi})\right|^{-\frac{1}{2}} \cdot \left|\det\left(V_{\pi}^{*}\right)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes (BB')^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\pi}^{*\prime} V_{\pi}^{*-1} \boldsymbol{\mu}_{\pi}^{*} + \boldsymbol{\mu}_{\pi}' V_{\pi}^{-1} \boldsymbol{\mu}_{\pi}\}}. \tag{22}$$

The terms $\tilde{\mathbf{y}}$, $\boldsymbol{\mu}_{\pi}^{*}$, and V_{π}^{*} are defined in Section S.3.1.1 of the Supplementary Material, where $(\boldsymbol{\mu}_{\pi}^{*}, V_{\pi}^{*})$ are a function of B, even though not made explicit by the notation. This is the joint distribution of interest that we aim to draw from. However, $p(B|Y)$ is not the density function of a known family of probability distributions. Conditioning on B, draws from $p(\boldsymbol{\pi}|Y, B)$ are straightforward, while mapping draws from $p(\boldsymbol{\pi}, B|Y)$ into $\tilde{p}(\boldsymbol{\pi}, B|Y)$ requires an accept/reject procedure. The posterior (21)–(22) implies a posterior on $(\boldsymbol{\pi}, \Sigma, Q)$,

$$\tilde{p}(\boldsymbol{\pi}, \Sigma, Q|Y) = \mathrm{I}\{\boldsymbol{\pi}, \Sigma, Q\} \cdot p(\boldsymbol{\pi}, \Sigma, Q|Y), \tag{23}$$

$$p(\boldsymbol{\pi}, \Sigma, Q|Y) = p(\boldsymbol{\pi}|Y, \Sigma, Q) \cdot p(Q|Y, \Sigma) \cdot p(\Sigma|Y),$$

[3]Note that, in general, $\tilde{p}(\boldsymbol{\pi}|B) = \frac{\tilde{p}(\boldsymbol{\pi}, B)}{\int_{\boldsymbol{\pi}} \tilde{p}(\boldsymbol{\pi}, B)\, d\boldsymbol{\pi}} \neq p(\boldsymbol{\pi}|B)$, $\tilde{p}(B) = \int_{\boldsymbol{\pi}} \tilde{p}(\boldsymbol{\pi}, B)\, d\boldsymbol{\pi} \neq p(B)$, and $\tilde{q}(Q) = \int_{\boldsymbol{\pi}} \int_{\Sigma} \tilde{q}(\boldsymbol{\pi}, \Sigma, Q)\, d\Sigma\, d\boldsymbol{\pi} \neq p_{NiWU,i}(Q)$, where the difference depends on terms that depend on B (for the first 2 inequalities) and on Q (for the last one). It is important to work with the distributions $p(\cdot)$ rather than $\tilde{p}(\cdot)$ because the algorithm requires the analytical evaluation of the implied prior and posterior distributions, which in turns allows for reweighting in the importance sampler. We refer to Section S.2 in the Supplementary Material (Bruns and Piffer (2023)) for a further discussion.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$
\begin{aligned}
p(\boldsymbol{\pi}|Y,\Sigma,Q) &= \phi\big(\boldsymbol{\mu}_{\pi}^{*},V_{\pi}^{*}\big),\\
p(Q|Y,\Sigma) &\propto \frac{p_B\big(h(\Sigma)Q\big)}{\int_{O(k)} p_B\big(h(\Sigma)Q\big)\,dQ},\\
p(\Sigma|Y) &\propto \big|\det(\Sigma)\big|^{-\frac{T+1}{2}}\cdot\big|\det(V_{\pi})\big|^{-\frac{1}{2}}\cdot\big|\det\big(V_{\pi}^{*}\big)\big|^{\frac{1}{2}}\\
&\quad\cdot e^{-\frac{1}{2}\{\tilde{y}'(I_T\otimes\Sigma^{-1})\tilde{y}-\boldsymbol{\mu}_{\pi}^{*\prime}V_{\pi}^{*-1}\boldsymbol{\mu}_{\pi}^{*}+\boldsymbol{\mu}_{\pi}'V_{\pi}^{-1}\boldsymbol{\mu}_{\pi}\}}\cdot\int_{O(k)} p_B\big(h(\Sigma)Q\big)\,dQ, \qquad (24)
\end{aligned}
$$

with $p_B(\cdot)=p(B)$; see the Supplementary Material for the derivations. $p(B|Y)$ and $p(\Sigma,Q|Y)$ are effectively two sides of the same posterior. One can draw from $p(B|Y)$ indirectly by drawing from $p(\Sigma|Y)$, $p(Q|Y,\Sigma)$, and then mapping draws from (Σ,Q) into B.

The posterior from equations (23)–(24) admits the posterior associated with the independent NiWU prior as a special case. When $p(B)$ is restricted to the special case implied by the independent NiWU prior, the posterior on $(\boldsymbol{\pi},\Sigma,Q)$ reduces to

$$
\begin{aligned}
\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y) &= \mathrm{I}\{\boldsymbol{\pi},\Sigma,Q\}\cdot p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y),\\
p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q|Y) &= p_{NiWU,i}(\boldsymbol{\pi}|Y,\Sigma,Q)\cdot p_{NiWU,i}(Q|Y,\Sigma)\cdot p(\Sigma|Y),\\
p_{NiWU,i}(\boldsymbol{\pi}|Y,\Sigma,Q) &= \phi\big(\boldsymbol{\mu}_{\pi,i}^{*},V_{\pi}^{*}\big), \qquad (25)\\
p_{NiWU,i}(Q|Y,\Sigma) &= p_{NiWU,i}(Q)=U_{O(k)}\propto 1,\\
p_{NiWU,i}(\Sigma|Y,\boldsymbol{\pi}) &= iW\big(S^{*},d^{*}\big). \qquad (26)
\end{aligned}
$$

The posterior moments $\boldsymbol{\mu}_{\pi,i}^{*}$, V_{π}^{*}, S^{*}, d^{*} are defined in Section S.2 of the Supplementary Material, where S^{*} depends on $\boldsymbol{\pi}$ and $\boldsymbol{\mu}_{\pi,i}^{*}$, V_{π}^{*} depend on Σ. In general, draws from $p_{NiWU,i}(\Sigma|Y)$ require running a Gibbs sampler on (25) and (26). Section S.2 of the Supplementary Material shows that $p_{NiWU,i}(\Sigma|Y)$ is an inverse Wishart distribution in the special case in which $p_{NiWU,i}(\boldsymbol{\pi}|\Sigma,Q)\propto 1$, further simplifying the analysis. We omit the posterior $p_{NiWU,i}(\boldsymbol{\pi},B|Y)$ for convenience. Contrary to $p_{NiWU,i}(\boldsymbol{\pi},B)$ and $p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q)$, the distributions $\tilde{p}_{NiWU,i}(\boldsymbol{\pi},B)$ and $\tilde{p}_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q)$ only attach positive probability mass to the part of the parameter space in which the sign restrictions are satisfied. Since Q does not enter the likelihood function and since $p_{NiWU,i}(\boldsymbol{\pi},\Sigma,Q)$ does not necessarily satisfy the sign restrictions, it holds that $p_{NiWU,i}(Q|Y,\Sigma)=p_{NiWU,i}(Q)$.

2.3 *The posterior sampler*

We aim to draw from $p(B|Y)$ from equation (22) using an importance sampler that draws from $p(\Sigma|Y)$, $p(Q|Y,\Sigma)$ using proposal draws from $p_{NiWU,i}(\Sigma|Y)$, $p_{NiWU,i}(Q|Y,\Sigma)$. At a general level, our importance sampler can be summarized in the following steps:

Stage A: draws from $p(\Sigma|Y)$:

(i) Generate proposal draws from $p_{NiWU,i}(\Sigma|Y)$;

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(ii) evaluate the weights

$$w^{\text{stage A}} = \frac{p(\Sigma|Y)}{p_{NiWU,i}(\Sigma|Y)},$$

(iii) if the effective sample size associated with $w^{\text{stage A}}$ is satisfactory, reweight the draws from $p_{NiWU,i}(\Sigma|Y)$ into draws from $p(\Sigma|Y)$;

Stage B: from $p(\Sigma|Y)$ *to* $p(B|Y)$:

(iv) for each draw from $p(\Sigma|Y)$ generate proposal draws from $p_{NiWU,i}(Q|Y,\Sigma)$;

(v) evaluate weights

$$w^{\text{stage B}} = \frac{p(Q|Y,\Sigma)}{p_{NiWU,i}(Q|Y,\Sigma)},$$

(vi) if the effective sample size associated with $w^{\text{stage B}}$ is satisfactory, reweight the draws from $p_{NiWU,i}(Q|Y,\Sigma)$ into draws from $p(Q|Y,\Sigma)$;

(vii) map the draws from $p(\Sigma|Y)$, $p(Q|Y,\Sigma)$ into a draw from $p(B|Y)$;

(viii) draw from $p(\boldsymbol{\pi}|Y,B)$.

To become operational, the above algorithm requires specifying the distributions generating proposal draws in stages A and B.

Stage A of the algorithm requires proposal draws from $p_{NiWU,i}(\Sigma|Y)$, that is, the marginal posterior distribution associated with the NiWU prior. One can in principle adjust the specification of the NiWU prior behind $p_{NiWU,i}(\Sigma|Y)$ to improve the behavior of $w^{\text{stage A}}$. In doing so, one can select whether the NiWU prior is specified in the independent form (as in equation (20)) or in its conjugate form (as in the previous illustration, equations (13)–(16)), adjust the hyperparameters of the Normal and the inverse Wishart distributions, or employ improper priors. The only requirements are to be able to evaluate $p_{NiWU,i}(\Sigma|Y)$ and to draw from it, either using a Gibbs sampler (in the independent case) or using direct sampling (in the conjugate case or in the independent case under a flat prior on $\boldsymbol{\pi}$). In addition, the specification of the NiWU prior behind the posterior proposal function $p_{NiWU,i}(\Sigma|Y)$ can change depending on the specific prior $p(\boldsymbol{\pi},B)$ used in the analysis. We will focus our analysis on two commonly used special cases, which are

$$\textit{Case 1}: \quad p(\boldsymbol{\pi},B) = p(\boldsymbol{\pi})\cdot p(B), \qquad p(\boldsymbol{\pi}) \propto 1, \tag{27}$$

$$\textit{Case 2}: \quad p(\boldsymbol{\pi},B) = p(\boldsymbol{\pi})\cdot p(B), \qquad p(\boldsymbol{\pi}) = \phi(\boldsymbol{\mu}_{\pi}, V_{\pi}). \tag{28}$$

In the Appendix of the paper, we show that it is possible to specify the NiWU prior behind $p_{NiWU,i}(\Sigma|Y)$ differently for *Case 1* and *Case 2* such that, for both cases, $w^{\text{stage A}} = \frac{p(\Sigma|Y)}{p_{NiWU,i}(\Sigma|Y)} = \int_{O(k)} p_B(h(\Sigma)Q)\,dQ$, which can be computed numerically. Note that the computation of $w^{\text{stage A}}$ has simplified since several terms in $p(\Sigma|Y)$, $p_{NiWU,i}(\Sigma|Y)$ cancel out. This considerably simplifies the computation of the weights in Stage A of the sampler and increases the corresponding effective sample size. In both cases, the NiWU prior for the proposal draws is specified in its independent form.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Stage B of the algorithm maps draws for $p(\Sigma|Y)$ into B using draws for Q. However, draws from Q must reflect $p(Q|Y,\Sigma)$, not $p_{NiWU,i}(Q|Y,\Sigma)$. Since $p_{NiWU,i}(Q|Y,\Sigma)$ coincides with the uniform distribution in the space of orthogonal matrices, it is potentially a suitable proposal distribution for the target distribution $p(Q|Y,\Sigma)$. In Section S.3 of the Supplementary Material, we show that, conditioning on some $\Sigma=\bar{\Sigma}$ and for two generic matrices Q_1, Q_2 associated with $B_1=h(\bar{\Sigma})Q_1$ and $B_2=h(\bar{\Sigma})Q_2$, it holds that $\frac{p(Q_2|Y,\bar{\Sigma})}{p(Q_1|Y,\bar{\Sigma})}=\frac{p(B_2)}{p(B_1)}$, that is, the ratio of the *prior* distribution on B. For each draw from $p(\Sigma|Y)$, one can use an importance sampler that draws Q from the Haar measure and reweights using weights that only require evaluating $p(B)$. The weights for Stage B are also necessary to evaluate the integral in equation (24) in Stage A (see Section S.3 in the Supplementary Material). Last, an accept/reject adjustment can be added to ensure that only draws satisfying the restrictions are stored, accounting also for possible sign restrictions on future horizons of the impulse response.

All in all, the algorithm to draw from $\tilde{p}(\boldsymbol{\pi}, B|Y)$ is

Algorithm: Sign restrictions
Stage A: draws from $p(\Sigma|Y)$:

(i) Generate m_1 proposal draws $\{\Sigma^{(d)}\}_{d=1}^{m_1}$ representative of $p_{NiWU,i}(\Sigma|Y)$ through either direct sampling or Gibbs sampling, using

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

through *Algorithms A* or *C* discussed in Section S.2 of the Supplementary Material, with $(S^*, d^*, V_{\pi}^*, \boldsymbol{\mu}_{\pi}^*)$ specified in Table A.1 in the Appendix of the paper, with S^*, d^* differing depending on *Case 1* or *Case 2* and S^* a function of $\boldsymbol{\pi}$ only under *Case 2*;

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

(ii,a) extract one matrix Q_c using the algorithm by Rubio-Ramirez, Waggoner, and Zha (2010);

(ii,b) compute the candidate matrix $B_c=h(\Sigma^{(d)})Q_c$. If B_c satisfies the sign restrictions on B, store

$$\big(Q_i\big(\Sigma^{(d)}\big),B_i\big(\Sigma^{(d)}\big),w_i\big(\Sigma^{(d)}\big)^{\text{stage B}}\big)=\big(Q_c,B_c,w_{dc}^{\text{stage B}}\big),$$

with $w_{dc}^{\text{stage B}}$ equal to $p(B)$ evaluated at B_c, otherwise move back to step (ii,a);

(ii,c) repeat steps (ii,a)–(ii,b) until m_2 draws $\{Q_i(\Sigma^{(d)}), B_i(\Sigma^{(d)}), w_i(\Sigma^{(d)})^{\text{stage B}}\}_{i=1}^{m_2}$ are stored. Store the number of attempts $m_3(\Sigma^{(d)})$ required to generate m_2 successful draws. Compute the absolute effective sample size

$$\text{ESS}_{\text{abs}}^{\text{B}}\big(\Sigma^{(d)}\big)=\left(\sum_i\big(w_i\big(\Sigma^{(d)}\big)^{\text{stage B}}\big)^2\right)^{-1};$$

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

(ii,d) compute

$$w\big(\Sigma^{(d)}\big)^{\text{stage A}} = \frac{p\big(\Sigma^{(d)}|Y\big)}{p_{NiWU,i}\big(\Sigma^{(d)}|Y\big)} = \int_{O(k)} p_B\big(h\big(\Sigma^{(d)}\big)Q\big)\,dQ \approx \frac{\sum_{i=1}^{m_2} w_i\big(\Sigma^{(d)}\big)^{\text{stage B}}}{m_3\big(\Sigma^{(d)}\big)},$$

and compute the relative effective sample size

$$\text{ESS}^{\text{A}}_{\text{rel}} = \left(\sum_d \left(w\big(\Sigma^{(d)}\big)^{\text{stage A}} / \sum_d \big(w\big(\Sigma^{(d)}\big)^{\text{stage A}}\big)\right)^2\right)^{-1}; \tag{29}$$

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

Stage B: from $p(\Sigma|Y)$ to $p(\boldsymbol{\pi}, B|Y)$:

(iv) randomly select a candidate draw Σ_c from $\{\Sigma^{(d)}\}_{d=1}^{m_4}$ generated in step (iii);

(v) randomly select one matrix B_c out of the set $\{B_i(\Sigma_c)\}_{i=1}^{m_2}$ stored from step (ii) using weights $\{w_i(\Sigma_c)^{\text{stage B}}\}_{i=1}^{m_2}$;

(vi) compute $(\boldsymbol{\mu}^*_\pi, V^*_\pi)$ associated with B_c and generate one draw $\boldsymbol{\pi}_c$ from $p(\boldsymbol{\pi}|Y, B) = \phi(\boldsymbol{\mu}^*_\pi, V^*_\pi)$;

(vii) if $\text{I}\{\boldsymbol{\pi}_c, B_c\} = 1$, store $(\boldsymbol{\pi}_c, B_c)$, otherwise move back to step (iv);

(viii) repeat until m_5 draws are successfully generated.

We design the algorithm in two stages to stress that Stage B should be carried out only if the $1 + m_1$ Effective Sample Sizes $\text{ESS}^{\text{A}}_{\text{rel}}$, $\{\text{ESS}^{\text{B}}_{\text{abs}}(\Sigma^{(d)})\}_{d=1}^{m_1}$ are sufficiently high. Importance sampler diagnostics can then be used to assess the performance of the sampler. Figure S.1 in the Supplementary Material reports the statistics that document the performance of our sampler in the applications of the paper, and Section S.3.2 discusses a number of additional diagnostic tests.

2.3.1 *A discussion of the sampler* The literature has explored how to sample numerically from the posterior of Bayesian SVAR models when the prior distribution differs from the NiWU prior. Kociecki, Rubaszek, and Ca'Zorzi (2012) show that for recursive Bayesian SVARs, the posterior distribution can be decomposed in a way that allows for direct sampling. However, this is not the case for the more general non-recursive (or even unrestricted) SVAR, for which a numerical procedure is required to sample from the posterior. This latter point has been known at least since Sims and Zha (1998) and is due to the Jacobian term involving the determinant of the contemporaneous structural parameters ($|\det(B)|^{-T}$ in equation (22)). This term drops out under the recursive structure with normalized diagonal entries, as in Kociecki, Rubaszek, and Ca'Zorzi (2012), but it does not in a more general framework.

Importance samplers have been tried in the literature to tackle the above problem. Early contributions take as proposal distribution a Normal or t-distribution centered at

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

the mode of the nonstandard posterior; see Sims and Zha (1998), Leeper et al. (1996), and Zha (1999). However, Waggoner and Zha (2003) show that this type of importance samplers work inefficiently in SVAR models. We differ from the above specifications of an importance sampler by setting the proposal distribution equal to the posterior distribution associated with the special case of an independent NiWU prior. Our algorithm effectively resamples the posterior draws from the NiWU approach and makes them representative of the posterior distribution associated with the prior beliefs $p(\boldsymbol{\pi}, B)$ from our approach. By construction, when $p(B)$ is restricted to the value associated with the independent NiWU prior, all weights in the importance sampler equal 1, leading to the special case of the posterior sampler for the NiWU approach.

We found that a similar idea is used in Canova and Pérez Forero (2015). They use a Metropolis–Hastings sampler rather than an importance sampler, but they too generate proposal draws by exploiting the information of a posterior distribution. While we use the exact posterior distribution associated with the independent NiWU prior, they use an approximate posterior that arises in the artificial special case in which the Jacobian term is left out from the joint posterior distribution. Similar to us, they do not approximate the target distribution with a Normal or t-distribution centered at the mode of the target distribution. An alternative to the approach by Canova and Pérez Forero (2015) is to use a standard random walk Metropolis–Hastings (Baumeister and Hamilton (2015, 2018, 2019)).

Our approach is also related to Algorithm 3 in Arias, Rubio-Ramírez, and Waggoner (2018). The importance sampler in Arias, Rubio-Ramírez, and Waggoner (2018) is only required for the case in which zero restrictions are introduced. When only sign restrictions are used, their approach converges back to the NiWU prior, from which, instead, we aim to depart in the case of sign restrictions. We also use an independent NiWU prior, while they use a conjugate NiWU prior. While the case of both sign and zero restrictions has attracted interest (see, for instance, Binning (2013) and Kociecki (2017)), our paper concentrates on the case of sign restrictions, and moves beyond the NiWU prior. Another difference is that the weights in Arias, Rubio-Ramírez, and Waggoner (2018) are computed directly on the posterior densities on the structural parameters (in line with the importance sampler by Sims and Zha (1998), Leeper et al. (1996), and Zha (1999)), while we prefer building the sampler in two steps.

Contrary to all the contributions mentioned above, we also use the more general Dynamic Striated Metropolis–Hastings (DSMH) algorithm by Waggoner, Wu, and Zha (2016) to study if our algorithm correctly samples the posterior distribution of interest. The DSMH algorithm is much more demanding both in terms of coding and in terms of computational time. However, it is designed to handle potentially irregularly shaped posterior distributions and a large number of parameters. We use the DSMH algorithm to approximate the true posterior distribution in each application, and use it as a benchmark to assess the performance of our sampler. We refer to Waggoner, Wu, and Zha (2016) for a detailed discussion of the DSMH algorithm and to Table S.1 in the Supplementary Material for the values of the tuning parameters we use in our applications. In our applications, our algorithm takes around 6 minutes for the bivariate applications

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

and 30 minutes for the five-variate applications. The DSMH algorithm took 2 hours for the bivariate applications and up to 8 hours in the five-variate applications.[4]

2.4 *Proposing one possible prior* $p(B)$

The previous section develops an approach that is general in the prior distribution $p(B)$ on contemporaneous impulse responses. We conclude the section on the methodology by discussing one possible prior specification for $p(B)$. This can be used in applications in which sign restrictions are either introduced only on contemporaneous impulse responses (B), or in applications that combine these restrictions with prior beliefs on the contemporaneous relation among variables (B^{-1}), following Baumeister and Hamilton (2018).

Specifying prior beliefs $p(B)$ is challenging. While a researcher might have beliefs about the sign of the entries of B, the actual prior $p(B)$ will effectively model beliefs also about the magnitude of the responses, which can be harder to express. 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. If, instead, one does have beliefs on both the sign and the magnitudes of B, it is not yet clear which type of prior is more suitable to model such beliefs.

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 using a training sample, setting s_i 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 (see, e.g., the discussion in Canova (2007) and Kilian and Lütkepohl (2017)). Bayesian shrinkage is then introduced through a set of hyperparameters that shrink the elements in $\boldsymbol{\pi}$ toward 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

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

with Σ_{ii} the ith element of the diagonal of Σ.[5] Given $\{\gamma_i\}_{i=1}^{k}$, we set $p(B) = \prod_i \prod_j p(b_{ij}|\gamma_i, \psi_1, \psi_2)$, with $p(b_{ij}|\gamma_i, \psi_1, \psi_2) = \phi(\mu_{ij}, \sigma_{ij}^2)$. 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 $(-\psi_2\gamma_i,\ \psi_2\gamma_i)$. If b_{ij} is restricted to be positive (or negative),

[4]All codes are run on Matlab on a computer with an Intel i7-7700K 4.2 GHz Quad Core processor and 64 GB RAM.

[5]Given $\Sigma = BB'$, the equations corresponding to the diagonal elements of Σ are $\Sigma_{ii} = b_{i1}^2 + b_{i2}^2 + \cdots + b_{ik}^2$. Since Σ_{ii} is nonnegative 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).

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

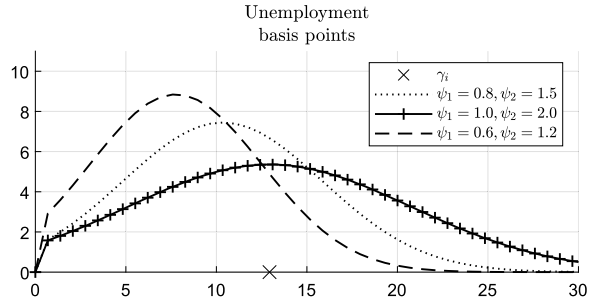

FIGURE 1. Illustration of our prior $\tilde{p}(B)$. *Note*: Three examples of our prior using different hyperparameters ψ_1, ψ_2.

start from a Normal distribution with $\mu_{ij} = \psi_1 \gamma_i$ (or $\mu_{ij} = -\psi_1 \gamma_i$) and calibrate the variance such that the truncated Normal has 95% prior mass in the space $(0,\ \psi_2 \gamma_i)$ (or $(-\psi_2 \gamma_i,\ 0)$). Our prior can be viewed as a generalization of the generalized Normal prior in Arias, Rubio-Ramírez, and Waggoner (2018); see Section S.5 of the Supplementary Material.

The advantage of our prior is that 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 ψ_2. ψ_1 controls for the first moment of the prior, ψ_2 controls for the second moment. This grants flexibility not only on the sign restrictions introduced, but also on the expected magnitude of the response. Figure 1 shows, as an illustration, three separate priors used in the application from Section 4. It reports the prior on the impact effect of a one-standard-deviation monetary shock on the unemployment rate. The hyperparameters (ψ_1, ψ_2) give the researcher flexibility on the prior introduced, taking into account the reasonable scale γ_i.[6]

3. MONTE CARLO SIMULATION

In this section, we provide further clarification on how our sampler works. We inspect graphically the proposal and the target distributions of each stage of our importance sampler. We study how the posterior distribution sampled with our sampler approximates the true posterior distribution, which is proxied using the DSMH algorithm.

[6]Other papers in the literature introduce priors directly on the impulse responses rather than the VAR parameters. Plagborg-Møller (2019) uses DSGE models to inform a joint prior over the structural moving average representation of the model. Kociecki (2010) proposes a set of joint Normal priors for the impulse responses at several horizons, where the prior mean and variance are informed either by previous results in the empirical literature or by DSGE models. Ferreira, Miranda-Agrippino, and Ricco (2023) develop a Bayesian local projection approach using priors centered around VAR estimates from a training sample. Baumeister and Hamilton (2018) partly use prior beliefs on impulse responses modelled with asymmetric t-distributions. Canova, Kociecki, and Piffer (2023) express beliefs about the shape of the impulse responses across multiple horizons.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

3.1 *The setup of the exercise*

We build our illustrative simulation exercise on the bivariate model by Baumeister and Hamilton (2015), who identify the effects of labor demand and supply shocks in the US. We use ordinary least squares estimates $(\hat{\boldsymbol{\pi}}, \hat{\Sigma})$ in a reduced form VAR model to generate the pseudo dataset Y^d. We then specify the prior distribution $p(\boldsymbol{\pi}, B)$ and explore the posterior distribution $p(\boldsymbol{\pi}, B|Y^d)$ using two alternative algorithms: the DSMH algorithm and our algorithm. We then compare the two posterior distributions graphically, taking the distribution sampled with the DSMH algorithm as the true posterior distribution associated with Y^d. We first use a single pseudo dataset Y^d to illustrate most of the intuitions behind our sampler, and then repeat the analysis over multiple simulated data $\{Y^d\}_{d=1}^{D}$.[7]

The estimates $(\hat{\boldsymbol{\pi}}, \hat{\Sigma})$ are obtained from the dataset and model specification by Baumeister and Hamilton (2015), who use data on the growth rates of the US real labor compensation and of total employment. The model includes a constant and 8 lags, and covers the period 1970Q1 through 2014Q4. For each replication, we generate 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 use simulated datasets of different size to study the performance of our sampler also in relatively small samples.

We specify the joint prior distribution as follows. We use an independent Normal prior on $\boldsymbol{\pi}$, using conventional values for the hyperparameters of the variance from Canova (2007) (*Case 2* from equation (28)). We specify $p(B)$ as from Section 2.4, setting $\psi_1 = 0.8$ and $\psi_2 = 1.5$, using $\gamma_i = \hat{\Sigma}^{0.5}_{ii,\text{training}}$ with $\hat{\Sigma}_{\text{training}}$ the estimate of Σ on the training sample. We then introduce sign restrictions on the contemporaneous impulse responses. We identify labor demand and supply shocks as the structural shocks that move wages and employment in the same and in the opposite direction, respectively. All estimated models include a constant term and 8 lags, in accordance with the data generating process.

3.2 *Simulation results*

A key part of our sampler builds on the intuition that different prior beliefs on Σ lead to potentially very similar posterior distributions on Σ, provided that the sample size

[7]Alternatively, one could specify the data generating process at the structural level rather than the reduced form level, and compare the estimated impulse responses with the underlying true impulse responses. This alternative exercise starts from a true matrix $\hat{B} = h(\hat{\Sigma})\tilde{Q}$ and generates data using covariance matrix $\hat{B}\hat{B}' = \hat{\Sigma}$. However, while the true underlying impulse responses change in $\tilde{Q}$, any alternative orthogonal matrix $Q \neq \tilde{Q}$ implies the same dataset; hence, it leads to the same estimated impulse responses. This makes the comparison of the true and the estimated impulse responses arbitrary due to the flexibility in the selection of $\tilde{Q}$. The aim of our exercise is to document the performance of our sampler, which only requires generating data, and not taking a stand on the true structure behind the data.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

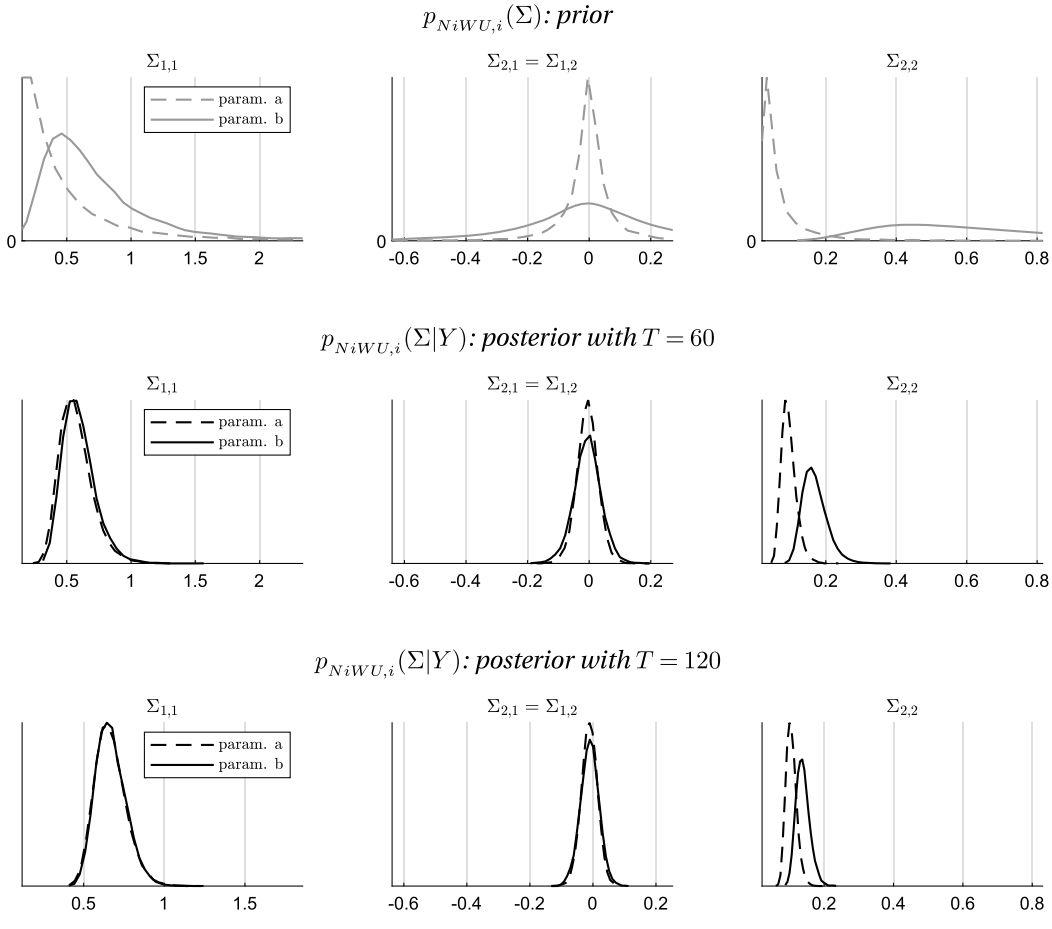

FIGURE 2. Prior beliefs and identified parameters.

is not too small. Figure 2 illustrates this point graphically for two parametrizations of an inverse Wishart distribution prior on Σ. *Parametrization a* corresponds to the one suggested by Kadiyala and Karlsson (1997), while *parametrization b* is an arbitrary parametrization that helps the illustration.[8] The posterior distributions are sampled using *Algorithm B* from Section S.2.1 of the Supplementary Material, which is associated with the independent NiWU prior $p_{NiWU,i}(\boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(\boldsymbol{\pi}) \cdot p_{NiWU,i}(\Sigma)$, $p_{NiWU,i}(\boldsymbol{\pi}) \propto 1$, $p_{NiWU,i}(\Sigma) = iW(S, d)$. The figure shows the results associated with a single pseudo dataset, although the results are unchanged when repeating the exercise.

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 (top panel). The posteriors, instead, are very close to each other already for $T = 60$ (middle panel), a dif-

[8]The inverse Wishart distribution $p_{NiWU,i}(\Sigma) \propto |\det(\Sigma)|^{-\frac{d+k+1}{2}} \cdot e^{-\frac{1}{2}\,\mathrm{trace}[\Sigma^{-1}S]}$ requires specifying the hyperparameters (S, d). Following Kadiyala and Karlsson (1997), *parametrization a* sets $d = k + 2$ and $S = (d - k - 1)\,\mathrm{diag}([s_1, \ldots, s_k])$, with s_i the standard deviation of the error term in univariate AR(p) processes estimated on the training sample. By contrast, *parametrization b* sets $S = 4 \cdot I_k$ and $d = 4 \cdot k$.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

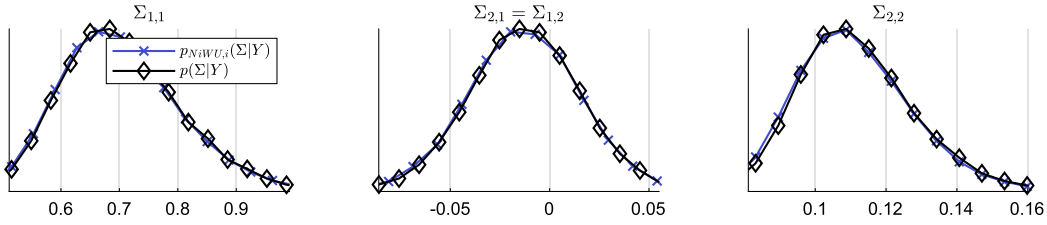

FIGURE 3. Stage A of our sampler: $p_{NiWU,i}(\Sigma|Y)$ *and* $p(\Sigma|Y)$.

ference that becomes even more negligible for $T = 120$ (bottom panel). This confirms the well-known result that, for identified parameters, differences in prior beliefs vanish asymptotically, provided the prior gives non-zero mass in the area that is relevant asymptotically. This suggests exploring the generic distribution $p(\Sigma|Y)$ using proposal draws from the convenient distribution $p_{NiWU,i}(\Sigma|Y)$.

Figure 3 builds on the above intuition and reports the update from stage A of our sampler for $T = 120$. A total of $m_1 = 25{,}000$ proposal draws for stage A are first generated from the posterior distribution $p_{NiWU,i}(\Sigma|Y)$ associated with the proposal prior $p_{NiWU,i}(\boldsymbol{\pi}) \propto 1$, $p_{NiWU,i}(\Sigma) = |\det(\Sigma)|^{-\frac{1}{2}}$. The proposal distribution $p_{NiWU,i}(\Sigma|Y)$ is already very close to the target distribution $p(\Sigma|Y)$, a finding further supported by the relative effective sample size for stage A ($\text{ESS}^{\text{A}}_{\text{rel}}$) being 0.9849. Table 1 generalizes the analysis by progressively increasing the sample size, and then repeating the analysis 50 times. As the table shows, $\text{ESS}^{\text{A}}_{\text{rel}}$ increases as the sample size increases, which is to be expected.

We now discuss stage B of the sampler. In stage B, posterior draws from $p(\Sigma|Y)$ are converted into draws from $p(B|Y)$ using draws from $p(Q|Y,\Sigma) = p(Q|\Sigma)$, before introducing sign restrictions. Since $p(Q|Y,\Sigma)$ cannot be drawn from directly, for each draw of Σ we draw multiple times from the special case $p_{NiWU,i}(Q|Y,\Sigma) = p_{NiWU,i}(Q|\Sigma)$ and reweigh in an importance sampler. The bivariate application used in this illustration helps clarify this step. When Q is of dimension 2×2, distributions $p(Q|\Sigma)$, $p_{NiWU,i}(Q|\Sigma)$ can be shown graphically as the distribution on the rotation angle θ corresponding to the Givens transformations matrix Q (see, e.g., Fry and Pagan (2011) and Baumeister and Hamilton (2015)). Figure 4 illustrates this point conditioning on the MLE estimation for Σ. $p_{NiWU,i}(Q)$ refers to the distribution on θ associated with the unconditional Haar measure, which is known to deliver $\theta \sim U[-\pi/2, \pi/2]$. $\tilde{p}_{NiWU,i}(Q|\hat{\Sigma})$ refers to the

TABLE 1. Relative effective sample size from Stage A.

	$T = 60$	$T = 120$	$T = 240$	$T = 480$
A single dataset	0.9607	0.9902	0.9790	0.9937
Minimum across 50 replications	0.8920	0.9727	0.9870	0.9917

Note: The relative effective sample size is defined in equation (29). See Table A.1 in the Appendix of the paper for the proposal distributions $p_{NiWU,i}(\Sigma|Y)$ used in the sampler, and Table S.1 in the Supplementary Material for how we set the tuning parameters.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

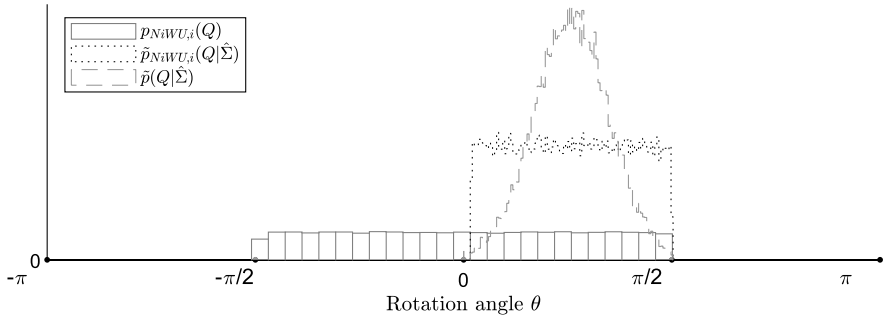

FIGURE 4. Stage B of our sampler: $p_{NiWU,i}(Q|Y,\Sigma)$ and $p(Q|Y,\Sigma)$. *Note*: θ computed as $\theta = \text{atan}(\frac{Q_{21}^{(d)}}{Q_{11}^{(d)}})$, with $Q_{ij}^{(d)}$ the (i,j) entry of $Q_{11}^{(d)}$. See Algorithm F in Section S.3 in the Supplementary Material for the details of the exercise.

distribution on θ associated with Haar distribution conditioning on $\hat{\Sigma}$ and the sign restrictions. The distribution $\tilde{p}(Q|\hat{\Sigma})$ refers to the distribution on θ associated with our prior $p(B)$, conditioning on $\hat{\Sigma}$ and the sign restrictions. $\tilde{p}(Q|\hat{\Sigma})$ is not flat in Q. However, since $\tilde{p}_{NiWU,i}(Q|\hat{\Sigma})$ is flat, it gives positive probability mass to the part of the parameter space that is relevant for $\tilde{p}(Q|\hat{\Sigma})$. The relative ESS associated with Figure 4 is 0.7040, and builds on 500,000 draws of Q, effectively using 35,202 draws.

Stage B of our sampler applies the logic behind Figure 4 to all posterior draws from $p(\Sigma|Y)$. For each draw $\Sigma^{(d)}$, we store and reweight $m_2 = 100$ draws of Q. The distribution of $\{\text{ESS}_{\text{abs}}^{B}(\Sigma^{(d)})\}_{d=1}$ is concentrated around 70 (see Figure S.1 in the Supplementary Material). The multiple importance samplers on Stage B then lead to the update shown in Figure 5. The comparison to the distribution sampled with the DSMH algorithm confirms that our algorithm samples the posterior distribution $\tilde{p}(B|Y)$ correctly. We found this result to hold for every generated dataset. On average, our algorithm runs in 6 minutes, while the DSMH runs in around 2 hours. Figure S.1, Figure S.3, and Table S.6 in the Supplementary Material report some informative statistics on the performance of our sampler and the results from standard convergence diagnostics.

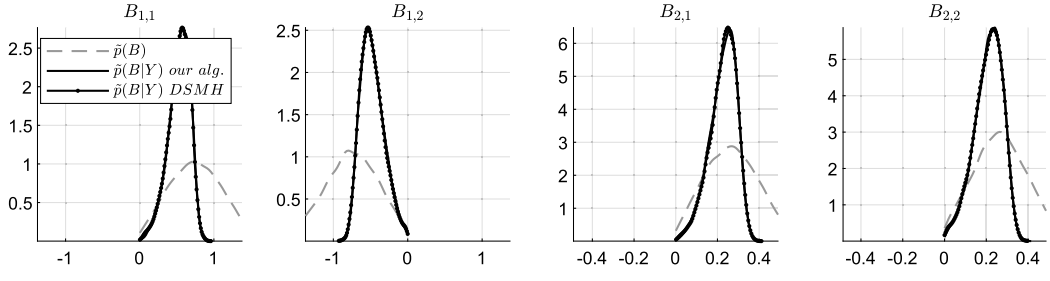

FIGURE 5. Stage B of our sampler: $\tilde{p}(B)$ and $\tilde{p}(B|Y)$.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

4. Application to monetary policy shocks

In this section, we apply our methodology to real data. We investigate how US monetary policy shocks affect the US economy. To do so, we use a SVAR model composed of five variables, following the model by Caldara and Herbst (2019) closely.

4.1 *Model, prior specification, and posterior sampling*

The model includes 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, and the Baa spread. A constant and 12 lags of the endogenous variables are included. The model uses the period 1990M1 through 1993M12 as training sample and the period 1994M1 through 2007M7 as estimation sample.

We set the prior in accordance with *Case 1* from equation (27), that is, $p(\boldsymbol{\pi}) \propto 1$. We then add dummy variables to model the Minnesota Prior on $\boldsymbol{\pi}$ as discussed in Del Negro and Schorfheide (2011). We follow our approach and specify the prior directly 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 3 months after the shock. We model the prior $p(B)$ as discussed in Section 2.4. We set $\gamma_i = \hat{\Sigma}^{0.5}_{ii,\text{training}}$ and use $\psi_1 = 0.8$ and $\psi_2 = 1.5$ in the baseline specification. 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 nonrepetition 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 using both the dynamic striated Metropolis–Hastings algorithm and our algorithm, always checking graphically that the results for $\tilde{p}(B|Y)$ are virtually identical. The sampler delivers a relative effective sample size for Stage A of 0.8013, and the distribution of $\{\text{ESS}^B_{\text{abs}}(\Sigma^{(d)})\}_{d=1}$ is concentrated around 40 draws. The computational time of our sampler is 30 minutes, while the DSMH algorithm took approximately 8 hours. Figure S.1, Figure S.3, and Table S.6 in the Supplementary Material report selected statistics on the performance of our sampler and the results from convergence diagnostics.

4.2 *Results*

Figure 6 shows the impulse responses to a contractionary monetary policy shock, which has been normalized to increase the federal funds rate by 25 basis points. We report the pointwise median and 68% credible band up to 4 years after the shock. The exogenous increase in the policy rate generates a hump-shaped response of industrial production and unemployment. Industrial production drops by a maximum of 1% 16 months from the shock, while the unemployment rate increases by up to 17 basis points 15 months from the shock. Both variables then progressively revert back. Prices decrease, with a

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

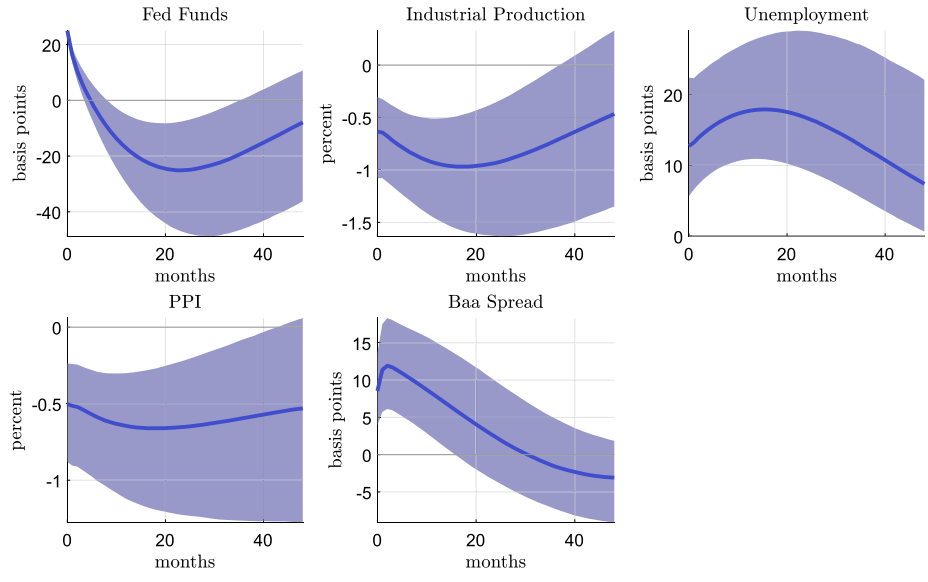

FIGURE 6. Monetary policy shocks: Impulse responses. *Note*: Pointwise median and 68% credible bands. The monetary shock is normalized to imply an impact 25 basis points increase in the policy rate.

maximum drop in PPI of 0.65% 17 months from the shock. The response of the Baa spread is, instead, more sudden. The peak increase in the spread is reached already 2 months from the shock, with an increase of up to 12 basis points. As the systematic component of monetary policy then generates a monetary expansion, the economy reverts back and spreads decrease.

The results from Figure 6 are broadly in line with the existing literature, yet with some differences. As also in Caldara and Herbst (2019) and Miranda-Agrippino and Ricco (2021), industrial production and unemployment display a hump-shaped response. This is remarkable, given that both papers identify the monetary policy shock using an external instrument, while we use a very different identification approach. Nevertheless, some differences emerge in the timing of the response. Caldara and Herbst (2019) document that both variables reach a peak around 25 months from the shock, while Miranda-Agrippino and Ricco (2021) find that the peak effects are reached around 11 months from the shock. Our estimates lie in-between, with the peaks reached around 17 months from the shock. Quantitatively, our estimates are larger than Caldara and Herbst (2019) and Miranda-Agrippino and Ricco (2021). Caldara and Herbst (2019) document a peak variation of industrial production and unemployment of 0.5% and 7 basis points, respectively, while Miranda-Agrippino and Ricco (2021) of 0.37% and 8 basis points, respectively. All papers document that the Baa spread responds mainly in the short horizon after the shock.

Figure 7 extends the analysis to forecast error variance decompositions. It reports the pointwise median and 68% confidence bands up to 4 years from the monetary policy shocks. We find that monetary policy shocks contribute to the variance of the forecast

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

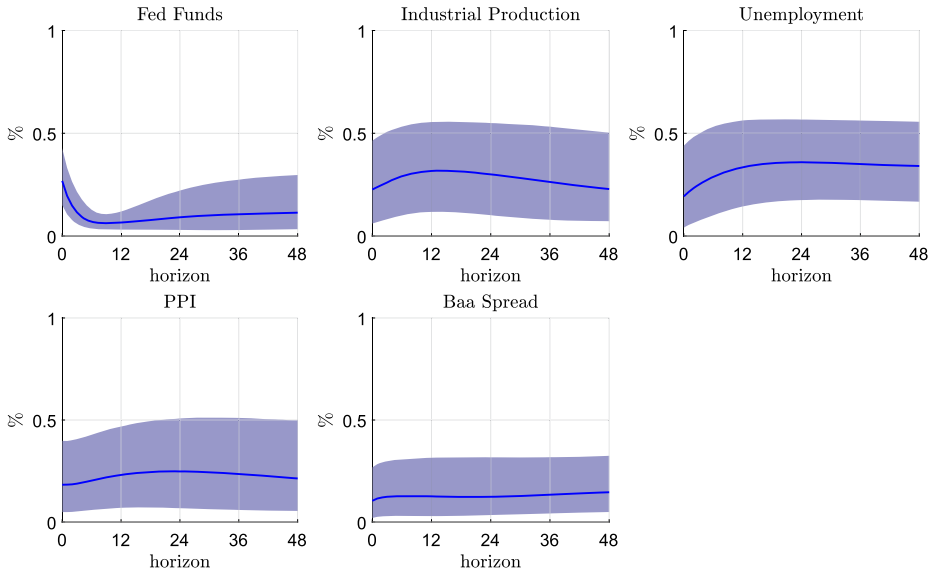

FIGURE 7. Monetary policy shocks: Forecast error variance decomposition.

error of the policy rate in a quantitatively important manner only in the very short term, with an impact value of 26%. At longer horizons, the variance decomposition never lies above 10%. The real economy is affected by monetary policy shocks more in that the forecast error variance decomposition reaches up to 30% of industrial production and unemployment rate. The result on PPI is milder, never exceeding 24%. Last, the Baa spread is found to be only weakly driven by monetary policy shocks in that the corresponding error decomposition never increases above 12%

Caldara and Herbst (2019) document that monetary shock play a nonnegligible role in explaining the forecast error variance of industrial production and unemployment. Our approach suggests an even stronger effect. According to Caldara and Herbst (2019), up to 20% of the forecast error variance of industrial production and unemployment is explained by monetary shocks, an effect taking place for horizons up to 36 months from the shock. We find that the highest share of such forecast errors are explained by the monetary policy shock around 12 months from the shock, with the share increasing up to 30%. As also in Caldara and Herbst (2019), we stress that posterior uncertainty on forecast error variance decompositions is high.

4.3 *Comparison to NiWU prior*

The sign restrictions that we introduce can be modeled using more than one prior distribution. In this last section, we document what happens if the same sign restrictions are modeled using the NiWU prior. Baumeister and Hamilton (2015) and Wolf (2020) argue that the NiWU prior can introduce undesired features on key model statistics into the analysis. Inoue and Kilian (2020) argue that the specific role played by the NiWU prior

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

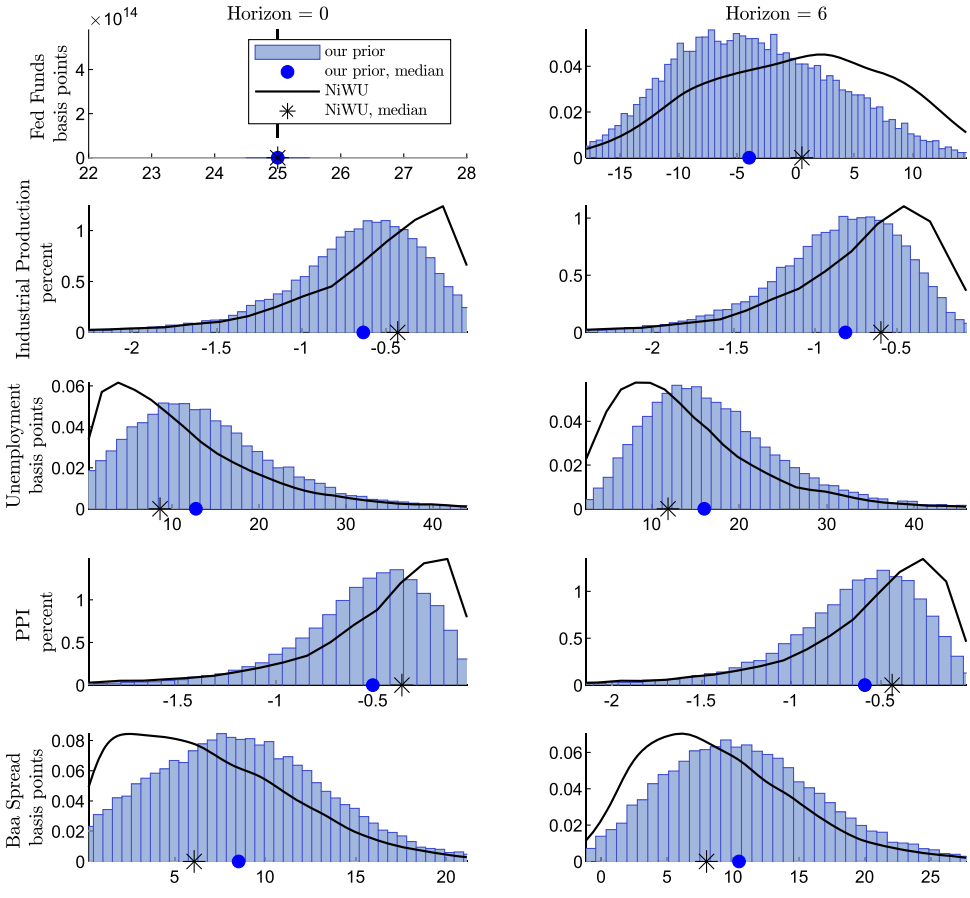

FIGURE 8. Marginal posterior distribution of the impulse responses. *Note*: The monetary shock is normalized to imply an impact 25 basis points increase in the policy rate. See Figure S.4 in the Supplementary Material for the comparison of pointwise impulse responses.

is irrelevant when the identified set is very tight. We now study how the results from our application to monetary policy shocks change when introducing the same sign restrictions using the NiWU prior. We use the very same model specification and use the same restrictions, except that $p(B)$ is now effectively replaced with $p_{NiWU,i}(B)$, the prior distribution implied by the NiWU prior. We calibrate the hyperparameters of the inverse Wishart distribution as in Kadiyala and Karlsson (1997). Posterior draws are generated using Algorithm B from Section S.2.1 of the Supplementary Material.

Figure 8 reports the marginal distribution of the normalized impulse responses on impact as well as 6 months after the shock. Each plot reports the posterior distribution associated with our prior, together with the distribution associated with the NiWU prior and the two corresponding pointwise median values. By construction, the impact effect on the federal funds rate is the same. The remaining plots suggest a nonnegligible dif-

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 2. Comparison of impulse responses (pointwise median).

	0	1	2	3	4	5	6	7	8	9	10	11	12
						Using the NiWU prior, i_0							
Fed F.	25.00	17.76	13.03	9.29	6.08	3.17	0.52	−1.92	−4.06	−6.13	−8.00	−9.69	−11.23
IP	−0.43	−0.45	−0.47	−0.51	−0.54	−0.57	−0.60	−0.62	−0.65	−0.67	−0.68	−0.70	−0.72
Unempl.	8.60	9.16	9.85	10.45	11.00	11.48	11.88	12.26	12.58	12.88	13.21	13.41	13.59
PPI	−0.35	−0.36	−0.37	−0.39	−0.40	−0.42	−0.43	−0.45	−0.46	−0.47	−0.48	−0.49	−0.50
Baa S.	6.09	8.33	8.85	8.76	8.51	8.26	8.00	7.70	7.41	7.13	6.86	6.53	6.22
						Using our prior, i_1							
Fed F.	25.00	16.21	10.62	6.18	2.38	−0.99	−3.99	−6.72	−9.22	−11.52	−13.63	−15.48	−17.17
IP	−0.63	−0.64	−0.68	−0.72	−0.75	−0.78	−0.81	−0.84	−0.87	−0.89	−0.91	−0.92	−0.94
Unempl.	12.73	13.22	13.94	14.57	15.12	15.58	16.01	16.39	16.75	17.05	17.30	17.50	17.68
PPI	−0.50	−0.51	−0.52	−0.54	−0.55	−0.57	−0.59	−0.60	−0.61	−0.62	−0.63	−0.64	−0.64
Baa S.	8.54	11.38	11.92	11.72	11.33	10.89	10.46	10.04	9.62	9.18	8.74	8.26	7.78
						Ratio of our prior to NiWU prior, $100 \cdot (i_1/i_0 - 1)$ (percent)							
Fed F.	0	−9	−18	−33	−61	−131	−870	250	127	88	70	60	53
IP	47	44	43	42	39	38	37	35	34	33	33	32	31
Unempl.	48	44	42	39	38	36	35	34	33	32	31	30	30
PPI	42	41	41	39	37	36	35	34	33	32	31	30	29
Baa S.	40	37	35	34	33	32	31	30	30	29	27	26	25

Note: Pointwise median impulse responses using the NiWU prior (top panel) or our prior (middle panel), together with the percent difference between the two (bottom panel).

TABLE 3. Comparison of forecast error variance decompositions (pointwise median).

	0	1	2	3	4	5	6	7	8	9	10	11	12
						Using the NiWU prior, f_0							
Fed F.	0.36	0.28	0.22	0.17	0.14	0.12	0.10	0.09	0.09	0.08	0.08	0.08	0.08
IP	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.20	0.21	0.21	0.22	0.22
Unempl.	0.11	0.13	0.14	0.15	0.16	0.17	0.18	0.19	0.20	0.21	0.22	0.22	0.23
PPI	0.12	0.12	0.12	0.12	0.13	0.13	0.14	0.14	0.15	0.15	0.16	0.16	0.16
Baa S.	0.06	0.07	0.07	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08	0.08
						Using our prior, f_1							
Fed F.	0.27	0.20	0.15	0.12	0.09	0.08	0.07	0.07	0.06	0.06	0.06	0.06	0.07
IP	0.23	0.24	0.25	0.26	0.27	0.28	0.29	0.30	0.30	0.31	0.31	0.31	0.31
Unempl.	0.20	0.22	0.24	0.25	0.27	0.28	0.29	0.30	0.31	0.32	0.33	0.33	0.34
PPI	0.18	0.18	0.19	0.19	0.19	0.20	0.20	0.21	0.21	0.22	0.22	0.22	0.23
Baa S.	0.10	0.11	0.12	0.12	0.12	0.13	0.13	0.13	0.13	0.13	0.13	0.13	0.13
						Ratio of our prior to NiWU prior, $100 \cdot (f_1/f_0 - 1)$ (percent)							
Fed F.	−24	−28	−30	−32	−33	−32	−30	−27	−26	−23	−20	−17	−14
IP	69	65	63	61	58	56	54	51	50	48	46	44	43
Unempl.	75	72	68	65	63	61	59	56	54	52	51	49	48
PPI	55	55	54	52	50	49	47	45	44	42	40	39	37
Baa S.	72	67	63	61	59	58	57	55	54	52	51	51	50

Note: Pointwise median forecast error decompositions using the NiWU prior (top panel) or our prior (middle panel), together with the percent difference between the two (bottom panel).

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

ference for the remaining variables, with the posterior associated with the NiWU prior being consistently closer to 0. For instance, the NiWU specification delivers an impact decrease in unemployment of around 9 basis points while our baseline specification finds around 13 basis points, suggesting that the effect is 50% stronger. The Baa spread is then predicted to increase on impact 40% more under the baseline prior than under the NiWU prior.

Table 2 further inspects these differences by reporting the pointwise median response of each variable for all horizons up to 1 year from the shock. The top panel shows the effect associated with the NiWU prior. The middle panel shows the effect from our baseline specification. The bottom panel reports the percentage difference between the two. The difference in the results equals up to 50% on impact, and decreases to around 40% in most variables, then stabilizing to around 30%. The sign of the difference remains unchanged for all variables and horizons. Some differences for the federal funds rate appear large when the federal funds rate response from the NiWU is close to 0.

Table 3 documents the same comparison between prior specifications by showing the forecast error variance decompositions. The first result is that modeling the very same sign restrictions with the NiWU approach leads to a larger estimated role for monetary policy shocks in driving the forecast error of the federal funds rate. From the NiWU approach, up to 36% of the forecast error variance of the federal funds rate is driven by exogenous variations in the federal funds rate. By contrast, our estimates point to a value that is lower by 27%, hence suggesting a stronger role for the systematic component of monetary policy. This difference quantitatively increases for future horizons. Consistent with the finding on the impulse responses, we then find that forecast error variance on most of the other variables is around 60% higher under our prior. The differences remain high also when considering alternative values of the hyperparameters for our prior, as we document in Table S.2–Table S.5 in the Supplementary Material. All in all, the analysis suggests that the specific prior distribution used to model a given set of sign restrictions can have quantitatively relevant implications.

5. Conclusions

The choice of prior distributions is particularly relevant in underidentified Bayesian structural vector autoregressions, for example, sign restricted SVARs, since the priors matter even in large samples for this case. In this study, we develop a new posterior sampler, which allows the use of a flexible set of priors in a computationally nonchallenging way. This sampler is a two-step importance sampler, which exploits the computational convenience of the NiWU prior to generate proposal draws and maps these into draws from the posterior of our model in two stages, first on the reduced-form parameters and then on the structural form. We show that we recover the true posterior, as approximated by a computationally much more challenging sampler, in a noncostly manner and illustrate the importance of the choice of priors for applied researchers in an application to US monetary policy shocks in the Great Moderation.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Appendix

Under *Case 1* considered in the paper (equation (27)), we generate proposal draws from

$$\begin{aligned}\Sigma|Y &\sim \text{iW}\big(S^*, d^*\big),\\ d^* &= T-m-c-k-1,\\ S^* &= S+\hat{\Sigma}_T(T-m).\end{aligned}$$

using *Algorithm A* discussed in Section S.2.1 of the Supplementary Material. This marginal posterior is associated with the independent NiWU joint prior

$$\begin{aligned}p_{NiWU,i}(\boldsymbol{\pi},\Sigma) &= p_{NiWU,i}(\boldsymbol{\pi})\cdot p_{NiWU,i}(\Sigma),\\ p_{NiWU,i}(\boldsymbol{\pi}) &\propto 1,\\ p_{NiWU,i}(\Sigma) &\propto \big|\det(\Sigma)\big|^{\frac{c}{2}}\cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}S]}.\end{aligned}$$

with generic hyperparameters (c, S). Despite the independent prior, the posterior $p_{NiWU,i}(\Sigma|Y)$ is inverse Wishart because $p_{NiWU,i}(\boldsymbol{\pi})\propto 1$; see Section S.2.1 of the Supplementary Material. The weights in Stage A are then equal to

$$\begin{aligned}w_d^{\text{stage A}} &= \frac{p(\Sigma|Y)}{p_{NiWU,i}(\Sigma|Y)}\\ &= \frac{v_{\{B\to\Sigma,Q\}}\cdot\big|\det(\Sigma)\big|^{-\frac{T-m}{2}}\cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}\hat{\Sigma}_T(T-m)]}\cdot\int_{O(k)}p_B\big(h(\Sigma)Q\big)\,dQ}{\big|\det(\Sigma)\big|^{-\frac{T-m-c}{2}}\cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}(S+\hat{\Sigma}_T(T-m))]}}\\ &= \frac{\big|\det(\Sigma)\big|^{-\frac{1}{2}}\cdot\big|\det(\Sigma)\big|^{-\frac{T-m}{2}}\cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}\hat{\Sigma}_T(T-m)]}\cdot\int_{O(k)}p_B\big(h(\Sigma)Q\big)\,dQ}{\big|\det(\Sigma)\big|^{-\frac{T-m-c}{2}}\cdot e^{-\frac{1}{2}\text{trace}[\Sigma^{-1}(S+\hat{\Sigma}_T(T-m))]}}\\ &= \big|\det(\Sigma)\big|^{-\frac{c+1}{2}}\cdot e^{\frac{1}{2}\text{trace}[\Sigma^{-1}S]}\cdot\int_{O(k)}p_B\big(h(\Sigma)Q\big)\,dQ\\ &= \int_{O(k)}p_B\big(h(\Sigma)Q\big)\,dQ,\end{aligned}\tag{30}$$

where the last step follows by setting $(c, S)=(-1, 0\cdot I_k)$ in order to reduce the volatility of the weights. Under *Case 2* (equation (28)), we generate proposal draws using a Gibbs sampler on

$$\begin{aligned}\boldsymbol{\pi}|Y,\Sigma &\sim \text{N}\big(\boldsymbol{\mu}_{\pi}^*, V_{\pi}^*\big),\\ \Sigma|Y,\Pi &\sim \text{iW}\big(S^*, d^*\big),\\ V_{\pi}^* &= \big(V_{\pi}^{-1}+WW'\otimes\Sigma^{-1}\big)^{-1},\end{aligned}$$

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\begin{aligned}\boldsymbol{\mu}_{\pi}^{*} &= V_{\pi}^{*}\left(V_{\pi}^{-1}\boldsymbol{\mu}_{\pi} + \left(WW' \otimes \Sigma^{-1}\right)\hat{\boldsymbol{\pi}}_{T}\right),\\ d^{*} &= T - c - k - 1,\\ S^{*} &= S + (Y - \Pi W)(Y - \Pi W)',\end{aligned}$$

via *Algorithm C* discussed in Section S.2.1 of the Supplementary Material. These conditional posteriors are associated with the NiWU joint prior

$$\begin{aligned}p_{NiWU,i}(\boldsymbol{\pi}, \Sigma) &= p_{NiWU,i}(\boldsymbol{\pi}) \cdot p_{NiWU,i}(\Sigma),\\ p_{NiWU,i}(\boldsymbol{\pi}) &= \phi(\boldsymbol{\mu}_{\pi}, V_{\pi}),\\ p_{NiWU,i}(\Sigma) &\propto \left|\det(\Sigma)\right|^{\frac{c}{2}} \cdot e^{-\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]}.\end{aligned}$$

where the prior mean and variance equal the ones used in our prior $p(\boldsymbol{\pi})$ in equation (28). As shown in Section S.2.2 of the Supplementary Material, the proposal draws are representative of the marginal posterior distribution

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

The weights in Stage A are then equal to

$$\begin{aligned}&w_{d}^{\text{stage A}}\\ &= \frac{p(\Sigma|Y)}{p_{NiWU,i}(\Sigma|Y)}\\ &= \frac{v_{\{B\to\Sigma,Q\}} \cdot \left|\det(\Sigma)\right|^{-\frac{T}{2}} \cdot \left|\det\left(V_{\pi}^{*}\right)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\pi}^{*\prime} V_{\pi}^{*-1} \boldsymbol{\mu}_{\pi}^{*}\}} \int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ}{\left|\det\left(V_{\pi}^{*}\right)\right|^{\frac{1}{2}} \cdot \left|\det(\Sigma)\right|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2}\{\operatorname{trace}[\Sigma^{-1}S] + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\pi}^{*\prime} V_{\pi}^{*-1} \boldsymbol{\mu}_{\pi}^{*}\}}}\\ &= \frac{\left|\det(\Sigma)\right|^{-\frac{1}{2}} \cdot \left|\det(\Sigma)\right|^{-\frac{T}{2}} \cdot \left|\det\left(V_{\pi}^{*}\right)\right|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\{\tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\pi}^{*\prime} V_{\pi}^{*-1} \boldsymbol{\mu}_{\pi}^{*}\}} \int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ}{\left|\det\left(V_{\pi}^{*}\right)\right|^{\frac{1}{2}} \cdot \left|\det(\Sigma)\right|^{-\frac{T-c}{2}} \cdot e^{-\frac{1}{2}\{\operatorname{trace}[\Sigma^{-1}S] + \tilde{\mathbf{y}}'(I_T \otimes \Sigma^{-1})\tilde{\mathbf{y}} - \boldsymbol{\mu}_{\pi}^{*\prime} V_{\pi}^{*-1} \boldsymbol{\mu}_{\pi}^{*}\}}}\\ &= \left|\det(\Sigma)\right|^{-\frac{c+1}{2}} \cdot e^{\frac{1}{2}\operatorname{trace}[\Sigma^{-1}S]} \cdot \int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ\\ &= \int_{O(k)} p_B\left(h(\Sigma)Q\right) dQ,\end{aligned}$$

as in equation (30), under the hyperparameter values $(c, S) = (-1, 0 \cdot I_k)$. Table A.1 summarizes these results. The integral term is computed numerically as $\frac{\sum_{i=1}^{m_2} p_B(h(\Sigma)Q_i)}{m_3(\Sigma)}$, with $m_3(\Sigma_d)$ defined as the number of draws required to generate m_2 draws satisfying the candidate sign restrictions on B.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE A.1. Proposal prior and posterior distributions for our sampler.

Case 1) $p(\boldsymbol{\pi}) \propto 1$ (Flat prior or Minnesota prior via dummies observations)

Proposal prior: independent N(flat) + iW	Proposal draws for $p_{NiWU,i}(\Sigma\|Y)$: direct sampling
$p_{NiWU,i}(\boldsymbol{\pi}, \Sigma) = p_{NiWU,i}(\boldsymbol{\pi}) \cdot p_{NiWU,i}(\Sigma)$	$\Sigma\|Y \sim iW(S^*, d^*)$
$p_{NiWU,i}(\boldsymbol{\pi}) \propto 1$	$d^* = T - m - k$
$p_{NiWU,i}(\Sigma) \propto \|\det(\Sigma)\|^{-\frac{1}{2}}$	$S^* = \hat{\Sigma}_T(T-m)$
	needs $T \geq 2k + m$

Case 2) $p(\boldsymbol{\pi}|B) = p(\boldsymbol{\pi})\ N(\boldsymbol{\mu}_\pi, V_\pi)$ (Flexible Minnesota prior)

Proposal prior: independent N + iW	Proposal draws for $p_{NiWU,i}(\Sigma\|Y)$: Gibbs sampling
$p_{NiWU,i}(\boldsymbol{\pi}) \sim N(\boldsymbol{\mu}_\pi, V_\pi)$	$\Sigma\|Y, \Pi \sim iW(d^*, S^*)$
$p_{NiWU,i}(\Sigma) \propto \|\det(\Sigma)\|^{-\frac{1}{2}}$	$d^* = T - k$
	$S^* = (Y - \Pi W)(Y - \Pi W)'$
	needs $T \geq 2k$
	$\boldsymbol{\pi}\|Y, \Sigma \sim N(\boldsymbol{\mu}^*_\pi, V^*_\pi)$
	$V^*_\pi = (V^{-1}_\pi + WW' \otimes \Sigma^{-1})^{-1}$
	$\boldsymbol{\mu}^*_\pi = V^*_\pi(V^{-1}_\pi \boldsymbol{\mu}_\pi + (WW' \otimes \Sigma^{-1})\hat{\boldsymbol{\pi}}_T)$

Note: Our algorithm requires a proposal distribution $p_{NiWU,i}(\Sigma|Y)$. The table lists different possible proposal distributions depending on the corresponding proposal prior $p_{NiWU,i}(\Sigma)$, and reports the corresponding weights for step (ii,d) of our algorithm. Under both cases, the weights in Stage A of our algorithm equal $\propto \int_{O(k)} p_B(h(\Sigma)Q)\, dQ$.

REFERENCES

Amir-Ahmadi, Pooyan and Thorsten Drautzburg (2021), "Identification and inference with ranking restrictions." *Quantitative Economics*, 12 (1), 1–39. [1223]

Amisano, Gianni and Carlo Giannini (2012), *Topics in Structural VAR Econometrics*. Springer Science & Business Media. [1224]

Antolín-Díaz, Juan and Juan F. Rubio-Ramírez (2018), "Narrative sign restrictions for SVARs." *American Economic Review*, 108 (10), 2802–2829. [1223]

Arias, Jonas E., Juan F. Rubio-Ramírez, and Daniel F. Waggoner (2018), "Inference based on structural vector autoregressions identified with sign and zero restrictions: Theory and applications." *Econometrica*, 86 (2), 685–720. [1223, 1224, 1226, 1233, 1235]

Baumeister, Christiane and James D. Hamilton (2015), "Sign restrictions, structural vector autoregressions, and useful prior information." *Econometrica*, 83 (5), 1963–1999. [1221, 1222, 1223, 1225, 1233, 1234, 1236, 1238, 1242]

Baumeister, Christiane and James D. Hamilton (forthcoming), "Advances in using vector autoregressions to estimate structural magnitudes." *Econometric Theory*. [1223]

Baumeister, Christiane J. S. and James D. Hamilton (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. [1223, 1233, 1234, 1235]

Baumeister, Christiane J. S. and James D. Hamilton (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. [1233]

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Binning, Andrew (2013), "Underidentified SVAR models: A framework for combining short and long-run restrictions with sign restrictions." [1233]

Bruns, Martin, and Michele Piffer (2023), "Supplement to 'A new posterior sampler for Bayesian structural vector autoregressive models'." *Quantitative Economics Supplemental Material*, 14, https://doi.org/10.3982/QE2207. [1228]

Caldara, Dario and Edward Herbst (2019), "Monetary policy, real activity, and credit spreads: Evidence from Bayesian proxy SVARs." *American Economic Journal: Macroeconomics*, 11 (1), 157–192. [1222, 1240, 1241, 1242]

Canova, Fabio (2007), *Methods for Applied Macroeconomic Research*, Vol. 13. Princeton University Press. [1234, 1236]

Canova, Fabio, Andrzej Kociecki, and Michele Piffer (2023), "Flexible prior beliefs on impulse responses in Bayesian vector autoregressive models." [1223, 1235]

Canova, Fabio and Fernando J. Pérez Forero (2015), "Estimating overidentified, nonrecursive, time varying coefficients structural VARs." *Quantitative Economics*, 6 (2), 309–358. [1233]

Chan, Joshua C. C. (2022), "Asymmetric conjugate priors for large Bayesian VARs." *Quantitative Economics*, 13 (3), 1145–1169. [1223]

Del Negro, Marco and Frank Schorfheide (2011), "Bayesian macroeconometrics." In *The Oxford Handbook of Bayesian Econometrics* (Gary Koop, Herman van Dijk, and John Geweke, eds.), 293–389, Oxford University Press. [1240]

Fry, Renee and Adrian Pagan (2011), "Sign restrictions in structural vector autoregressions: A critical review." *Journal of Economic Literature*, 49 (4), 938–960. [1238]

Giacomini, Raffaella, Toru Kitagawa, and Harald Uhlig (2019), "Estimation under ambiguity." Technical report, CEMMAP working paper. [1223]

Giacomini, Raffaella and Toru Kitagawa (2021), "Robust Bayesian inference for set-identified models." *Econometrica*, 89 (4), 1519–1556. [1223]

Inoue, Atsushi and Lutz Kilian (2020), "The role of the prior in estimating VAR models with sign restrictions." Technical report, Federal Reserve Bank of Dallas. [1242]

Kadiyala, K. Rao and Sune Karlsson (1997), "Numerical methods for estimation and inference in Bayesian VAR-models." *Journal of Applied Econometrics*, 12, 99–132. [1237, 1243]

Kilian, Lutz and Helmut Lütkepohl (2017), *Structural Vector Autoregressive Analysis*. Cambridge University Press. [1223, 1234]

Kociecki, Andrzej (2010), "A prior for impulse responses in Bayesian structural VAR models." *Journal of Business & Economic Statistics*, 28 (1), 115–127. [1223, 1235]

Kociecki, Andrzej (2017), "Fully Bayesian analysis of SVAR models under zero and sign restrictions." Report. [1226, 1233]

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Kociecki, Andrzej, Michał Rubaszek, and Michele Ca'Zorzi (2012), "Bayesian analysis of recursive SVAR models with overidentifying restrictions." ECB Working Paper Series No. 1492. [1223, 1232]

Korobilis, Dimitris (2022), "A new algorithm for sign restrictions in vector autoregressions." *European Economic Review*, 148, 104241. [1223]

Leeper, Eric M., Christopher A. Sims, Tao Zha, Robert E. Hall, and S. Ben Bernanke (1996), "What does monetary policy do?" *Brookings papers on economic activity*, 1996 (2), 1–78. [1222, 1233]

Miranda-Agrippino, Silvia and Giovanni Ricco (2021), "The transmission of monetary policy shocks." *American Economic Journal: Macroeconomics*, 13 (3), 74–107. [1241]

Ferreira, Leonardo N., Silvia Miranda-Agrippino, and Giovanni Ricco (2023), "Bayesian Local Projections." *The Review of Economics and Statistics*, 1–45. [1235]

Plagborg-Møller, Mikkel (2019), "Bayesian inference on structural impulse response functions." *Quantitative Economics*, 10 (1), 145–184. [1235]

Poirier, Dale J. (1998), "Revising beliefs in nonidentified models." *Econometric Theory*, 14 (4), 483–509. [1227]

Rubio-Ramirez, Juan F., Daniel F. Waggoner, and Tao Zha (2010), "Structural vector autoregressions: Theory of identification and algorithms for inference." *The Review of Economic Studies*, 77 (2), 665–696. [1221, 1224, 1231]

Sims, Christopher A. and Tao Zha (1998), "Bayesian methods for dynamic multivariate models." *International Economic Review*, 39, 949–968. [1222, 1224, 1232, 1233]

Uhlig, Harald (2005), "What are the effects of monetary policy on output? Results from an agnostic identification procedure." *Journal of Monetary Economics*, 52 (2), 381–419. [1221, 1224]

Volpicella, Alessio (2021), "SVARs identification through bounds on the forecast error variance." *Journal of Business & Economic Statistics*, 40 (3), 1291–1301. [1223]

Waggoner, Daniel F., Hongwei Wu, and Tao Zha (2016), "Striated Metropolis–Hastings sampler for high-dimensional models." *Journal of Econometrics*, 192 (2), 406–420. [1222, 1233]

Waggoner, Daniel F. and Tao Zha (2003), "A Gibbs sampler for structural vector autoregressions." *Journal of Economic Dynamics and Control*, 28 (2), 349–366. [1222, 1233]

Wolf, Christian K. (2020), "SVAR (mis)identification and the real effects of monetary policy shocks." *American Economic Journal: Macroeconomics*, 12 (4), 1–32. [1221, 1242]

Zha, Tao (1999), "Block recursion and structural vector autoregressions." *Journal of Econometrics*, 90 (2), 291–316. [1222, 1233]

Co-editor James D. Hamilton handled this manuscript.

Manuscript received 18 July, 2022; final version accepted 23 June, 2023; available online 6 July, 2023.

17597331, 2023, 4, Downloaded from https://onlinelibrary.wiley.com/doi/10.3982/QE2207 by University Of East Anglia, Wiley Online Library on [06/02/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License