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

Martin Bruns & Michele Piffer

Working paper No. 878 November 2018 ISSN1473-0278

School of Economics and Finance

Queen Mary University of London

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

Martin Bruns† Michele Piffer‡

November 23, 2018

Please click this link for the most recent version of the paper

Abstract

Structural VAR models are frequently identified using sign restrictions on impulse responses. Moving beyond the popular but restrictive Normal-inverse-Wishart-Uniform prior, we develop a methodology that can handle almost any prior distribution on contemporaneous responses. We then propose a new sampler that explores the posterior just as efficiently as done by the existing algorithm for the Normal-inverse-Wishart-Uniform case. We use this flexible and tractable framework to combine sign restrictions with information on the volatility of the data, giving less prior mass to impulse effects that are inconsistent with the data from a training sample. This approach sharpens posterior bands and makes sign restrictions more informative. We apply the methodology to the oil market and show that oil supply shocks have a non-negligible effect on oil price dynamics.

*We are thankful to Dario Caldara, Lutz Kilian, Helmut Lütkepohl, Haroon Mumtaz, Edoardo Palombo, Gabor Pinter and Malte Rieth for helpful comments and suggestions. Many thanks to Taiki Yakamura for excellent research assistance. Michele Piffer thanks for the financial support received from the European Union's Horizon 2020 research and innovation program, Marie Sklodowska-Curie grant agreement number 744010.

†Freie Universität Berlin and German Institute for Economic Research (DIW Berlin).

‡Queen Mary, University of London. Corresponding author e-mail: m.b.piffer@gmail.com

JEL classification: C32, C11, E50, H62.

Keywords: Sign restrictions, Bayesian inference, Oil market.

1 Introduction

Structural Vector Autoregressive models (SVARs) are extensively used in applied Macroeconomics. To provide results that can be interpreted economically, SVARs require identifying restrictions. It has become popular to introduce sign restrictions on selected structural statistics using Bayesian informative priors that reflect the intended signs (Uhlig, 2005, Baumeister and Hamilton, 2015, Arias et al., 2018).

This paper studies how to implement sign restrictions on the most frequently used statistic, namely impulse responses. The literature often limits prior beliefs to the relatively narrow family of Normal-inverse-Wishart-Uniform prior beliefs, in order to ensure a particularly tractable posterior distribution (Uhlig, 2005, Rubio-Ramirez et al., 2010, Arias et al., 2018). Yet, as pointed out by Baumeister and Hamilton (2015), since this traditional approach does not express prior beliefs directly on the structural parameters of interest, it can introduce features that go beyond the intention of the researcher. This limitation is important because, even in large sample, information from the data does not fully dominate the prior, making the results dependent on the prior distribution used. We complement the work by Sims and Zha (1998) and Baumeister and Hamilton (2015) and study the popular case of prior beliefs on impulse responses, rather than on the elasticities between variables.

The first contribution of the paper consists of developing a framework that can handle almost any prior distribution on the contemporaneous impulse responses. We parametrize the model in the reduced form autoregressive elements and in the contemporaneous impulse responses, as in Uhlig (2005), but express prior beliefs directly on contemporaneous responses rather than on reduced form covariances. In being general about prior beliefs on contemporaneous impulse responses, we offer a framework that gives prior flexibility on the impulse response horizon where flexibility is needed the most, as argued by Canova and Pina (2005) and Canova and Paustian (2011). We then retain the normal prior distribution on the reduced form autoregressive component of the model in order to ensure a more tractable posterior distribution, a restriction that

is without loss of generality in a large sample. Overall, our approach offers a balance between prior flexibility on the key structural parameters and conditional conjugate priors on all the remaining parameters. We also extend our framework to allow for shape restrictions on future horizons of the responses.

The second contribution of the paper consists of developing a posterior sampler that ensures that the additional flexibility provided by our methodology does not come at a computational cost. We show that while an importance sampler cannot be applied directly on the structural parameter space (except for special cases, as for example in Arias et al., 2018), it can be applied when sampling the posterior distribution on the structural parameters in two stages, using two separate importance functions. The key intuition is that if the sample size is not too small, the posterior distribution on the reduced form covariances depends only mildly on the prior beliefs used, making the Normal-inverse-Wishart familiy a suitable importance function. In addition, the algorithm by Rubio-Ramirez et al. (2010) is a suitable importance function for the mapping from the reduced form to the structural parameters, because it explores the full space of orthogonal matrices. In resampling posterior draws from the Normal-inverse-Wishart-Normal approach to make them consistent with a less restrictive prior distribution, our approach shows that there is no trade-off between the flexibility advocated by Baumeister and Hamilton (2015) and the numerical efficiency guaranteed by Rubio-Ramirez et al. (2010). To further stress the efficiency of this sampler we show that the results are the same when exploring the posterior using the more time-consuming sequential approach by Waggoner et al. (2016).

Having developed a framework that combines posterior tractability with flexibility on the prior beliefs on the key structural parameters, we illustrate how the results can depend on the prior distribution used, further suggesting that prior beliefs on structural parameters should be selected carefully in applied work. When mapping reduced form parameters into structural parameters, the traditional approach does not take into account the volatility of the data. Hence, it can for example imply that the

impact of a one standard deviation shock on a variable is as big as the largest variation ever observed in that variable. We propose a new prior specification that combines sign restrictions with information on the scaling of the variables, based on a training sample. We show that this new feature is enough to tighten posterior bands, making sign restrictions more informative, with potentially wide consequences in applied work. Alternative prior specifications are, of course, possible.

We first develop the intuition behind our methodology using data simulated from the An and Schorfheide (2007) model. This exercise discusses in detail prior specification, posterior sampling, and the key driver of the difference between our approach and the traditional approach. We then apply our methodology to revisit a long lasting debate on the effect of oil price dynamics on the real economy. In his seminal contribution, Kilian (2009) shows how variations in the price of oil affect the US economy differently depending on the underlying source of the oil price variation. Kilian and Murphy (2012) revisit the results by Kilian (2009) by replacing the recursive identification scheme with sign restrictions on the effect of the shocks. However, they conclude that sign restrictions alone are not enough to disentangle the different channels driving oil price dynamics. We show that applying the same sign restrictions using our methodology does deliver results on the relative importance of different structural shocks. We confirm the result by Kilian (2009) that demand shocks have a strong and long lasting effect on the price of oil, but also show that supply shocks have a non-negligible effect on the price of oil.

From the applied point of view, the paper relates to the wide literature on the oil market. Starting from the aforementioned papers by Kilian (2009) and Kilian and Murphy (2012), many studies have looked for identifying restrictions to study how oil price dynamics affect the US economy. Kilian and Murphy (2012) restrict the price elasticity of oil supply. Antolín-Díaz and Rubio-Ramírez (2018) restrict the sign of the estimated structural shocks and the historical decompositions of the variables. Baumeister and Hamilton (2017) express sign restrictions on the price elasticity of oil

supply and oil demand. Last, Caldara et al. (2018) develop an exact-identification approach that relies on external estimates of the elasticities. Interestingly, several of our results support the conclusions reached by these papers, despite using a different methodology. For example, the result that oil supply shocks do have an effect on the price of oil is consistent with Baumeister and Hamilton (2017), Caldara et al. (2018). Our approach is relatively agnostic compared to Caldara et al. (2018), Antolín-Díaz and Rubio-Ramírez (2018) and Baumeister and Hamilton (2017) in that we only introduce sign restrictions on the impact effect of the structural shocks. Yet, our results are consistent with the additional restrictions used in some of these contributions. For example, we find that oil supply shocks were indeed the prevailing driver of the drop in oil production during the first Gulf War, a feature that Antolín-Díaz and Rubio-Ramírez (2018) introduce as an identifying restriction.

From the methodological point of view, we follow Baumeister and Hamilton (2015) and express prior beliefs directly on the structural parameters of interest. However, we study the case of beliefs on contemporaneous impulse responses rather than on the contemporaneous relation among variables, as the former are more frequent in applied work (Kilian and Lütkepohl, 2017). Baumeister and Hamilton (forthcoming) combine prior beliefs on contemporaneous relations and contemporaneous impulse responses. Relative to Baumeister and Hamilton (forthcoming) and Arias et al. (2018), we focus on impulse responses and propose a different prior specification and posterior sampler. We also depart from Plagborg-Møller (forthcoming), who expresses prior beliefs on impulse responses at any horizon at the cost of working with a less tractable posterior distribution. Kociecki (2010) departs from prior beliefs on impulse responses, but works with the recursive identification scheme. As we document in the appendix, his approach can be used to extend our methodology to shape restrictions on impulse responses beyond the contemporaneous effect. Last, we relate to Giacomini and Kitagawa (2015) in stressing the mapping from reduced form to structural parameters, but we concentrate on a single prior.

The traditional approach to sign restricted SVARs frequently implies relatively wide posterior bands for impulse responses, and for other statistics. For this reason, many studies combine sign restrictions with additional information on other statistics, as in Kilian and Murphy (2012), Antolín-Díaz and Rubio-Ramírez (2018) and Amir-Ahmadi and Drautzburg (2018). We argue that the traditional approach to sign restrictions can be improved upon by introducing explicit information on which region of the structural parameter space is plausible when mapping reduced form parameters into structural parameters. We show that augmenting sign restrictions on impulse responses with information on the volatility of the data is sufficient to deliver sharper inference, possibly to the point that no additional information is needed to interpret the results. Our tighter posterior bands are not trivially driven by tighter prior beliefs on impulse responses, but by the fact the researcher can inform the mapping between reduced and structural parameters with his knowledge of the scale of the data.

The paper is organized as follows. Section 2 outlines the methodology proposed and discusses its relation to the existing literature. Section 3 shows the illustrative example on simulated data from the An and Schorfheide (2007) model. Section 4 reports the application to the oil market. Section 5 concludes.

2 The methodology

In this section we first highlight why the data allows for different structural parametrizations within the same SVAR model. We then summarize the approach to sign restrictions that is widely used in the literature. Last, we outline our methodology, introduce an efficient sampler to make the methodology computationally unchallenging, and propose a prior specification.

2.1 Parametrizations of the structural model

Structural VAR models can be written in different forms. A popular specification is

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

where $\boldsymbol{y}_t$ is a $k \times 1$ vector of endogenous variables, $\boldsymbol{\epsilon}_t$ is a $k \times 1$ vector of structural shocks, and $\boldsymbol{w}_t = (1, \boldsymbol{y}'_{t-1}, .., \boldsymbol{y}'_{t-p})'$ is an $m \times 1$ vector of the constant and p lags of the variables, with $m = kp + 1$. The $k \times k$ matrices $A, A_1, .., A_p$ are gathered in $A_+ = [\boldsymbol{a}_0, A_1, .., A_p]$, which is of dimension $k \times m$. We normalize the covariance matrix of $\boldsymbol{\epsilon}_t$ to the identity matrix.[1]

Equation (1) highlights the structural nature of the model. Yet, as it is already known, one can rewrite equation (1) after premultiplying both sides of the equality by $B = A^{-1}$, obtaining

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

with $\boldsymbol{\pi}_0 = A^{-1}\boldsymbol{a}_0$, $\Pi_l = A^{-1}A_l$ and $\Pi = [\boldsymbol{\pi}_0, \Pi_1, .., \Pi_p] = A^{-1}A_+$. Matrix B captures the contemporaneous effects of one standard deviation shocks, while future horizons of the impulse responses are calculated using model (2) recursively. Equivalently, one can compute impulse responses using the structural moving average representation of the data,

$$\boldsymbol{y}_t = \boldsymbol{b}_0 + \sum_{l=1}^{\infty} B_l \boldsymbol{\epsilon}_{t-l} + B_0 \boldsymbol{\epsilon}_t, \qquad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, I_k), \tag{3}$$

[1] In doing so, we depart from Baumeister and Hamilton (2015), who exploit conjugate priors for the variance of the structural shocks. We apply this normalization because it is frequently used in applications that employ sign restrictions on impulse responses, see for example Uhlig (2005) and Arias et al. (2018).

where it holds that $B_0 = B$. For the full mapping of (Π, B) into $\{B_l\}_{l=0}^{\infty}$ see, for example, Kilian and Lütkepohl (2017), chapter 2.

Specification (1) and (2) of the structural model stress different features. Equation (1) highlights the contemporaneous relations among the variables in the system, as captured by matrix A. Because of these relations, a structural shock that hits one variable potentially affects contemporaneusly all variables, in a way that is captured by matrix B. Whether the model is more conveniently expressed as model (1) or (2) (or even as a combined form) depends on whether the identifying restrictions introduced by the researcher are more naturally expressed on A or B.[2]

The reduced form representation of the data is

$$\begin{aligned} \boldsymbol{y}_t &= \boldsymbol{\pi}_0 + \sum_{l=1}^{p} \Pi_l \boldsymbol{y}_{t-l} + \boldsymbol{u}_t, \\ &= \Pi \boldsymbol{w}_t + \boldsymbol{u}_t, \qquad\qquad \boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma), \end{aligned} \tag{4}$$

where it holds that $\boldsymbol{u}_t = B\boldsymbol{\epsilon}_t$ and $\Sigma = BB' = A^{-1}A'^{-1}$. Matrices A, A_+ and B contain structural parameters while Π and Σ represent reduced form parameters. Orthogonal matrices Q, which by construction satisfy $QQ' = I_k$, allow for the mapping from

[2]For example, the literature on the identification of monetary policy shocks employs restrictions either on B, as in Uhlig (2005), or on A, as in Arias et al. (forthcoming) and Baumeister and Hamilton (forthcoming). In the terminology used by Lütkepohl (2005), model (1) represents the A form of the SVAR while model (2) represents the B form. Alternatively, Amisano and Giannini (2012) refer to models (1) and (2) as the K and C form of the SVAR, respectively. To appreciate the importance of this distinction, note that restrictions imposed on one form might not be apparent in the other form, due to the nonlinearities in the mapping from one to another. As an example, the A model by Sims and Zha (2006) imposes the non-recursive zero restrictions $A = \begin{pmatrix} a_{11} & a_{12} & 0 \\ a_{21} & a_{22} & 0 \\ a_{31} & 0 & a_{33} \end{pmatrix}$, which implies $B = A^{-1} = \begin{pmatrix} b_{11} & b_{12} & 0 \\ b_{21} & b_{22} & 0 \\ b_{31} & b_{32} & b_{33} \end{pmatrix}$. A well-known special case is the recursive structure of A, which implies a recursive structure of B. Going through the publications of all top-five journals and the Journal of Monetary Economics since 1998, we found that around 13% of the total number of issues checked included at least one application of Structural Vector Autoregressive models. Of the total number of SVAR applications that we found, approximately 15% specifies the model in the A form, 76% specifies the model in the B form, and 9% specifies the model in the hybrid AB form. The detailed list is available at this link.

reduced form to structural parameters, with

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

and $h(\Sigma)$ any decomposition of Σ satisfying $h(\Sigma)h(\Sigma)' = \Sigma$, for example the Cholesky decomposition. See Arias et al. (2018) for a thorough discussion.

2.2 The traditional NiWU prior used in the literature

Define $\boldsymbol{\pi} = vec(\Pi)$ as the vector of dimension $km \times 1$ that stacks the columns of Π. The most popular approach for sign restrictions in SVAR models starts from prior beliefs on the parameters $(\boldsymbol{\pi}, \Sigma, Q)$, equations (4) and (5). In fact, as already known in the literature, when $p(\boldsymbol{\pi}, \Sigma)$ falls within the independent Normal-inverse-Wishart family, the joint posterior distribution $p(\boldsymbol{\pi}, \Sigma|Y)$ can be conveniently explored using a Gibbs sampler.[3] More formally, with prior beliefs

$$\boldsymbol{\pi} \sim N(\boldsymbol{\mu}_\pi, V_\pi), \tag{6}$$

$$\Sigma \sim iW(d, S), \tag{7}$$

with $p(\boldsymbol{\pi}, \Sigma) = p(\boldsymbol{\pi})p(\Sigma)$, it holds that

$$\boldsymbol{\pi}|Y, \Sigma \sim N(\boldsymbol{\mu}_\pi^*, V_\pi^*), \tag{8}$$

$$\Sigma|Y, \Pi \sim iW(d^*, S^*), \tag{9}$$

see Section B of the Appendix for the full derivations. Matrices Q, required in equation (5), are drawn from a distribution that is uniform in the parameter space of orthogonal

[3] To make the analysis comparable to the direct approach discussed in Section 2.3, we do not restrict the indirect approach to the more tractable case of conjugate priors, which impose $V_\pi = V \otimes \Sigma$, but consider the more general independent Normal-inverse-Wishart prior specification. By not imposing a Kronecker structure on V_π, $p(\boldsymbol{\pi})$ allows for the popular prior by Litterman (1986), treating the variance on 'own lags' and 'lags on other variables' differently. See Koop et al. (2010) for a discussion and Arias et al. (2018) for the analysis using conjugate priors.

matrices, a feature satisfied when drawing Q from the popular sampler by Rubio-Ramirez et al. (2010). The algorithm typically used to conduct posterior analysis with this Normal-inverse-Wishart-Uniform prior specification is summarized in Algorithm 1.

Algorithm 1 (traditional Normal-inverse-Wishart-Uniform approach):

1. draw $\boldsymbol{\pi}_i, \Sigma_i$ from $p(\boldsymbol{\pi}, \Sigma|Y)$ using a Gibbs sampler based on (8) and (9);
2. draw an orthogonal matrix Q_i using the method by Rubio-Ramirez et al. (2010);
3. map $\boldsymbol{\pi}_i, \Sigma_i, Q_i$ into the structural parameters of interest;
4. keep $\boldsymbol{\pi}_i, \Sigma_i, Q_i$ if the sign restrictions on the selected parameters of interest are satisfied;
5. repeat steps 1-4 until the required nnumber of draws is obtained.[4]

The convenience of the above approach is that efficient algorithms like Algorithm 1 exist for the sampling of the posterior distribution. In addition, this approach allows for the introduction of sign restrictions on a wide range of statistics, for example on the contemporaneous relation among variables (in step 3, mapping $\{\Sigma_i, Q_i\}$ into $\{A_i\}$, model (1)), on the contemporaneous impulse responses (mapping $\{\Sigma_i, Q_i\}$ into $\{B_i\}$ model (2)) or on future impulse responses (mapping $\{\boldsymbol{\pi}_i, \Sigma_i, Q_i\}$ into $\{B_{i,l}, \}_{l=1}^{\infty}$ from model (3)). The inconvenience is that prior beliefs are not directly specified on the structural parameters of interest (A, B or $\{B_l\}_{l=0}^{\infty}$), but on reduced form parameters and on orthogonal matrices (Π, Σ and Q). Importantly, since the structural model is not exact-identified, prior beliefs on the structural parameters play a role also in

[4]To ensure a higher efficiency of the algorithm, the sign restrictions are assessed after taking into account that shocks are identified only up to sign and ordering. The algorithm changes if zero restrictions are introduced, see Arias et al. (2018) and Binning (2013).

large sample (see Baumeister and Hamilton, 2015, and the illustrative example in Section 3.3). Hence, it is important to minimize the unwanted curvature in the prior that $p(\boldsymbol{\pi}, \Sigma, Q)$ implies for the prior beliefs on the structural parameters of interest, and on the results.[5]

The above intuition can be expressed more formally using analytical derivations. This paper studies the case in which the researcher has prior beliefs on contemporaneous impulse responses, namely on matrix B from model (2). Section B of the Appendix derives the prior and posterior marginal distributions for B implied by the prior distribution (6)-(7) and by the algorithm for Q by Rubio-Ramirez et al. (2010). For simplicity we consider the case when no zero restrictions are introduced. These distributions equal

$$p(B)_{NiWU} \propto I\{\text{sign}\} \cdot |det(B)|^{-(d+k)} \cdot e^{-\frac{1}{2}\left\{vec(B^{-1})'(S\otimes I_k)vec(B^{-1})\right\}}, \tag{10}$$

$$p(B|Y)_{NiWU} \propto I\{\text{sign}\} \cdot |det(B)|^{-(d+k+T)} \cdot e^{-\frac{1}{2}\left\{vec(B^{-1})'(S\otimes I_k)vec(B^{-1})+\tilde{\boldsymbol{y}}'\left(I_T\otimes(BB')^{-1}\right)\tilde{\boldsymbol{y}}-\boldsymbol{\mu}_\pi^{*\prime}V_\pi^{*-1}\boldsymbol{\mu}_\pi^*\right\}}, \tag{11}$$

with $I\{\text{sign}\}$ an indicator function equal to unity if B satisfies the sign restrictions and the suffix $_{NiWU}$ indicating that the distribution refers to the Norma-inverse-Wishart-Uniform case.[6] From equation (10) we can appreciate to what extent the indirect approach can be unintentionally restrictive. If there are values of the hyperparameters d, S such that $p(B)_{NiWU}$ sufficiently approximates the resercher's prior beliefs on contemporaneous impulse responses, then the traditional approach offers a convenient way to implement the analysis. Yet, as we outline in the illustrative example in Section 3.3, $p(B)_{NiWU}$ can introduce undesirable features, because it is specified indirectly

[5]That prior beliefs on one parametrization imply questionable or unintended features on some other parametrization is to some extent inevitable, as remarked, for instance, by Baumeister and Hamilton (2015). We agree with the authors that prior beliefs should be judged relative to the structural parametrization of interest, which in our application is B.

[6]The derivations of (10) build on Arias et al. (2018), who specify the model as in equation (1) and derive the prior distribution that the prior (7) and the algorithm by Rubio-Ramirez et al. (2010) imply for A.

via $p(\Sigma, Q)$, because $p(\Sigma)$ is restricted to the Inverse-Wishart family, and because the algorithm for Q does not necessarily ensure uniformity in the dimension of interest. In addition, the popular hyperparameter specification $d = 0$, $S = 0$ (Uhlig, 2005, Arias et al., 2018) implies an improper prior, which makes it difficult to appreciate the functional form that $p(B|Y)_{NiWU}$ inherits from the prior.[7]

2.3 The approach proposed in this paper

To overcome the above limitations we propose to express prior beliefs directly on B. To do so we parametrize the model as in equation (2) and express independent prior beliefs

$$p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}) \cdot p(B). \tag{14}$$

Since $\boldsymbol{\pi}$ is identified, $p(\boldsymbol{\pi})$ matters less compared to $p(B)$, as long as the sample has sufficient length. Hence, as also in the traditional approach from Section 2.2, we restrict $p(\boldsymbol{\pi})$ to

$$\boldsymbol{\pi} \sim N(\boldsymbol{\mu}_\pi, V_\pi), \tag{15}$$

with $\boldsymbol{\mu}_\pi$ and V_π independent on B. By contrast, $p(B)$ can fall within a wide range of prior distributions, granting the researcher flexibility on the prior beliefs used to express sign restrictions on structural parameters.[8]

[7] The parametrization $d = 0$, $S = 0$ is frequently used in applied work, because, once combined with the conjugate prior specification, it implies $E(\Sigma|Y) = \frac{T-k}{T-k-1} \cdot \hat{\Sigma}_{OLS} \approx \hat{\Sigma}_{OLS}$, with $\hat{\Sigma}_{OLS}$ the least squared estimate of Σ. With $d = 0$, $S = 0$, $p(B)_{NiWU}$ and $p(B|Y)_{NiWU}$ become

$$p(B)_{NiWU} \propto I\{\text{sign}\} \cdot |det(B)|^{-k}, \tag{12}$$

$$p(B|Y)_{NiWU} \propto I\{\text{sign}\} \cdot |det(B)|^{-(k+T)} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T \otimes (BB')^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^*\right\}}. \tag{13}$$

[8] The approach discussed in this section can be extended without difficulty to the case in which $\boldsymbol{\mu}_\pi$ and V_π depend on B. Indeed, the derivations in Section C of the Appendix consider the general case in which the independence restriction is not introduced. Here, we assume independence between B and $\boldsymbol{\pi}$ to facilitate the comparison to the traditional approach, for which we use the more general independent prior specification rather than the conjugate priors. Regarding $p(B)$, as in Baumeister and Hamilton (2015) and Baumeister and Hamilton (2017) the only requirement is that $p(B)$ is everywhere nonnegative, and when integrated over the set of all values of B it produces a finite

As we show in Section C of the Appendix, the posterior distribution satisfies

$$p(\boldsymbol{\pi}, B|Y) = p(\boldsymbol{\pi}|B, Y) \cdot p(B|Y), \tag{16}$$

$$\boldsymbol{\pi}|B, Y \sim N(\boldsymbol{\mu}_\pi^*, V_\pi^*), \tag{17}$$

$$p(B|Y)_{Np(B)} \propto p(B) \cdot |det(B)|^{-T} \cdot |det(V_\pi^*)|^{\frac{1}{2}} \cdot e^{-\frac{1}{2}\left\{\tilde{\boldsymbol{y}}'\left(I_T \otimes (BB')^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1} \boldsymbol{\mu}_\pi^*\right\}}, \tag{18}$$

with

$$V_\pi^* = [V_\pi^{-1} + (WW' \otimes (BB')^{-1})]^{-1}, \tag{19}$$

$$\boldsymbol{\mu}_\pi^* = V_\pi^* \cdot \left[V_\pi^{-1}\boldsymbol{\mu}_\pi + \left[W \otimes (BB')^{-1}\right]\tilde{\boldsymbol{y}}\right], \tag{20}$$

and $\tilde{\boldsymbol{y}} = vec([\boldsymbol{y}_1, ..., \boldsymbol{y}_t, ..., \boldsymbol{y}_T])$, $W = [\boldsymbol{w}_1, ..., \boldsymbol{w}_t, ..., \boldsymbol{w}_T]$. The suffix $_{Np(B)}$ indicates that the distribution refers to our approach, which uses the normal distribution for $\boldsymbol{\pi}$ and the general prior $p(B)$ for B. The analysis of the joint posterior distribution then requires a suitable posterior sampling procedure for the k^2 elements in $p(B|Y)_{Np(B)}$, or even for fewer parameters in case zero restrictions are introduced on B. Draws for the km elements in $\boldsymbol{\pi}|B, Y$ can instead be obtained with a standard random number generator.

The above approach strikes a balance between flexibility and tractability. On the one hand, it grants the researcher flexibility on impulse responses at the horizon where flexibility is needed the most. Since sign restrictions on impulse responses are typically introduced contemporaneously rather than on future horizons, we do not view our framework as particularly restrictive. On the other hand, it exploits the normal prior distribution on $\boldsymbol{\pi}$ that the indirect approach employs to make posterior sampling more tractable. Plagborg-Møller (forthcoming) develops a general approach that works directly on the moving average representation of the data, model (3).

positive number.

His approach is more challenging because a numerical simulator is required for all parameters. Alternatively, one can follow Kociecki (2010) and derive the posterior distribution for the SVAR model implicit in prior beliefs expressed on the moving average representation of the data. His approach is numerically challenging when the recursive structure by Kociecki (2010) is abandoned, as we discuss in Section C of the Appendix.

The approach that we propose is also convenient for a number of additional reasons. First, contrary to Arias et al. (2018), it makes it easier to introduce zero restrictions, as long as such restrictions are expressed on B. Second, by parametrizing the model in Π rather than in $A_+ = B\Pi$, it makes it straightforward to use the prior by Litterman (1986) (which is applied directly on $\boldsymbol{\pi}$), simplifying the analysis compared to Sims and Zha (1998) and Baumeister and Hamilton (2015).[9]

2.4 The new posterior sampler proposed in this paper

To make our approach viable in applied work we require an efficient algorithm that explores the posterior distribution $p(B|Y)_{Np(B)}$ from equation (18). We saw that when prior beliefs $p(B)$ take the special case implicit in the traditional approach (equation (10)) the posterior distribution $p(B|Y)_{Np(B)}$ can be explored using an existing and very popular algorithm, which we summarized in Algorithm 1 from Section 2.2. We now develop an extension of Algorithm 1 to allow for the wider class of prior beliefs on B.

We build our sampling procedure on the importance sampling techniques. Suppose

[9]Parametrizing the model in Π rather than in A_+ simplifies the comparison of the direct and the indirect approach, as $p(\boldsymbol{\pi})$ appears in both parametrizations. The parametrization in Π rather than A_+ does not affect the number of parameters for which the algorithm is required, because both $p(\Pi|B,Y)$ (when parametrizing the model in Π) and $p(A_+|B,Y)$ (when parametrizing the model in A_+) have a common form that can be drawn from using available random number generators. In the special case of the A form, Sims and Zha (1998) and Baumeister and Hamilton (2015) show that the parametrization in A, A_+ together with prior independence across the parameters in the different structural equations of the model allow for posterior simulation equation by equation. In the more general framework considered in this paper, the presence of $B \neq I_k$ prevents from breaking the analysis equation by equation.

we are interested in sampling from the target distribution $p(\boldsymbol{\theta})^{target}$, and suppose we cannot draw from $p(\boldsymbol{\theta})^{target}$ directly, but can only evaluate it. In addition, suppose that we can draw from the importance function $p(\boldsymbol{\theta})^{importance}$. To the extent that the importance function fully covers the support of $p(\boldsymbol{\theta})^{target}$, we can obtain draws from $p(\boldsymbol{\theta})^{target}$ by resampling the draws $\{\boldsymbol{\theta}_i\}$ from the importance function with replacement using weigths $w(\boldsymbol{\theta}_i) = \frac{p(\boldsymbol{\theta}=\boldsymbol{\theta}_i)^{target}}{p(\boldsymbol{\theta}=\boldsymbol{\theta}_i)^{importance}}$ (see for example Koop, 2003, chapter 4). One can assess the performance of the sampler by ensuring that the effective sample size $ESS = \left(\sum_i \left(w_i / \sum_i (w_i)\right)^2\right)^{-1}$, which captures the effective number of draws used to represent the target probability, is not excessively small relative to the number of draws from the importance function. If the importance function sufficiently covers the support of the target function, a small effective sample size only suggests increasing the number of draws from the importace function. If, instead, we cannot ensure that the importance function gives sufficient mass to the support of the target function, a low effective sample size suggests that the importance function must be changed.

In principle, one could use the posterior distribution from the traditional approach as an importance function to study the posterior distribution associated with our more flexible approach, setting $p(\boldsymbol{\theta})^{target} = p(B|Y)_{Np(B)}$ and $p(\boldsymbol{\theta})^{importance} = p(B|Y)_{NiWU}$. This procedure does not work in a general framework, because one cannot ensure that $p(B|Y)_{NiWU}$ sufficiently covers the support of $p(B|Y)_{Np(B)}$, except for special cases.

We circumvent the above challenge by exploring $p(B|Y)_{Np(B)}$ indirectly. Call $p(\Sigma|Y)_{NiWU}$ the posterior distribution on Σ from the Normal-inverse-Wishart-Uniform approach, which can be sampled efficiently using Algorithm 1. Then call $p(\Sigma|Y)_{Np(B)}$ the posterior distribution implicit in $p(B|Y)_{Np(B)}$. The first stage of our procedure consists of converting draws from $p(\Sigma|Y)_{NiWU}$ into draws from $p(\Sigma|Y)_{Np(B)}$. Since Σ is identified, $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ are close to each other as long as the sample size is not excessively small. This makes $p(\Sigma|Y)_{NiWU}$ a viable importance function to explore $p(\Sigma|Y)_{Np(B)}$, a conjecture that can be directly verified from the corresponding effective sample size.

The second stage of our procedure consists of mapping posterior draws from $p(\Sigma|Y)_{Np(B)}$ into posterior draws from $p(B|Y)_{Np(B)}$. Call $p(Q|\Sigma)_{NiWU}$ the uniform distribution on Q used in the traditional approach given the sign restrictions used and call $p(Q|\Sigma)_{Np(B)}$ the distribution on Q implicit in $p(B)$ given Σ. Characterizing $p(Q|\Sigma)_{Np(B)}$ analytically can be challenging, but not numerically. The mapping from Σ to B requires orthogonal matrices Q from $p(Q|\Sigma)_{Np(B)}$, not $p(Q|\Sigma)_{NiWU}$. However, by sampling the parameter space of orthogonal matrices uniformly, the algorithm by Rubio-Ramirez et al. (2010) fully explores the parameter space for Q, reducing to zero the probability that $p(Q|\Sigma)_{NiWU}$ does not explore the relevant parameter space covered by $p(Q|\Sigma)_{Np(B)}$. For the second stage, a low effective sample size only suggests the need to increase the number of draws from the importance function, but not that the importance function is not covering the relevant support of the target funcion.

In Section D of the Appendix we show that the above procedure can be made operational through the following algorithm:

Algorithm 2 (this paper):

1. extract a wide number of draws from $p(B)$ and compute numerically the implicit value of $E(\Sigma)_{Np(B)}$;
2. set $d = k + 2$ and $S = E(\Sigma)_{Np(B)} \cdot (d - k - 1)$, which ensures that $E(\Sigma)$ from $\Sigma \sim iW(d, S)$ is not far away from $p(\Sigma)_{Np(B)}$;
3. run Algorithm 1 from Section 2.2 and store a large number of draws of Σ_i and B_i (ensuring $\Sigma_i = B_i B_i'$), which represent draws from $p(\Sigma|Y)_{NiWU}$ and $p(B|Y)_{NiWU}$, respectively;
4. for each Σ_i compute weigths

$$w_i^{\text{stage A}} = |det(\Sigma_i)|^{\frac{d+k}{2}} \cdot e^{\frac{1}{2}tr[\Sigma_i S]}, \tag{21}$$

and assess that the effective sample size $ESS^{\text{stage A}} = \left(\sum_i \left(w_i^{\text{stage A}} / \sum_i (w_i^{\text{stage A}})\right)^2\right)^{-1}$

is sufficiently high;

5. compute the weights

$$w_i^{\text{stage B}} = p(B = B_i); \tag{22}$$

6. generate draws $\tilde{B}_i$ by resampling $\{B_i\}$ with replacement using weigths $\{w_i\}$ with $w_i = w_i^{\text{stage A}} \cdot w_i^{\text{stage B}}$. If the effective sample size computed on $\{w_i\}$ is large enough, the draws $\{\tilde{B}_i\}$ represent draws from $p(B|Y)_{Np(B)}$. If not, repeat the exercise by increasing the number of draws in step 3.

Algorithm 2 resamples the posterior draws from the traditional approach by making them representative of the posterior distribution associated with the generic prior beliefs $p(B)$ from our approach. In Section 3 we document that the sampling time of Algorithm 2 is very close to the computational time from Algorithm 1, in that steps 4-6 of Algorithm 2 require a minimal running time. It is in this sense that the posterior sampler proposed in this section is considered as efficient as Algorithm 1. Step 1 and 2 of Algorithm 2 can be changed to a different selection of d and S. We use this approach to further limit the distance of $p(\Sigma|Y)_{NiWU}$ from $p(\Sigma|Y)_{Np(B)}$ and improve the weigths in step 4.

To further assess the effectiveness of our algorithm we also explore $p(B|Y)_{Np(B)}$ using the Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016). This altrnative algorithm is computationally more demanding, but can handle postentially irreguarly shaped posterior distributions and a large number of parameters. Using the posterior distribution from this algorithm, we show that the sampling procedure proposed in this section does a good job in exploring $p(B|Y)_{Np(B)}$, provided that the sample size is not too small. We discuss how we implement the algorithm by Waggoner et al. (2016) in Section E of the Appendix.

2.5 Proposing one possible prior

So far we have developed an approach that does not restrict prior beliefs to the Normal-inverse-Wishart-Uniform prior beliefs, while still allowing for a fast and efficient posterior sampler. We conclude the section on the methodology by discussing one possible prior specification for $p(B)$, which we have so far treated as general. We do so with two goals in mind. First, to highlight that the prior distributions expressed on structural parameters implicit in the traditional approach matter for the results. Second, to propose an alternative specification to the applied researcher.

Specifying prior beliefs $p(B)$ is challenging, because the literature still provides limited guidance on explicit prior beliefs on structural parameters. Baumeister and Hamilton (2015) work with model (1) and use the existing literature to form prior beliefs on the contemporaneous elasticities among variables. However, as discussed by Kilian and Lütkepohl (2017), researchers may well lack explicit prior information on the contemporaneous relationship among variables. Instead, they frequently have prior beliefs that do not go beyond the sign of contemporaneous impulse responses. 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.[10]

To overcome this challenge, we propose a prior specification for $p(B)$ that builds on a conventional prior specification used in the literature for $p(\boldsymbol{\pi})$. The proposed prior aims to combine sign restrictions with information on the volatility of the data. To do so, the crucial step is to take a reasonable stand on the scale of the parameters. With the Minnesota Prior on $\boldsymbol{\pi}$, one first associates each variable with a reasonable scale capturing the volatility of the variables. This is usually implemented by estimating the variance σ_i of the residual on univariate AR processes on each variable. Then, Bayesian

[10]The same challenge is present also in Plagborg-Møller (forthcoming) and in Kociecki (2010). Plagborg-Møller (forthcoming) calibrates the prior distribution on the impulse responses using information from a DSGE model. Kociecki (2010) uses a normal distribution on B, but his recursive, exact-identifying framework implies that the prior beliefs on structural parameters vanish asymptotically.

shrinkage is introduced through a set of hyperparameters that shrink the parameters in $\boldsymbol{\pi}$ towards the random walk or the white noise process, taking the relative scale of the variables into account (see Canova, 2007 and Kilian and Lütkepohl, 2017).

We extend the above procedure as follows. Call b_{ij} the entry of B capturing the effect of a one standard deviation shock j to variable i. It can be shown that the covariance restrictions $\Sigma = BB'$ imply

$$-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}, \tag{23}$$

with Σ_{ii} the $i-th$ element of the diagonal of Σ.[11] Accordingly, $\gamma_i = \hat{\Sigma}_{ii}^{0.5}$ provides a candidate assessment of the reasonable scale for b_{ij}, where $\hat{\Sigma}$ is an estimate based on a training sample. We then introduce two hyperparameters ψ_1 and ψ_2 that control for the location and the spread of $p(b_{ij})$. We use independent, untruncated normal distributions $N(\mu_{ij}, \sigma_{ij})$ as follows: 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, and 95% of the prior mass is in the space $(-\psi_2\gamma_i, \psi_2\gamma_i)$; if b_{ij} is restricted to be positive, start from a normal distribution with $\mu_{ij} = \psi_1\gamma_i$ and calibrate the variance such that the distribution truncated on the positive support has 95% prior mass in the space $(0, \psi_2\gamma_i)$ (or $(-\psi_2\gamma_i, 0)$); if b_{ij} is restricted to be negative, start from a normal distribution with $\mu_{ij} = -\psi_1\gamma_i$ and calibrate the variance such that the distribution truncated on the negative support has 95% prior mass in the space $(-\psi_2\gamma_i, 0)$. Last, for the shocks that remain unidentified, we numerically introduce the restriction that they do not replicate the sign restrictions of the identified shocks.

The convenience of the above approach is that the researcher sets a reasonable scale for the effect of the shocks by selecting γ_i, and then introduces Bayesian shrinkage through the hyperparameters ψ_1 and ψ_2. In this paper we use the parametrizations $\psi_1 = 0.8$ and $\psi_2 = 1.2$, which ensure that the prior mass is given close to the estimated

[11] Given $\Sigma = BB'$, the restrictions corresponding to the diagonal elements of Σ are $\Sigma_{ii} = b_{i1}^2 + b_{i2}^2 + ... + b_{ik}^2$. Since Σ_{ii} is nonnegative and since $b_{ij}^2 \geq 0$, each element b_{ij} must satisfy $-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}$.

reasonable scale of each variable. The rest of the paper uses the prior specification for B discussed in this section, and provides graphical intuition for it in Section 3.3. Alternative specifications are also possible. We limit the analysis to an illustration that employs these values of the hyperparameters, as developing an assessment of the most robust way of specifying prior beliefs on structural parameters goes beyond the purpose of the paper.

3 An illustrative example based on data simulated from the model by An and Schorfheide (2007)

We first apply the methodology discussed in Section 2.3 to simulated data, and provide further intuition on how the procedure works.

3.1 The data generating process and the model

We generate data using the linearized DSGE model by An and Schorfheide (2007), which is given by the following set of equations:

$$x_t = E_t x_{t+1} + g_t - E_t g_{t+1} - \frac{1}{\tilde{\tau}}(r_t - E_t \pi_{t+1} - E_t z_{t+1}), \tag{24a}$$

$$\pi_t = \beta E_t \pi_{t+1} + \tilde{\kappa}(x_t - g_t), \tag{24b}$$

$$r_t = \rho_r r_{t-1} + (1-\rho_r)\eta_1 \pi_t + (1-\rho_r)\eta_2 (x_t - g_t) + \epsilon_{rt}, \tag{24c}$$

$$g_t = \rho_g g_{t-1} + \epsilon_{gt}, \tag{24d}$$

$$z_t = \rho_z z_{t-1} + \epsilon_{zt}. \tag{24e}$$

The variables of the model are the output gap (x_t), inflation (π_t), the interest rate (r_t), productivity (z_t) and government spending (g_t). They are driven by the technology shock $\epsilon_{zt} \sim N(0, \sigma_z^2)$, the government spending shock $\epsilon_{gt} \sim N(0, \sigma_g^2)$, and an interest rate shock $\epsilon_{rt} \sim N(0, \sigma_r^2)$. Equations (24a) to (24e) summarize a model economy

of a representative household, perfectly competitive intermediate good producers, a final good producer, a fiscal authority, and the central bank. Households gain utility from consumption and real money balances, and disutility from labour. They supply labour on competitive markets to intermediate good producers, who face a production function subject to technology shocks and adjust prices after incurring a cost. The fiscal authority consumes a stochastic fraction of final output, raises lump-sum taxes and issues bonds. The interest rate is set by the central bank according to a Taylor rule subject to monetary policy shocks.

The parameters of the model are $\boldsymbol{\iota} = (\tilde{\tau}, r^A, \tilde{\kappa}, \rho_r, \rho_g, \rho_z, \psi_1, \psi_2, \sigma_r, \sigma_g, \sigma_z)'$, with $\beta = \frac{1}{1+r^A/400}$.[12] We calibrate $\boldsymbol{\iota}$ using the parameter values that An and Schorfheide (2007) employ for their data generating process, as summarized in Table G4 in the Appendix. We then use the solution method by Sims (2002) to solve the model and the factorization by Fernandez-Villaverde et al. (2007) to compute the associated VAR representation. The DGP has the exact, reduced form VAR(1) representation

$$\underbrace{\begin{pmatrix} r_t \\ x_t \\ \pi_t \end{pmatrix}}_{\boldsymbol{y}_t} = \underbrace{\begin{pmatrix} 0.6048 & 0 & 0.9017 \\ 0.0616 & 0.95 & -1.9302 \\ 0.0103 & 0 & 0.4452 \end{pmatrix}}_{\Pi(\boldsymbol{\iota})} \underbrace{\begin{pmatrix} r_{t-1} \\ x_{t-1} \\ \pi_{t-1} \end{pmatrix}}_{\boldsymbol{y}_{t-1}} + \underbrace{\begin{pmatrix} u_{zt} \\ u_{gt} \\ u_{rt} \end{pmatrix}}_{\boldsymbol{u}_t}, \tag{25}$$

with $\boldsymbol{u}_t \sim N\big(\mathbf{0}, \Sigma(\boldsymbol{\iota})\big)$ and

$$\Sigma(\boldsymbol{\iota}) = \begin{pmatrix} 0.0011 & -0.004 & -0.0001 \\ -0.0004 & 0.0095 & 0.0004 \\ -0.0001 & 0.0004 & 0.0001 \end{pmatrix}. \tag{26}$$

The above model is used to generate data, which we then use to estimate a SVAR model identified with sign restrictions.[13]

[12] To calibrate the model as in An and Schorfheide (2007), we treat β as a function of the fundamental parameter r^A, which determines the steady state interest rate.

[13] While the model has a unique structural form, we emphasize its reduced form and avoid running

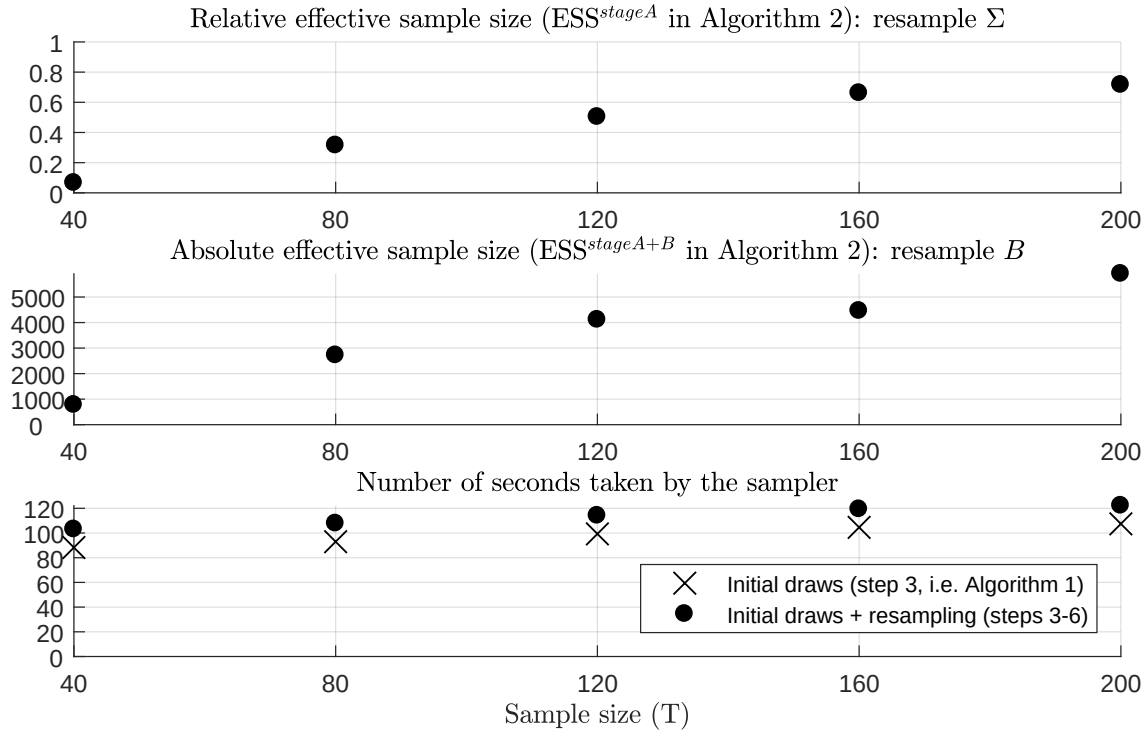
3.2 Illustration of our approach

We use the data generating process to generate a dataset of 300 draws, and discard the first 100 draws to make the data less dependent on the initial point, which we set equal to the unconditional means of the variables. We then divide the remaining 200 draws into five different sub-datasets, including up to the first 40, 80, 120, 160 and 200 observations. For each sub-dataset, we estimate a model with a constant and four lags. We parametrize $\boldsymbol{\mu}_\pi$ and V_π from equation (15) as in the Minnesota Prior (Litterman, 1986) and use the hyperparameter values discussed by Canova (2007), chapter 10. We identify two structural shocks using the following sign restrictions: that a technology shock that increases output increases the interest rate and inflation, and that a monetary policy shock that increases the interest rate decreases output and inflation. We set $p(B)$ using the non-hierarchical specification discussed in Section 2.5. The hyperparameter γ_i is set equal to the estimate $\hat{\Sigma}_{ii}^{0.5}$ obtained on a training sample that employs the first 20% of the sample, as in Primiceri (2005), except for the sub-dataset including 40 observations, for which the sample size requires estimating $\hat{\Sigma}_{ii}^{0.5}$ based on the full sub-dataset.

For each sub-dataset, we explore $p(B|Y)_{Np(B)}$ using Algorithm 2 proposed in Section 2.4. We first extract 100,000 from Algorithm 1 (step 3 of Algorithm 2) and then apply the importance sampler from steps 4-6 of Algorithm 2. Figure 1 assesses the performance of our posterior sampler. As should be expected, the smaller the size of the sample, the lower the effective sample size when converting draws from $p(\Sigma|Y)_{NiWU}$ into draws from $p(\Sigma|Y)_{Np(B)}$ (top panel of the figure). In this application, $T = 40$ leads us to effectively use only 6% observations of the initial 100,000, while $T = 200$ leads us to use as many as 72%. This confirms the intuition that as long as the sample size is large enough, an importance sampler can be used to explore $p(\Sigma|Y)_{Np(B)}$. Ac-

a horse race on the ability of different approaches to better approximate the true data generating process. The conclusion from such an analysis would be specific to the application used, and hard to generalize to a more realistic framework. We use the model only for illustrative purposes, as it helps clarify how our approach differs from the existing literature.

Figure 1: Assessment of our posterior sampler using simulated data: effective sample size (ESS) and running time

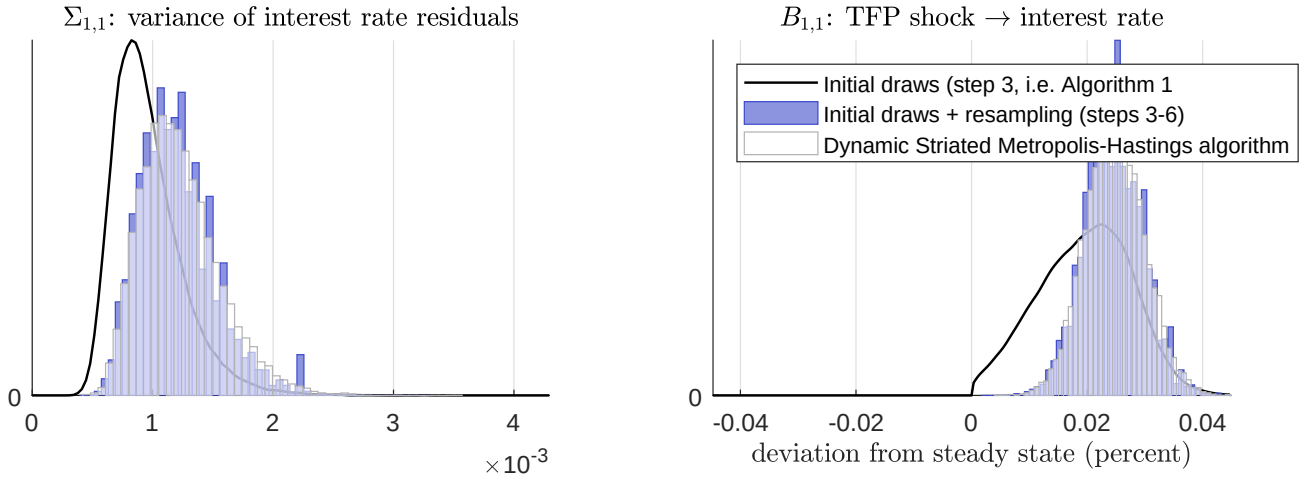


counting then also for step B of the algoritm (middle panel), we find that for $T = 200$ the posterior $p(B|Y)_{Np(B)}$ is effectively explored using 5,910 draws, a number that can be increased by increasing the number of initial draws from step 3 of the algorithm. The computational time of Algorithm 2 effectively coincides with the computational time of Algorithm 1 (bottom panel). In this application the resampling discussed in steps 4-6 from Algorithm 2 takes around 14s irrespectively of the sample size considered, for a total computational time of less than 2m. Running the Dynamic Striated Metropolis-Hastings algorithm took from 7m36s for $T = 40$ to 48m49s for $T = 200$. See Section E of the Appendix for how we set the tuning parameters of the algorithm.

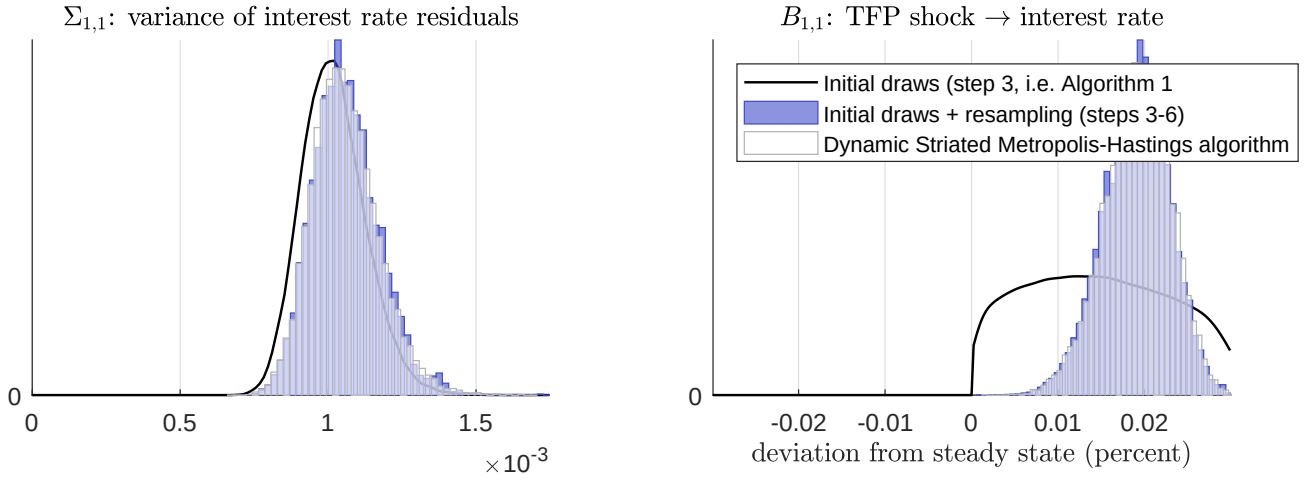
Figure 2 further provides intuition for our sampler by showing the posterior distribution for the $(1, 1)$ entry of Σ and B. The solid line shows the posterior distributions $p(\Sigma|Y)_{NiWU}$ and $p(B|Y)_{NiWU}$ sampled from step 3 of Algorithm 2. The blue shaded area shows $p(\Sigma|Y)_{Np(B)}$ and $p(B|Y)_{Np(B)}$ obtained from Algorithm 2, while

Figure 2: Assessment of our posterior sampler using simulated data: Σ and B

A) Low sample size: $T = 40$



B) High sample size: $T = 200$



Note: The grey shaded area shows the 'correct' posterior distribution $p(B|Y)_{Np(B)}$, as explored by the Dynamic Striated Metropolis-Hastings algorithm. The solid lines on the left panels show the importance function used to explore $p(\Sigma|Y)_{Np(B)}$. The blue shaded area shows the posterior distribution $p(B|Y)_{Np(B)}$ sampled from Algorithm 2. The higher is the sample size, the closer the posterior distribution from Algorithm 2 is to the 'correct' posterior distribution detected by the Dynamic Strated Metropolis-Hastings algorithm by Waggoner et al. (2016).

the light shaded area shows the same distributions sampled with the Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016). The latter posteriors

can be thought of as capturing the 'correct' posterior distribution, which Algorithm 2 aims to explore. For $T = 40$, $p(\Sigma|Y)_{NiWU}$ is still partly away from $p(\Sigma|Y)_{Np(B)}$. This leads to a low effective sample size and a less precise sampling of $p(B|Y)_{Np(B)}$. As the sample size increases, $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ are closer to each other, which makes $p(\Sigma|Y)_{NiWU}$ a valid importance function and improves the sampling of $p(B|Y)_{Np(B)}$.

Figure 3: $p(B|Y)_{Np(B)}$, update of prior beliefs on B

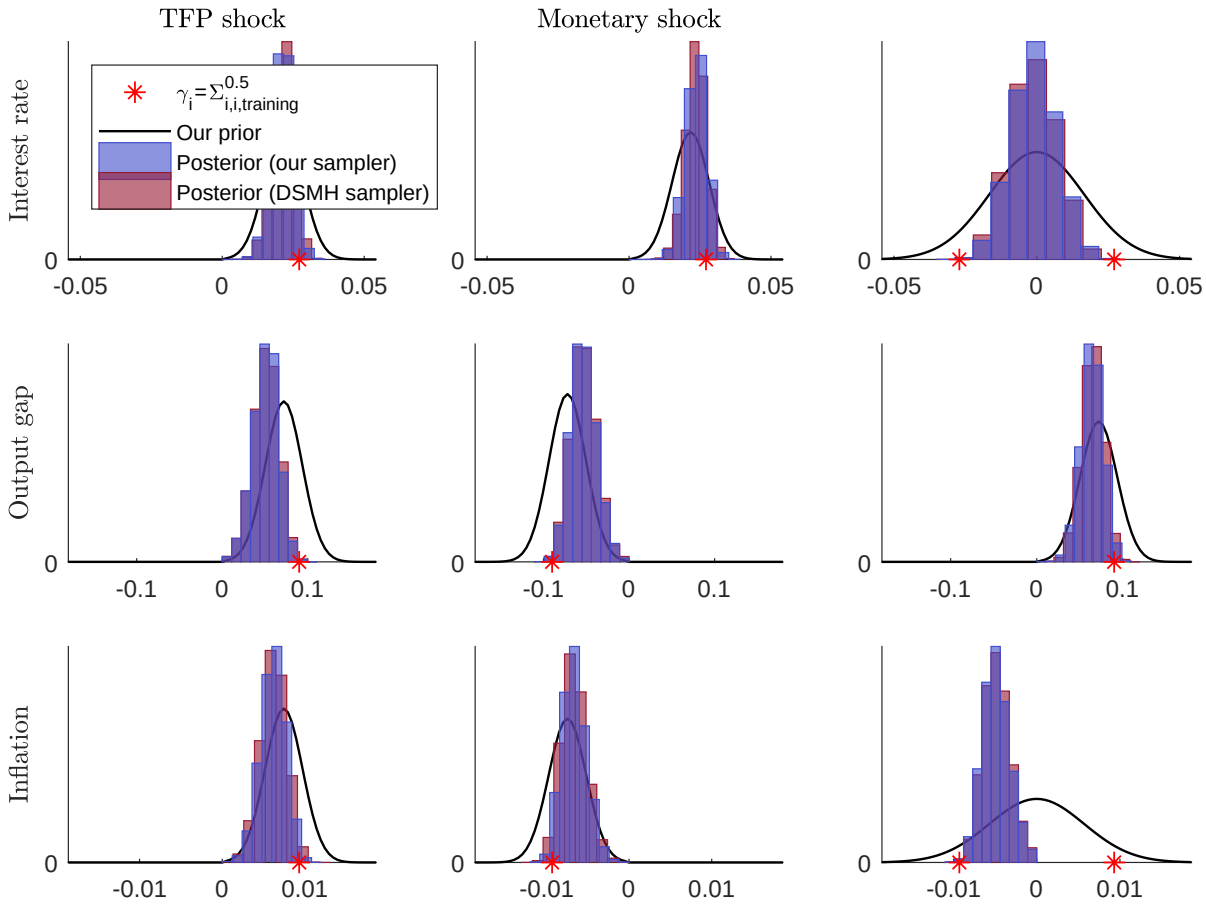


Figure 3 shows the prior and posterior beliefs on B corresponding to the sub-sample with $T = 120$. The stars in the figure indicate the estimated reasonable scale γ_i. Given γ_i, the mode of the marginal prior distribution $p(b_{ij})$ equals $\psi_1\gamma_i = 0.8\gamma_i$,

while 95% of the prior mass is concentrated in the support $[0, \psi_2\gamma] = [0, 1.2\gamma]$ or $[0, -\psi_2\gamma] = [0, -1.2\gamma]$, depending on the sign restriction. The figure also shows the posterior distribution explored using Algorithm 2 and using the algorithm by Waggoner et al. (2016). The posterior distributions are very similar, suggesting that the sample size is already sufficiently high for Algorithm 2 to efficiently explore $p(B|Y)_{Np(B)}$.

3.3 Comparison to the NiWU approach

We conclude the illustrative example by comparing our approach with the traditional Normal-inverse-Wishart-Uniform approach. We make the two approaches more easily comparable using decomposition (5), which we rewrite here for convenience:

$$B = h(\Sigma)Q. \tag{27}$$

The two approaches differ in how they draw Σ and Q. The traditional approach draws Σ from either $p(\Sigma)_{NiWU}$ or $p(\Sigma|Y)_{NiWU}$, and draws Q from $p(Q|\Sigma)_{NiWU}$. As outlined by Algorithm 2, our approach effectively draws Σ from the distribution on Σ implicit in either $p(B)_{Np(B)}$ or $p(B|Y)_{Np(B)}$, and draws Q from $p(Q|\Sigma)_{Np(B)}$ implicit in $p(B)$. Since Σ is identified, differences in the prior distributions on Σ vanish asymptotically. This is not true for the distribution used for Q, implying that the difference between the two approaches mainly relies on the difference in the stochastic process generating Q (see Giacomini and Kitagawa, 2015).

As discussed, for example, by Baumeister and Hamilton (2015, forthcoming), the traditional approach treats candidate draws of Q as equally plausible, irrespectively of the curvature implied on the parameter space of B. Conditioning on a draw Σ, only a subset of candidate orthogonal matrices is consistent with the sign restrictions. Yet, all matrices Q within this subset are treated as equally plausible. By contrast, while this subset is the same across the two approaches, the direct approach does not treat Q matrices as equally plausible, but takes into account the part of the parameter

space of B that they lead to, as reflected by $p(B)$.

Figure 4: An illustration using a two variable model

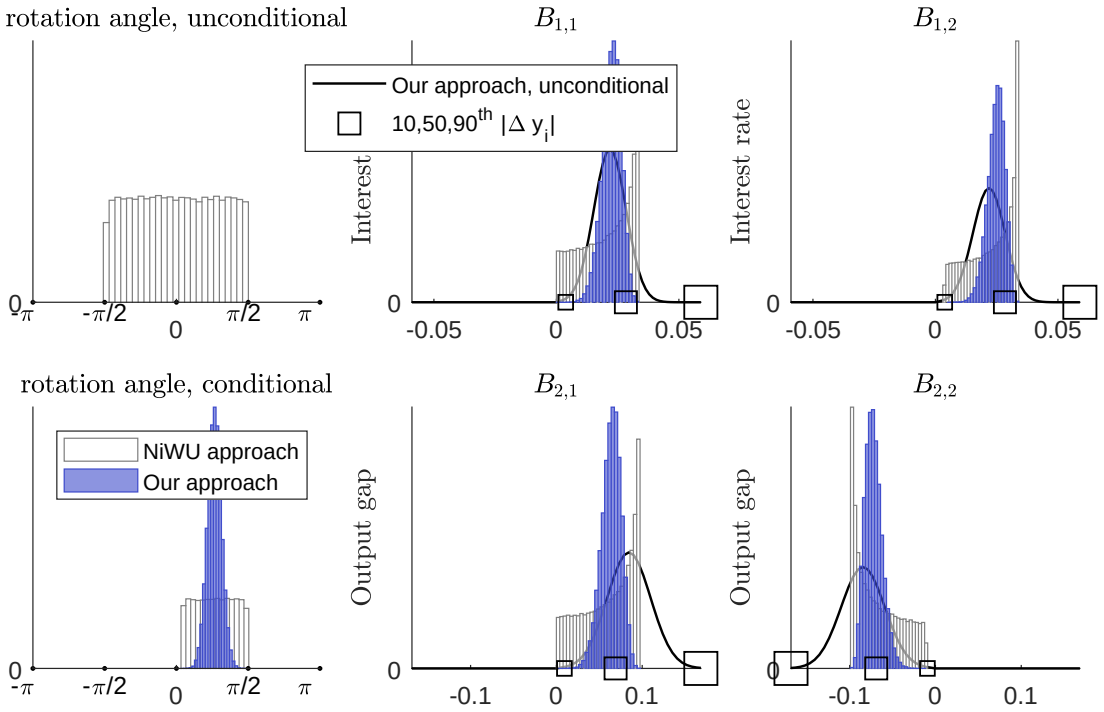
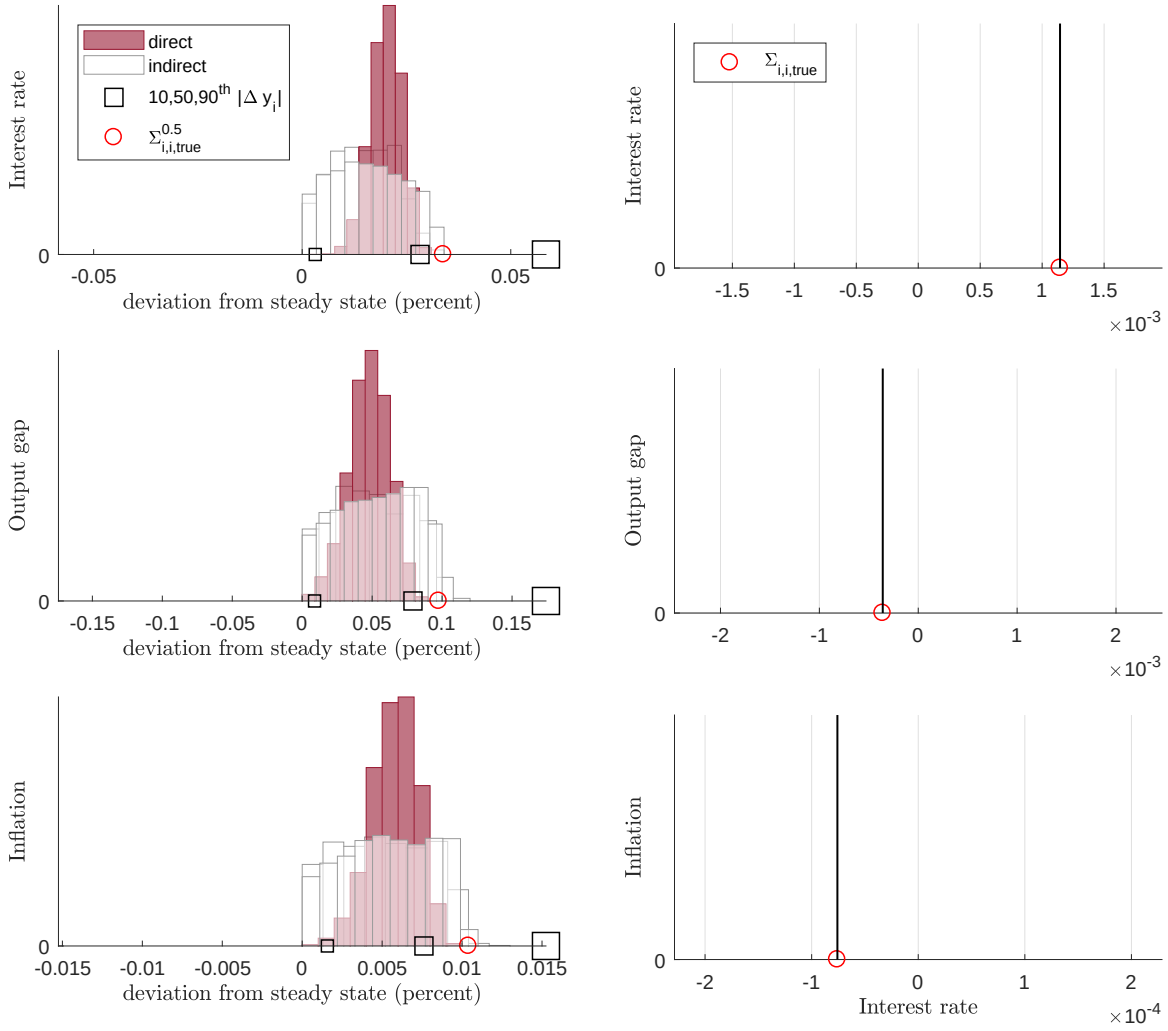


Figure 4 develops the above intuition using a simplified bivariate VAR that recomputes the exercise using the interest rate and the output gap from the simulated data. A bivariate case facilitates the illustration, as orthogonal matrices Q can be generated using 2×2 Givens transformations. The top-left plot of Figure 4 shows the distribution of the angle of the rotation matrices that replicate draws of orthogonal matrices from a uniform distributions of Q (see Section D.2 of the Appendix for the discussion of the sampler). The distribution of the rotation angle is uniform in the support $[-\pi/2, \pi/2]$ (Fry and Pagan, 2011, Baumeister and Hamilton, 2015). The rest of the figure conditions the analysis on a draw of Σ, which we set equal to the true 2×2 upper left block of $\Sigma(\boldsymbol{\iota})$. Conditioning on such Σ, the rotation angles consistent with the sign restrictions are the subset shown in the bottom-left plot of the figure. While the traditional approach treats such angles as equally plausible, the direct approach does not. The middle and the right panels show the implied distribution on B.

Figure 5: Compare asymptotic posterior distributions on B and Σ

$B(:,1)$ - Effect of TFP shock

$\Sigma(:,1)$ - Covariance of the interest rate residuals



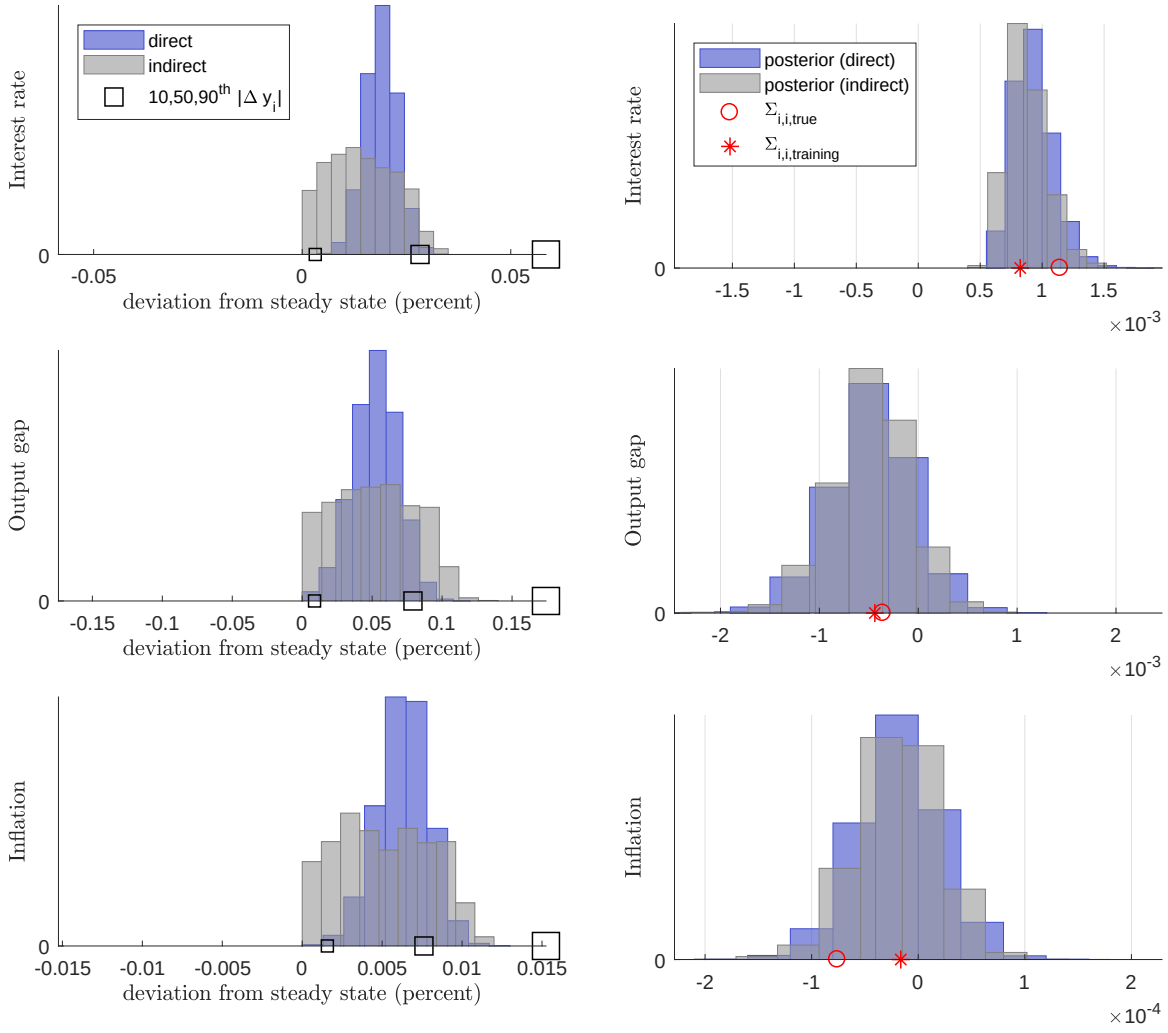
Note: Figure G5 in the Appendix complements the results from this figure.

Given the normalization used, the distributions on B reflect the effect of one standard deviation shocks. We provide a sense of the volatility of the data by showing the 10^{th}, 50^{th} and 90^{th} percentile of the absolute value in the period-by-period variation in the data. The traditional approach introduces a wide and asymmetric distribution on all four entries on B, with a strong spike above the median value of the period-by-period variations in the variables. This distribution shows that since the traditional approach

Figure 6: Compare finite sample posterior distributions on B and Σ

$B(:,1)$ - Effect of TFP shock

$\Sigma(:,1)$ - Covariance of the interest rate residuals



Note: Figure G6 and Figure G7 complement the results from this figure.

does not express beliefs directly on the parameters of interest, it can introduce unintended features into the prior on such parameters. By contrast, the direct approach takes an explicit stand on the plausible shape implied on B.

Moving back to the three-variate simulation exercise, we first compare the posterior distributions in the hypothetical case in which the researcher has access to a dataset of infinite size. Figure 5 makes the differences more visible by showing the results on the

first columns of Σ and B, leaving the full comparison to Section G of the Appendix. We do as in Figure 4 and report the magnitude of the volatility of the variables showing the 10^{th}, 50^{th} and 80^{th} percentiles in the absolute value of the variations in the data. The asymptotic scenario makes the prior distributions on Σ irrelevant and isolates differences between $p(Q|\Sigma)_{NiWU}$ and $p(Q|\Sigma)_{Np(B)}$ (see Section D.3 in the Appendix for the details on the sampler we use, and Baumeister and Hamilton, 2015 for a discussion of the asymptotic posterior distribution in this class of models). Since Σ is identified, as the data increases in size the posterior distribution on Σ collapses to a point mass in $\Sigma_{true} = \Sigma(\boldsymbol{\iota})$. Yet, the posterior distribution on B does not collapse to a mass point, but reflects the different distributions for Q.

To conclude the analysis, Figure 6 shows the posterior distribution in the finite sample used. In order to assess this distribution, we need a parametrization of the Inverse-Wishart distribution for Σ. We set $d = 0$ and $S = 0$, as discussed in Section 2.2. Other parametrizations are possible, but they imply very small differences, given that Σ is identified. The figure shows that the sample is already large enough for the posterior distributions on Σ to be very similar. Consistent with this result, the features of the asymptotic results for B dominate on the prior beliefs, as $p(B|Y)_{NiWU}$ is wider than $p(B|Y)_{Np(B)}$ despite $p(B)_{NiWU}$ being tighter than $p(B)$. This suggests that inference from the direct approach is sharper despite making use of the same sign restrictions. Figure G8 in the Appendix shows that this result holds also in a sample of only 20 observations.

4 An application to real data: demand and supply shocks in the oil market

We now illustrate our methodology by revisiting the model of the oil market by Kilian and Murphy (2012). This model is particularly suited for our application because it employs sign restrictions on the contemporaneous impulse response to the shocks of

interest. We show that inference becomes sharper when applying the sign restrictions on the impulse responses by Kilian and Murphy (2012) using our methodology and our prior specification, allowing for sharper results.

4.1 The model

We use the three-variate reduced form model by Kilian (2009) and Kilian and Murphy (2012), which has become standard in the literature. The model includes the percentage variation in global crude oil production, the detrended index of global real economic activity developed by Kilian (2009), and the log of the real price of oil, multiplied by 100. We use the data updated by Antolín-Díaz and Rubio-Ramírez (2018), which covers the period from January 1971 to December 2015. To improve the comparability with Antolín-Díaz and Rubio-Ramírez (2018) we add a constant and 24 lags in the model, and use a flat prior on $\boldsymbol{\pi}$, setting $V_{\pi}^{-1} = 0$ in equation (6).

Table 1: Sign restrictions on the contemporaneous impulse responses

	oil supply shock	aggregate demand shock	oil demand shock
oil production	–	+	+
economic activity	–	+	–
real price of oil	+	+	+

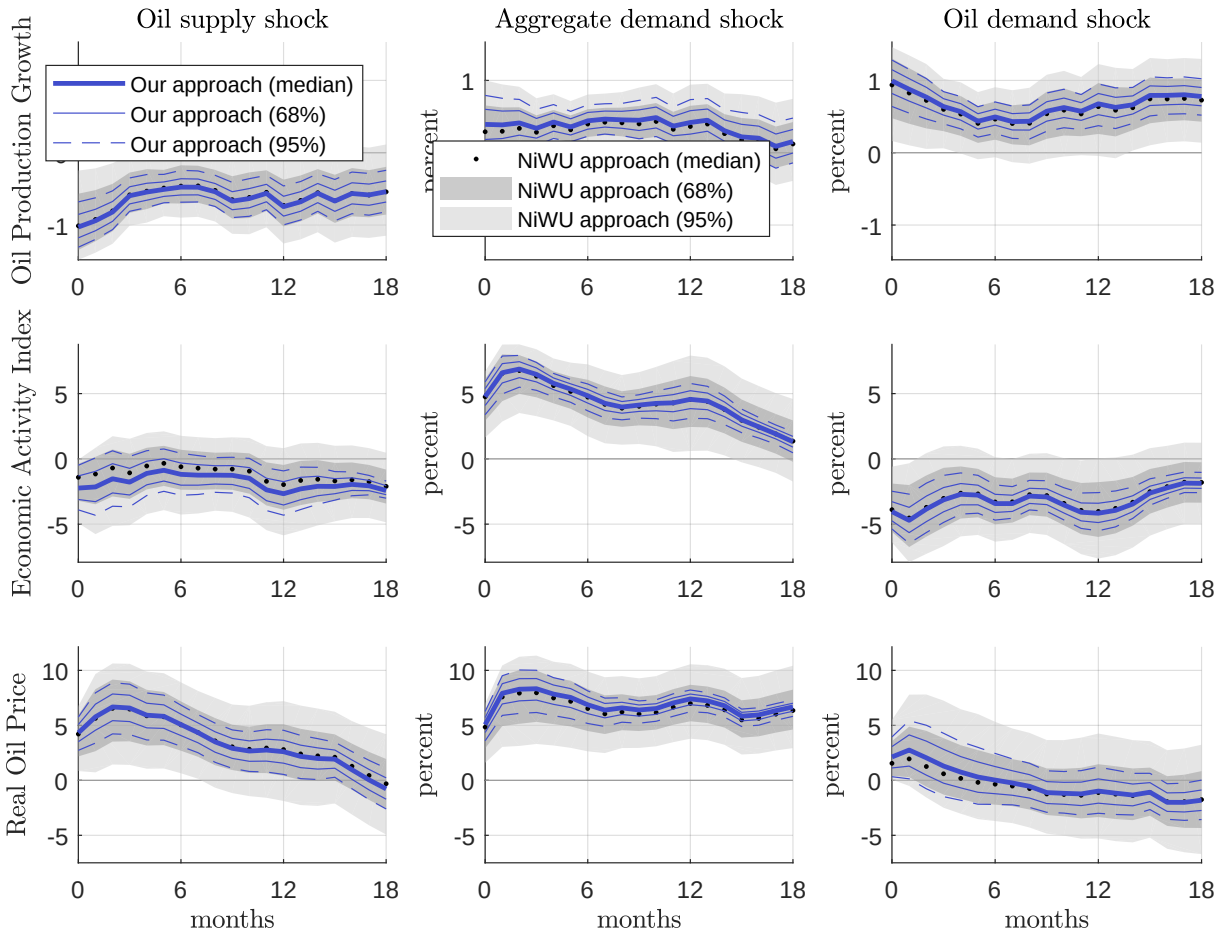
We label the structural shocks by introducing the sign restrictions on the contemporaneous impulse responses used by Kilian and Murphy (2012) and summarized in Table 1. We then depart from Kilian and Murphy (2012) and Antolín-Díaz and Rubio-Ramírez (2018) and do not introduce restrictions on elasticities, nor on the sign of the structural shocks, nor on the historical decomposition. By contrast, we model the sign restrictions from Table 1 using the methodology discussed in Section 2.3. We do so in order to inspect to what extent a more informed way of introducing the same sign restrictions can affect the results. We first set the hyperparameters $\psi_1 = 0.8$ and $\psi_2 = 1.2$ as in the illustrative example from Section 3.

4.2 Results

The posterior distribution of the model is explored using the new algorithm proposed in Section 2.4. We extract 100,000 draws from $p(B|Y)_{NiWU}$ and apply the proposed modification of the importance sampling procedure to convert such draws into draws from $p(B|Y)_{Np(B)}$. The effective sample size in stage A equals 88.70% of the initial draws, suggesting that the sample size, which equals 540, is sufficiently large to allow for the use of our procedure. Given the initial 100,000 draws, $p(B|Y)_{Np(B)}$ is effectively sampled using 11,942 draws when accounting also for stage B of the procedure. On Matlab, the algorithm takes 12m25s to extract draws from $p(B|Y)_{NiWU}$ and 18s to resample the draws and make them representative of $p(B|Y)_{Np(B)}$. We also run the Dynamic Striated Metropolis-Hastings to further show that our sampler correctly explores the posterior distribution. The Dynamic Striated Metropolis-Hastings takes 20h15m23s to run on Fortran, and suggests that the 11,942 effective draws from our sampler are enough to recover the 'correct' posterior distribution. The details of how we set the tuning parameters are discussed in Section E of the Appendix.

Figure 7 shows the results for the impulse responses to one standard deviation shocks, normalized such that the price of oil increases. The solid and dashed lines display the 68 and the 95% pointwise credible sets and the pointwise median when applying our methodology. The shaded areas and the dotted line show the same posterior moments when applying the same sign restrictions using the indirect approach. Consistent with Kilian and Murphy (2012), the indirect approach delivers relatively wide posterior bands that do not allow for an assessment of whether supply or demand shocks are more effective in driving the real price of oil. However, when applying the same sign restrictions using our methodology, the posterior bands tighten, sharpening inference. Consistent with Kilian (2009), oil demand shocks generate an immediate and persistent increase in the price of oil, an increase that then progressively declines, while aggregate demand shocks produce stronger effects also at longer horizons. However, while we confirm the results by Kilian (2009) that demand shocks are important

Figure 7: IRFs, posterior, compare direct and indirect approach



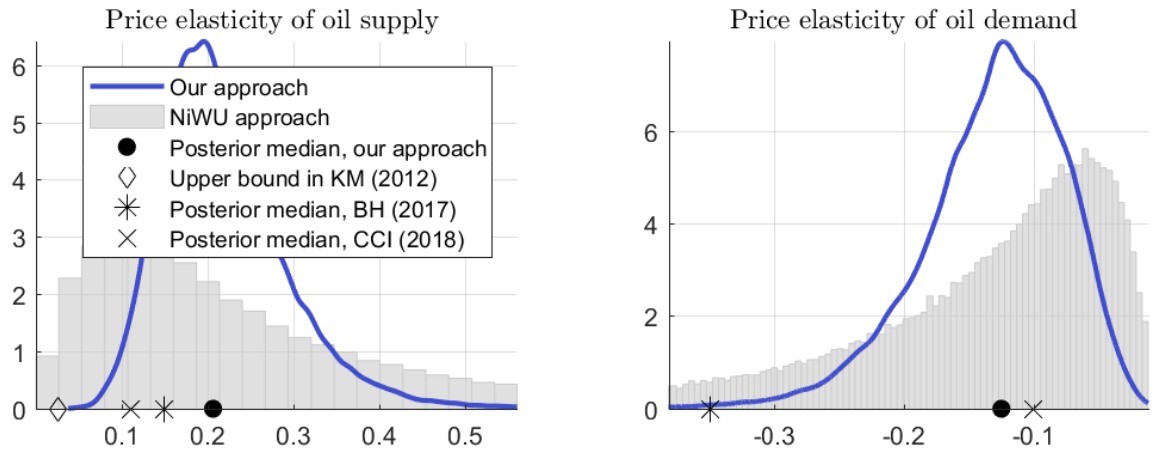
Note: The grey areas and the dotted line show the 68 and 95% pointwise credible sets and the pointwise median under the indirect approach. The blue solid and dashed lines show the same posterior moments under the direct approach proposed in this paper.

drivers of oil price responses, we find that this is more so for aggregate demand shocks rather than oil specific demand shocks.

A notable result from Figure 7 is that oil supply shocks generate sizeable effects on the price of oil, although with smaller effects when focusing on the longer horizon of the response. We find it interesting that the result on the importance of supply shocks is in line with the results by Caldara et al. (2018) and Baumeister and Hamilton (2017), despite the different methodologies used. Caldara et al. (2018) build their analysis on a point-identified model that minimizes the distance between the elasticities implied by

the VAR model and external estimates. Yet, as they show, the actual parametrization of the elasticities have an important effect on the results. Baumeister and Hamilton (2017) also build their analysis on external information on price elasticities on oil, and use a sign restricted framework. They then add information on the dynamics in inventories and measurement error, weigh data differently depending on the period that they correspond to, and combine sign restrictions on elasticities with sign restrictions on the contemporaneous impulse responses. We show that results from Caldara et al. (2018) and Baumeister and Hamilton (2017) are robust to a framework that focuses on the sign restrictions on the contemporaneous impulse responses. Figure 8 shows that the posterior distribution on the price elasticities of oil supply and oil demand implicit in our approach are consistent with the estimates by Baumeister and Hamilton (2017) and Caldara et al. (2018), especially for the supply elasticity.

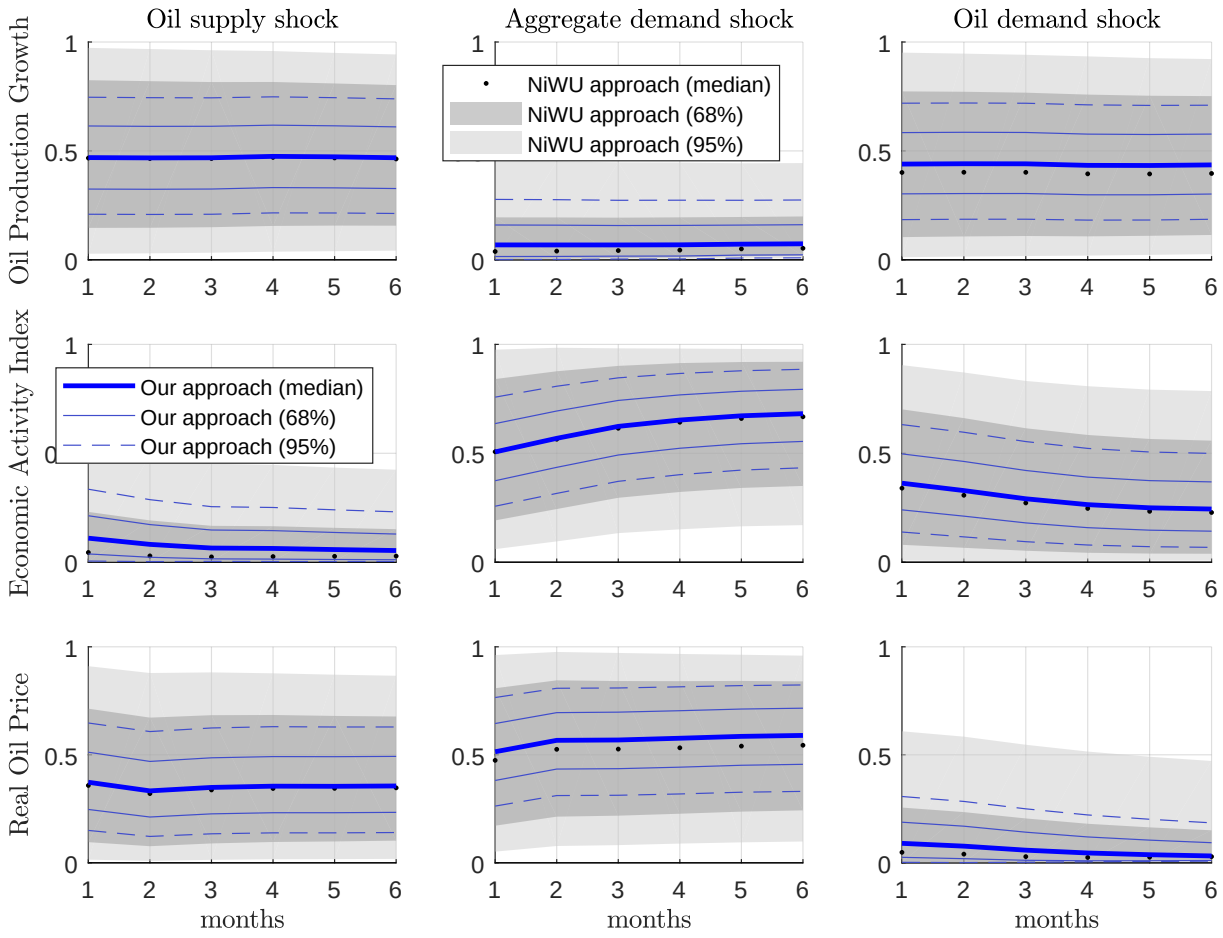
Figure 8: Elasticities, compare direct and indirect approach



Note: Elasticities are computed from the A form of the structural VAR, equation (1). The figure reports the upper limit used by Kilian and Murphy (2012) on the supply elasticity, and the posterior medians estimated by Baumeister and Hamilton (2017), by Caldara et al. (2018) and in our paper. The distribution of the demand elasticity under the traditional approach has some mass where the support takes positive values. This mass is however negligible and equals 0.13%. With our approach this mass is 0.

The analysis of forecast error variance decomposition, shown in Figure 9, confirms that the traditional approach can deliver credible bands that are too wide to imply

Figure 9: Forecast error variance decomposition, compare direct and indirect approach



Note: The grey areas and the dotted line show the 68 and 95% pointwise credible sets and the pointwise median under the indirect approach. The blue solid and dashed lines show the same posterior moments under the direct approach proposed in this paper.

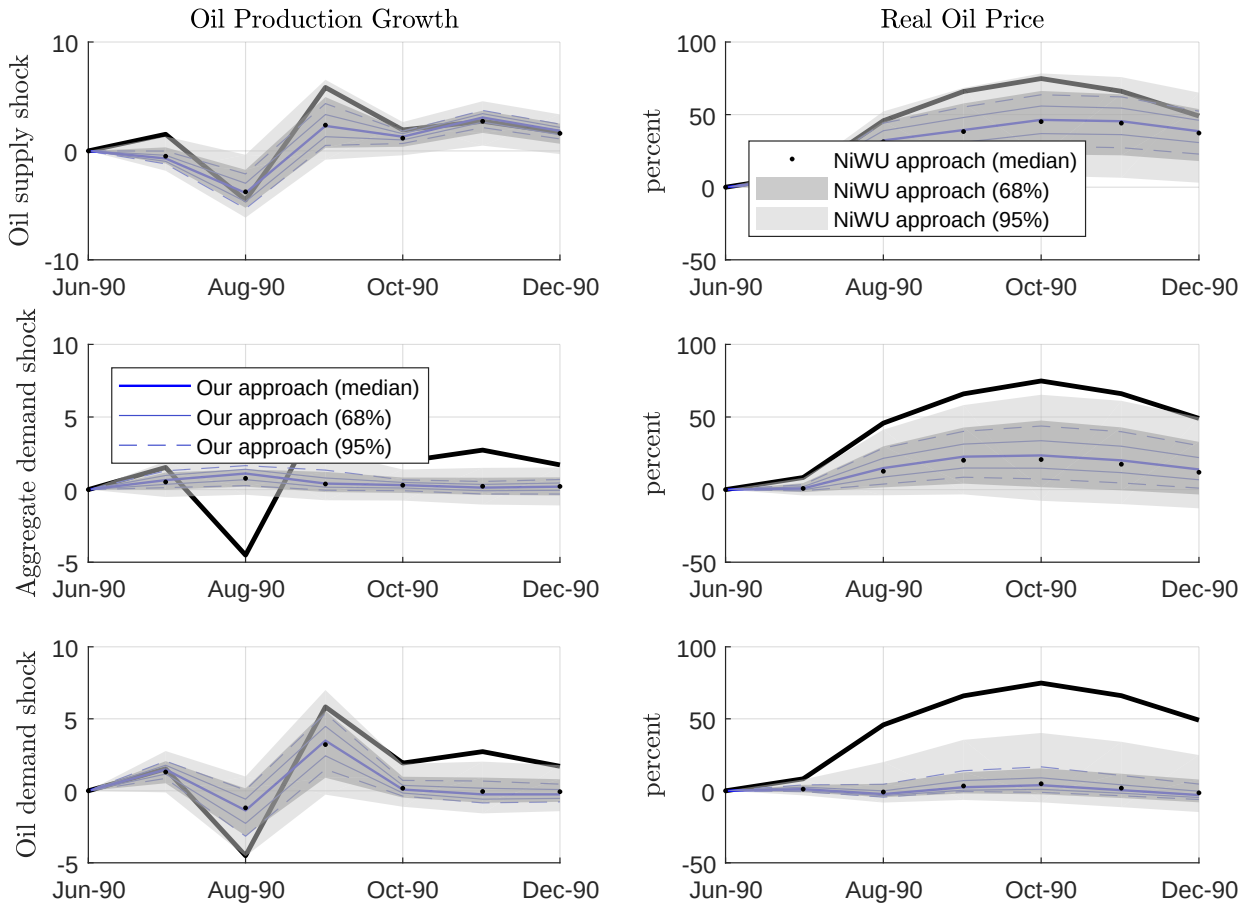
results that can be interpreted. The 95% pointwise credible band can go from close to 0 to close to 1, essentially failing to disclose the role of the structural shocks in driving the variance of forecast errors. By contrast, the inference is much sharper when applying the same sign restrictions using our methodology. We find that the unexpected variations in the price of oil are mainly driven by supply shocks and aggregate demand shocks in similar proportions, while oil demand shocks have a more subdued effect. The result that supply shocks have an important role in driving unexpected variations in the price of oil is consistent with Caldara et al. (2018). As for the forecast errors in

oil production, we find that aggregate demand shocks matter only very little, a result consistent with Antolín-Díaz and Rubio-Ramírez (2018), while oil demand shocks and oil supply shocks matter in similar proportions. The result that unexpected variations in oil production are largely driven by supply rather than aggregate demand shocks is again consistent with Caldara et al. (2018).

We conclude the analysis by displaying the historical decomposition associated with selected periods. Figure 10 and Figure 11 show the historical decomposition for the period corresponding to the Gulf War and for the slump in oil prices from 2014-2015. The figures reports the contribution that the cumulative estimated shocks of each structural shock had on each variable, starting from the beginning of the sample. The figures highlight the results by reporting the decomposition of only oil production growth and the price of oil, while reporting the full analysis in Section G of the Appendix. See Kilian and Lütkepohl (2017), chapter 4, for a detailed discussion of historical decompositions, as well as Antolín-Díaz and Rubio-Ramírez (2018) for an application in sign restricted models.

Figure 10 highlights an interesting result. Antolín-Díaz and Rubio-Ramírez (2018) achieve a sharpening of the credible sets by introducing the restriction that oil supply shocks matter significantly in driving the drop in oil production in August 1990. Indeed, this is the key event in their application, as they discuss. Figure 10 shows that our approach delivers this feature as a result, rather than as a restriction. The credible sets associated with our methodology leaves very little doubt that oil supply shocks were the main drivers of the drop in the oil production, while the indirect approach still leaves considerable uncertainty, leading Antolín-Díaz and Rubio-Ramírez (2018) to introduce the restriction. For the same period, Antolín-Díaz and Rubio-Ramírez (2018) introduce the restriction that aggregate demand shocks had a small role in driving the price of oil. Our results are less consistent with this restriction. As also in Caldara et al. (2018), the initial drop in oil production was driven by supply shocks. We find that the price increase was due to oil supply shocks, followed by aggregate

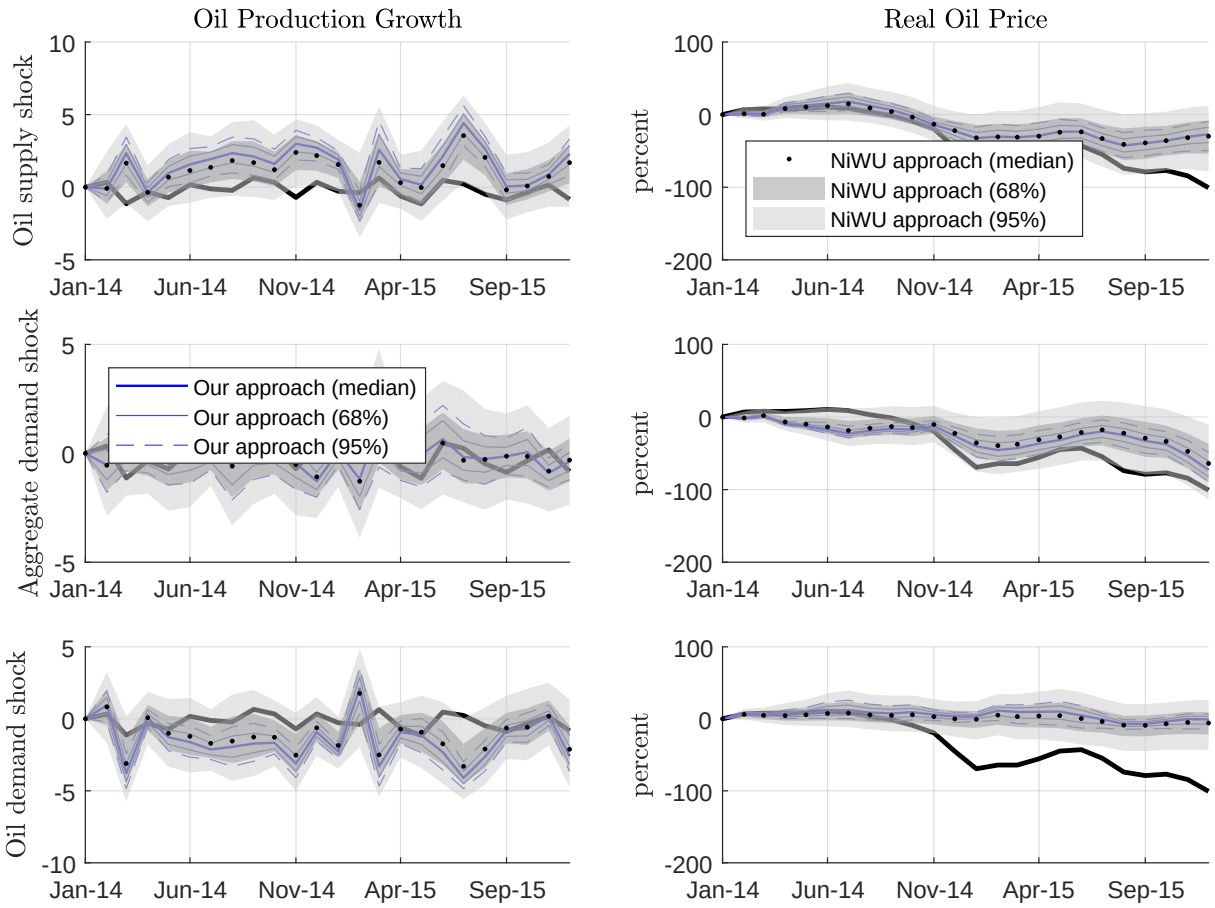
Figure 10: Historical decomposition, Gulf War, January-December 1990



Note: The solid black line shows the data. At each period t and for each structural shock, the data is decomposed into the cumulative contribution of the estimated structural shocks from the beginning of the sample until period t. The grey areas and the dotted line show the 68 and 95% pointwise credible sets and the pointwise median under the indirect approach. The blue solid and dashed lines show the same posterior moments under the direct approach. We subtract from the data and from the decompositions the value corresponding to June 1990, and then compute the pointwise statistics. The figure can be interpreted as percent relative to June 1990. Figure G11 in the Appendix reports the results for all variables in the model.

demand shocks. Figure G9 to Figure G13 in the Appendix complement the analysis by reporting the historical decomposition for the remaining periods for which Antolín-Díaz and Rubio-Ramírez (2018) introduce identifying restrictions. Last, Figure 11 confirms the result from Caldara et al. (2018) and Baumeister and Hamilton (2017) that oil supply shocks contributed to the temporary decrease in the price of oil in

Figure 11: Historical decomposition, January 2014 - December 2015



Note: The solid black line shows the data. At each period t and for each structural shock, the data is decomposed into the cumulative contribution of the estimated structural shocks from the beginning of the sample until period t. The grey areas and the dotted line show the 68 and 95% pointwise credible sets and the pointwise median under the indirect approach. The blue solid and dashed lines show the same posterior moments under the direct approach. We subtract from the data and from the decompositions the value corresponding to January 2014, and then compute the pointwise statistics. The figure can be interpreted as percent relative to January 2014. Figure G14 in the Appendix reports the results for all variables in the model.

2014. Compared to Antolín-Díaz and Rubio-Ramírez (2018), we find that oil demand shocks mattered less. Consistent with Caldara et al. (2018), we find that the initial decrease was generated by oil supply shocks.

5 Conclusions

Structural Vector Autoregressive models are frequently identified using sign restrictions on the impulse response of selected structural shocks of interest. However, it is not clear how this identification approach should be implemented in practice. On the one hand, it is convenient to start from a specification on reduced form parameters, as this makes posterior sampling highly tractable. On the other hand it is important to retain flexibility on the prior beliefs implied for the key structural parameters of interest, since such prior affects the posterior distribution even in a large sample.

We propose an approach that offers flexibility for the prior specification on the impulse response horizon that matters the most, while ensuring that the joint posterior distribution is tractable. We apply this approach to data simulated from the An and Schorfheide (2007) model. We then develop an application to the oil market and show that our approach delivers sharper inference. Consistent with Baumeister and Hamilton (2017) and Caldara et al. (2018), we find that oil supply shocks have a comparable role in explaining oil price dynamics relative to oil demand shocks.

References

Amir-Ahmadi, P. and Drautzburg, T. (2018), 'Identification and inference with ranking restrictions'.

Amisano, G. and Giannini, C. (2012), *Topics in structural VAR econometrics*, Springer Science & Business Media.

An, S. and Schorfheide, F. (2007), 'Bayesian analysis of DSGE models', *Econometric reviews* **26**(2-4), 113–172.

Antolín-Díaz, J. and Rubio-Ramírez, J. F. (2018), 'Narrative sign restrictions for SVARs', *American Economic Review* **108**(10), 2802–29.

Arias, J., Caldara, D. and Rubio-Ramírez, J. F. (forthcoming), 'The systematic component of monetary policy in SVARs: an agnostic identification procedure', *Journal of Monetary Economics* .

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

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

Baumeister, C. J. and Hamilton, J. D. (2017), 'Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks'.

Baumeister, C. J. and Hamilton, J. D. (forthcoming), '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* .

Binning, A. (2013), 'Underidentified SVAR models: A framework for combining short and long-run restrictions with sign restrictions'.

Caldara, D., Cavallo, M. and Iacoviello, M. (2018), 'Oil price elasticities and oil price fluctuations', *Journal of Monetary Economics* .

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

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

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

Fernandez-Villaverde, J., Rubio-Ramírez, J. F., Sargent, T. J. and Watson, M. W. (2007), 'ABCs (and Ds) of understanding VARs', *American Economic Review* **97**(3), 1021–1026.

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

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

Giannone, D., Lenza, M. and Primiceri, G. E. (2015), 'Prior selection for vector autoregressions', *Review of Economics and Statistics* **97**(2), 436–451.

Kilian, L. (2009), 'Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market', *American Economic Review* **99**(3), 1053–69.

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

Kilian, L. and Murphy, D. P. (2012), 'Why agnostic sign restrictions are not enough: understanding the dynamics of oil market VAR models', *Journal of the European Economic Association* **10**(5), 1166–1188.

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

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

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

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

Lütkepohl, H. (2005), *New introduction to multiple time series analysis*, Springer Science & Business Media.

Plagborg-Møller, M. (forthcoming), 'Bayesian inference on structural impulse response functions', *Quantitative Economics* .

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

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

Sims, C. A. (2002), 'Solving linear rational expectations models', *Computational economics* **20**(1), 1–20.

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

Sims, C. A. and Zha, T. (2006), 'Were there regime switches in US monetary policy?', *American Economic Review* **96**(1), 54–81.

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

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

School of Economics and Finance

This working paper has been produced by
the School of Economics and Finance at
Queen Mary University of London

School of Economics and Finance
Queen Mary University of London
Mile End Road
London E1 4NS
Tel: +44 (0)20 7882 7356
Fax: +44 (0)20 8983 3580
Web: www.econ.qmul.ac.uk/research/workingpapers/