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

Martin Bruns[†]   Michele Piffer[‡]

May 20, 2019

Abstract

Structural VAR models are frequently identified using sign restrictions on contemporaneous impulse responses. We develop a methodology that can handle a set of prior distributions that is much larger than the one currently allowed for by traditional methods. We then develop an importance sampler that explores the posterior distribution just as conveniently as with traditional approaches. This makes the existing trade-off between careful prior selection and tractable posterior sampling disappear. We use this framework to combine sign restrictions with information on the volatility of the variables in the model, and show that this sharpens posterior inference. Applying the methodology to the oil market, we find that supply shocks have a strong role in driving the dynamics of the price of oil and in explaining the drop in oil production during the Gulf war.

[†]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

1 Introduction

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

Implementing sign restrictions presents the researcher with a trade-off. There exist infinitely many prior probability distributions that reflect a desired set of sign restrictions. Out of this large class of priors, the literature often limits the analysis to the independent or the conjugate Normal-inverse-Wishart-(Haar)Uniform priors (hereafter NiWU) in order to simplify the analysis of the posterior distribution (Uhlig, 2005, Rubio-Ramirez et al., 2010). However, this constrains the type of prior information introduced by the researcher to the one that can be modelled by the NiWU prior. This is an important limitation, given that, even in a large sample, the results are affected by the specific probability distribution used to model the desired sign restrictions. Yet, moving beyond the NiWU prior makes the posterior distribution (and hence the results) more challenging to analyse (Arias et al., 2018). A trade-off hence emerges between the flexibility in the selection of the prior distribution used, advocated by Baumeister and Hamilton (2015), and the tractability of the posterior distribution, favoured by Rubio-Ramirez et al. (2010).

The first contribution of the paper consists in developing a methodology that makes the above trade-off disappear. We build our methodology on a new importance sampler that uses the posterior distribution of the convenient NiWU case as an importance distribution. While relatively unchallenging to implement, importance samplers require that the importance distribution covers the relevant support of the target distribution (Creal, 2012). When working directly on structural parameters, this condition can be argued to hold only for prior beliefs that do not differ considerably from the

NiWU prior, a case explored by Arias et al. (2018). We build on their work and show that this condition holds for a much wider class of prior beliefs if one builds the importance sampler in two separate stages: first on the reduced form parameters, and second on the mapping into structural parameters. We show that, after acknowledging this point, the trade-off mentioned above disappears: one can follow Baumeister and Hamilton (2015) and use prior beliefs that differ considerably from the ones implied in the NiWU approach, but the sampling of the corresponding posterior distribution does not become technically more involved compared to the techniques developed by Rubio-Ramirez et al. (2010) for the NiWU approach. Accordingly, the methodology offers the most desirable scenario, as it allows for prior flexibility at no additional computational cost. To further confirm the effectiveness of our sampler, we show that the results of the applications in this paper are the same when exploring the posterior distribution using the sequential approach by Waggoner et al. (2016), which is more time-demanding, but also suitable to explore potentially ill-shaped distributions. We first develop our methodology by focusing on the case of only sign restrictions, and then discuss an extension that combines sign and zero restrictions.

The second contribution of the paper consists in proposing a new approach for sign restrictions on impulse responses, which are arguably the most important statistic of SVAR models. On the one hand, starting from prior beliefs directly on impulse responses makes it technically demanding to explore the posterior distribution (see Kociecki, 2010, Barnichon and Matthes, 2018 and Plagborg-Møller, 2019). On the other hand, as discussed above, the use of the NiWU approach reduces the flexibility on the actual prior probability distribution introduced on the parameters of interest. We propose a compromise that parametrizes the structural VAR model as in Uhlig (2005), hence in the reduced form autoregressive elements and in the *contemporaneous* impulse responses. We then depart from Uhlig (2005) and do not restrict the prior on the contemporaneous impulse responses to the one implied by the NiWU prior. Instead, we allow for a general prior distribution. In offering prior flexibility on the

impulse response horizon where flexibility is needed the most (Canova and Pina, 2005 and Canova and Paustian, 2011), our approach provides a balance between prior flexibility on the key structural parameters, and conditionally conjugate priors on all the remaining parameters. We then explore the posterior distribution of the remaining parameters using the newly developed importance sampler.

Having developed a tractable framework that can handle a wide class of prior distributions on the contemporaneous impulse responses, we illustrate that indeed the results in applied work can be sensitive to the prior distribution used. When mapping reduced form parameters into structural parameters, the criterion used in the NiWU approach focuses on orthogonal matrices, namely that orthogonal matrices are drawn from the uniform (or Haar) distribution. This approach can unintentionally treat as equally plausible orthogonal matrices that imply an impact of a one-standard-deviation shock as big as a multiple of the standard deviation of a variable of the model. We propose a prior specification that ensures that the prior mass associated with one-standard-deviation shocks is in line with the scaling of the variables, in a way modelled explicitly by the researcher through a training sample and a set of hyperparameters. We show that this new feature can tighten posterior bands considerably. Compared to the NiWU approach, the tighter posterior bands do not trivially come from tighter priors (and indeed we show that the opposite holds). They come from the fact that the mapping from reduced form to structural parameters is made consistent with the volatility of the variables. Alternative prior specifications are, of course, possible. All in all, the paper suggests that prior beliefs on structural parameters should be selected carefully, as advocated by Baumeister and Hamilton (2015), but also that the NiWU approach advocated by Rubio-Ramirez et al. (2010) offers the required point of departure to explore the posterior distribution associated with this more general approach.

Since the traditional NiWU approach to sign restricted SVARs frequently implies relatively wide posterior bands on impulse responses, many studies have proposed

to combine sign restrictions with additional restrictions on other statistics (see, for example, Kilian and Murphy, 2012, Antolín-Díaz and Rubio-Ramírez, 2018 and Amir-Ahmadi and Drautzburg, 2018). We argue that taking into account the scaling of the variables when forming prior beliefs to model sign restrictions on impulse responses is sufficient to deliver sharper inference, to the point that no additional restrictions are needed to interpret the results. We show this by applying our methodology to the long lasting debate on what drives the unexpected variations in the price of oil and the associated effects on the US economy. Kilian and Murphy (2012) address this question using sign restrictions on contemporaneous impulse responses applied in a setting close to the NiWU approach. They show that sign restrictions alone deliver posterior bands that are too wide to disentangle the different channels driving oil price dynamics. They propose to add restrictions on the elasticity of oil supply, and find that oil demand shocks are the main drivers of oil price dynamics. We show, instead, that applying the same initial sign restrictions in a way that is more consistent with the scaling of the variables can tighten posterior bands considerably without need for restrictions on elasticities (as in Kilian and Murphy, 2012) nor on estimated shocks and historical decompositions (as in the extension of the model by Kilian and Murphy, 2012 proposed by Antolín-Díaz and Rubio-Ramírez, 2018).

More precisely, we construct our application to the oil market as follows. We use a prior probability distribution that treats different structural shocks symmetrically, ensuring that the prior does not favour one shock over the other as drivers of the variables in the model. We then show that the wide posterior bands in Kilian and Murphy (2012) can be replicated using prior beliefs that attach 80% prior probability mass to large effects of one-standard-deviation shocks on the variables of the model, based on an initial quantitative assessment from a training sample. Last, we tighten the prior mass by making it more consistent with the scaling of each variable of the model. While confirming the initial results by Kilian and Murphy (2012) on the importance of oil demand shocks, we find that oil supply shocks have a considerable

effect on oil price dynamics. Quantitatively, we find that as much as 30-40% of the forecast error variance of the real price of oil can be explained by oil supply shocks. Our results confirm the findings by Baumeister and Hamilton (forthcoming) and Caldara et al. (2018). We also 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 complement the work by Sims and Zha (1998) and Baumeister and Hamilton (2015) and study the case of beliefs on contemporaneous impulse responses rather than on the contemporaneous relation among variables. Baumeister and Hamilton (2018) combine prior beliefs on contemporaneous relations and contemporaneous impulse responses. Relative to Baumeister and Hamilton (2018), we focus on impulse responses and propose a different prior specification and posterior sampler. Last, we relate to Giacomini and Kitagawa (2015) in stressing the mapping from reduced form to structural parameters, but we concentrate on a single prior.

The paper is organized as follows. Section 2 outlines the methodology proposed and discusses its relation to the existing literature. Section 3 shows an illustrative example on simulated data based on the estimated bivariate VAR model by Baumeister and Hamilton (2015). Section 4 reports the application to the oil market. Section 5 concludes.

2   The methodology

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

2.1   The model

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

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

$$= \Pi \boldsymbol{w}_t + B\boldsymbol{\epsilon}_t, \qquad\qquad \boldsymbol{\epsilon}_t \sim N(\boldsymbol{0}, I_k), \qquad\qquad (1)$$

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

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

---

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

[2]Whether the model is more conveniently expressed in $A = B^{-1}$ or B (or even in a combined form) depends on whether the identifying restrictions introduced by the researcher are more naturally expressed on the contemporaneous relation among variables or on the contemporaneous effects of the shocks, respectively. Restrictions imposed on one form might not be apparent in the other form, due to the nonlinearities in the mapping from one to another. Going through the publications of all top-five journals and the Journal of Monetary Economics 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.

The reduced form representation of the structural model is

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

$$= \Pi \boldsymbol{w}_t + \boldsymbol{u}_t, \qquad\qquad \boldsymbol{u}_t \sim N(\boldsymbol{0}, \Sigma), \qquad\qquad (2)$$

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

$$B = h(\Sigma)Q, \qquad\qquad (3)$$

and $h(\Sigma)$ a factorization of Σ satisfying $h(\Sigma)h(\Sigma)' = \Sigma$, for example the Cholesky factorization.

2.2   The NiWU approach used in the literature

The most popular approach for sign restricted SVAR models expresses prior beliefs on the parameters $(\boldsymbol{\pi}, \Sigma, Q)$, with $\boldsymbol{\pi} = vec(\Pi)$ the $km \times 1$ vector that stacks the columns of Π. As already discussed in the literature, when $p(\boldsymbol{\pi}, \Sigma)$ falls within either the independent or the conjugate Normal-inverse-Wishart prior, drawing from the joint posterior distribution $p(\boldsymbol{\pi}, \Sigma|Y)$ is technically convenient (see, for example, Koop et al., 2010). One can then extract Q matrices from the Haar-uniform distribution on the parameter space $\mathcal{Q}_\Sigma$, defined as the set of orthogonal matrices such that the sign restrictions on the structural parameters are satisfied, given a draw of Σ. Draws of Q are retained if the sign restrictions are satisfied, and are discarded otherwise.

The convenience of the NiWU approach lies in the existence of efficient algorithms for the sampling of the posterior distribution, developed for example by Rubio-Ramirez et al. (2010). In addition, the possibility of discarding undesired draws allows for the straightforward introduction of sign restrictions not only on contemporaneous impulse responses, but also on future horizons. The inconvenience is that the prior probability

distribution is not directly specified on the structural parameters of interest, namely the impulse responses, but on reduced form parameters and on orthogonal matrices. Since impulse responses are not identified parameters (or in the terminology by Rubio-Ramirez et al., 2010, are not exactly identified parameters), the implicit prior distribution also matters in a large sample and must be selected carefully (Baumeister and Hamilton, 2015).

To appreciate the importance of the above point, consider for simplicity the case of sign restrictions on the contemporaneous impulse responses. The restrictions can be modelled with a probability distribution $p(B)$ that attaches zero mass to the values that do not satisfy the restrictions. However, there are infinitely many probability distributions $\{p(B)_1, p(B)_2, p(B)_2, ...\}$ that reflect the same candidate sign restrictions. Since B is not identified, the posterior distributions $\{p(B|Y)_1, p(B|Y)_2, p(B|Y)_3, ...\}$ differ even in a large sample. Accordingly, not only the sign restrictions are important, but also the actual probability distribution used to model them (Baumeister and Hamilton, 2015). Under the NiWU approach, the flexibility on $p(B)$ is constrained by the fact that it is expressed indirectly through $p(\Sigma, Q)$, that $p(\Sigma)$ must be the inverse Wishart probability distribution, and that $p(Q|\Sigma)$ is uniform in the space $\mathcal{Q}_\Sigma$.[3]

2.3 The Np(B) approach proposed in this paper

To overcome the limitations discussed in the previous section, we propose expressing prior beliefs directly on B. We then develop an importance sampler that ensures that the additional flexibility on the prior specification does not come at a computational cost.

---

[3] That prior beliefs on one parametrization imply questionable or unintended features on some other parametrization is to some extent inevitable. Baumeister and Hamilton (2015) argue that prior beliefs should be judged relative to the structural parametrization of interest, which in our application is B. Arias et al. (2018) derive analytically the distribution implied by the NiWU approach on structural parameters. See also Section B in the Appendix.

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

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

$$p(\boldsymbol{\pi}, B) = p(\boldsymbol{\pi}|B) \cdot p(B). \tag{4}$$

Since $\boldsymbol{\pi}$ is identified, $p(\boldsymbol{\pi})$ matters less compared to $p(B)$, as long as the sample is sufficiently long. Hence, as also in the NiWU approach, we restrict $p(\boldsymbol{\pi})$ to

$$\boldsymbol{\pi} \sim N(\boldsymbol{\mu}_\pi, V_\pi), \tag{5}$$

where $\boldsymbol{\mu}_\pi$ and V_π can be a function of B. By contrast, $p(B)$ allows for a large class of prior distributions, granting the researcher flexibility on the prior beliefs used to express sign restrictions on key structural parameters.[4]

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

$$p(\boldsymbol{\pi}, B|Y) = p(\boldsymbol{\pi}|B, Y) \cdot p(B|Y), \tag{6}$$

where

$$\boldsymbol{\pi}|B, Y \sim N(\boldsymbol{\mu}_\pi^*, V_\pi^*), \tag{7}$$

$$p(B|Y) \propto p(B) \cdot |det(B)|^{-T} \cdot |det(V_\pi)|^{-\frac{1}{2}} \cdot |det(V_\pi^*)|^{\frac{1}{2}} \cdot$$
$$\cdot e^{-\frac{1}{2}\left\{ \tilde{\boldsymbol{y}}'\left(I_T \otimes (BB')^{-1}\right)\tilde{\boldsymbol{y}} - \boldsymbol{\mu}_\pi^{*\prime} V_\pi^{*-1}\boldsymbol{\mu}_\pi^* + \boldsymbol{\mu}_\pi' V_\pi^{-1}\boldsymbol{\mu}_\pi \right\}}, \tag{8}$$

with $\tilde{\boldsymbol{y}}$, W, $\boldsymbol{\mu}_\pi^*$ and V_π^* defined in the Appendix. Extracting from $p(\boldsymbol{\pi}, B|Y)$ requires a suitable posterior sampling procedure for the k^2 elements in $p(B|Y)$, or even for fewer parameters in case zero restrictions are introduced on B. Draws for the km elements

---

[4]As in Baumeister and Hamilton (2015) and Baumeister and Hamilton (forthcoming), we require that $p(B)$ is everywhere nonnegative, and when integrated over the set of all values of B, it produces a finite positive number. If the posterior distribution is then explored with our importance sampler, an additional requirement is that the variance of the weights in Stage A of our algorithm is finite (Geweke, 1989). We return to this point below, as well as in Section C.2 of the Appendix.

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, which is the horizon of the impact effect. On the other hand, as also the NiWU approach, it makes the analysis more tractable by using a normal prior distribution on $\boldsymbol{\pi}$. The normality on $\boldsymbol{\pi}$ is not restrictive except in small samples, given that $\boldsymbol{\pi}$ is identified. Since sign restrictions on impulse responses are frequently introduced contemporaneously rather than on future horizons, we do not view our framework as particularly restrictive. In addition, by parametrizing the model in $\boldsymbol{\pi}$, our approach makes it straightforward to use the prior by Litterman (1986), which is applied directly on $\boldsymbol{\pi}$.

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

To make our approach viable in applied work, we require an efficient algorithm that explores the posterior distribution $p(B|Y)$ from equation (8). When prior beliefs $p(B)$ take the special case implied by the NiWU approach, the posterior distribution $p(B|Y)$ can be explored using existing algorithms for the NiWU approach (Section B of the Appendix). We now develop an extension of such algorithms to allow for a wider class of prior beliefs on B.

We build our sampling procedure on importance sampling techniques. Consider a parameter vector of interest, $\boldsymbol{\theta}$. Suppose 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 evaluate it. In addition, suppose that we can extract proposal draws 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 with replacement the draws $\{\boldsymbol{\theta}_i\}$ obtained from the importance distribution using weights $w(\boldsymbol{\theta}_i) = \frac{p(\boldsymbol{\theta}=\boldsymbol{\theta}_i)^{target}}{p(\boldsymbol{\theta}=\boldsymbol{\theta}_i)^{importance}}$ (see for example Koop, 2003, chapter 4). A popular diagnostic metric is 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, given an initial number of proposal draws. If the importance function sufficiently covers the support of the target function, a small effective sample size suggests increasing the number of draws from the importance function. If, instead, we cannot ensure that the importance function gives sufficient mass to the support of the target function, the importance function must be changed irrespectively of the effective sample size (see the simulation exercise in Section 3).

Define $p(B|Y)_{Np(B)}$ as the posterior distribution associated with the general prior $p(B)$ from our Np(B) approach (equation 8), and $p(B|Y)_{NiWU}$ as the posterior distribution associated with the NiWU approach. Since sampling from $p(B|Y)_{NiWU}$ is not challenging, in principle one could set $\boldsymbol{\theta} = B$, $p(\boldsymbol{\theta})^{target} = p(B|Y)_{Np(B)}$ and $p(\boldsymbol{\theta})^{importance} = p(B|Y)_{NiWU}$. Arias et al. (2018) show that this approach works successfully if the target distribution $p(B|Y)_{Np(B)}$ does not differ too much from the tractable distribution $p(B|Y)_{NiWU}$. However, this procedure does not work in a general framework, because one cannot ensure that $p(B|Y)_{NiWU}$ sufficiently covers the support of $p(B|Y)_{Np(B)}$, except for special cases in which $p(B)_{Np(B)}$ is close to the prior on B implied in the NiWU approach.

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

- $p(\Sigma|Y)_{Np(B)}$: posterior distribution on Σ implied by $p(B|Y)_{Np(B)}$ from equation (8), corresponding to our Np(B) approach;

- $p(\Sigma|Y)_{NiWU}$: posterior distribution on Σ corresponding to the NiWU approach;

- $p(Q|\Sigma)_{Np(B)}$: conditional distribution on Q implicit in the prior $p(B)$ from our approach;

- $p(Q|\Sigma)_{NiWU}$: conditional distribution on Q employed in the NiWU approach, which coincides with a uniform distribution on the space $\mathcal{Q}_\Sigma$;

11

- $p(B)_{Np(B)}$: prior distribution on B used in our Np(B) approach.

Then, notice that drawing from $p(B|Y)_{Np(B)}$ is equivalent to drawing from $p(\Sigma|Y)_{Np(B)}$ and mapping Σ into B using draws of Q from $p(Q|\Sigma)_{Np(B)}$. Accordingly, consider the following importance sampling procedure. First, explore $p(\Sigma|Y)_{Np(B)}$ using $p(\Sigma|Y)_{NiWU}$ as an importance function. Since Σ is identified, $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ are close to each other except in small samples, making $p(\Sigma|Y)_{NiWU}$ a candidate importance function for $p(\Sigma|Y)_{Np(B)}$. Then, use $p(Q|\Sigma)_{NiWU}$ as a proposal function for $p(Q|\Sigma)_{Np(B)}$ to map draws from $p(\Sigma|Y)_{Np(B)}$ into draws from $p(B|Y)_{Np(B)}$. Since $p(Q|\Sigma)_{NiWU}$ is conditionally uniform, it fully explores the parameter space $\mathcal{Q}_\Sigma$, reducing to zero the probability that $p(Q|\Sigma)_{NiWU}$ does not explore the relevant parameter space covered by $p(Q|\Sigma)_{Np(B)}$. In the first stage, a low effective sample size suggests that the sample is too short to imply that $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ are similar distributions, a conjecture that can be verified indirectly by computing the effective sample size and employing existing diagnostic procedures. In the second stage, a low effective sample size only suggests to increase the number of draws from the importance function.

Section C.2 of the Appendix provides a further discussion of the sampler. It gives the analytical form for $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$, and shows that numerically evaluating $p(Q|\Sigma)_{Np(B)}$ only requires evaluating the prior $p(B)_{p(B)}$. When only sign restrictions are introduced on B, when $\boldsymbol{\mu}_\pi$ and V_π are not a function of B, and when the NiWU approach employed to obtain proposal draws is used in its independent prior specification, the sampler can then be implemented using the following algorithm:

Our Algorithm (sign restrictions):

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

1. run a Gibbs sampler to explore $p(\boldsymbol{\pi}, \Sigma|Y)_{NiWU}$ using m_1 burn-in replications and m_2 retained replications. Store the retained draws in $\{\Sigma_d\}_{d=1}^{m_2}$, which by construction represent draws from $p(\Sigma|Y)_{NiWU}$;

2. for each Σ_d compute weights

$$w_d^{\text{stage A}} = \frac{p(\Sigma = \Sigma_d|Y)_{Np(B)}}{p(\Sigma = \Sigma_d|Y)_{NiWU}} \propto \frac{\int_{\mathcal{B}(\Sigma_d)} p(B)_{Np(B)} dB}{|det(\Sigma_d)|^{-\frac{d+k}{2}} \cdot e^{-\frac{1}{2}tr[\Sigma_d^{-1}S]}}, \quad (9)$$

with $\int_{\mathcal{B}(\Sigma_d)} p(B)_{Np(B)} dB$ the integral of $p(B)_{Np(B)}$ along the space of B that implies $BB' = \Sigma_d$. Assess if the effective sample size $ESS^{\text{A}} = \left(\sum_d \left(w_d^{\text{stage A}} / \sum_d (w_d^{\text{stage A}})\right)^2\right)^{-1}$ is sufficiently high;

Stage B: map draws from $p(\Sigma|Y)_{Np(B)}$ into draws from $p(B|Y)_{Np(B)}$:

3. randomly select Σ_d from $\{\Sigma_d\}_{d=1}^{m_2}$ with replacement using weights $w_d^{\text{stage A}}$;

4. draw an orthogonal matrix (Q_d) using the method by Rubio-Ramirez et al. (2010) and compute $B_d = h(\Sigma_d)Q_d$;

    5a. if (B_d) satisfies the sign restrictions (up to sign and ordering of the shocks), store (B_d, Q_d) and proceed to Step 6;

    5b. if (B_d) does not satisfy the sign restrictions, repeat Step 4 up to m_4 times. Stop as soon as (B_d) satisfies the sign restrictions and proceed to Step 6, otherwise discard (Σ_d) and move back to Step 3;

6. repeat Steps 3 to 5 until m_5 draws are stored;

7. for all draws $\{B_d, Q_d\}_{d=1}^{m_5}$ compute weights

$$w_i^{\text{stage B}} = \frac{p(Q = Q_d|\Sigma_d)_{Np(B)}}{p(Q = Q_d|\Sigma_d)_{NiWU}} \propto p(B = B_d)_{Np(B)}. \quad (10)$$

Assess if the effective sample size $ESS^{\text{B}} = \left(\sum_i \left(w_d^{\text{stage B}} / \sum_i (w_i^{\text{stage B}})\right)^2\right)^{-1}$ is larger than a desired minimum number m_6. If so, proceed to Step 8, otherwise move back to Steps 3 to 6 and increase m_5;

13

8. generate $\left\{B_d\right\}_{d=1}^{ESS^\mathrm{B}}$ by resampling the draws $\left\{B_d\right\}_{d=1}^{m_5}$ from Step 7 with replacement using weights $w_d^{\text{stage B}}$.[5]

Our algorithm resamples the posterior draws from the NiWU approach and makes them representative of the posterior distribution associated with the generic prior beliefs $p(B)$ from our approach. In the rest of the paper we document that the sampling time of our algorithm is roughly 6 minutes in the simulation exercise and 30 minutes in the oil application. In Section C.2 of the Appendix we also argue that the size of the dataset beyond which the effective sample size in Stage A is sufficiently high is relatively small in both models considered. The only computationally demanding term to evaluate for our algorithm is $\int_{\mathcal{B}(\Sigma_d)} p(B)_{Np(B)} dB$, which we evaluate numerically as discussed in Section C of the Appendix. Yet, the Appendix shows that an approximate algorithm that sets $\int_{\mathcal{B}(\Sigma_d)} p(B)_{Np(B)} dB = 1$ reaches an almost identical approximation of the posterior distribution, further reducing the computational time.

Our algorithm offers a way of implementing sign restrictions. Two modifications allow extending the algorithm to also account for zero restrictions on B. First, the computation of $p(\Sigma|Y)_{Np(B)}$ (required for the weights in Stage A, Step 2) and the evaluation of $p(Q|\Sigma)_{Np(B)}$ (required for the weights in Stage B, Step 7) must now account for the fact that the mapping from B to Σ features zero restrictions. Accordingly, a numerical approach must be used to compute the corresponding Jacobian transformation, and can be done, for example, using the method developed by Arias et al. (2018). Second, the algorithm generating candidate Q matrices (required in Stage A for the computation of $\int_{\mathcal{B}(\Sigma_d)} p(B)_{Np(B)} dB$ and in Stage B to generate proposal draws for Q), must now be replaced with the methods by either Binning (2013) or Arias et al. (2018). The existing version of the algorithm can then be applied to the case in which zero restrictions are introduced on one structural shock of interest.[6]

---

[5]Since ESS^B can be much smaller than m_5, we resample $\left\{B_d\right\}_{d=1}^{m_5}$ only ESS^B times (or the closest integer), rather than m_5 times. This avoids unnecessary repetitions.

[6]In our algorithm, extracting Q matrices from the algorithm by Rubio-Ramirez et al. (2010) ensures that the orthogonal parameter space $\mathcal{Q}$ is explored uniformly, and hence is fully explored.

To further assess whether the algorithm correctly samples from the posterior, we also explore $p(B|Y)_{Np(B)}$ using the Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016). This alternative algorithm is computationally more demanding, but can handle potentially irregularly shaped posterior distributions and a large number of parameters. Using the posterior distribution from this algorithm, we use simulations to show that the sampling procedure proposed in this section does a good job in exploring $p(B|Y)_{Np(B)}$, even in relatively small samples. Section D of the Appendix discusses how we implement the algorithm by Waggoner et al. (2016).

2.4   Proposing one possible prior $p(B)$

The paper has so far developed an approach that uses a general prior distribution $p(B)_{Np(B)}$ for the contemporaneous impulse responses, while still allowing for fast and efficient posterior sampling. We conclude the section on the methodology by discussing one possible prior specification for $p(B)$. Other prior beliefs are also possible, and must ultimately be chosen by the applied researcher.

Specifying prior beliefs $p(B)_{Np(B)}$ is challenging, because the literature still provides limited guidance on explicit prior beliefs on structural parameters. Baumeister and Hamilton (2015) impose restrictions on B^{-1} rather than on B, and use the existing literature to form prior beliefs on the contemporaneous elasticities among variables. However, as discussed by Kilian and Lütkepohl (2017) and Uhlig (2017), researchers may 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

---

When zero restrictions are introduced, the uniformity in the extraction of Q is lost, except in the case in which zero restrictions are introduced on only one structural shock of interest. In this case, the full relevant orthogonal parameter space is still explored. When zero restrictions are introduced on the effects of more than one shock of interest, the distribution $p(Q|\Sigma)_{NiWU}$ must be evaluated numerically, and the possibility that the relevant part of the orthogonal space is not explored must be addressed.

inflation, but lacks prior beliefs on the scale of such a decrease.

To overcome this challenge, we propose a prior specification for $p(B)_{Np(B)}$ that builds on a conventional prior specification used in the literature for $p(\boldsymbol{\pi})$ known as the Minnesota prior (see, for example, the discussion in Canova, 2007 and Kilian and Lütkepohl, 2017). The crucial step is to take a stand on what is considered a reasonable scale, or magnitude, for the parameters. With the Minnesota prior, 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, using a training sample. Then, Bayesian 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.

We propose to 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

$$-\Sigma_{ii}^{0.5} \leq b_{ij} \leq \Sigma_{ii}^{0.5}, \tag{11}$$

with Σ_{ii} the $i-th$ element of the diagonal of Σ.[7] Accordingly, $\gamma_i = \hat{\Sigma}_{ii}^{0.5}$ provides a candidate assessment of the upper bound 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 width of $p(b_{ij})$. We use independent normal distributions $N(\mu_{ij}, \sigma_{ij})$ as follows:

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

---

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

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

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

Put differently, since the mode of $p(b_{i,j})$ equals $\psi_1 \gamma_i$ (if $b_{i,j}$ is restricted to be positive) or $-\psi_1 \gamma_i$ (if $b_{i,j}$ is restricted to be negative), ψ_1 controls for the first moment of the prior. The hyperparameter ψ_2 then controls for the second moment of the prior, given that ψ_2 is positively related to the probability mass attached to $|b_{i,j} > \gamma_i|$. The convenience of the above approach is that the researcher sets a plausible upper bound for the effect of the shocks by selecting γ_i, and then explicitly introduces Bayesian shrinkage through the hyperparameters ψ_1 and ψ_2. If sign restrictions do not identify all shocks, we suggest numerically introducing the restriction that the non-identified shocks do not replicate the sign restrictions of the identified shocks. Alternative specifications are also possible. Note that the prior proposed in this section ensures that the results do not depend on the ordering of the variables, a convenient property also featured by the NiWU approach.

3 An illustrative example

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

3.1 Simulation exercise

We build the simulation exercise on the model estimated by Baumeister and Hamilton (2015). We first employ ordinary least squares to estimate their bivariate reduced form VAR model, which uses data on the growth rates of the US real labour compensation and total employment from 1970Q1 through 2014Q4, adding a constant and 8 lags. We then use the estimated reduced form VAR as the data generating process. We generate a dataset of 680 draws initializing the data from the estimated unconditional mean. We discard the first 100 draws to make the data less dependent on the initial point, and store the next 100 draws to use as a training sample. We then divide the remaining 480 draws into five pseudo datasets, including up to the first 30, 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.

For each pseudo dataset, we estimate the structural VAR model from equation (1) by introducing sign restrictions on the contemporaneous impulse responses. We identify the demand shock and the supply shock as the structural shocks that move wages and employment in the same and in the opposite direction, respectively. While the models employ the same sign restrictions, we model such restrictions using different prior probability distributions. For the NiWU approach, Section 2.2, we use the independent prior specification and specify the inverse Wishart distribution using the two most popular parametrizations, which are either the improper prior specification or the specification by Kadiyala and Karlsson (1997).[8] For the Np(B) approach, Section 2.3, we introduce prior independence between $\boldsymbol{\pi}$ and B and specify $p(B)$ as discussed in Section 2.4, setting $\psi_1 = 0.8$ and $\psi_2 = 1.5$ for the illustration. For all models estimated, we set $\boldsymbol{\mu}_\pi = \mathbf{0}$ and $V_\pi^{-1} = 0$ for both the NiWU and the Np(B). All

---

[8]The improper prior specification sets $d = 0$ and $S = 0 \cdot I_k$. The parametrization by Kadiyala and Karlsson (1997) sets $d = k + 2$, and sets S such that $E(\Sigma)$ equals the diagonal matrix displaying, on the diagonal, the variance of the residuals in univariate autoregressive processes, estimated on a training sample.

models include a constant term and 8 lags, as in the DGP.

3.2 The intuition behind our importance sampler

We illustrate the intuition behind our posterior sampler by showing the different probability distributions involved in our algorithm. Figure 1 shows the results for the $(1, 2)$ entry of Σ and B for some of the datasets considered (see Section E of the Appendix for the full illustration). The left column of Figure 1 shows the results for Σ and displays the marginal distributions of $\Sigma_{1,2}$ associated with $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$. These are the importance density and the target density in Stage A of the algorithm, respectively. $p(\Sigma|Y)_{Np(B)}$ is sampled using both our algorithm and the Dynamic Striated Metropolis-Hastings algorithm by Waggoner et al. (2016). The closer these two empirical distributions are, the more the algorithm successfully explores the posterior distribution of interest. The right column of Figure 1 reports the equivalent distributions for B. It shows the marginal distribution of $p(B_{1,2}|Y)_{Np(B)}$ explored using either our algorithm or the Dynamic Striated Metropolis-Hastings algorithm, and the proposal distribution obtained when mapping draws from $p(\Sigma|Y)_{Np(B)}$ into B using draws from $p(Q|\Sigma)_{NiWU}$. See Table C2 in the Appendix for how we set the tuning parameters required in our algorithm, and Table E4 for the diagnostics on the importance weights.

As we see from the left column of Figure 1, the dataset with $T = 30$ observations is still too small for $p(\Sigma|Y)_{NiWU}$ (dashed line) to be similar to $p(\Sigma|Y)_{Np(B)}$ (sampled by the DSMH, dotted-dashed line), making the reweighted draws a poor approximation of $p(\Sigma|Y)_{Np(B)}$. However, as the sample size increases, the likelihood dominates. This makes $p(\Sigma|Y)_{NiWU}$ an excellent importance function for $p(\Sigma|Y)_{Np(B)}$ already for $T = 60$, such that the reweighted draws now well approximate $p(\Sigma|Y)_{Np(B)}$ from the DSMH sampler. The right column of the figure displays how successful our algorithm is in sampling the posterior $p(B|Y)_{Np(B)}$. While for $T = 30$ the distribution $p(\Sigma|Y)_{Np(B)}$ is still quite different from the distribution generating proposal draws, the associated

19

Figure 1: Illustration of our algorithm
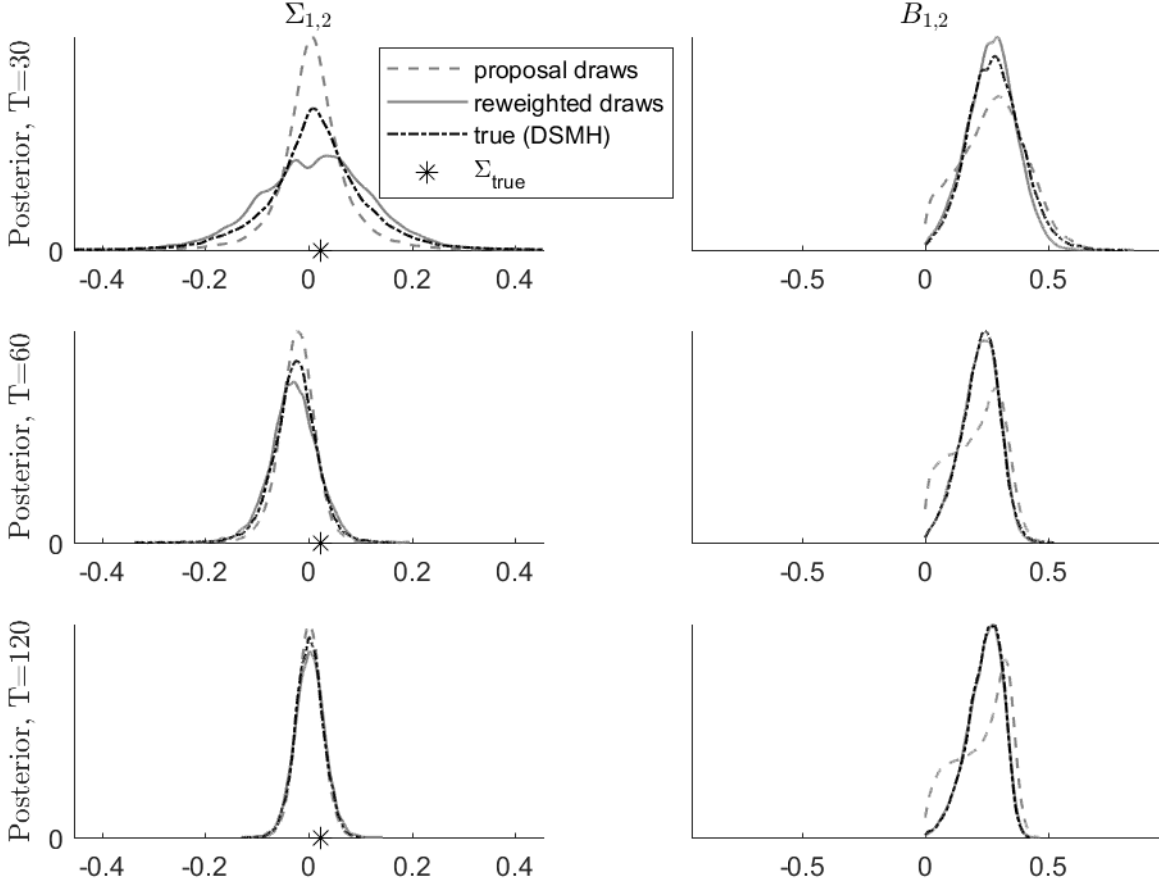
Note: In the left column, the proposal draws are obtained from Step 1 in our algorithm, the reweighted draws correspond to the same draws reweighted using weights from Step 2, while the draws associated with the Dynamic Striated Metropolis-Hastings algorithm are obtained indirectly after running such algorithm on $p(B|Y)_{Np(B)}$. In the right column, the proposal draws correspond to draws obtained from Step 6 of our algorithm, the reweighted draws are obtained from Step 8, and the remaining draws are associated with the Dynamic Striated Metropolis-Hastings algorithm run on $p(B|Y)_{Np(B)}$. See Figure E1 to Figure E5 in the Appendix for the full illustration.

sampling of $p(B|Y)_{Np(B)}$ is already close to what is detected by the Dynamic Striated Metropolis-Hastings algorithm. For the remaining datasets the approximation improves even further, showing that our sampler is successful in sampling the posterior distribution of interest.

Table 1 provides additional intuition for our algorithm by reporting relevant metrics from the sampler. As should be expected, the higher the sample size of the dataset,

Table 1: Performance of our algorithm

		Stage A			Stage B		
T	$T-p$	proposal draws	effective sample size	relative effective sample size	proposal draws	effective sample size	relative effective sample size
		m_2	ESS^A	$\frac{ESS^A}{m_2}$	m_5	ESS^B	$\frac{ESS^B}{m_5}$
30	22	50,000	1,894	0.0379	80,000	45,322	0.5665
60	52	50,000	27,594	0.5519	80,000	56,260	0.7033
120	112	50,000	38,853	0.7771	80,000	53,740	0.6718
240	232	50,000	44,156	0.8831	80,000	52,522	0.6565
480	472	50,000	46,672	0.9334	80,000	51,858	0.6482

the higher the effective sample size in Stage A, further confirming that the importance function tends to coincide with the target distribution. By contrast, in Stage B a high effective sample size is not required for the sampler to successfully explore the posterior distribution.

3.3   Comparison to the NiWU approach

Having discussed the key intuition of the sampler, we now illustrate what drives the difference between the Np(B) and the NiWU approach, and compare the computational time. To improve the comparison, for each dataset we run the NiWU approach to generate the same number of draws that are effectively obtained from the Np(B).

Figure 2 shows the equivalent of Figure 1 by reporting prior and posterior distributions associated with our Np(B) approach and with the traditional NiWU approach.[9] As we see from Figure 2, the prior distributions on $\Sigma_{1,2}$ are quite different, but the associated posterior distributions are very similar already for $T = 60$. By contrast, since B is not identified, differences in prior beliefs on B between the NiWU and the Np(B) approach remain present in the posterior distributions also in a large sample, as $p(Q|\Sigma)_{NiWU}$ and $p(Q|\Sigma)_{Np(B)}$ differ. Figure 2 also shows that the posterior distributions associated with the two parametrizations of the NiWU approach are quite

---

[9]The prior distribution in the NiWU case with the improper prior specification is approximated using $d = k + 2$ and $S = 0.01 \cdot I_2$, see Figure E7 in the Appendix.

Figure 2: Comparison to the NiWU approach
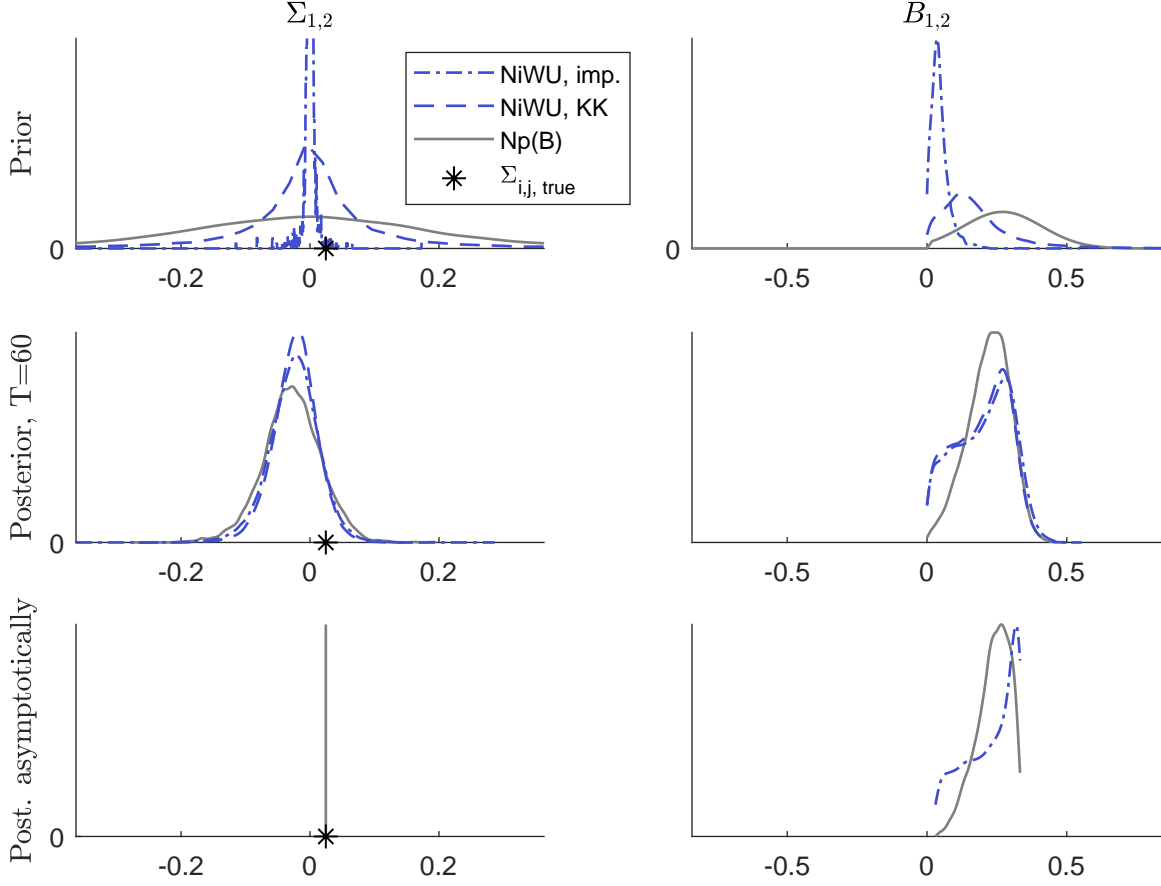
Note: See Figure E6 to Figure E13 in the Appendix for the full illustration.

similar already for $T = 60$. This occurs because, upon learning from the data about Σ, the remaining posterior uncertainty on B largely comes from $p(Q|\Sigma)_{NiWU}$, which is the same irrespectively of the parametrization of the inverse Wishart prior. Note also that $p(B|Y)_{Np(B)}$ is tighter than $p(B|Y)_{NiWU}$, despite $p(B)_{Np(B)}$ being wider than $p(B)_{NiWU}$. This happens because $p(B|Y)_{NiWU}$ inherits posterior uncertainty from $p(Q|\Sigma)_{NiWU}$, which does not take an explicit stand on which part of the structural parameter space the researcher considers more reasonable, but accepts what is implied by the uniformity on $\mathcal{Q}_{\Sigma}$.

The intuition behind the differences between the Np(B) and the NiWU approach

Figure 3: Rotation angle implicit in $p(Q|\Sigma)_{NiWU}$ and $p(Q|\Sigma)_{Np(B)}$
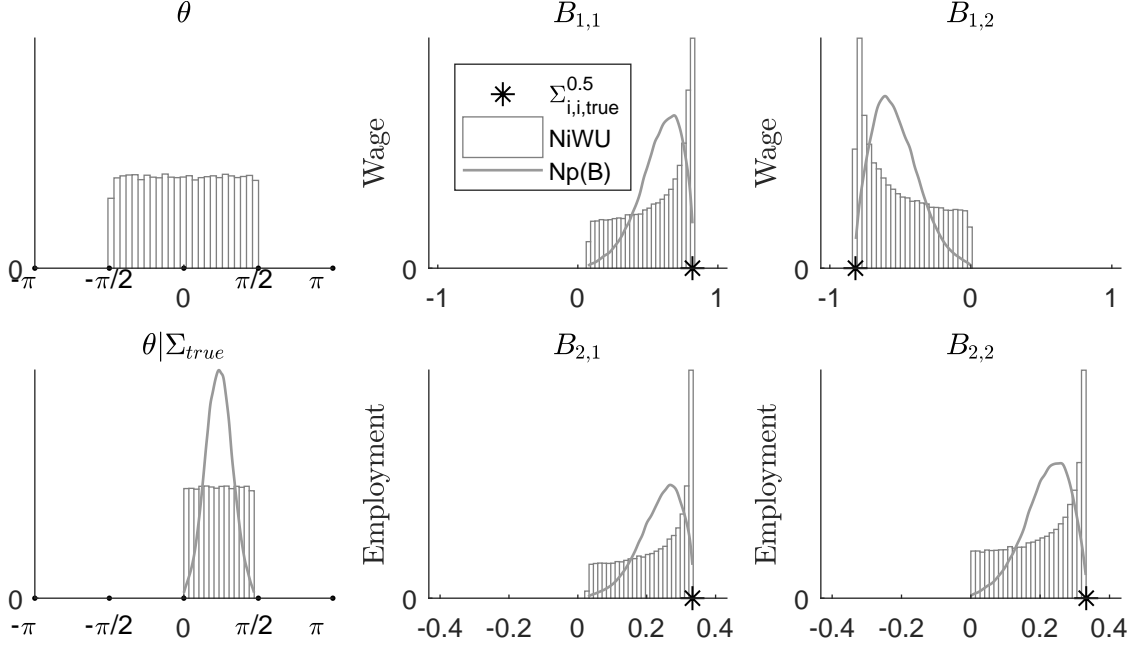
Note: The figure shows the distribution of the ration angle that ensures $\tilde{Q} = Q(\theta)$, with $Q(\theta)$ the Givens transformations matrix $\begin{pmatrix} cos(\theta) & -sin(\theta) \\ sin(\theta) & cos(\theta) \end{pmatrix}$ and Q a draw from either $p(Q|\Sigma)_{NiWU}$ or $p(Q|\Sigma)_{Np(B)}$. See Section C.2 of the Appendix for further details.

can be further clarified by abstracting from estimation uncertainty and comparing $p(Q|\Sigma)_{NiWU}$ to $p(Q|\Sigma)_{Np(B)}$. In the bivariate case, distributions on Q can be shown graphically as the distribution on the corresponding rotation angle θ of Givens transformations matrices (see, for example, Fry and Pagan, 2011, as well as the analysis in Baumeister and Hamilton, 2015). Uniformity on Q is equivalent to uniformity on θ. The top-left plot of Figure 3 shows that indeed the angle of the rotation matrices that replicate draws of Q from the algorithm by Rubio-Ramirez et al. (2010) is uniformly distributed in the support $[-\pi/2, \pi/2]$. Conditioning on Σ_{true}, the rotation angles consistent with the sign restrictions are the subset shown in the bottom-left plot of the figure, which correspond to $\mathcal{Q}_{\Sigma_{true}}$. While the NiWU approach treats such angles as equally plausible, the Np(B) approach does not, taking instead an explicit stand

on the part of the structural parameter space that is considered more in line with the scaling of the variables. The remaining panels of Figure 3 show the implied distribution on B. Given the constraint from equation (11), no draw of b_{ij} is obtained outside of the interval $[-\Sigma_{i,i,true}^{0.5}, +\Sigma_{i,i,true}^{0.5}]$, as displayed in the figure. The NiWU approach implies a distribution that is skewed towards such bounds (see also equation (33) in Baumeister and Hamilton, 2015 and their Figure 1), while the Np(B) approach implies a distribution that reflects $p(B)_{Np(B)}$. As the sample size increases, the posterior distributions $p(B|Y)_{NiWU}$ and $p(B|Y)_{Np(B)}$ approach the ones displayed in Figure 3.[10]

Table 2: Comparison of the computational time

	Np(B) approach													NiWU approach				
	Our algorithm									DSMH algorithm			with improper prior			with KK(1997) prior		
	Stage A			Stage B			Total											
T	h	m	s	h	m	s	h	m	s	h	m	s	h	m	s	h	m	s
30		5	52			20		6	12		3	40			24			22
60		5	32			17		5	50		4	35			30			29
120		5	34			17		5	52		7	5			32			31
240		5	37			17		5	55		28	8			34			32
480		5	52			17		6	9	2	25	12			37			37

Note: All codes are run on Matlab, except for the Dynamic Striated Metropolis-Hastings algorithm, which we coded on Fortran to reduce computational time. The codes run on a computer with an Intel i7-7700K 4.2GHz Quad Core processor and 64 GB RAM.

Table 2 shows the computational time of the Np(B) approach, the NiWU approach, and of the Dynamic Striated Metropolis-Hastings algorithm. All applications of the Np(B) approach and the NiWU approach take only a few minutes to run on Matlab.

---

[10]Figure 3 shows the analysis conditioning on Σ_{true}. As the sample size increases, both $p(\Sigma|Y)_{NiWU}$ and $p(\Sigma|Y)_{Np(B)}$ collapse to a point mass at Σ_{true}, making the analysis conditioning on Σ_{true} relevant as a discussion of the posterior distributions $p(B|Y)_{NiWU}$ and $p(B|Y)_{Np(B)}$. Within the NiWU approach, the fact that the prior beliefs on B differ across parametrizations, while still leading to almost identical posteriors, suggests that it can be misleading to inspect prior beliefs on structural parameters to study what information the NiWU approach introduces to the results. It is, instead, best to consider the analysis conditioning on Σ_{true}. Note also that the uniform distribution in the full space $\mathcal{Q}$ explored by the algorithm by Rubio-Ramirez et al. (2010) is used in our algorithm as a device to explore the distribution that is uniform in $\mathcal{Q}_{\Sigma_{true}}$. The marginal distributions $p(Q)_{NiWU} = \int p(Q|\Sigma)_{NiWU}\, p(\Sigma)_{NiWU}\, d\Sigma$ and $p(Q|Y)_{NiWU} = \int p(Q|\Sigma)_{NiWU}\, p(\Sigma|Y)_{NiWU}\, d\Sigma$ are not necessarily uniform, nor they have mass in the full space $\mathcal{Q}$ (Figure E16 in the Appendix) .

The Dynamic Striated Metropolis-Hastings algorithm takes longer to run due to its sequential nature.

4 Application to the oil market

We now apply our methodology to real data and revisit the model of the oil market by Kilian and Murphy (2012). We show that inference becomes sharper when taking into account the scaling of the variables in forming prior beliefs to introduce the same sign restrictions as Kilian and Murphy (2012). The exercise also illustrates to what extent the results are affected by the actual prior probability distribution used to express the sign restrictions.

4.1 The model

We use the three-variate 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. We use the independent NiWU specification and set $\boldsymbol{\mu}_\pi = \mathbf{0}$ and $V_\pi^{-1} = 0$ for both the NiWU and our Np(B) approach.

We label the structural shocks using the sign restrictions on the contemporaneous impulse responses employed by both Kilian and Murphy (2012) and Antolín-Díaz and Rubio-Ramírez (2018), see Table 3. However, we depart from both papers along two dimensions. First, we do not introduce explicit restrictions on elasticities, nor on the sign of the structural shocks and on the historical decompositions. Second we do not model the sign restrictions through the NiWU approach, but through the

Table 3: Sign restrictions on the contemporaneous impulse responses

A) Sign restrictions used

	oil supply shock	aggregate demand shock	oil-specific demand shock
Oil production	−	+	+
economic activity	−	+	−
real price of oil	+	+	+

B) Prior distributions modelling the sign restrictions

		ψ_1	ψ_2	$Prob(|b_{i,j}| > \gamma_i)$
prior I	*wide prior*	2	4	0.83
prior II	*medium prior*	1	2	0.53
prior III	*tight prior*	0.8	1.5	0.33

prior distribution proposed in Section 2.4. As discussed, this prior first uses a training sample to estimate an indicative upper bound γ_i for the elements $b_{i,j}$, and then allocates prior mass by selecting the hyperparameters ψ_1 and ψ_2. ψ_1 affects the first moment of the marginal prior distribution in $p(b_{i,j})$, whose mode is set equal to $\pm\psi_1\gamma_i$, while ψ_2 controls for the second moment of the prior and is positively related to the prior mass allocated to $|b_{i,j}| \geq \gamma_i$. We explore the role of prior beliefs by using the three separate specifications for ψ_1 and ψ_2 documented in Table 3. Prior I corresponds to a wide prior that attaches approximately 80% probability mass beyond the estimated upper bound γ_i. Priors II and III progressively tighten the prior to make it more consistent with the scaling of the variables, giving approximately 50% and 30% prior probability mass beyond γ_i, respectively. We favour Prior III, which implies a mode of the marginal prior slightly below the estimated upper bound, while still allowing for a non-negligible tail that gives prior mass above this point.[11]

---

[11]We estimate the scale γ_i using a training sample on the first 20% of the available observations, as in Primiceri (2005). The prior distributions, which are shown in Figure F20 of the Appendix, are such that the marginal prior on the effect of different shocks on each variable only differ potentially up to sign but not magnitude, in order not to introduce asymmetries in the results. See also Table F10 for an illustration of the distribution of the prior probability mass.

4.2 Results

To make the analysis more focused, we concentrate our discussion on the drivers of oil price variations, comparing the results from our Np(B) approach to the results from the NiWU approach, parametrized with the independent improper prior specification. We refer to Section F of the Appendix for the analysis of the other variables in the model, as well as for robustness checks and diagnostics on the performance of the sampler.

Figure 4 shows how one-standard-deviation shocks affect the price of oil, and compares the results from the NiWU approach and from our Np(B) approach. The columns of the figure differ for the structural shock considered, while the rows differ for whether prior specification I, II or III is used in the Np(B) approach. The figure shows that the pointwise posterior bands associated with the NiWU approach are quite wide. Indeed, it is this feature that led Kilian and Murphy (2012) and Antolín-Díaz and Rubio-Ramírez (2018) to introduce further restrictions on elasticities and/or shocks and historical decompositions. The posterior bands associated with the Np(B) approach can, instead, be tighter, depending on the prior specification used. We find that, irrespectively of the prior specification, the dataset is sufficiently large for the NiWU and the Np(B) approach to deliver nearly identical posteriors for the bounds $\pm\Sigma_{i,i}^{0.5}$ that constrain $b_{i,j}$ through equation (11) (see Figure F19, Figure F25-Figure F27 in the Appendix). Accordingly, the differences in $p(B|Y)_{NiWU}$ and $p(B|Y)_{Np(B)}$ are strongly influenced by differences in $p(Q|\Sigma)_{NiWU}$ (which is the uniform distribution in $\mathcal{Q}_\Sigma$) and $p(Q|\Sigma)_{Np(B)}$ (which is the distribution implied by the prior beliefs $p(B)_{Np(B)}$ used).

As shown in the first row of Figure 4, prior I from our Np(B) approach replicates the posterior bands from the NiWU approach up to a close approximation. Yet, prior I attaches as much as 80% prior mass to values of B above the estimated reasonable bound γ_i. We view this prior mass as too wide given the scaling of the variables. As shown with prior II and III, tightening $p(B)_{Np(B)}$ to make it more consistent with the

Figure 4: Posterior impulse responses for the real oil price, comparing NiWU and Np(B)
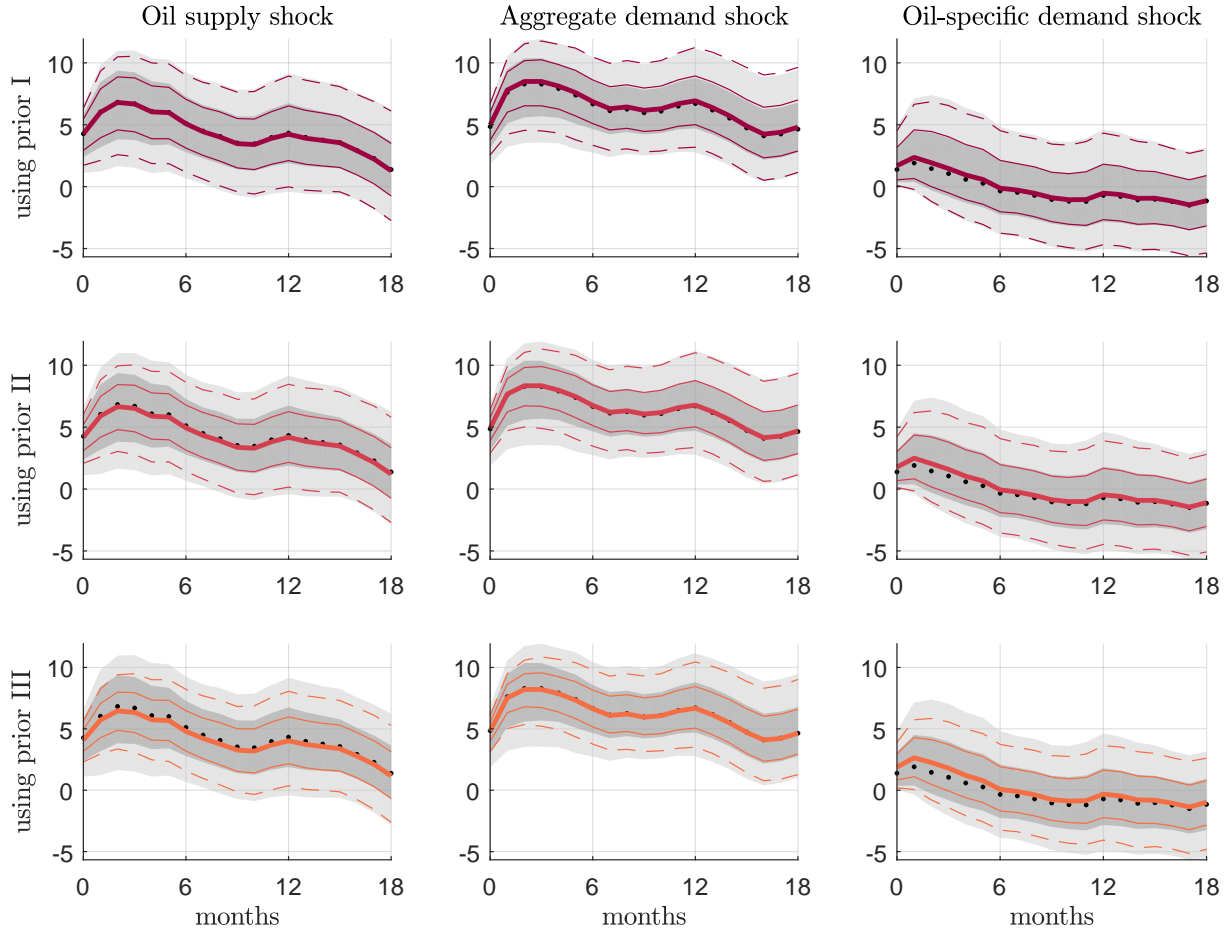
Note: The dotted line and the shaded areas show the pointwise median, 68 and 95% credible bands associated with the improper prior parametrization of the NiWU approach. The remaining solid and dashed lines show the same statistics associated with our Np(B) approach. The rows of the figure differ for the parametrization used for the prior distribution $p(B)_{Np(B)}$, as from Table 3. See Figure F28 to Figure F31 in the Appendix for the full analysis.

scaling of the variables tightens the posterior bands considerably. On the short horizon of the response, the 95% credible bands associated with the Np(B) approach under prior III are as tight as the 68% credible bands of the NiWU approach, making inference sharper. Using prior III, which attaches approximately 30% prior mass to values of B above the estimated upper bounds γ_i, we find that oil-specific demand shocks generate an immediate increase in the price of oil, an increase that then progressively declines,

while aggregate demand shocks produce stronger effects also at longer horizons. While this confirms the results by Kilian (2009) that demand shocks are important drivers of oil price responses, we find that this is more so for aggregate demand shocks rather than oil-specific demand shocks. In addition, we find 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.

The result on the importance of oil supply shocks in driving the price of oil is in line with the results by Caldara et al. (2018) and Baumeister and Hamilton (forthcoming) despite the different methodologies used. Caldara et al. (2018) use an exactly identified model that minimizes the distance between the elasticities implied by the VAR model and external estimates. As they show, the parametrization of the elasticities have an important effect on the results. Baumeister and Hamilton (forthcoming) 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, weight 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 the results in Caldara et al. (2018) and Baumeister and Hamilton (forthcoming) are robust to a framework that focuses on the sign restrictions on the contemporaneous impulse responses. Figure F42 in the Appendix shows that the posterior distributions on the price elasticities implicit in our approach are broadly consistent with the estimates by Caldara et al. (2018) and Baumeister and Hamilton (forthcoming).

The analysis of forecast error variance decompositions, displayed in Figure 5, shows that the NiWU approach can deliver credible bands that are too wide to imply results that can be interpreted. The 95% pointwise credible band can go from close to 0 to close to 1, failing to disclose the role of the structural shocks in driving the variance of forecast errors. By contrast, inference is much sharper when prior mass on key structural parameters is ensured to be in line with the scaling of the variables. As

29

Figure 5: Posterior forecast error variance decomposition for the real oil price, comparing NiWU and Np(B)
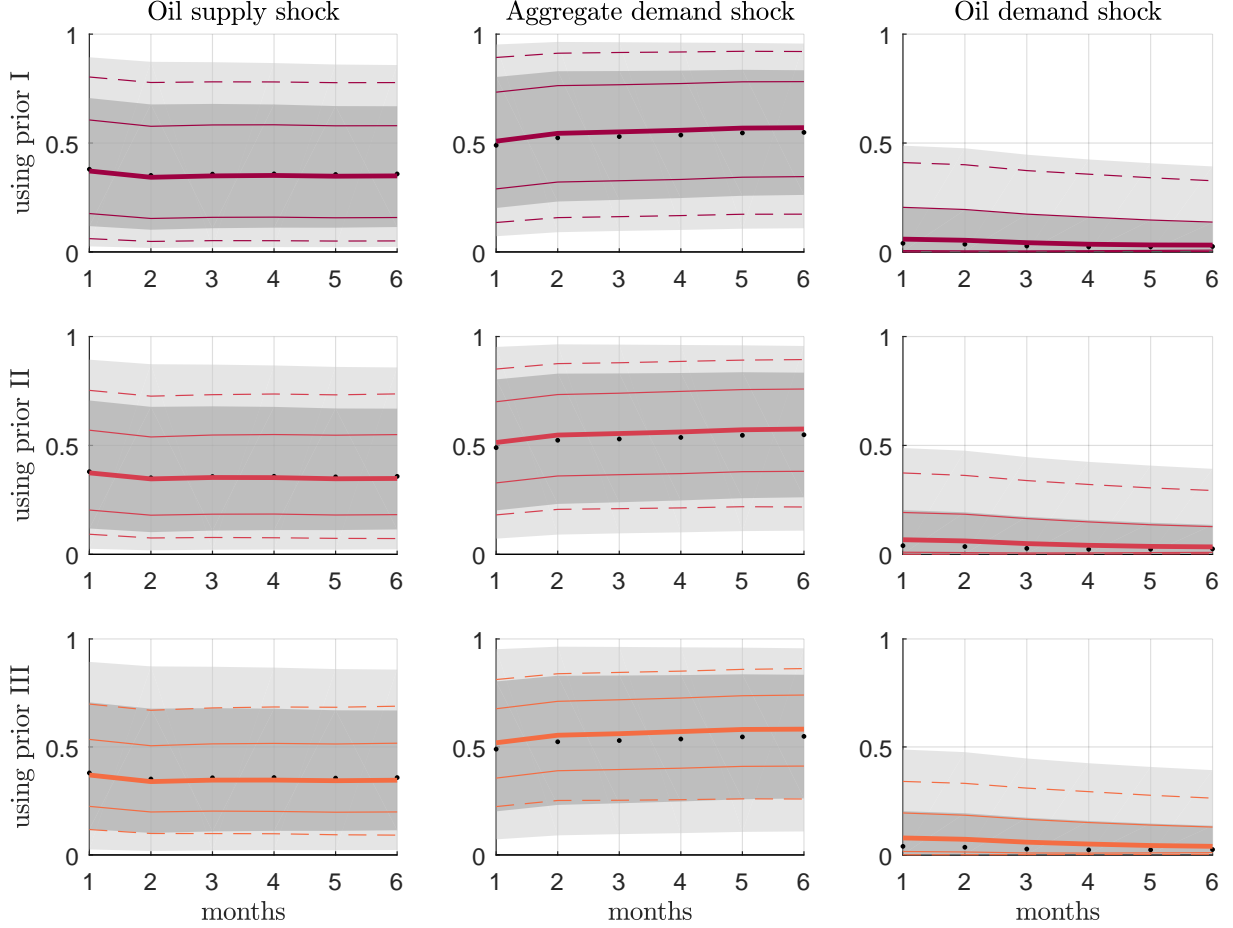
Note: The dotted line and the shaded areas show the pointwise median, 68 and 95% credible bands associated with the improper prior parametrization of the NiWU approach. The remaining solid and dashed lines show the same statistics associated with our Np(B) approach. The rows of the figure differ for the parametrization used for the prior distribution $p(B)_{Np(B)}$, as from Table 3. See Figure F32 to Figure F34 in the Appendix for the full analysis.

we move our prior from specification I to III, we find that the unexpected variations in the price of oil are mainly driven by supply shocks and aggregate demand shocks for approximately 20-50% and 30-60%, respectively, while oil-specific 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).

Figure 6: Historical decomposition, cumulative effects of the shocks
1990: oil supply shocks → oil production growth
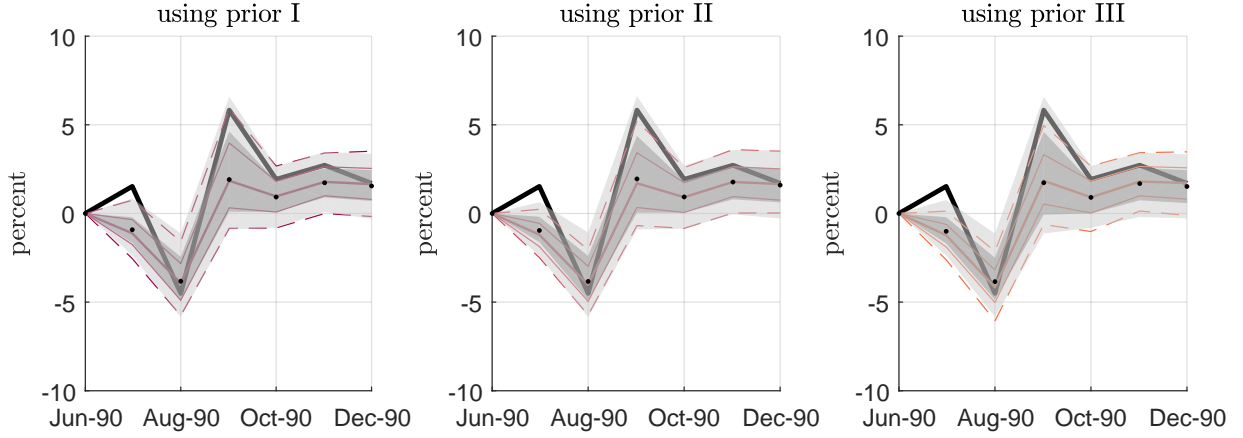
Note: The figure shows the data (solid black line) and its decomposition into the cumulative contribution of the estimated structural shocks from the beginning of the sample until period t. The dotted line and the shaded areas show the pointwise median, 68 and 95% credible bands associated with the improper prior parametrization of the NiWU approach. The remaining solid and dashed lines show the same statistics associated with our Np(B) approach. Having subtracted the value corresponding to June 1990 before computing pointwise statistics, the figure can be interpreted as percent relative to the initial point. See Figure F39 to Figure F41 in the Appendix for the full analysis.

We conclude the analysis by further relating our work to Antolín-Díaz and Rubio-Ramírez (2018). Antolín-Díaz and Rubio-Ramírez (2018) achieve a sharpening of the posterior 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 6 shows that our approach delivers this feature as a result, rather than as a restriction. As we make our prior more in line with the scale of the variables, the credible sets associated with our methodology leave little doubt that oil supply shocks were relevant drivers of the drop in the oil production. By contrast, the NiWU approach delivers wide posterior bands, leading Antolín-Díaz and Rubio-Ramírez (2018) to introduce the restriction. The result that supply shocks contributed to the decline in oil production in August 1990 is also reported by Caldara et al. (2018).

5 Conclusions

Structural Vector Autoregressive models are frequently identified using sign restrictions on the impulse response of selected structural shocks of interest. However, it is not clear how this identification approach should be implemented in practice. On the one hand, it is convenient to use the independent or the conjugate Normal-inverse-Wishart-Uniform prior currently employed in the literature, 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 statistics of interest 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 illustrate the intuition of our approach using simulations on the bivariate demand and supply model by Baumeister and Hamilton (2015). We then develop an application to the oil market and show that inference becomes sharper when prior beliefs are made consistent with the scaling of the variables. We find that oil supply shocks have a comparable role in explaining oil price dynamics relative to oil-specific demand shocks.

References

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

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

Arias, J. E., Caldara, D. and Rubio-Ramirez, J. F. (2019), 'The systematic component of monetary policy in SVARs: an agnostic identification procedure', *Journal of Monetary Economics* **101**, 1–13.

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.

Barnichon, R. and Matthes, C. (2018), 'Functional approximation of impulse responses', *Journal of Monetary Economics* **99**, 41–55.

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

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

Baumeister, C. J. and Hamilton, J. D. (forthcoming), 'Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks', *The American Economic Review* .

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

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

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

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

Creal, D. (2012), 'A survey of sequential Monte Carlo methods for economics and finance', *Econometric reviews* **31**(3), 245–296.

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

Geweke, J. (1989), 'Bayesian inference in econometric models using Monte Carlo integration', *Econometrica: Journal of the Econometric Society* pp. 1317–1339.

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

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

Kilian, L. (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.

Plagborg-Møller, M. (2019), 'Bayesian inference on structural impulse response functions', *Quantitative Economics* **10**(1), 145–184.

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

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

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

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